

**Fábio Macêdo Mendes**

*Entropias Generalizadas e os Fundamentos  
Estatísticos da Termodinâmica*

Brasília, 2006

Fábio Macêdo Mendes

*Entropias Generalizadas e os Fundamentos  
Estatísticos da Termodinâmica*

Sobre o papel das entropias generalizadas  
e equiprobabilidade no contexto de uma  
teoria de contagem a partir da fórmula de  
Boltzmann  $S = k \ln W$

Orientador:  
Anníbal Dias Figueiredo Neto

UNIVERSIDADE DE BRASÍLIA  
NÚCLEO DE FÍSICA MATEMÁTICA DO INSTITUTO DE FÍSICA

Brasília, 2006

Dissertação de Mestrado sob o título "*Entropias Generalizadas e os Fundamentos Estatísticos da Termodinâmica*", defendida por Fábio Macêdo Mendes e aprovada em 10 de Abril de 2006, em Brasília-DF, pela banca examinadora constituída pelos doutores:

Prof. Annibal Dias Figueiredo Neto  
Universidade de Brasília (orientador)

Prof. Chang Chung Yu Dorea  
Universidade de Brasília

Prof. Tarcísio Marciano da Rocha  
Universidade de Brasília

Prof. Ademir Eugênio de Santana  
Universidade de Brasília (suplente)



# *Resumo*

Estudo sobre os fundamentos estatísticos da termodinâmica com ênfase na fórmula da entropia de Boltzmann e os métodos de contagem também introduzidos por ele. Ao adotar um ponto de vista moderno sobre o significado das probabilidades, a termodinâmica segue, com pouquíssimas adições e de maneira bastante geral, apenas da identificação  $S(X) = \ln P(X)$ . Ao enfatizar o modelo estatístico em oposição a leis da física, nossa apresentação não fica restrita ao equilíbrio e nem a fórmulas predeterminadas para a entropia. Seguindo este quadro conceitual unificado, esboçamos um formalismo de não-equilíbrio bastante incipiente e introduzimos o conceito de entropia generalizada de maneira bastante natural, onde vemos que a forma usual decorre de uma hipótese com algum apelo físico, mas estritamente desnecessária.



# *Abstract*

An study on the statistical foundations of thermodynamics which emphasizes Boltzmann's counting procedures and entropy formula. Once one adopt a modern point of view regarding the meaning of probabilities, thermodynamics follows, with very few additions and in a very general fashion, just from the identification  $S(X) = k \ln P(X)$ . By choosing to focus on the statistical model instead of physical law, our description isn't stuck to equilibrium situations and neither to predefined entropy formulas. A seminal non-equilibrium formalism is presented and the concept of generalized entropies follows on a very natural fashion where the usual formula is implied by an hypothesis with some physical appeal, but strictly unnecessary.





# *Agradecimentos*

Aos meus pais, familiares e amigos pelo carinho, inspiração e paciência. À Mari por tudo isso e duplamente pela paciência. Aos mestres que me apresentaram este legado monumental. Ao meu orientador pelos votos de confiança e pela sua visão. Aos tantos outros físicos que me iluminaram, mas também (um pouco) àqueles que confundem. Ao Google e ao chimarrão, que são fundamentais. Aos bons momentos, besteiras e cachaças compartilhadas durante estes anos.

Um agradecimento especial ao CNPq e ao Instituto de Física da UnB sem os quais este trabalho nunca teria sido desenvolvido. Mas, principalmente, sou grato à Natureza pela indulgência para com as nossas pretensões desesperadas em compreendê-la e por fazer deste processo, a ciência, algo tão recompensador. E também aos que não entendem a Física, mas sabem tolerar esse bando de malucos.



# *Prólogo*

POEE is one manifestation of THE DISCORDIAN SOCIETY about which you will learn more and understand less.

We are a tribe of philosophers, theologians, magicians, scientists, artists, clowns, and similar maniacs who are intrigued with ERIS GODDESS OF CONFUSION and with Her Doings. (...) For further information, consult your pineal gland.

(MALACLYPSE, THE YOUNGER — Principia discordia)

Physical laws should have mathematical beauty and simplicity.

(PAUL A. M. DIRAC)

## *A Zen Story*

Texto extraído de THE MAGNUM OPIATE OF MALACLYPSE THE YOUNGER. Principia Discordia, OR, How I Found Goddess And What I Did To Her When I Found Her. Wherein is Explained Absolutely Everything Worth Knowing About Absolutely Anything — *by Camden Benares, The Count of Five Headmaster, Camp Meeker Cabal*

A serious young man found the conflicts of mid 20th Century America confusing. He went to many people seeking a way of resolving within himself the discords that troubled him, but he remained troubled.

One night in a coffee house, a self-ordained Zen Master said to him, "go to the dilapidated mansion you will find at this address which I have written down for you. Do not speak to those who live there; you must remain silent until the moon rises tomorrow night. Go to the large room on the right of the main hallway, sit in the lotus position on top of the rubble in the northeast corner, face the corner, and meditate."

He did just as the Zen Master instructed. His meditation was frequently interrupted by worries. He worried whether or not the rest of the plumbing fixtures would fall from the second floor bathroom to join the pipes and other trash he was sitting on. He worried how would he know when the moon rose on the next night. He worried about what the people who walked through the room said about him.

His worrying and meditation were disturbed when, as if in a test of his faith, ordure fell from the second floor onto him. At that time two people walked into the room. The first asked the second who the man was sitting there was. The second replied "Some say he is a holy man. Others say he is a shithead."

Hearing this, the man was enlightened.

# Sumário

## Introdução

<b>1</b>	<b>Fundamentos da estatística Bayesiana</b>	p. 23
1.1	Conceito de probabilidade . . . . .	p. 23
1.1.1	. . . . .	p. 26
1.1.2	Da atribuição de probabilidades . . . . .	p. 26
1.2	Distribuições de Probabilidade . . . . .	p. 28
1.2.1	Densidade de probabilidade contínuas . . . . .	p. 30
1.2.2	Entropia estatística . . . . .	p. 32
1.2.3	Entropia e inferência . . . . .	p. 34
1.3	O processo de inferência . . . . .	p. 36
1.3.1	Teorema de Bayes . . . . .	p. 36
1.3.2	Processo de inferência e determinação da verossimilhança . . .	p. 38
1.3.3	O significado da entropia inferencial . . . . .	p. 42
1.3.4	Entropias generalizadas: sub-localidade . . . . .	p. 43
<b>2</b>	<b>Física de sistemas macroscópicos</b>	p. 47
2.1	Indeterminismo macroscópico . . . . .	p. 48
2.1.1	Probabilidades associadas a processos macroscópicos . . . . .	p. 50
2.1.2	Ensemble . . . . .	p. 52
2.1.3	Como determinar a probabilidade anterior? . . . . .	p. 57
2.1.4	Teoria ergódica . . . . .	p. 63

2.1.5	Aproximação mecânica . . . . .	p. 66
2.2	Irreversibilidade . . . . .	p. 67
2.3	Entropia e termodinâmica . . . . .	p. 70
2.3.1	Seta do tempo e entropia . . . . .	p. 72
2.3.2	Determinação dos estados de equilíbrio . . . . .	p. 75
<b>3</b>	<b>Estrutura dinâmica</b>	p. 77
3.1	Distribuição número de partículas . . . . .	p. 77
3.2	Uso de integrais funcionais . . . . .	p. 79
3.2.1	Incorporando informação . . . . .	p. 83
3.2.2	Tempos assintóticos . . . . .	p. 84
3.3	Formalismo de equilíbrio . . . . .	p. 86
3.4	Entropias fisicamente plausíveis . . . . .	p. 89
3.5	Entropias para bósons e férmions . . . . .	p. 92
3.6	Aproximação de campo médio e as distintas fases . . . . .	p. 96
<b>4</b>	<b>Termodinâmica</b>	p. 99
4.1	Postulados termodinâmicos . . . . .	p. 99
4.2	O conceito de temperatura . . . . .	p. 101
4.3	Transformada de Legendre . . . . .	p. 102
4.4	Calor e trabalho . . . . .	p. 105
4.5	Entropia à temperatura zero . . . . .	p. 107
<b>5</b>	<b>Estatística de Gibbs</b>	p. 111
5.1	Equivalência entre os formalismos . . . . .	p. 112
5.1.1	Método do ponto de sela . . . . .	p. 114
5.1.2	Entropia de Gibbs . . . . .	p. 117
5.2	Teoria fora do equilíbrio . . . . .	p. 120

5.2.1	Médias de grão grosseiro . . . . .	p. 122
5.2.2	Formalismo MaxEnt fora do equilíbrio . . . . .	p. 123
<b>Conclusão</b>		p. 125
5.1	Uma pequena digressão . . . . .	p. 125
5.2	Probabilidades . . . . .	p. 127
5.3	Entropia . . . . .	p. 129
5.4	Perspectivas futuras . . . . .	p. 130
<b>Apêndice A - Regras de Cox</b>		p. 133
A.1	Cálculo de predicados . . . . .	p. 134
A.2	Raciocínio indutivo . . . . .	p. 136
A.2.1	Plausibilidade da conjunção . . . . .	p. 138
A.2.2	Valores de certeza e impossibilidade . . . . .	p. 141
A.2.3	Plausibilidade da negação . . . . .	p. 142
A.2.4	Plausibilidade da disjunção . . . . .	p. 144
A.2.5	Conjuntos exaustivos e mutualmente exclusivos (EME) . . . . .	p. 146
<b>Índice Remissivo</b>		p. 149





# *Introdução*

Dado uma inteligência que pudesse compreender todas as forças que animam a natureza e a situação respectiva dos seres que a compõem — uma inteligência suficientemente vasta para submeter todos esses dados a uma análise — englobaria na mesma fórmula os movimentos dos maiores corpos do universo e os do mais pequeno átomo; para ela, nada seria incerto e o futuro, tal como o passado, seriam presente aos seus olhos.

(LAPLACE — Essai philosophique sur les probabilités)

Devido à sua amplitude, física estatística é difícil de se definir. Este trabalho analisa o problema fundamental proposto e sistematizado pela geração de Maxwell, Boltzmann e Gibbs; *como explicar e compatibilizar a termodinâmica com as leis microscópicas de movimento*. O resultado amplamente conhecido destas investigações é categoricamente ilustrado pela fórmula

$$S = k \ln W(x) \tag{1}$$

que relaciona a entropia termodinâmica ( $S$ ) com as variáveis microscópicas ( $x$ ) através do peso estatístico dos estados  $W(x)$ .

Note que apesar de haver uma certa uniformidade do ponto de vista metodológico, a física estatística ainda não estabeleceu um consenso com relação à interpretação de seus mais importantes conceitos (probabilidades, entropia, irreversibilidade, etc). Em vista disso, é necessário esclarecer que seguimos uma linha que decorre conceitualmente da vertente desenvolvida por E. T. Jaynes [?, ?, ?] e outros colaboradores sob a égide da estatística Bayesiana. Por motivos que não interessam agora, nota-se uma certa reverência à figura de Josiah Willard Gibbs na literatura Bayesiana — sem querer diminuir suas contribuições fundamentais, neste trabalho nos voltamos a um outro mestre. Em grande medida, o conteúdo apresentado consiste em recuperar algumas idéias de Boltzmann em relativo desuso na prática comum enquanto rechaçamos outras de suas idéias com bastante suporte na literatura.

Neste ponto delineamos o que será apresentado nos capítulos seguintes e quais foram (possivelmente) as contribuições deste trabalho.

O primeiro capítulo consiste numa apresentação da estatística Bayesiana. Tomamos, no entanto, um ponto de vista ligeiramente diferente do usual ao interpretar a metodologia de maximização da entropia de Shannon explicitamente como um processo de inferência. Esta linha de raciocínio decorre de ([?]), ainda que modificamos ligeiramente o argumento original para explicitar a diferença entre a entropia de Shannon e outras entropias generalizadas no contexto puramente estatístico. Neste sentido introduzimos o conceito de sub-localidade, ou seja, entropias generalizadas estariam relacionadas ao fato de informação local possuir um efeito “ligeiramente” não-local no processo de inferência. Esta não-localidade estaria traduzida na existência de correlações entre os subdomínios e eventualmente se refletiria na não-aditividade das entropias consideradas (note que é difícil interpretar a aditividade como conceito primitivo no contexto de uma entropia informacional).

O segundo capítulo, intitulado “Física de sistemas macroscópicos”, lida com o aparato conceitual relacionado à física estatística. Nele tratamos de conceitos como irreversibilidade macroscópica e sua relação com o crescimento da entropia. Mais ainda, tentamos mostrar como a irreversibilidade surge naturalmente a partir duma lei de probabilidades simples, “a seta do tempo”, que relaciona a probabilidade de um processo ocorrer do passado para o futuro e o contrário. Mais ainda, tentamos livrar nossa apresentação de conceitos desnecessários como ergodicidade e uma interpretação excessivamente literal dos ensemble, discutindo criticamente estas questões. As contribuições (possivelmente) originais aparecem pontualmente e consistem num raciocínio por “regras de simetria” para fundamentar a equiprobabilidade do espaço de fase, a lei da seta do tempo, que apesar de ser um resultado trivial é de extrema importância conceitual e, por fim, a introdução do conceito de “aproximação mecânica”. *Aproximação mecânica* consiste na possibilidade de tratar um sistema estatístico como um sistema puramente mecânico. Por isso entendemos que a probabilidade que o sistema realize uma determinada evolução macroscópica  $F(t)$  é muito acentuada com relação a uma certa trajetória “mecânica”, de sorte que podemos ignorar as outras trajetórias.

Já o terceiro capítulo é dedicado a extrair resultados matemáticos mais concretos destas idéias. Nele introduzimos a representação da *distribuição de número de partículas* (DNP), para a qual pode valer a *aproximação mecânica* discutida anteriormente. Primeiramente desenvolvemos uma teoria para a evolução das DNPs que se apóia na idéia de integração sobre todas as trajetórias devido à Feynman. Diferente deste, que estava preocupado com a interpretação da mecânica quântica, nossa teoria diz respeito a probabilidades de forma que, literalmente, somamos sobre todas as possíveis trajetórias (a exponencial

decrecente da ação) no cômputo da probabilidade de uma transição entre dois estados da DNP. Feynman obviamente desenvolveu uma teoria para as amplitudes de probabilidades e tal “soma sobre todas as possibilidades” não possui um significado estatístico muito claro; é sim um artefato misterioso da natureza quântica da matéria. No nosso domínio, felizmente, não se verifica estas dificuldades.

Seguindo a mesma linha de raciocínio, mostramos que um sistema que evolui irreversivelmente para uma situação de equilíbrio permite a identificação de certos termos da sua Lagrangiana com sua entropia em tempos longos. Isto é a base para uma teoria de equilíbrio na qual a entropia como função da DNP exerce um papel fundamental. Tentamos encontrar qual seria a fórmula mais geral para a entropia em tal teoria a partir das propriedades esperadas para um sistema físico altamente idealizado sem interações entre suas partículas. A partir daí recuperamos as “entropias quânticas” que lidam com férmions e bósons além de também ser possível escrever outras entropias generalizadas (como Tsallis, Renyi). Neste ponto interpretamos as entropias generalizadas como entropias efetivas associadas a processos para o qual não se possui uma descrição completa em termo de todos invariantes de movimento relevantes. Daí, por exemplo, entendemos a necessidade de ajustar um certo parâmetro aos dados experimentais.

O capítulo seguinte desta apresentação lida com a questão da compatibilidade entre a física estatística e a termodinâmica. Neste mostramos, uma a uma, que as leis da termodinâmica são sempre respeitadas no que se refere a sistemas ideais mesmo se considerarmos as estatísticas generalizadas. Este é talvez o resultado original mais importante: a termodinâmica independe em grande maneira da forma específica da entropia que se utiliza. Aqui defendemos o ponto de vista que não há qualquer distinção de primeiros princípios entre física estatística e termodinâmica, a segunda sendo apenas uma coleção de teoremas úteis de “alto nível” sobre uma situação específica (de equilíbrio) da primeira teoria.

O último capítulo, por fim, apresenta o formalismo de Gibbs e Jaynes no equilíbrio e a sua relação com os procedimentos aqui apresentados. Mostramos como as entropias generalizadas se manifestariam neste esquema, que aparecem simplesmente como um peso de não-equiprobabilidade na entropia relativa. A partir disto, generalizamos o método de Darwin-Fowler que tradicionalmente é apresentado como uma justificativa para a eficácia dos métodos de Gibbs para lidar com o caso não-equiprovável (entropias generalizadas).

Por fim, expomos uma breve discussão sobre um ponto de vista mais geral adotado pelo autor. Para o bom entendimento das motivações que guiaram este trabalho é fundamental, antes de tudo, avaliar como o raciocínio estatístico se encaixa no esquema geral de uma teoria mecânica; aí se torna nebuloso o raciocínio de Boltzmann e de vários de seus contemporâneos. Adotamos o ponto de vista desenvolvido por Pierre Simon Laplace a mais de 200 anos no seu “Ensaio Filosófico sobre as Probabilidades” — é o quadro conceitual que funda a estatística Bayesiana e está exposto na introdução desta importante obra

(...) os mais importantes problemas da vida (...), em sua maioria, não são mais que problemas de probabilidades. Com rigor, também pode se dizer que quase todos os nossos conhecimentos são apenas prováveis, e do escasso número de coisas que podemos saber com certeza, (...) os principais meios de alcançar a verdade — a indução e a analogia —, se fundam nas probabilidades, de sorte que todo sistema de conhecimentos humanos se vincula com a teoria exposta neste ensaio.

A física, sendo um dos “mais importantes problemas da vida”, não é uma exceção. O trabalho do físico estatístico é essencialmente o mesmo que o trabalho do estatístico comum que, por exemplo, estuda uma pesquisa de opinião mercadológica. Na física estatística não há novas leis para serem descobertas — e nisto é radicalmente diferente da mecânica —, em física estatística, assim como outras formas de análise estatística, há padrões escondidos pelo fato de que um grande número de elementos operam de maneira essencialmente desconhecida, e queremos revelar estes padrões. Os resultados da disciplina, assim como os resultados de uma simples pesquisa de opinião, se fundam mais na “indução e analogia” que em leis imutáveis e irreduzíveis. Neste sentido, o crescimento da entropia, antes de expressar uma lei da natureza, é um juízo de bom senso. A única qualidade que distingue a física estatística de outros modelos menos sofisticados é, comparativamente, a enorme facilidade de incorporar informação uma vez que se trata de um corpo de conhecimento extremamente matematizado.

Ao considerar um sistema com um grande número de partículas, e sem saber exatamente as influências que animam seus movimentos, é necessário apelar para leis de probabilidades. Afora a dificuldade técnica de resolver as equações de movimento, o que pode ser aperfeiçoado pela construção sistemática de computadores mais velozes e métodos matemáticos mais eficientes, existe outro problema fundamental. O conhecimento acerca do estado de qualquer sistema físico pressupõe, em certa medida, o conhecimento sobre uma configuração anterior a partir da qual sua evolução possa ser calculada. Muitas vezes — e isso é *sempre* verdade quando se trata de sistemas macroscópicos —

o conhecimento completo sobre o estado inicial não está disponível. Esta ignorância pode ser atribuída tanto a uma pequena indeterminação nas medidas sobre as variáveis mecânicas, mas também à ausência absoluta de dados a respeito de alguns destes graus de liberdade e à impossibilidade prática de medi-los.

Um sistema macroscópico típico possui cerca de  $10^{23}$  graus de liberdade dos quais poucos são de fato acessíveis ao experimentalista. E para uma parcela grande de aplicações, apenas uma fração reduzida destas coordenadas é realmente relevante. Ademais, um simples catálogo dos  $10^{23}$  números subentendidos numa descrição mecânica está muito acima da capacidade técnica atual — a informação original impressa, digitalizada e transmitida na forma de texto e multimídia durante a existência da humanidade é avaliada em cerca de  $\sim 10^{19}$  bytes, um valor comparativamente modesto<sup>1</sup>.

Tamanho quantidade de informação não só é inacessível, mas também é excessiva. O esforço empregado para processá-la dificilmente se justificaria e várias conclusões de utilidade prática são obtidas mais facilmente por outros métodos. Nos restringiremos, numa primeira análise, a sistemas simples que podem ser integralmente especificados por umas poucas coordenadas (ex.: 1 *kg* de carvão para acender a churrasqueira). A questão fundamental é descobrir que tipo de inferências podem ser feitas a partir da física conhecida no nível microscópico na situação desfavorável em que não existe informação disponível a respeito de praticamente todas, a exceção de umas poucas, dentre as  $10^{23}$  coordenadas. Na opinião do autor esta é, fundamentalmente, a tarefa que se recai sobre a física estatística.

---

<sup>1</sup>Estudo feito pela Universidade de Berkeley em <http://www.sims.berkeley.edu/research/projects/how-much-info/>



# 1 *Fundamentos da estatística Bayesiana*

There are three kinds of lies: lies, damned lies, and statistics

(MARK TWAIN)

## 1.1 Conceito de probabilidade

Para que se proceda com a tarefa apresentada na seção anterior, é necessário definir um objeto matemático que capture formalmente a ignorância a respeito dos parâmetros que governam os processos físicos. Este objeto são probabilidades Bayesianas — elas codificam de maneira simples (números reais) o grau de crença racional com relação a qualquer proposição.

Ainda sem se prender aos detalhes técnicos, vale ressaltar que a história da teoria de probabilidades é bastante controversa; a própria definição sobre o que é uma probabilidade foi, e ainda é, motivo de debates entre os proponentes de diversas interpretações. A literatura clássica reconhece probabilidades como o limite de frequências estatísticas de um experimento aleatório. Autores com viés matemático tendem a ressaltar os aspectos formais da teoria da medida e os axiomas de Kolmogorov. A própria designação “Bayesiana” é utilizada indiscriminadamente para se referir a autores com concepções distintas sobre o significado das probabilidades. Mais notadamente, é duvidoso se o próprio Bayes seria Bayesiano em certas acepções do termo.

Adotamos aqui o ponto de vista dos físicos americanos Richard T. Cox e Edwin T. Jaynes exposto no apêndice A que formaliza de maneira unívoca o conceito de “ignorância” a respeito de uma afirmação. Cox deriva as regras de manipulação de probabilidade de maneira muito natural, independente de qualquer experimento de frequências ou estrutura axiomática *a posteriori*. Ainda que a concepção Bayesiana de

Cox apresente um paralelo formal com as outras escolas de probabilidade, já que as regras de manipulação são as mesmas, a visão aqui defendida dispensa qualquer uma destas apresentações.

A teoria de probabilidades Bayesiana, como é exposta no livro de E.T. Jaynes — “Probability Theory, the logic of science”[?], pode ser entendida como uma extensão para o cálculo de proposições relativa ao pensamento indutivo. Diferente da lógica clássica que abrange todas situações em que as proposições são falsas ou verdadeiras, mas com um valor de verdade conhecido, a teoria de probabilidades permite lidar com incerteza, levando o programa de formalização do raciocínio a um importante passo adiante. O tipo de raciocínio baseado em premissas incertas é chamado raciocínio indutivo para distinguir do raciocínio dedutivo formalizado anteriormente. Para ilustrar a diferença entre ambos, considere o silogismo

se  $A$  é verdadeiro, então  $B$  é verdadeiro

$A$  é verdadeiro (1.1)

.....

então,  $B$  é verdadeiro,

e ainda

se  $A$  é verdadeiro, então  $B$  é verdadeiro

$B$  é falso (1.2)

.....

então,  $A$  é falso.

Em ambos casos, o modelo que relaciona as duas afirmações ( $A \Rightarrow B$ ), permite *deduzir* o valor de verdade de  $A$  ou  $B$  a partir do conhecimento do valor de verdade da outra afirmação *se  $A$  e  $B$  possuírem um valor de verdade adequado*<sup>1</sup>. Um segundo tipo de raciocínio empregado extensivamente no senso comum e na prática científica não é, com efeito, captado por nenhuma das duas formas. É a indução, que corresponde aos

---

<sup>1</sup>Neste exemplo não é possível deduzir nada se  $A$  for verdadeiro ou  $B$  for falso.



“silogismos fracos”

se  $A$  é verdadeiro, então  $B$  é verdadeiro

$B$  é verdadeiro (1.3)

.....  
então,  $A$  se torna mais plausível,

e ainda

se  $A$  é verdadeiro, então  $B$  é verdadeiro

$A$  é falso (1.4)

.....  
então,  $B$  se torna menos plausível.

De fato, o primeiro tipo de raciocínio permite-nos fazer “inferências exatas” a respeito do valor de  $A$  ou  $B$  enquanto no segundo caso só é possível refinar nossas crenças ao levar em conta informação que pareça relevante. Não existe conflito entre as duas maneiras de pensar. Em particular, o silogismo lógico é recuperado pela teoria de probabilidades nos casos especiais em que as probabilidades tendem à certeza ou à impossibilidade.

Neste trabalho utilizamos a abordagem desenvolvida pelo físico americano R. T. Cox na qual define-se uma grandeza chamada plausibilidade que relaciona duas afirmações entre si. A plausibilidade que a assertiva  $A$  feita sobre uma hipótese  $H$  seja verdadeira é representada por

$$(A|H) \equiv \text{O quão } A \text{ é plausível dado } H, \quad (1.5)$$

onde as afirmações  $A$  e  $H$  são sentenças lógicas arbitrárias. A hipótese  $H$  descreve o que se sabe sobre o objeto sobre o qual se afirma  $A$  e corresponde à conjunção de todas hipóteses feitas no modelo empregado para descrevê-lo. A plausibilidade  $(A|H)$ , portanto, representa o grau de crença racional sobre uma assertiva  $A$  feita sobre certas hipóteses (que nem sempre precisam ser explicitadas). De fato, a medida de plausibilidade não é simétrica nas variáveis  $A$  e  $H$ ; sempre assume-se  $H$  verdadeiro já que não faz sentido especular sobre o valor de verdade de  $A$  quando se supõe que o modelo utilizado é falso.

Antes de se comprometer com valores específicos para a medida  $(A|H)$ , é possível assumir certas propriedades gerais que condicionam o cálculo de probabilidades segundo

a exigência de consistência lógica. Assumimos que a plausibilidade é representada por números reais para que exista uma cadeia de transitividade bem definida entre as diversas possibilidades. Utilizamos a notação para o cálculo de proposições:  $AB = "A \text{ e } B"$ ,  $A + B = "A \text{ e/ou } B"$  e  $\bar{A} = "n\tilde{a}o-A"$  sendo que a questão que se apresenta é, mais especificamente, como calcular as plausibilidades do tipo  $(AB|H)$ ,  $(A + B|H)$  e assim por diante, a partir das plausibilidades primitivas  $(A|H)$ ,  $(B|H)$ ,  $(A|BH)$  etc.

Partindo de uma certa escala de plausibilidades  $(A|H)$ , é possível mapeá-la em novas escalas  $(A|H)' = F(A|H)$ . Esta mudança, é claro, não altera o conteúdo de  $(A|H)$  desde que seja possível inverter a relação para calcular  $(A|H)$  em outras representações. Uma vez de acordo que a escolha da escala de plausibilidades  $F(A|H)$  é imaterial, importa apenas escolher a representação que proporcione a maior facilidade operacional. Neste espírito, é possível provar que existe uma certa escala de plausibilidades que obedece às regras de cálculo

$$(AB|H) = (B|H)(A|BH) \quad (1.6)$$

$$= (A|H)(B|AH) \quad (1.7)$$

$$(A + B|H) = (A|H) + (B|H) - (AB|H) \quad (1.8)$$

$$(\bar{A}|H) = 1 - (A|H). \quad (1.9)$$

A esta escala específica nos referimos como *probabilidade*. De agora em diante se utilizará a notação  $P(A|H)$  para ressaltar a escolha, entendendo que a liberdade que inicialmente existia na determinação da escala de plausibilidade foi perdida em prol de se utilizar um conjunto fixo de regras simples. A demonstração deste resultado foi realizada pela primeira vez por Cox e é ricamente reproduzida na literatura Bayesiana [?, ?] e no apêndice A. Estas regras se reduzem trivialmente às regras de cálculo booleano nos casos em que as probabilidades do lado direito correspondem à certeza ou impossibilidade representadas pelos números 1 e 0, respectivamente.

### 1.1.1

### 1.1.2 Da atribuição de probabilidades

Sabendo as regras de composição de probabilidades simples em outras complexas, resta agora descobrir como atribuir valores numéricos para as mesmas. As regras do cálculo de probabilidades, naturalmente, não fornecem uma prescrição universal para

fazê-lo, e portanto, poderia se pensar que os números atribuídos a  $P(A|H)$ ,  $P(B|H)$  etc. são totalmente arbitrários. Ainda que, rigorosamente, é isso que ocorre, pois diferentes indivíduos podem possuir um grau de conhecimento e familiaridade distintos sobre cada situação observada na natureza, existem princípios mais ou menos gerais em que se guiar. São princípios que tornam o formalismo adequado para o tratamento de problemas científicos.

O primeiro destes princípios gerais para a atribuição de probabilidades é fornecido pelo teorema de Bernoulli ou “lei dos grandes números”. Se as afirmações  $A_i$  expressam a ocorrência de um evento que pode ser repetido sistematicamente num experimento de freqüências estatísticas, o número de ocorrências  $\nu(A_i)$  de cada evento  $A_i$  determina assintoticamente a probabilidade de partida  $P(A_i|H)$ . Esta determinação corresponde ao procedimento associado à teoria clássica de probabilidades — fosse o único procedimento legítimo para se atribuir probabilidades, as probabilidades Bayesianas corresponderiam em escopo e significado à teoria clássica.

Existem, felizmente, outras maneiras de se atribuir probabilidades objetivas a um certo conjunto de possibilidades. Um princípio proposto por Laplace — comumente denominado *princípio da razão insuficiente* — diz respeito a uma seqüência de proposições mutuamente exclusivas  $A_1, A_2, \dots, A_n$  em que não há fundamento racional para se justificar a preferência de uma proposição  $A_i$  em comparação a outra  $A_j$ . Neste caso, a probabilidade atribuída a cada uma das proposições, por um argumento de honestidade intelectual, deve ser a mesma

$$P(A_i|H) = \frac{1}{n}; \quad i = 1, 2, \dots, n. \quad (1.10)$$

Casos simples em que o princípio da razão insuficiente se aplica são a atribuição de probabilidades para jogadas de um dado sem vício, mãos em um jogo de baralho, números em bilhetes de loteria etc. Algumas situações apresentam assimetrias onde se colocaria em dúvida a possibilidade de utilizar a atribuição equiprovável. De um modo geral, o princípio da razão insuficiente ilustra um princípio mais geral que é o dos chamados grupos de invariância. A metodologia consiste em identificar alguma operação sobre as funções  $f(A_i) \equiv P(A_i|H)$  que, por princípio, deve tornar invariante o valor das atribuições de probabilidade  $f(x)$ . A relevância deste princípio para a física é óbvia dado o papel fundamental que os grupos de simetria dinâmica possuem na descrição das leis mecânicas. Assim argumenta-se, por exemplo, que a medida de ignorância sobre o estado de um certo sistema deve ser invariante por mudança

de sistema de coordenadas, caso contrário observadores poderiam conhecer mais ou menos a respeito do sistema simplesmente mudando o estado de movimento relativo a uma situação inicial.

Um terceiro princípio, chamado de princípio da máxima entropia (MaxEnt) é uma forma generalizada de *princípio da razão insuficiente* que leva em conta certos tipos de informação disponível na forma grandezas testáveis  $F = F[P(A|H)]$ . Mesmo que esta informação não seja suficiente para determinar uma única possibilidade  $A_k$ , deve ser levada em conta na atribuição de probabilidade  $P(A_i|H)$ . É interessante, antes de tudo, especificar claramente o significado do conceito de entropia na teoria probabilística, assim como discutir outras grandezas utilizadas para descrever atribuições de probabilidade.

## 1.2 Distribuições de Probabilidade

O caso de maior interesse corresponde à situação em que existem várias afirmações  $A_i; i = [1, m]$  em que uma e apenas uma é verdadeira. Um exemplo simples é o sistema de dois níveis que pode se encontrar no estado de spin *up* ( $U$ ) ou *down* ( $D$ ). Neste caso, é possível afirmar:

$$U \equiv \text{''Sistema se encontra no estado up''} \quad (1.11)$$

$$D \equiv \text{''Sistema se encontra no estado down''}, \quad (1.12)$$

onde, naturalmente, ambas afirmações não podem ser verdadeiras simultaneamente

$$UD = \text{falso} \Rightarrow P(UD|H) = 0 \quad (1.13)$$

e ou  $U$  ou  $D$  deve ser verdadeiro em qualquer circunstância

$$U + D = \text{verdadeiro} \Rightarrow P(U + D|H) = 1.$$

De maneira geral, considere dois sistemas de afirmações  $\{A_i\}$  e  $\{B_j\}$  exaustivos e mutuamente exclusivos (EME)<sup>2</sup>. Estes sistemas possuem interesse especial pois é possível aplicar uma metodologia de cálculo muito conveniente. Note que qualquer sistema de afirmações pode ser representado por um sistema EME equivalente. Um conjunto não exaustivo formado por  $N$  afirmações  $\{C_i\}$  se torna exaustivo trivialmente pela adição da afirmação  $C_{N+1} = \overline{C_1 + C_2 + \dots + C_N}$ . De maneira semelhante, um conjunto não-

---

<sup>2</sup>Conjunto de afirmações em que sempre uma e apenas uma afirmação é verdadeira numa certa situação.

mutualmente exclusivo formado por duas afirmações  $X$  e  $Y$  é equivalente ao conjunto EME dado por  $C_{ME} = \{C_1 \equiv XY, C_2 \equiv X\bar{Y}, C_3 \equiv \bar{X}Y\}$ . É possível resgatar o problema original utilizando as relações  $X = C_1 + C_2, Y = C_1 + C_3$ , e utilizar o conjunto EME  $\{C_1, C_2, C_3\}$  em cálculos práticos. A generalização deste procedimento para um conjunto de afirmações arbitrário é óbvia.

A mais importante das propriedades relacionadas a aos sistemas EME é que eles obedecem à condição de normalização<sup>3</sup>

$$\underbrace{P\left(\sum_i A_i \middle| H\right)}_{\text{Verdade}} = \sum_i P(A_i|H) = 1, \quad (1.14)$$

que, obviamente, também se aplica às probabilidades condicionais

$$\sum_i P(A_i|B_jH) = 1. \quad (1.15)$$

Destes resultados, deriva-se a regra de normalização para sistemas compostos, ou seja,

$$\sum_{ij} P(A_i B_j|H) = \underbrace{\sum_i P(A_i|H)}_1 \underbrace{\sum_j P(B_j|A_iH)}_1 = 1, \quad (1.16)$$

e a importante regra de marginalização

$$\sum_j P(A_i B_j|H) = P(A_i|H) \sum_j P(B_j|A_iH) = P(A_i|H). \quad (1.17)$$

Estes são os resultados básicos da teoria. De agora em diante nos referimos a *distribuição de probabilidade* toda atribuição feita sobre um conjunto de afirmações EME. As regras simples a que eles obedecem não só facilitam os cálculos, mas também permitem generalizar o conceito de probabilidade para distribuições contínuas como será exposto na próxima seção. Em vertentes axiomáticas da teoria das probabilidades, estes resultados normalmente são entendidos como a definição do próprio conceito de

---

<sup>3</sup>Estamos usando explicitamente o fato que o sistema é mutuamente exclusivo. Em geral temos  $P(A + B|H) = P(A|H) + P(B|H) - P(AB|H)$ . Para um sistema mutuamente exclusivo,  $P(AB|H) = 0$  já que  $AB = \text{Falso}$ , portanto  $P(A + B|H) = P(A|H) + P(B|H)$ . Este argumento pode ser generalizado para o caso em que existam mais de 2 afirmações, considerando-as aos pares. Mostra-se, então, que

$$P\left(\sum_i A_i \middle| H\right) = \sum_i P(A_i|H).$$

probabilidade. Daí podem significar qualquer coisa: frequências estatísticas, variáveis aleatórias, propensões ou mesmo probabilidades Bayesianas. Acreditamos, no entanto, que o ponto de vista Bayesiano oferece uma visão mais profunda e mais unificada sobre o significado dos elementos matemáticos manipulados .

### 1.2.1 Densidade de probabilidade contínuas

Há certos problemas em que devemos especificar as probabilidades que um parâmetro  $\theta$  definido em um intervalo *contínuo*  $X \equiv [x_0, x_f]$  possua um determinado valor  $x$ . É lógico que, para dar um tratamento probabilístico adequado, é necessário traduzir este problema em afirmativas lógicas. À semelhança com o caso anterior, tenderíamos a definir

$$x \equiv \text{"o parâmetro } \theta \text{ possui o valor } x\text{"}. \quad (1.18)$$

Parâmetros definidos num suporte contínuo, implicam numa série de dificuldades relativas ao fato que se trata de um conjunto infinito de elementos não-enumeráveis. Neste caso, as somas de variáveis em (1.17) e (1.15) claramente não fazem qualquer sentido<sup>4</sup>; em especial, não há como definir a operação de negação para um conjunto não-enumerável. Um olhar mais atento mostra, então, que a afirmação (1.18) é, no mínimo, problemática. Uma maneira de proceder consistentemente é definir as afirmações

$$S(x) \equiv \text{"}\theta \text{ é menor ou igual a } x\text{"} \quad (1.19)$$

$$G(x) \equiv \text{"}\theta \text{ é maior que } x\text{"}. \quad (1.20)$$

Para um valor fixo  $x$ ,  $S$  e  $G$  representam um conjunto de afirmações EME perfeitamente bem definido, de forma que é possível se questionar sobre as probabilidades associadas à  $S$  ou  $G = \bar{S}$ . Para tanto, definimos a função  $F(x)$  dada por

$$F(x) = P(S(x)|H), \quad (1.21)$$

onde  $F(x)$  é comumente referida como a distribuição de probabilidades cumulativa (DPC). É óbvio que  $F(x)$  é uma função monotônica crescente de  $x$  com os valores extremos fixados em  $F(x_0) = 0$  e  $F(x_f) = 1$ . A partir do conhecimento da função  $F(x)$  é possível calcular a probabilidade de  $x$  estar em qualquer intervalo  $a, b$  dentro

---

<sup>4</sup>Há que se lembrar que a regra de normalização de probabilidades é consequência do cálculo de disjunção, não é, de maneira alguma, uma definição independente. Desta maneira, substituir as somas por integrais ainda é, neste ponto, uma extrapolação totalmente injustificada.

do domínio  $[x_0, x_f]$  usando as regras familiares. Assim definimos as afirmações

$$A \equiv x \leq a, \quad B \equiv x \leq b, \quad C \equiv a < x \leq b, \quad (1.22)$$

que obviamente se relacionam por  $B = A + C$  e  $A \Rightarrow \bar{C}$ , sendo que  $P(AC|H) = 0$ . Desta forma calculamos,

$$\begin{aligned} P(B|H) &= P(A + C|H) = P(A|H) + P(C|H) \\ &\Downarrow \\ P(a < x \leq b|H) &= P(B|H) - P(A|H) = F(b) - F(a). \end{aligned} \quad (1.23)$$

Nos casos especiais em que  $F$  é uma diferencial exata, ou seja,  $F(x) = \int_{x_0}^x dy f(y|H)$ , o resultado anterior pode ser escrito de maneira mais simples como

$$P(a < x \leq b|H) = \int_a^b dx f(x|H), \quad (1.24)$$

onde  $f(x)$  é conhecida como função de *distribuição de densidade de probabilidade* (DDP) para  $x$ . De maneira geral, podemos associar probabilidades a intervalos arbitrariamente pequenos em torno de um ponto específico  $x$ , como

$$P(x|H) \equiv \text{"probabilidade do parâmetro } \theta \text{ estar entre } x \text{ e } x + dx\text{"}. \quad (1.25)$$

Desta forma, a distribuição de probabilidades pode ser escrita formalmente como

$$P(x|H) = f(x|H)dx, \quad (1.26)$$

entendendo que a probabilidade se refere ao intervalo limite em que  $dx \rightarrow 0$ . É necessário prestar atenção a alguns detalhes técnicos. A começar, não existe operação de negação bem definida para variáveis contínuas, de modo que  $1 - f(x|H)dx$  rigorosamente não diz nada. Mais ainda, o objeto matemático associado ao formalismo contínuo, ou seja, as distribuição de probabilidades cumulativas, carrega informação importante sobre o sistema de coordenadas, algo que pode parecer contra-intuitivo se deseja-se interpretar as DDPs de maneira muito literal. Desta forma, as densidades de probabilidades  $f(x|H) = \frac{\partial F(x|H)}{\partial x}$  se modificam por  $f(x'|H) \mapsto f(x|H)J(x, x')$  onde  $J(x, x')$  é o Jacobiano da transformação de coordenadas de  $x$  para  $x'$ . Do ponto de vista formal, as regras para o tratamento das densidades de probabilidades se relacionam com as regras utilizadas no caso discreto passando os limites de soma de Riemann que, efetivamente, correspondem a substituir as somas em  $P(x|H)$  por integrais de  $f(x|H)$ . Note que uma passagem descuidada destes limites ou de mudanças de coordenadas podem levar a todo tipo de

inconsistências.

## 1.2.2 Entropia estatística

“(...) it will give you a great edge in debates because nobody really knows what entropy is anyway”

(História apócrifa em que J. von Neumann sugere um nome para o funcional  $-\sum_i p_i \ln p_i$  a C. Shannon)

Dada uma certa distribuição de probabilidades associada a um conjunto EME de afirmações  $\{A_i\}$  é importante saber se a teoria de probabilidades pode fornecer inferências conclusivas e ainda, se possível, quantificar o quão boas elas seriam. Em uma certa situação, por exemplo, a atribuição de probabilidades poderia ser descrita como: *“Existem várias possibilidades, mas apenas algumas poucas são de fato prováveis”*, ou ainda *“Cada possibilidade é aproximadamente tão provável quanto as outras, de forma que muito pouco pode se dizer sobre o resultado esperado”*. Estas observações particulares podem ser quantificadas pelo conceito de entropia estatística. Aqui se considera que os esquemas do primeiro tipo correspondem a um baixo valor de entropia enquanto os do segundo tipo são esquemas de alta entropia.

Um exemplo extremo de esquema de baixa entropia, é caso em que uma probabilidade se reduz à certeza,  $P(A_i|H) = 1$ , de forma que não há incerteza e convencionam-se que a entropia é mínima. Esquemas de baixa entropia representam as situações em que a teoria de probabilidades pode fornecer indicações muito significativas sobre o que se esperar, tanto pela eliminação de alternativas impossíveis e extremamente improváveis, quanto pela eleição de um ou poucos estados muito mais prováveis que os outros. Este é, por exemplo, o motivo que um homem de negócios sensato jamais investiria seu dinheiro na loteria apesar da possibilidade de um ganho considerável com baixíssimo investimento: é praticamente certo que todo dinheiro ali empregado não trará retorno algum. É lógico que não há certezas que todo dinheiro teria o mesmo destino — ele *pode* comprar o bilhete premiado —, mas certamente esta corresponde à situação plausível.

Em algumas casos, medidas padronizadas feitas sobre uma atribuição de probabilidades podem fornecer informações relevantes sobre a quantidade de incerteza que se encerra em um determinado esquema. Voltando ao exemplo do homem de negócios, é correto imaginar que, mesmo que o lucro não seja garantido, um bom empresário aceitaria negociar se a expectativa de lucro for positiva e o risco moderado. Estes dois parâmetros



são capturados pelos valores médios e desvio padrão da distribuição de probabilidades para os lucros obtidos em um certo empreendimento.

De maneira mais geral, é interessante definir uma grandeza que descreva o quão informativas podem ser as previsões resultantes de um certo esquema. Isto é feito pelos funcionais *informação* ou *entropia*. Intuitivamente, esperamos que esta medida seja mínima para o caso em que há certeza:  $p_i = 1$ ; e máxima para o caso “totalmente aleatório”:  $p_1 = \dots = p_m = \frac{1}{m}$ . A forma explícita do funcional de entropia pode ser obtida a partir de um conjunto de exigências elaborado por Claude Shannon.

1. Existe uma medida numérica  $H(p_1, p_2, \dots, p_m)$  representada por números reais que associa uma quantidade de “incerteza” a cada esquema de probabilidades.
2.  $H(p_1, \dots, p_m)$  é uma função contínua dos argumentos  $p_i$ . De outra forma, uma mudança arbitrariamente pequena na distribuição de probabilidades resultaria numa mudança grande na quantidade de “incerteza”.
3. O acréscimo de afirmações com probabilidade nula não altera o valor de  $H$ , ou seja:  $H(p_1, \dots, p_m) = H(p_1, \dots, p_m, 0)$ . A justificativa para isto é que a “incerteza” de um esquema de probabilidades não deve aumentar simplesmente considerando novas hipóteses impossíveis.
4. Se existir mais de um modo de obter  $H(p)$ , ambos devem coincidir. Em especial, seja um esquema composto por 3 afirmações associadas às probabilidades  $p_1, p_2, p_3$ ; exigimos que, ao compor a afirmação  $\bar{A} = A_2 + A_3$  resulte, por consistência, que

$$H(p_1, p_2, p_3) = H(p_1, \bar{p}) + \bar{p}H\left(\frac{p_2}{\bar{p}}, \frac{p_3}{\bar{p}}\right), \quad (1.27)$$

onde  $\bar{p} \equiv p_2 + p_3$  é a probabilidade  $P(\bar{A}|H) \equiv P(A_2|H) + P(A_3|H)$  do evento “1 e/ou 2”.

5. A medida  $H(p)$  deve respeitar a condição de máximo para o caso equiprovável  $H\left(\frac{1}{m}, \dots, \frac{1}{m}\right) \equiv h(m)$  e  $h(m)$  deve ser uma função crescente de  $m$  de forma que um esquema com muitas afirmações igualmente prováveis seja mais incerto que outro com poucas informações igualmente prováveis.

Aceitando-se estas 5 exigências, o funcional  $H(p)$  fica unicamente determinado pela forma

$$H(p) = - \sum_i p_i \log p_i, \quad (1.28)$$

onde a base do logaritmo não é especificada, indicando a liberdade de escolha da escala multiplicativa que convier a cada problema. Note que até este momento não existe *nenhuma* razão para identificar o funcional  $H(p)$  com a entropia termodinâmica — o nome *entropia* é uma convenção particularmente infeliz resultante do fato que ambos respeitam a mesma relação funcional com as distribuições de probabilidades. De fato, o conceito de entropia informacional, longe de ser um conceito relacionado especificamente à física, tampouco é adequadamente capturado pelos adjetivos simples normalmente empregados para descrevê-lo: “(des)informação”, “incerteza”, “desordem”, “aleatoriedade”, etc. De fato, a interpretação mais correta para o funcional  $H(p)$  provavelmente vai de encontro com as indagações informais apresentadas no início da seção. Mesmo neste contexto restrito, ainda poderia se argumentar que a solução obtida não é completamente satisfatória. Uma discussão mais detalhada a este respeito será feita posteriormente.

O funcional de entropia pode ser generalizado para casos em que as afirmações  $A_i$  não formam um conjunto de afirmações mutualmente exclusivas, i.e.:  $A_i$  e  $A_j$  podem ser simultaneamente verdadeiras. Neste caso, a entropia do sistema de probabilidades se escreve como

$$\begin{aligned}
 H(A_1, A_2, \dots, A_m|H) &= - \sum_i P(A_i|H) \log P(A_i|H) \\
 &+ \sum_{i,j>i} P(A_i A_j|H) \log P(A_i A_j|H) - \dots \\
 &\pm \sum_{i,j>i,\dots} P(A_i A_j \dots A_m|H) \log P(A_i A_j \dots A_m|H)
 \end{aligned} \tag{1.29}$$

### 1.2.3 Entropia e inferência

Uma conotação particularmente interessante a respeito do funcional  $H(p)$  sugere que a entropia mede o grau de enviesamento de uma distribuição de probabilidades em respeito a qualquer elemento de um conjunto de afirmações EME. Uma distribuição em que uma das afirmativas  $A_k$  é bastante mais provável que as outras (baixa entropia) é vista como uma distribuição muito enviesada em relação à  $A_k$  enquanto que o contrário — a ausência de qualquer forma de viés — corresponde às entropias máximas.

Neste sentido, uma boa inferência estatística consiste em escolher a distribuição com mínimo viés e, portanto, máxima entropia. Este modo de pensar generaliza o raciocínio por trás do *princípio da razão insuficiente* de Laplace ao fornecer uma prescrição para atribuir probabilidades quando *há* uma razão para preferir certas possibilidades.

A maximização da entropia sem vínculos define, por construção, uma atribuição equiprovável. Ao quantificar o viés estatístico, podemos generalizar o argumento de Laplace pela prescrição que produz a distribuição menos enviesada que leve em conta um determinado tipo de informação.

Em muitas situações a informação disponível consiste apenas *estimativas* sobre os valores exatos de uma grandeza por estar sujeita a erros ou flutuações experimentais. Por exemplo, uma vez que se mediu a posição de uma determinada partícula, não há garantias que o valor apontado em laboratório seja, de fato, o valor correto devido à imprecisão intrínseca dos instrumentos de medição. Deste modo, qualquer informação adquirida por meio desta medição é encarada apenas como uma *inferência* sobre o valor exato do parâmetro que se desejava obter. É típico representar esta situação relacionando  $F$  com o valor médio de  $f_i$  — a justificativa é que a média de uma grandeza é o funcional que minimiza o desvio padrão,  $\langle (f(A) - F)^2 \rangle$ , estando associado a um tipo de inferência que privilegia o valor mais próximo dos pontos estatisticamente relevantes da distribuição.

Outro procedimento típico de inferência é associar a informação experimental com o valor mais provável da distribuição de probabilidades para  $f(A)$ , a ver,  $F = \max f(A)$ . A menos que a distribuição de probabilidades represente uma situação de certeza, nenhum argumento completamente conclusivo indicará a melhor estratégia de inferência. A estratégia da valor máximo corresponde à expectativa de se acertar o valor correto com *maior frequência*, enquanto a escolha pelo valor médio representa a tentativa de *minimizar a diferença* entre o valor exato e a estimativa. Em situações não muito distintas, as duas estratégias podem ser particularmente desastrosas<sup>5</sup>.

Ao optar pela inferência do tipo valor médio,  $F \equiv \langle f(A_i) \rangle$ , é possível utilizar o método da máxima entropia para incorporar esta informação na atribuição de probabilidades. A distribuição menos enviesada em respeito ao conjunto de possibilidades  $\{A_i\}$  é dada pelas probabilidades que maximizam o funcional  $-\sum_A p(A) \ln p(A)$  sujeitas ao vínculo  $\sum_A p(A)f(A) = F$ . Explicitamente se escrevem como

$$p(A) = \frac{e^{-\beta f(A)}}{Z(\beta)}; \quad Z(\beta) \equiv \sum_A e^{-\beta f(A)}. \quad (1.30)$$

---

<sup>5</sup>Um caso típico em que a inferência por valor máximo falha é se houverem mais de um máximo associado a uma determinada grandeza. Uma situação igualmente dramática ocorre na inferência por valor médio quando a melhor expectativa é um valor impossível (algo como a família brasileira típica possuir 2,5 filhos). A não ser que a distribuição esteja relativamente bem concentrada sobre um único pico de probabilidades, ambas estratégias de inferência levam invariavelmente a resultados esdrúxulos.

Uma extensão natural deste método diz respeito à maximização da entropia relativa, ou distância de Kullback-Leibler, dada por

$$H(p|\omega) \equiv - \sum_A p(A) \ln \frac{p(A)}{\omega(A)}. \quad (1.31)$$

O funcional  $H(p|\omega)$  pode ser interpretado como o viés da distribuição  $p(A)$  em relação a uma atribuição de referência denotada por  $\omega(A)$ . A maximização de  $H(p|\omega)$  com vínculos corresponde ao procedimento que encontra a distribuição menos enviesada em relação a um conhecimento prévio representado pela distribuição de probabilidades  $\omega(A)$ .

O funcional que descreve a entropia relativa pode ser obtido formalmente ao modificar a condição 5 dos axiomas de Shannon pela exigência que o máximo da entropia recupere uma certa distribuição  $\omega_i$ . Isto introduz um viés explícito em relação às possibilidades encerradas no conjunto  $\{A_i\}$  que pode ser utilizado para expressar diferentes tipos de informação. No sentido estrito que a maximização da entropia relativa incorpora informação nova à uma situação anterior representada por  $\omega(A)$ , nos referimos a este procedimento como uma forma de *atualização de probabilidades*.

## 1.3 O processo de inferência

### 1.3.1 Teorema de Bayes

Mais que pareça plausível que a entropia estatística mede o grau de enviesamento de uma distribuição, não é claro que a prescrição de maximização produz resultados universalmente consistentes com as regras do cálculo de probabilidades. Mais ainda, também é pouco claro se a entropia de Shannon é o único funcional a que pode ser dada esta interpretação. A verdade é que a axiomatização feita por Shannon não visa construir um esquema de inferência estatística, mas sim definir uma quantidade informacional para outros usos mais específicos em Teoria da Comunicação. De fato, propostas de entropias alternativas aparecem com freqüência na literatura. Vale citar as entropias de Renyi e Tsallis que podem ser derivadas a partir de axiomas semelhantes aos de Shannon, com um apelo estatístico apenas ligeiramente diferente.

Um problema ainda mais dramático é observado nas entropias definidas sobre probabilidades de variáveis contínuas em que o funcional  $-\int dx \rho \ln \rho$  está associado a

um processo de limite divergente. Neste caso existem *infinitos*<sup>6</sup> funcionais que obedecem aos axiomas de Shannon sendo que o já mencionado *não* é um deles. Parece imprescindível reescrever a prescrição de atualização de probabilidades como um processo de inferência explícito. Nisto abdicamos temporariamente do conceito de entropia para elucidar melhor o que se entende por *inferência estatística*.

O processo de inferência consiste em incorporar nova informação sobre certas variáveis, digamos  $x$ , à distribuição de probabilidades  $P(x|H)$  em razão de alguma constatação a respeito de outras variáveis — muitas vezes, dados experimentais — que denotamos por  $\theta$ . Este processo é caracterizado pelo teorema de Bayes que calcula a “probabilidade posterior”  $P(x|\theta H)$  como sendo

$$P(x|\theta H) = P(x|H) \frac{P(\theta|xH)}{P(\theta|H)} = \frac{P(x|H)P(\theta|xH)}{\sum_x P(x|H)P(\theta|xH)}. \quad (1.33)$$

A atribuição  $P(x|H)$  é conhecida como “probabilidade anterior” (*prior*) e  $P(\theta|xH)$  como “verossimilhança” (*likelihood*). Note que o teorema de Bayes é uma consequência trivial da regra de cálculo para probabilidades conjuntas

$$P(x\theta|H) = P(x|H)P(\theta|xH) = P(\theta|H)P(x|\theta H), \quad (1.34)$$

de forma que reproduz o processo de inferência de uma maneira absolutamente fundamental.

Situações típicas em que se emprega o teorema de Bayes são situações em que há uma expressão disponível para a verossimilhança. Note que a probabilidade anterior é considerada um parâmetro do algoritmo, de tal maneira que, espera-se que em cada caso ela possa ser fornecida de antemão. Mais ainda, a distinção entre probabilidade anterior e posterior é simplesmente convencional e uma aplicação adicional do teorema de Bayes pode tratar a probabilidade posterior como probabilidade anterior de um problema de inferência que incorpore outro tipo de informação.

Para ilustrar os usos comuns do teorema de Bayes, considere um caso típico: a inferência do valor de uma quantidade física a partir da medição em laboratório sujeita a erros. Imagine que se meça repetidamente a quantidade unidimensional  $x$  com os resultados  $y_1, y_2, \dots$  ligeiramente diferentes entre si. Neste caso, assumimos que os

---

<sup>6</sup>O primeiro exemplo a aparecer na literatura é a informação de Fisher definida para qualquer espaço de parâmetros associados à uma distribuição  $\rho_A(x) \equiv P(x|\{A_i\}H)dx$  dada por

$$I_{ij} = \int dx \rho_A(x) \left( \frac{\partial}{\partial A_i} \ln \rho_A(x) \frac{\partial}{\partial A_j} \ln \rho_A(x) \right) \quad (1.32)$$

onde o traço da matriz  $I$  define uma informação no sentido de Shannon.

erros experimentais se distribuem de forma gaussiana, mais explicitamente, segundo a distribuição

$$P(y_1|x\sigma H) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-y_1)^2}{2\sigma^2}} dy_1. \quad (1.35)$$

Considerando medidas sucessivas independentes de forma que  $P(y_i|y_jxH) = P(y_i|xH)$ , ficamos com

$$P(y_1y_2 \dots y_N|xH) = P(y_1|xH)P(y_2 \dots y_N|y_1xH) \quad (1.36)$$

$$= P(y_1|xH)P(y_2|xH) \dots P(y_N|xH) \quad (1.37)$$

$$= \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - x)^2 \right] d^N y. \quad (1.38)$$

Neste ponto importa a experiência prévia de cada um, onde diferentes indivíduos podem chegar a conclusões diferentes de acordo com a familiaridade e pressupostos com relação ao problema apresentado. Isto é ilustrado pela indeterminação sobre o que seria a probabilidade anterior, ou seja, a probabilidade que codifica a informação pregressa ao experimento. Uma escolha usual é uma probabilidade uniforme que expressa a ausência de qualquer conhecimento prévio que favoreça um valor em relação aos outros. Considerando o intervalo de equiprobabilidade  $[-\frac{1}{2}\Delta x, \frac{1}{2}\Delta x]$  suficientemente grande, o teorema de Bayes determina que

$$P(x|y_1y_2y_3 \dots H) = \frac{1}{\Delta x} \theta \left( |x| - \frac{\Delta x}{2} \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-y_1)^2+(x-y_2)^2+\dots}{2\sigma^2}} dx,$$

de forma que o valor mais provável para  $x$  é aquele que minimiza  $\sum_i (x - y_i)^2$ . Este raciocínio simples deduz o método dos mínimos quadrados como um resultado trivial do processo de inferência. Podemos refiná-lo ao considerar outros tipos de informação pregressa, como por exemplo, uma probabilidade anterior gaussiana. Neste caso, se modifica a regra dos mínimos quadrados, substituindo-a por outra possivelmente mais complicada. Se ganha no que ao considerar o conhecimento anterior, podemos reduzir em várias vezes o esforço experimental necessário para se obter  $x$  dentro de uma boa margem de tolerância (variância de  $x$ ).

### 1.3.2 Processo de inferência e determinação da verossimilhança

Existem várias situações em que o teorema de Bayes não é facilmente aplicável devido à ausência de uma fórmula conhecida para a verossimilhança. Podemos imaginar, no entanto, que em algumas situações ela possa ser derivada sistematicamente a partir

de algum tipo de informação pré-definida. Consideramos o caso de informação testável, ou seja, informação do tipo  $A[P(x_i|aH)] = \text{cte}$  onde  $A[p]$  é um funcional da distribuição posterior que testa — seleciona — se a mesma obedece à alguma propriedade determinada. Nesta seção nos baseamos na metodologia desenvolvida em [?].

Neste ponto é conveniente abandonar a notação  $P(x_i|AH)$ , onde  $x_i$  e  $A$  são interpretados as afirmações “o estado do sistema é  $x_i$ ” e “sabe-se que um certo parâmetro  $\langle a_i \rangle$  vale  $A$ ”, e definir as funções  $p_i \equiv P(x_i|aH)$ ,  $\omega_i \equiv P(x_i|H)$  e  $L_i(A) \equiv P(A|x_iH)$ . Na construção do processo de inferência, nos restringimos especificamente ao caso em que  $A[p]$  é um valor médio como em

$$A[p] = \sum_i p_i a_i, \quad (1.39)$$

onde os  $a_i$ 's são arbitrários.

Existem, é claro, infinitas distribuições que fazem com que  $A[p] = A$  seja obedecido. De fato, algumas delas parecem intuitivamente mais plausíveis que outras na medida que descrevem o estado de conhecimento de maneiras mais simples ou mais complicadas. Em vista disso, postulamos a existência de um funcional  $\eta[p|\omega]$  que ordena as distribuições de probabilidade de forma que os maiores  $\eta[p|\omega]$  sejam as melhores inferências e menores  $\eta[p|\omega]$ , as piores. Este funcional se chama entropia inferencial, para distinguir da entropia informacional de Shannon.

A entropia inferencial é concebida para ser maximizada. Nesta expectativa, exige-se algumas propriedades para que o processo de inferência corresponda a uma aplicação específica do teorema de Bayes. A condição de máximo pode ser determinada pelo método dos multiplicadores de Lagrange que se escreve como

$$\frac{\partial \eta(p, \omega)}{\partial p_i} \equiv g_i(p_i, \omega) = \beta a_i + \lambda \Rightarrow p_i = g_i^{-1}(\beta a_i + \lambda, \omega_i), \quad (1.40)$$

onde os parâmetros  $\beta$  e  $\lambda$  foram introduzidos para fixar o vínculos em  $\langle a_i \rangle$  e a normalização. Neste ponto vale reescrever o teorema de Bayes na notação padrão para explicitar o tipo de dependência de cada variável com os parâmetros do algoritmo;  $A$ ,  $\{\omega_i\}$ ,  $\{a_i\}$  etc. Desta forma, temos

$$P(x_i|AH) = P(x_i|H) \frac{P(A|x_iH)}{P(A|H)} \Leftrightarrow p_i = \text{cte} \times \omega_i L_i(A). \quad (1.41)$$

Desejando precisar melhor a prescrição que realiza a inferência, é necessário fazer suposições sobre propriedades das probabilidades resultantes do algoritmo. Do fato que  $P(A|x_iH)$  carrega índices em apenas um  $x_i$  específico, é razoável supor que a

dependência nas variáveis  $\{a_i\}$  ocorra apenas neste índice. Desta forma, a verossimilhança  $P(A|x_i H)$  se escreve simplesmente como  $L(a_i, A)$ . Isso pode ser expresso de maneira mais clara pela *condição de localidade*.

**Condição de localidade:** *Informação local possui efeitos locais.* Sempre que  $\langle a_i \rangle$  fornecer informação a respeito de um subdomínio  $D = \{i, i', \dots\}$ , mas não de outros subdomínios disjuntos  $\bar{D}$ , a atualização das probabilidades procede de forma que apenas as probabilidades  $p_{i \in D}$  são alteradas. As probabilidades de estados em  $\bar{D}$  podem ser revisadas por um fator multiplicativo global já que a atualização de  $D$  pode, em princípio, modificar a normalização naquele domínio.

Em especial, exige-se que na ausência de informação adicional, o algoritmo não modifique a atribuição inicial dada por  $\{\omega_i\}$ . Em linguagem matemática isso equivale à exigência que a maximização de  $\eta[p|\omega]$  sem vínculos forneça  $p_i = \omega_i$ .

É necessário, então, que as probabilidades  $p_i = g_i^{-1}$  não tenham qualquer dependência em  $p_{j \neq i}$  fora a normalização controlada pelo parâmetro  $\lambda$ . A independência de  $g_i^{-1}$  com  $p_{j \neq i}$  obviamente reflete na independência de  $g_i$  com estas variáveis. Isto implica que

$$\frac{\partial^2 \eta}{\partial p_i \partial p_j} = 0; \quad \text{se } i \neq j. \quad (1.42)$$

A forma mais geral que satisfaz a equação acima é

$$\eta[p|\omega] = \sum_i f_i(p_i, \omega_i), \quad (1.43)$$

onde a maximização deste funcional fornece uma probabilidade posterior função da probabilidade anterior e verossimilhança. Para que o teorema de Bayes seja respeitado, a dependência com relação a estes parâmetros deve ser linear como a forma  $p_i \propto \omega_i L_i(A)$ . Fosse o resultado da maximização  $p_i = \omega_i^2 L_i(A)$ , ou qualquer situação igualmente esdrúxula, o funcional específico que o determinou seria imediatamente descartado. Para satisfazer esta condição, basta que a função  $g_i \equiv \frac{\partial \eta}{\partial p_i} = f_i$  dependa de  $p_i$  e  $\omega_i$  através da razão entre as duas probabilidades  $\frac{p_i}{\omega_i}$ . Desta forma temos

$$g_i(p_i, \omega_i) = g_i \left( \frac{p_i}{\omega_i} \right) \Rightarrow \frac{p_i}{\omega_i} = g_i^{-1}(\beta a_i + \lambda), \quad (1.44)$$

ou seja,

$$\eta[p|\omega] = \sum_i \omega_i f_i \left( \frac{p_i}{\omega_i} \right). \quad (1.45)$$



A condição de localidade impõe que informação adicional sobre um sub-domínio disjunto apenas renormalize as probabilidades dos outros sub-domínios. Se dois problemas são idênticos a exceção de certa informação diferente sobre um  $x_k$  específico, os quais distinguimos pelos valores  $a_k$  e  $a'_k$ , é necessário que ambos possuam as mesmas atribuições de probabilidade para  $p_{i \neq k}$  a menos da normalização do subdomínio. Esta modificação implica que o multiplicador de Lagrange  $\lambda$  que controla a normalização é alterado de um problema para o outro, apesar de manter inalterados os  $p_i \neq p_k$ . Desta forma que escrevemos  $\lambda' = \lambda + \delta\lambda$ , onde a condição anteriormente exposta significa que

$$\frac{p_i}{\omega_i} = g_i^{-1}(\beta a_i + \lambda) \propto g_i^{-1}(\beta a_i + \lambda + \delta\lambda); \quad i \neq k, \quad (1.46)$$

que, em geral, corresponde a propriedade de invariância pela transformação

$$g_i^{-1}(x) = \text{cte} \times g_i^{-1}(x + a), \quad (1.47)$$

obedecida apenas pela família das exponenciais

$$g_i^{-1}(x) = q_i^x \Rightarrow g_i(x) = \log_{q_i}(x). \quad (1.48)$$

A redefinição da constante de normalização para os níveis diferentes de  $k$  é feita de maneira uniforme de sorte que  $q_i^{\delta\lambda} = \text{cte}$ . Isto implica que o fator  $q_i$  é o mesmo para todos  $i \neq k$ . Integrando  $g(x) = K \ln(x)$ , se obtêm  $f(x) = K(x \ln(x) - x)$ , e notando que a escolha do nível  $k$  específico é imaterial no decorrer do argumento, escrevemos o resultado final como

$$\eta[p|\omega] = K \sum_i p_i \ln \frac{p_i}{\omega_i} + \text{cte}. \quad (1.49)$$

O fato de que a extremização de  $\eta$  corresponde a um máximo também implica que

$$\frac{\partial^2 \eta}{\partial p_i^2} = K \frac{1}{p_i} < 0, \quad (1.50)$$

de forma que a escolha de um  $K < 0$  determina completamente o processo de inferência. Fazemos, sem perda de generalidade,  $K = -1$  já que uma mudança de escala em  $K$  reflete apenas na escala de  $\beta$ . Desta forma, o processo aqui descrito implica no resultado

$$P(x_i|A(\beta)H) = \frac{\omega_i}{Z(\beta)} e^{-\beta a_i}; \quad Z(\beta) \equiv \sum_i \omega_i \frac{e^{-\beta a_i}}{Z(\beta)}, \quad (1.51)$$

onde  $\beta$  é uma função implícita de  $A$  tal que  $\beta(A)$  e  $A(\beta)$  especificam a probabilidade  $P(x_i|AH)$  dada pela fórmula acima.

### 1.3.3 O significado da entropia inferencial

Numa retrospectiva crítica, revisemos se a prescrição anterior corresponde a uma inferência compatível com o teorema de Bayes. A resposta obtida não é inteiramente adequada: a verossimilhança implícita em (1.51) é função do parâmetro  $\beta$ , não do valor de  $A$ , como se queria inicialmente. A crítica procede no que, mesmo existindo uma relação implícita entre  $A$  e  $\beta$  dada pela condição de vínculo  $\langle a_i \rangle = A$ , a função  $\beta(A)$  depende dos valores específicos da probabilidade anterior  $\{\omega_i\}$ , refletindo numa dependência implícita da verossimilhança com a probabilidade anterior. Isto, obviamente, não faz sentido no contexto do teorema de Bayes.

A razão por trás deste problema, no entanto, não está relacionada especificamente à forma da entropia, mas deveria ser esperada desde o começo quando especificou-se que o conhecimento a ser incorporado nas atribuições de probabilidade posteriores faz referência a informação testável sobre a própria distribuição posterior. Neste sentido, qualquer algoritmo que gere a função verossimilhança a partir de informação testável compartilha o mesmo problema: a probabilidade anterior aparece explicitamente na distribuição de teste, sem, no entanto, ser possível separá-la da verossimilhança que se deseja determinar. Para o mesmo vínculo, por exemplo  $\langle a_i \rangle = A$ , a verossimilhança muito provavelmente seria diferente em situações que correspondem a probabilidades anteriores diferentes. Ao manter os mesmos valores nos dois problemas, as alterações em  $P(x|H)$  e, portanto, em  $P(x|AH)$  implicariam na violação do vínculo em pelo menos um dos casos.

É possível contornar o problema trabalhando com a situação correlata que visa atualizar a probabilidade anterior a partir do conhecimento da variável  $\beta$  definida em (1.40). Desta forma, substituímos o problema original pelo problema equivalente

$$P(x_i|\beta H) = P(x_i|H) \frac{P(\beta|x_i H)}{P(\beta|H)} \quad (1.52)$$

em que o verossimilhança  $P(\beta|x_i H)$  é, de fato, independente de  $\{\omega_i\}$ . Esta mudança de perspectiva torna a inferência estatística bem definida a partir de informação testável, às custas, obviamente, de introduzir um parâmetro desconhecido. É importante notar que, em física, muitas vezes estes parâmetros assumem um papel importante definindo conceitos como “temperatura”, “pressão”, “potencial químico” etc.

O modo como este problema se relaciona com o primeiro é ditado pela forma funcional que existe (implicitamente) entre  $A$  e  $\beta$ . Uma propriedade fundamental

no esquema apresentado é que as variáveis se relacionam por uma transformada de Legendre. Isto determina a equivalência entre o conteúdo matemático expresso pela probabilidade posterior  $P(x_i|AH)$  com a probabilidade  $P(x_i|\beta H)$ . Para verificar este fato, escrevemos a entropia inferencial avaliada no máximo como função de  $A$ , ou seja,  $\bar{\eta}(A) \equiv \max \eta[p|\omega]$ . Daí se obtêm a relação

$$\frac{\partial \bar{\eta}}{\partial A} = \sum_I \frac{\partial \eta}{\partial p_i} \frac{\partial p_i}{\partial A} = \beta \sum_i \frac{\partial F}{\partial p_i} \frac{\partial p_i}{\partial A} + \lambda \frac{\partial}{\partial A} \sum_i p_i = \beta. \quad (1.53)$$

A partir deste resultado, escrevemos a transformada de Legendre de  $\bar{\eta}(A)$  como

$$\bar{\eta}(A) = \beta A + \ln Z(\beta), \quad (1.54)$$

onde mostra-se de maneira semelhante que  $\frac{\partial}{\partial \beta} \ln Z(\beta) = -A$ . Este fato eleva a importância da entropia inferencial de um mero artifício para gerar funções *verossimilhança* associadas a uma classe de informação testável, para o papel mais fundamental de um objeto que conecta o problema de inferência para  $P(x_i|AH)$  com a inferência de  $P(x_i|\beta H)$  rigorosamente equivalente ao teorema de Bayes. A partir deste ponto, não podemos nos livrar do conceito já que a habilidade de conectar os dois problemas se revela fundamental para qualquer uso consistente deste método.

### 1.3.4 Entropias generalizadas: sub-localidade

Um fato que restringe a generalidade do resultado acima é que as funções de verossimilhança obtidas pelo algoritmo são invariavelmente exponenciais de  $a_i$ . O argumento pode ser generalizado para lidar com outras distribuições afrouxando a condição de localidade. Deste modo, especifica-se

**Condição de sub-localidade:** Ao fornecer informação na forma  $\langle a_i \rangle$  que diz respeito a um subdomínio  $D = \{i, i', \dots\}$ , as probabilidades referentes aos outros domínios são fracamente revisadas. Se entende por “fracamente revisadas” a possibilidade que tais probabilidades se alterem de acordo com uma prescrição global que não corresponde necessariamente a uma renormalização. A única exigência, e daí o nome sub-localidade, é que tal prescrição seja independente de que subdomínio foi atualizado. A motivação por trás da sub-localidade é lidar com situações em que existam correlações que quebrem a linearidade do sistema, desta forma, ao realocar probabilidade de um subdomínio para outro, pode ser necessário retirar ou acrescentar probabilidades de forma desigual entre os subdomínios.

Existe pelo menos uma aplicação óbvia em física. É o caso das chamadas probabilidades a 1-partícula dos modelos cinéticos. Neste caso, a alocação de probabilidades também corresponde à alocação de partículas entre os estados  $q, p$ . Incorporar vínculos/informação poderia modificar a dinâmica no espaço  $q, p$  de maneiras inesperadas devido à ação das interações entre partículas.

Mesmo trabalhando com esta condição mais geral, ainda é necessário que as probabilidades  $p_i = g_i^{-1}$  não dependam explicitamente de  $p_{j \neq i}$  pelo critério que prescrição de atualização das probabilidades não discrimina nenhum subdomínio. De maneira idêntica ao caso anterior, concluímos que

$$\eta[p|\omega] = \sum_i f_i(p_i, \omega_i). \quad (1.55)$$

Utilizando o mesmo argumento a respeito da compatibilidade com o teorema de Bayes, obtemos a fórmula geral para a entropia sub-local

$$\eta[p|\omega] = \sum_i \omega_i f_i \left( \frac{p_i}{\omega_i} \right), \quad (1.56)$$

onde ainda é necessário fazer algumas exigências a respeito da forma funcional de  $f(x)$ .

Para que o algoritmo forneça uma resposta única, exige-se que a seja possível inverter a relação

$$\frac{\partial f}{\partial p_i} \equiv g_i \left( \frac{p_i}{\omega_i} \right) = \beta a_i + \lambda, \quad (1.57)$$

o que implica que as funções  $g_i(x)$  são inversíveis. Disto segue que existe um sinal bem definido para a derivada de  $g_i(x)$ , garantindo que  $f_i(x)$  seja côncava ou convexa. Para satisfazer o critério de maximização só resta a primeira escolha.

No caso especial que a probabilidade anterior é equiprovável e, considerando a ausência de vínculos, teríamos que

$$g_i(p_i) = g_i(m^{-1}) = \lambda, \quad (1.58)$$

onde a constante  $m$  é o inverso do número de possibilidades consideradas. Dada a arbitrariedade deste número conclui-se que

$$g_i(m^{-1}) = g_j(m^{-1}) \Rightarrow f_i(m^{-1}) = f_j(m^{-1}) + \text{cte}. \quad (1.59)$$

Esta propriedade significa que as funções  $f_i(n_i)$  são essencialmente as mesmas a menos de uma constante aditiva imaterial. Colecionando-as numa constante global,  $\eta[p|\omega]$  pode ser escrita como

$$\eta[p|\omega] = \sum_i \omega_i f\left(\frac{p_i}{\omega_i}\right) + \text{cte.} \quad (1.60)$$

Aderimos à convenção que  $f(0) = 0$ , o que possui o apelo de que considerar novos eventos impossíveis não contribui para a entropia.



## 2 *Física de sistemas macroscópicos*

Executam, segundo as regras herméticas  
Desde a trituração, a fixação, a destilação e a coagulação

Trazem consigo cadinhos,  
Vasos de vidro, copos de louça, todos bem, e iluminados

(JORGE BEN — Os alquimistas estão chegando)

Uma das constatações mais impressionantes a respeito do comportamento de sistemas macroscópicos é a razoável simplicidade com que podem ser descritos. Isto não é dizer que as leis que determinam este comportamento sejam particularmente simples: sistemas biológicos, a dinâmica atmosférica e mesmo a matéria comum sob condições especiais de temperatura, pressão etc, exibem uma fenomenologia absolutamente não trivial. O modo como estes sistemas operam, no entanto, pode ser considerado bastante regular quando se avalia que a descrição fundamental a partir das partículas que os compõe envolve algo como  $10^{23}$  graus de liberdade.

A razão para tal regularidade é a enorme robustez associada à descrição macroscópica. Uma afirmação simples do tipo “*quilograma é a massa contida em 1l de água*”, determina uma variedade enorme de modos como se pode arrumar um certo número de moléculas de água que constituem um litro. É lógico que nem sempre é necessário estar atento aos detalhes microscópicos do objeto físico que se quer estudar; mesmo no âmbito puramente teórico é importante entender se existem e quais são as leis que determinam o comportamento das variáveis simplificadas *quilos, litros, densidade de partículas* etc.

Uma característica singular das variáveis macroscópicas é seu caráter contextual; Jaynes e Wigner utilizam a denominação “antropomórfica”, no sentido específico que o critério de escolha do estado macroscópico é *utilidade* segundo a perspectiva humana<sup>1</sup>.

---

<sup>1</sup>Como “*utilidade*” considera-se a capacidade de aperceber e manipular tais variáveis. Isto inclui, em geral, considerações acerca escolhas de caráter pragmático (é possível montar uma teoria simples para estas variáveis), técnico (é possível medi-las e controlá-las em laboratório) e epistemológico (são

Assim, não se busca um conjunto “fundamental” de parâmetros macroscópicos: a escolha depende das motivações e da engenhosidade do cientista na tarefa de descrever a Natureza. Na física elementar, alguns objetos (pontos no espaço de fase ou vetores no espaço de Hilbert, por exemplo) são eleitos para identificar *univocamente* qualquer sistema físico, encerrando uma descrição irreduzível. Não se espera explicá-los em termo de outras coordenadas mais fundamentais, a não ser, é claro, quando admitimos que determinada teoria não passa de uma aproximação efetiva da teoria *realmente* fundamental, i.e. como a mecânica quântica seria para a teoria de cordas. No âmbito macroscópico, conjuntos de coordenadas distintas representam o mesmo sistema físico, mas nem sempre há uma equivalência ou um mapeamento que conecte as diferentes descrições. As várias maneiras de se caracterizar o mesmo objeto podem estar associadas a fenomenologias e procedimentos experimentais distintos (ex.: as coordenadas mecânicas como posição, momento, tensão-torção etc e as coordenadas termodinâmicas que eventualmente descrevem o mesmo objeto físico).

O primeiro passo para relacionar as descrições macroscópica e microscópica é determinar a função  $\mathcal{F}(x)$  que mapeia as coordenadas microscópicas  $\{x\}$  nas coordenadas macroscópicas  $\{F\}$ . Note que, fora os processos relativamente simples normalmente tratados pela física, o quadro geral de descrição é extremamente complicado; imagine, por exemplo, coordenadas que quantificariam os aspectos relevantes de sistemas como “estado cerebral”, “ser vivo”, “sociedade” etc. Uma vez que a relação entre os níveis descritivos seja estabelecida, é possível tirar proveito do conhecimento acerca das leis mecânicas a fim de demarcar a dinâmica macroscópica associada — o intuito é fornecer uma equação de movimento para  $F(t)$ , dado um problema de valor inicial  $F_0$  e um modelo de interações fundamentais. A nosso proveito está o fato que em sistemas de alta dimensionalidade, considerações simples de natureza estatística condicionam mais fortemente alguns aspectos da dinâmica macroscópica que as próprias leis de movimento das partículas.

## 2.1 Indeterminismo macroscópico

“We are all under an ego-driven temptation to project our private thoughts onto the real world, by supposing that the creations of one’s own imagination are real properties of Nature, or that one’s own ignorance signifies some kind of indecision on the part of Nature.

(E.T JAYNES)

---

conceitos importantes na descrição de mundo).



Muitas vezes é tentador acreditar que as propriedades da evolução macroscópica espelham a diretamente a dinâmica microscópica; mais precisamente, poderia se pensar que as simetrias dinâmicas obedecidas pelas leis microscópicas tais como a reversibilidade temporal e conservação de volume no espaço de estados se espelhariam imediatamente no mundo macroscópico. Esta expectativa pode se confirmar eventualmente, mas em muitas situações é falsa. É necessário, antes de tudo, especificar precisamente o que se entende por “leis para o comportamento macroscópico”.

Ao contrário do que acontece na mecânica, um problema macroscópico de valor inicial admite múltiplas soluções. O motivo para isso é que a dinâmica macroscópica é determinada por um dos vários estados microscópicos compatíveis com a condição inicial, e alguns destes estados podem fornecer uma evolução macroscópica distinta que outros. Como se assume, desde o início, que o problema é especificado integralmente por variáveis macroscópicas, requerimentos muito fortes sobre o estado mecânico devem ser evitados e é necessário apelar para uma linguagem que lide com esta indeterminação. Enfatizamos que, como nos lembra Laplace, estas probabilidades não “são mais que a expressão da nossa ignorância em respeito às verdadeiras causas”, elas não implicam na existência de qualquer traço de aleatoriedade. Os objetos matemáticos que formulam a nossa “lei para o comportamento macroscópico” são, desta forma, as probabilidades

$$P(F_f|F_0) \equiv \text{"probabilidade de estar em } F_f \text{ dado que estava em } F_0\text{"}, \quad (2.1)$$

e ainda

$$P(F_f \dots F_1|F_0) \equiv \text{"probabilidade de, partindo de } F_0, \text{ seguir por } F_1, F_2, \dots, F_f\text{"}, \quad (2.2)$$

onde se entende por  $F_i \equiv$  “sistema está no estado  $F_i$  no instante  $t_i$ ”. As duas probabilidades se relacionam pela regra de marginalização

$$P(F_f|F_0) = \sum_{F_1} \sum_{F_2} \dots \sum_{F_{f-1}} P(F_f \dots F_1|F_0H), \quad (2.3)$$

de forma que, conhecendo as probabilidades de cada trajetória (2.2), é possível determinar a probabilidade de cada transição (2.1).

Possivelmente a maior vantagem do esquema Bayesiano em relação a outras teorias de probabilidades é a facilidade em se incorporar informação explicitamente. Isto, como já se devia suspeitar, é vital para construir boas atribuições de probabilidades já que é necessário levar em conta todo o conhecimento sobre a dinâmica microscópica a nosso favor. Neste espírito, uma distribuição de probabilidades para variáveis macroscópicas

$P(F)$  pode ser encarada segundo várias perspectivas. Caso o “desconhecimento” sobre o estado macroscópico consista numa indeterminação experimental, uma limitação que sempre existe nos aparatos de medição, possivelmente a melhor atribuição seria uma Gaussiana,  $P(F) = \frac{1}{\sqrt{2\pi}\Delta F} e^{-\frac{1}{2\Delta F^2}(F-F_{lab})^2}$ . Em uma abordagem teórica, é necessário, por exemplo, relacionar a distribuição de probabilidades para o estado macroscópico com a distribuição microscópica. Desta maneira, utilizamos a regra de marginalização para obter<sup>2</sup>

$$P(F) = \sum_x P(F|x) = \sum_x P(F|x)P(x). \quad (2.4)$$

Note que, segundo a metodologia Bayesiana de Cox, os argumentos de  $P(\dots|\dots)$  são afirmações lógicas. Neste caso, a afirmação  $x' \equiv$  “O estado microscópico é  $x'$ ” implica em  $F' \equiv$  “O estado macroscópico é  $F' \equiv \mathcal{F}(x')$ ”. Ou, se quiser,  $x' \Rightarrow F'$  e, para todo  $F \neq F'$ ,  $x' \Rightarrow \bar{F}$ . Desta maneira, conhecendo um estado microscópico  $x$ , a probabilidades condicional  $P(F|x)$  é 1 se  $\mathcal{F}(x) = F$  e 0, caso contrário — isto também pode ser expresso como  $P(F|x) = \delta(\mathcal{F}(x) - F)$ . A especificação  $P(F)$  possui uma interpretação simples, que é a parcela do volume no espaço de estados macroscópicos correspondente a uma determinada configuração macroscópica e por isso nos referimos a ela como *representatividade* de  $F$ . Em geral tanto o espaço de estados microscópicos quanto macroscópicos não podem ser propriamente normalizados, mas mesmo assim é legítimo comparar diferentes representatividades sem atribuir um valor absoluto à cada uma delas. Isto não é um problema pois inferências geralmente são feitas em um domínio de discurso restrito, i.e.: conhecendo  $A, B, C$ , o que se espera de  $D$ ? Em se tratando da descrição macroscópica, estas especificações adicionais consistem em informação (parcial ou completa) sobre o estado *macroscópico* do sistema em algum instante do tempo. A próxima seção lida parcialmente com este problema, a solução que pode-se dizer completa ainda depende de uma prescrição adequada para determinar  $P(x)$ .

### 2.1.1 Probabilidades associadas a processos macroscópicos

O conhecimento explícito do operador da evolução dinâmica, aqui denotado por  $\mathcal{U}(t)$ <sup>3</sup>, permite determinar as probabilidades de um estado macroscópico mudar de  $F_0$

<sup>2</sup>É válido apontar que nem todas grandezas macroscópicas serão escritas desta maneira. Temperatura, pressão, potencial químico, calor específico, etc não escritos como as médias em (2.4). No entanto, todas elas se relacionam com a entropia de um estado dada por  $k \ln P(F)$ . Dizer que estas grandezas não admitem uma expressão no modelo microscópico é como dizer que a força também não possui sentido mecânico porque é a derivada do potencial.

<sup>3</sup>Este operador, obviamente, está relacionado ao operador Liouvilliano  $\mathcal{L}$  pela relação  $\mathcal{U}(t) = e^{it\mathcal{L}}$ , onde o Liouvilliano pode ser representado de várias maneiras, dependendo em se tratar de uma

em  $t_0$  para o estado final  $F_1$  em  $t_1$ . Considere uma probabilidade do tipo

$$P(F_1|F_0) \equiv \text{ "probabilidade do sistema estar em } F_1 \text{ em } t_1 \text{ sendo que estava em } F_0 \text{ no instante } t_0 \text{ "}$$
 (2.5)

que incorpora informação dinâmica a respeito dos estados iniciais. Neste caso, é fácil mostrar que

$$P(F_1|F_0) = \sum_{x_0} P(x_0|F_0)P(F_1|x_0F_0).$$
 (2.6)

Conquanto  $F_0$  seja compatível com o estado microscópico  $x_0$ , o que é garantido pela presença do termo  $P(x_0|F_0)$ , o aparecimento de  $F_0$  em  $P(F_1|x_0F_0)$  é redundante e podemos omití-lo<sup>4</sup>. Aplicando o teorema de Bayes no outro termo da somatória, ficamos com

$$P(F_1|F_0) = \frac{1}{P(F_0)} \sum_{x_0} P(x_0)P(F_0|x_0)P(F_1|x_0).$$
 (2.7)

Dado que o operador  $\mathcal{U}(t)$  é reversível, o conhecimento de uma condição inicial implica no conhecimento de uma condição posterior assim como o oposto. Em símbolos, é o mesmo que  $x_0 \Leftrightarrow \mathcal{U}(t) \cdot x_0$ . Desta forma, introduzindo o estado  $x(t_1) \equiv \mathcal{U}(t_1) \cdot x(0)$ , a dinâmica microscópica nos diz que as afirmações lógicas se relacionam por  $x_1 \Leftrightarrow x_0$ . Fazendo uso disto em  $P(F_1|x_0) = P(F_1|x_1)$ , ficamos com

$$P(F_1|F_0) = \frac{1}{P(F_0)} \sum_{x_0} P(x_0) \delta(\mathcal{F}(x_0) - F_0) \delta(\mathcal{F}(\mathcal{U}(t_1) \cdot x_0) - F_1).$$
 (2.8)

Intuitivamente esta atribuição consiste em contar a fração das trajetórias microscópicas saindo do estado  $F_0$  em  $t_0$  que chegam no estado  $F_1$  em  $t_1$  ponderadas pelo peso estatístico dos estados iniciais, o que não é nada surpreendente.

Note que o aparecimento de termos do tipo  $P(x_0|F_i)$  justifica que se aceite as implicações  $x_0 \Leftrightarrow x_i$  e portanto  $x_0 \Rightarrow F_i$  pois restringe as condições iniciais  $\{x\}_0$  àquelas que, evoluídas de  $t_i$ , correspondam ao estado macroscópico  $F_i$ . Utilizando este fato é possível

---

mecânica clássica ou quântica.

<sup>4</sup>Ao considerar a probabilidade de uma afirmação sob uma hipótese,  $P(A|H)$ , restrições adicionais que sejam uma implicação de  $H$  são redundantes. Usando o teorema de Bayes  $P(A|hH) = P(A|H) \frac{P(h|AH)}{P(h|H)}$  e se  $H \Rightarrow h$ , as probabilidades  $P(h|H)$  e  $P(h|AH)$  devem corresponder necessariamente à certeza e, portanto, são iguais a um. Isto mostra que se  $H \Rightarrow h$ ,  $P(A|hH) = P(A|H)$ , e um corolário interessante é que no caso especial em que  $H \Leftrightarrow h$ , para qualquer  $A$  vale que  $P(A|H) = P(A|h)$ .

Note que um resultado similar vale para  $P(AB|H)$  quando  $A \Rightarrow B$ . Neste caso,  $P(AB|H) = P(A|H)$  e a bi-implicação garante que  $P(AB|H) = P(A|H) = P(B|H)$ .

calcular a probabilidade de uma trajetória, que é simplesmente

$$P(F_f \dots F_1 | F_0) = \frac{1}{P(F_0)} \sum_{x_0} P(x_0) P(F_0 | x_0) P(F_1 | x_0) \dots P(F_f | x_0), \quad (2.9)$$

onde os condicionais  $P(F_i | x_0)$  são entendidos como  $\delta(\mathcal{F}(\mathcal{U}(t_i) \cdot x_0) - F_i)$ . Novamente a interpretação do resultado é bastante intuitiva, correspondendo à fração de trajetórias que saem de  $F_0$  passando por  $F_1, F_2, F_3$ , etc. Estes resultados justificam a utilização de uma teoria de *ensemble*, já que todas as probabilidades aqui calculadas correspondem exatamente ao que se esperaria ao reunir um grande número de “cópias mentais” do sistema e considerar as respectivas evoluções Hamiltonianas. Discutimos a origem histórica do conceito e como ele pode ser utilizado construtivamente (e destrutivamente) na próxima seção.

### 2.1.2 Ensemble

(...) We may imagine a great number of *systems of the same nature* [grifo do autor], but differing in the configurations and velocities which they have at a given instant; (...) it may be so as to embrace every conceivable combination of configurations and velocities. And here we may set the problem, not to follow a particular system through its succession of configurations, but to determine how the whole number of systems will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given form some one time. (...)

Such inquiries have been called by Maxwell *statistical*.

(J. W. GIBBS — Elementary Principles in Statistical Mechanics)

Sobre uma coisa todo mundo concorda. Jogue uma moeda sem vício 1.000 vezes e é bem provável que as frequências relativas para cara e coroa sejam semelhantes. Isso é válido para repetições independentes de qualquer evento aleatório: no limite que o número de repetições tende ao infinito, as frequências convergem para uma determinada distribuição de probabilidades. Isto é uma consequência da chamada “lei dos grandes números”, que identifica frequência estatística com probabilidade no limite de infinitas repetições. Podemos utilizar esta “lei” a nosso favor de basicamente duas maneiras: a primeira consiste na formulação original do teorema, ou seja, determinar uma probabilidade a partir das frequências estatísticas. A outra é o caminho inverso; define-se um conjunto de elementos distribuídos em razões idênticas às respectivas atribuições de probabilidade, e para o qual calcula-se as grandezas estatísticas levando em conta a soma das configurações individuais de cada elemento. Sempre que procedemos

da segunda maneira, estamos fazendo uma teoria de *ensemble*.

Em física, o conceito é particularmente útil porque permite visualizar a evolução de uma distribuição de probabilidades definida, por exemplo, no espaço de grandezas microscópicas. Este é o conteúdo da afirmação de Gibbs, que é conhecido como substituição estatística. Uma vez de posse de uma distribuição de probabilidades, tome  $P(x)$  definido em (2.4) como exemplo, podemos nos perguntar sobre várias questões correlatas à ela — a maneira que se modifica no tempo, as médias de grandezas mecânicas, a evolução destas grandezas, etc. A lei dos grandes números assegura que é possível simular todas as propriedades da probabilidade, prescrevendo as frequências de ocupação corretas para um conjunto muito grande, i.e.: infinito, de elementos. Isso, em essência, é o que Gibbs e Maxwell tinham em mente ao introduzir o conceito de *ensemble*.

Ainda que represente uma metáfora útil e uma ferramenta valorosa para o cálculo e simulação de processos físicos, a terminologia geralmente envolvida na descrição de ensembles muitas vezes insinua toda sorte de erros, confusão e mistificações em pessoas, de outro modo, bastante racionais. Na verdade, uma avaliação crítica mais consistente é, se não impossível, mas muito difícil já que cada autor utiliza terminologia e se apóia em conceitos radicalmente diferentes que preenchem todo espectro entre o correto e o absurdo<sup>5</sup>. O site *wikipedia.org*, que pela sua proposta de autoria coletiva provavelmente representa uma espécie de “compreensão média” dos assuntos gerais diz que “*ensemble formalises the notion that a physicist can imagine repeating an experiment again and again under the same macroscopic conditions, but, unable to control the microscopic details, may expect to observe a range of different outcomes.*”. Queremos mostrar o quão vazio é este tipo de afirmação.

Um erro comum é interpretar os ensembles objetivamente, ou seja, que a distribuição de elementos no ensemble seria obtida a partir de experimentos reais que, mais ainda, corroboram a hipótese da equiprobabilidade *a priori*<sup>6</sup>. Este tipo de afirmação dificilmente

---

<sup>5</sup>Como passagem do segundo tipo, vale assinalar a referência [?, p. 46] em que o autor, após assinalar a crença que a probabilidade só pode ser entendida objetivamente, ou seja, como o limite de frequências de uma série infinita, refere positivamente o leitor à “(Hajek 1996) for fourteen other arguments against finite frequentism)”. Sendo que estas observações arbitrarias, restringem consideravelmente o domínio de aplicabilidade da teoria de probabilidades (i.e. a zero), o autor necessariamente adota uma interpretação do tipo *gedankenexperiment* para os ensembles de Gibbs e assinala posteriormente “the frequency interpretation can be twisted to become an inter-subjective interpretation [como a Bayesiana], if the ensembles are thought of as *mental copies* of a single system of interest, and not as a collection of systems that all exist in the real world. ” Apesar destes tropeços, o autor assinala problemas reais e muito pertinentes com algumas interpretações comuns dos ensembles de Gibbs.

<sup>6</sup>Diz que os elementos são distribuídos com a mesma frequência pelo espaço de fase. Em mecânica quântica é necessário supor que, além da distribuição nos auto-estados de uma base que comuta ser equiprovável, as fases das funções de onda também o seriam. Consideramos que esta afirmação possui

seria feita por um experimentalista: o resultado de um experimento macroscópico é (por definição) uma variável macroscópica; nunca foi, e nem nunca será, o estado mecânico definido por  $x$ , ou seja, a informação necessária para determinar assintoticamente a probabilidade  $P(x)$ . As probabilidades  $P(x)$ ,  $P(x|U)$  etc, estão veladas à qualquer experimento macroscópico factível, de forma que a prescrição para criar um determinado ensemble, tal como o ensemble canônico, *não se baseia na experiência, mas é uma simples idealização tão boa quanto outras*. Uma segunda crítica à viabilidade deste procedimento é que, mesmo que um super-experimentalista desenvolva um método para determinar o estado mecânico de um sistema macroscópico, o espaço de fase associado é tão vasto, mas tão inimaginavelmente vasto, que nunca seria escrutinado com uma resolução boa o suficiente para fundamentar objetivamente qualquer forma de distribuição (sem falar que dificilmente um financiador público patrocinaria tamanha empreitada).

Este tipo de expectativa está relacionada à falsa identificação entre probabilidade e frequência estatística. Não adequa aos físicos, naturalmente, cometer nenhuma violência com relação à estatística e muito menos com as matemáticas. Neste ponto podemos ter problemas pois há várias maneiras de interpretar uma probabilidade; autores com posições frequentistas muito radicais tendem a ver mais ontologia nos ensembles que realmente a razão justifica. Para ilustrar alguns pontos da discussão, considere um ensemble com um grande número de moedas obtido aleatoriamente. No limite de infinitas jogadas, esperamos obter 50% de caras  $C$  e 50% de coroas ( $\bar{C}$ ), isso estabelece univocamente a atribuição de probabilidades para os eventos  $P(C|H) \rightarrow 0.5$  e  $P(\bar{C}|H) \rightarrow 0.5$ . Em qualquer série finita, os resultados raramente serão exatamente  $N/2$  caras e  $N/2$  coroas, de modo que a distribuição de frequências não determina a “probabilidade correta” ; poucos argumentariam que, numa série de 1.000.000 de jogadas em que se obtêm 500.001 caras e 499.999 coroas, a moeda utilizada certamente seria viciada por um fator  $\pm 10^{-6}$ . É lógico que a escolha desta probabilidade não é rigorosamente incorreta, mas os dados experimentais também não descartam a equiprobabilidade que se apresenta como a hipótese mais razoável.

O argumento rigoroso que nos leva a concluir que a atribuição equiprovável não é descartada, de uma maneira ou de outra, envolve o cálculo da probabilidade que a atribuição  $P(C|H) = P(\bar{C}|H) = 0.5$  esteja correta. Numa notação Bayesiana, isto é expresso por  $P(p_c|n_c N)$ , que é a probabilidade que a atribuição  $p_c$  seja correta dado

---

o mesmo conteúdo da anterior (e as mesmas dificuldades), mas apenas leva em conta as especificidades da descrição quântica da natureza. No entanto, nos restringimos aos ensembles clássicos por questões de simplicidade.

que se obteve  $n_c$  caras em  $N$  jogadas. O valor exato envolve hipóteses adicionais que não universalmente aceitas, mas em qualquer situação, converge assintoticamente para uma gaussiana com variância igual à  $\sqrt{\frac{p_c(1-p_c)}{n_c}}$ . No exemplo atual, este valor seria  $\sigma = 2.000^{-1}$ , ou seja, poderíamos nos sentir perfeitamente confortáveis com a hipótese da equiprobabilidade em qualquer situação em que o número de caras ou coroas estivesse na faixa  $n_c = 500.000 \pm 500$  (e com graus menores de conforto na medida que se avança por estes limites).

Um ponto importante na discussão é que todos os chamados “testes de confiança estatística” podem ser traduzidos, de alguma maneira, em afirmações do tipo: “a probabilidade que o modelo estatístico representa os dados obtidos é  $XX\%$ ”. Estes testes podem ser interpretados como uma adição conceitualmente independente à teoria, ou como uma mera consequência do cálculo usual de probabilidades. Muitos preferem pensar da segunda maneira, que também acreditamos ser a correta. Mas a partir do momento em que se faz uma identificação estrita entre probabilidade e frequência, não há como atribuir legitimamente probabilidades deste tipo. Para um freqüentista rigoroso, o significado dos testes de confiança (que são importantíssimos dado a impossibilidade de executar séries aleatórias infinitas), permanece essencialmente misterioso, assim como o significado dos teoremas de convergência (o que significa a frequência relativa convergir para uma probabilidade se a probabilidade é esta frequência?) e a maioria das aplicações de probabilidades em problemas reais. Neste sentido, se a física busca compatibilizar o ensemble com a teoria de probabilidades, é melhor que seja com a parte da teoria que se refere à realidade, de forma que não precisa adotar uma interpretação ontológica para os ensembles (ensembles como séries reais). Os ensembles são criações formais, eles não existem na Natureza (e se estão ninguém nunca viu um deles!)<sup>7</sup>.

Mesmo aceitando a idéia que os ensembles residem na mente dos cientistas, e não em seus laboratórios, esperamos intuitivamente que propriedades objetivas dos sistemas descritos reflitam, de algum modo, na determinação das distribuições mais adequadas. Isto é verdade no sentido que o conhecimento disponível (macroscópico e microscópico) permite eliminar uma série de ensembles esdrúxulos que, de outra maneira, poderiam ser considerados. No entanto, é totalmente falacioso pensar que

---

<sup>7</sup>Vale ressaltar que estas críticas não são direcionadas exclusivamente por Bayesianos. Várias correntes diferentes dariam o mesmo veredito com relação à interpretação freqüentista. Vale citar correntes formalistas (para as quais não há interpretação pre-estabelecida), os que, como Popper, alegam que probabilidades refletem objetivamente uma aleatoriedade intrínseca aos objetos, os que aceitam, de maneira limitada, a identificação entre probabilidade e conhecimento mas defendem que o experimento de frequências é o único meio de capturá-las objetivamente, entre outros.

o sucesso das nossas teorias implique, ou mesmo dependa, da escolha do *Ensemble Correto*. A questão é: não existe ensemble correto. Todo modelo estatístico, e isso é válido para a mecânica estatística, representa um certo estado de conhecimento acerca dos objetos que se estuda, mas não representa, de maneira alguma, todas as propriedades relevantes destes objetos. A incerteza não pode ser eliminada e não há garantias que a informação descartada é realmente irrelevante para o tipo de inferência que queremos fazer. Ao não considerar devidamente um certo tipo de informação, um bom modelo (que parta de pressupostos claros e bem fundamentados) pode fornecer previsões consideravelmente piores que um mau modelo (confuso, artificial, que exige o ajuste de vários parâmetros arbitrários etc). Dado que a ciência só avança seguindo modelos do primeiro tipo, restringimos a escolha dos nossos ensembles por um critério estético. Estes critérios não são constatações de novas leis da natureza e muito menos de fatos experimentais; existem infinitos ensembles possíveis para cada conjunto de parâmetros experimentais e vários critérios razoáveis para escolher entre vários deles. Qualquer prescrição que elimine o excesso sempre está sob o risco de também eliminar outras informações importantes, e não há como realmente contornar isto.

As grandezas que normalmente se mede em laboratório obedecem a uma relação de um-para-infinito com relação às possíveis distribuições de probabilidades (ensembles). Desta forma, se uma distribuição  $P(x)$  fornece os valores corretos  $F_i = \sum_x \mathcal{F}_i(x)P(x)$ , podemos dizer que  $P(x)$  é adequada; mas ainda assim existem infinitas outras distribuições que estariam perfeitamente de acordo com a informação disponível. Deste modo, o fato que a distribuição canônica,  $P(x|U) = \frac{1}{Z(\beta)} \exp(-\beta(U)H(x))$ , é adequada para descrever processos de equilíbrio, não implica que tenhamos qualquer compromisso com ela, nem que ela *defina* o equilíbrio de alguma forma<sup>8</sup> já que equilíbrio só pode ser expresso macroscopicamente. Uma característica que imediatamente a eliminaria de um quadro descritivo mais geral é o fato que esta distribuição é estacionária. Ainda que isto seja uma característica aceitável na descrição de situações de equilíbrio, um pouco de reflexão sobre a equação de Liouville mostra que as coisas não podiam ser piores em outras situações. Nenhuma distribuição de probabilidades jamais evoluirá para a distribuição canônica, a não ser, é claro, que se tratasse dela desde o começo. Para lidar com processos fora do equilíbrio é necessário utilizar outros ensembles; e novamente nenhum deles evoluirá para o ensemble canônico, mas se espera que todos eles, após um certo período de evolução, se encontrem na classe infinita de ensembles

---

<sup>8</sup>Ainda assim, é possível dizer legitimamente que a distribuição canônica é a atribuição de probabilidades obtida segundo a prescrição mais simples que ainda descreve adequadamente o equilíbrio. Note que, aqui, a razão para seu uso é pragmática, não é um compromisso ontológico.



que fornecem os mesmos valores médios para as grandezas macroscopicamente acessíveis — isso sim, define o equilíbrio.

É interessante, se não necessário, possuir alguma prescrição para se criar os ensembles compatíveis com uma certa configuração macroscópica. O tipo de metodologia comumente utilizado em estatística Bayesiana é reduzir os problemas ao menor número de atribuições de probabilidade das quais admitidamente não se possui controle, elas são as chamadas probabilidades anteriores ou *priors*. Este é o espírito da passagem  $P(F) = \sum_x P(F|x)P(x)$ , onde efetivamente transferimos o problema do domínio macroscópico, que talvez seja intratável, para o domínio microscópico, sobre o qual temos melhor conhecimento e intuições mais confiáveis. O modelo estatístico fica escorado na determinação de um único parâmetro,  $P(x)$ ; de tal modo que não se baseia em escolhas levianas motivadas exclusivamente pela adequação experimental a todo custo. Caso apareça uma falha, é possível ter clareza para detectar os possíveis culpados.

A postura pragmática tomada por Gibbs, que os ensembles seriam igualmente representados por todos os estados microscópicos compatíveis com uma configuração macroscópica, não resiste a um escrutínio mais profundo. Seria tentador, por exemplo, interpretar este tipo de atribuição pelo o argumento da razão insuficiente; mas em espaços contínuos, a equiprobabilidade não consiste numa definição operacional muito clara: o que é equiprovável num sistema de coordenadas pode não ser em outro, de forma que uma justificativa mais adequada seria altamente desejável. Caso isso não seja possível, a equiprobabilidade seria definitivamente o calcanhar de Aquiles da física estatística e a menor discrepância com o experimento imediatamente a colocaria sob suspeita. Neste sentido, discutimos duas maneiras de tentar colocá-la em bases mais adequadas; na próxima seção, expomos a abordagem que nos parece correta, enquanto, na seção posterior, faremos uma discussão sobre a abordagem mais tradicional com suas devidas críticas.

### 2.1.3 Como determinar a probabilidade anterior?

Jaynes considera o problema da determinação da probabilidade anterior (prior) a metade incompleta da teoria de probabilidades e estatística. Em muitos problemas existem regras mais ou menos bem estabelecidas para determinar a probabilidade associada a um certo evento (ou afirmação que representamos por  $E$ ) condicionada a ( $C$ ) — esta probabilidade representada como  $P(E|CH)$ . Estas regras passam por métodos de máxima entropia, princípios de contagem e procedimentos mais ou menos

tácitos que selecionam certas distribuições padronizadas (Gaussianas, Poissonianas, Deltas de Dirac etc). Note que, em geral, também é necessário conhecer a probabilidade correlata de que, sabendo da ocorrência de um certo evento (ou uma seqüência deles), o quão provável é a condição  $C$  em comparação a outros possíveis fatores  $C'$ ,  $C''$  etc. Estas probabilidades descrevem o grau de confiança depositado no modelo  $C$  condicionado à resposta de ( $E$ ). São representadas por  $P(C|EH)$  e se relacionam com a atribuição anterior pelo teorema de Bayes

$$P(C|EH) = P(C|H) \frac{P(E|CH)}{P(E|H)}. \quad (2.10)$$

O maior problema, na maioria das vezes, consiste em determinar os valores de  $P(C|H)$  — ou seja, os priors. Os priors, obviamente, não possuem um status especial no formalismo que os distinguem das demais atribuições de probabilidade, a dificuldade está em que, enquanto no caso dos condicionais temos uma articulação clara entre duas afirmações que permite elaborar argumentos do tipo “*se  $C$  está certo, então  $E$  blah, blah, blah*”, no caso dos priors o discurso se situa necessariamente nas hipóteses tácitas representadas por  $H$ . Desta forma é difícil conduzir argumentos convincentes que formalizem intuições como “*hmm, diria 95% de chance da teoria Bayesiana ser uma furada...*”.

Em várias situações, existem dois princípios que podem guiar uma determinação mais racional do prior: um é o teorema de Bernoulli, que serve no caso em que existe repetição de variáveis aleatórias e o outro é o Princípio da razão insuficiente, que, por exemplo, nos permite atribuir uma probabilidade de  $\frac{1}{6}$  para cada face de um dado comum. O problema é que nenhum destes princípios funcionam em variáveis contínuas. No caso do teorema de Bernoulli<sup>9</sup>, a dificuldade está em que o espaço de possíveis eventos é não-enumerável, já o caso do princípio de Laplace, é necessário que se responda à pergunta: *atribuição equiprovável em que coordenadas?*.

Ao apresentar o problema, Jaynes utiliza uma metáfora bastante adequada que um espaço contínuo é como um molusco escorregadio que, por mais que tentemos colocá-lo sobre uma estrutura rígida, sempre existe uma transformações de variáveis capaz

---

<sup>9</sup>A referência [?] propõe um formalismo muito interessante que pode, ao menos parcialmente, estender a identificação de frequências estatísticas com probabilidade em espaços contínuos. É claro que o uso do termo frequência seria enganador já que não há como definir a repetição de um mesmo evento num espaço contínuo. O autor se refere como “on-line learning” o processo em que, a partir de um conjunto de dados  $y_1, y_2, \dots, y_N$ , se determinaria a distribuição de probabilidade mais adequada na medida que o número de repetições aumenta. Note que o processo não envolve binagem ou recursos semelhantes.

de trasmutá-lo em qualquer outra forma concebível. Naturalmente precisamos de um jeito de domá-lo e colocá-lo em moldes fixos que eliminem esta arbitrariedade. A resposta para isto não é necessariamente resgatar uma parametrização em que o espaço amostral seja equiprovável — nenhuma prescrição será geral caso este espaço não seja determinado a partir de um limite óbvio a partir do discreto —, mas a resposta está em escrutinar as afirmações tácitas em  $H$  em busca de simetrias. Sobre que transformações de simetria  $P(x)$  deve ficar invariante? Por quais transformações poderíamos modificar as variáveis do problema sem alterar nenhum aspecto fundamental do mesmo?

Talvez seja apropriado recapitular a questão mais específica da física estatística em uma outra linguagem. O que nos leva a crer, de todas as possíveis parametrizações do espaço de fase, que a equiprobabilidade vale apenas em coordenadas canônicas? Clamar que estas coordenadas representam a maneira natural de descrever os processos físicos é uma falsa saída, mas dizer que outras escolhas eventualmente resultem em absurdos certamente é um passo na direção correta. Precisamos determinar precisamente que absurdos são estes e em que situações eles aparecem. Mais ainda, é necessário mostrar se tais imposições físicas selecionam uma única distribuição.

Para E. Wigner, assim como as condições iniciais são o como cimento e a areia para quem trabalha com leis da natureza, as próprias leis estariam em posição similar com relação aos princípios de simetria. Desta forma, o nosso modelo estatístico, além de respeitar qualquer restrição que eventualmente lhe imponha a mecânica, também, e antes de tudo, deve estar de acordo com os princípios de simetria que condicionam estas leis. Como diz Wigner, *“Uma lei da natureza (e aqui também incluímos um modelo estatístico) pode ser considerada válida somente se as correlações que ela postula são consistentes com princípios de invariância aceitos.”* Estes princípios são:

1. Deslocamentos temporais
2. Transformações espaciais
3. Mudança por movimento uniforme

O que Wigner quis dizer mais precisamente é que as leis da natureza expressam um encadeamento lógico entre eventos. Coisas do tipo: se ocorreram os eventos  $A$  e  $B$ , então  $C$  segue imediatamente ( $AB \Rightarrow C$ ). As leis de invariância expressam o fato que ao modificarmos  $A$ ,  $B$  e  $C$  por qualquer transformação (ou composição de transformações) acima, os eventos resultantes  $A'$ ,  $B'$  e  $C'$  obedecem necessariamente a

mesma relação  $A'B' \Rightarrow C'$ . Desta forma voltamos a nossa atenção às probabilidades mencionadas em (2.4), exigindo uma invariância se dê de tal forma que

$$P(F) = \sum_x P(xF) = \sum_x P(F|x)P(x) \quad (2.11)$$

$$= \sum_x P(x'F') = \sum_x P(F'|x')P(x'), \quad (2.12)$$

ou seja, as nossas conclusões e o grau de conhecimento sobre as afirmações  $F$ ,  $x$ ,  $F'$  e  $x'$  devem se manter os mesmos em qualquer situação. Mais ainda, queremos que o conhecimento anterior de diferentes observadores que situem nos sistemas de coordenadas  $S$  e  $S'$ , expresso respectivamente por  $P(x)$  ou  $P(x')$  seja consistente com a exigência imposta acima.

Considere dois observadores, o Sr. X num ponto do espaço representado por  $x$  e o Sr. Y em outro ponto representado por  $x'$ . Pedimos para que ambos expressem as probabilidades de onde irão detectar uma determinada partícula e, talvez demonstrando mais otimismo que justificaria as leis da física, ambos atribuem probabilidades Gaussianas centradas em suas respectivas origens. Como não há nada nas leis da física que distinguem sistema de referência, ou as inferências feitas pelos dois independem do ponto central da Gaussiana ou suas conclusões seriam inconsistentes entre si e com a física.

Precisamos de uma atribuição de probabilidade que, para qualquer observador, em qualquer sistema de referência válido, seja consistente com as previsões dos outros observadores. Neste sentido, a maneira de expressar o fato que a posição espacial de cada um dos distintos senhores X e Y não consiste numa informação privilegiada é garantir que o prior  $P(x)$  seja invariante pela transformação que leva o sistema  $S$  em  $S'$ . Ou seja, sendo  $P(x) = f(x)dx$  e  $P(x') = g(x')dx'$ , exigimos que

$$f(x)dx = g(x')dx', \quad (2.13)$$

já que o problema de inferência se apresenta como o mesmo nas coordenadas  $x$  e  $x'$ . Se  $f(x)dx \neq g(x')dx'$ , o estado de conhecimento sobre a posição da partícula discriminaria um sistema de coordenada  $S$  de outro  $S'$ . Em outras palavras, olhando apenas para as probabilidades teríamos informação sobre o sistema de coordenadas, o que é um absurdo já que  $P(x)$  descreve um estado de ignorância total sobre a posição da partícula.

Uma vez que se identifica as probabilidades  $P(x)$  e  $P(x')$ , é fácil determiná-la procurando pelas atribuições  $f(x)$  invariantes por translação. Dado que o Jacobiano

relacionado à esta operação é 1, ficamos com

$$f(x) = g(x - A) = f(x - A), \quad (2.14)$$

que obviamente só é satisfeito se  $f(x) = \text{cte}$ .

Uma vez de acordo com este exemplo, voltamos ao problema mais complicado que é determinar a probabilidade anterior para os estados microscópicos, ou seja,  $P(x)$ , onde  $x$  representa um sistema de  $N$  partículas dadas por posições  $q_1, \dots, q_N$  e velocidades  $\dot{q}_1, \dots, \dot{q}_N$ . O princípio de invariância acima não é o suficiente para determinar completamente  $P(x) = \int f(q, \dot{q}) dq d\dot{q}$  já que, considerando  $q$ 's e  $p$ 's cartesianos, implica nas equações diferenciais parciais

$$\sum_{i=1}^N \frac{\partial}{\partial q_i} f(q, \dot{q}) = 0; \quad \sum_{i=1}^N \frac{\partial}{\partial \dot{q}_i} f(q, \dot{q}) = 0, \quad (2.15)$$

que admitem múltiplas soluções. É lógico que a atribuição equiprovável não foi eliminada já que é uma solução trivial destas equações, mas até o momento não é a única solução compatível com a informação que a física nos fornece.

Claramente precisamos de um princípio adicional que restrinja um pouco mais o tipo de atribuição fisicamente aceitável. Desta maneira instituímos o princípio de democracia entre partículas:

#### 4 Permutação de partículas do mesmo tipo não altera a física.

Isto implica em certas restrições sobre a atribuição de probabilidades  $P(x)$  já que ela deve ser simétrica por permutações de partículas. A probabilidade reduzida à uma partícula, escrita como  $P(x_i) = \int f_1(x_i) dx_i$  deve ser a mesma para todas as partículas, ou seja

$$P(x_1) \equiv \sum_{x_2} \dots \sum_{x_N} P(x) = P(x_2) = \dots = P(x_N). \quad (2.16)$$

Podemos expressar isto, impondo que  $P(x)$  seja um funcional simétrico da probabilidade reduzida a uma partícula,

$$P(x) = \gamma [f_1(x_1), \dots, f_1(x_N)] dx_1 \dots dx_N, \quad (2.17)$$

com

$$\gamma [f_1(x_1), \dots, f_1(x_i), \dots, f_1(x_j), \dots, f_1(x_i), \dots, f_1(x_N)] = \gamma [f_1(x_1), \dots, f_1(x_j), \dots, f_1(x_i), \dots, f_1(x_N)]. \quad (2.18)$$

Desta forma substituímos o funcional  $P(x) = \gamma [f_1] dx_1 \dots dx_N$  na equação (2.15) para concluir que ou  $\gamma[f] = \text{cte}$  ou  $f_1(x_i) = \text{cte}$ . Dado que o argumento de invariância por transformação de sistema de coordenadas ainda se aplica às probabilidades reduzidas à 1-partículas (ou 2, ou 3, ...), optamos pela segunda escolha. Na realidade não importa, porque em tese, a atribuição anterior  $P(x)$  deve ser constante devido à sua dependência em  $x_1, x_2, \dots, x_N$  estar toda contida em  $f_1(x_i) = \text{cte}$ . Note que na medida que o número de partículas cresce, as frequências de ocupação de cada estado a 1-partícula tendem assintoticamente para  $f_1(x_i) dx_i$ , isto será importante para uma discussão adiante.

Apesar do argumento mostrar de maneira bem convincente que a atribuição equiprovável no espaço cartesiano de posições e velocidades consiste na melhor representação de “ignorância total” de um sistema físico, é fácil conceber modelos em que algum dos princípios de invariância mencionados acima possa ser violado. Não é que esperamos que sistemas reais possam efetivamente violá-los, o problema é que tais princípios de simetria são válidos se aplicados a um sistema fechado para o qual se conhece todos graus de liberdade e a forma de todas interações. Considere, por exemplo, um sistema simples como Sol-Terra. Ele obedece a uma mecânica em que rigorosamente valem os princípios de 1 a 3. Esconda metade destes graus de liberdade, digamos, o Sol e não é possível concluir as mesmas coisas: afaste a Terra do seu devido lugar e ela sairá de órbita; desloque sua posição no tempo e o Sol, percorrendo um arco na Galáxia, esquecerá do nosso planeta; ao mudar sua velocidade as conseqüências seriam igualmente desastrosas.

Existem duas maneiras em que estas simetrias se romperiam em um modelo estatístico e as duas envolvem o mesmo problema: informação. Se, por exemplo, existirem certos graus de liberdade escondidos como o sistema Sol-Terra mencionado anteriormente, a relevância destas simetrias (e a conseqüente equiprobabilidade) para os graus de liberdade conhecidos estaria sob suspeita. A boa notícia é que pequenas distorções em  $P(x)$  não afetam em nada as conseqüências macroscópicas (observáveis) da teoria<sup>10</sup>,

---

<sup>10</sup>E as vezes nem distorções grandes grandes produzem qualquer efeito. Tome um gás ideal clássico com  $10^{23}$  partículas como exemplo, de tal maneira que sua superfície de energia é uma  $(3 \times 10^{23} - 1)$  –esfera. Discretizamos este sub-espaco definindo, por exemplo, 100 células para cada ângulo  $[0 \dots \pi]$  necessário para especificar um ponto no mesmo. Selecione aleatoriamente 1 a cada, bem, digamos  $10^{10}$  pontos. Definimos um prior esdrúxulo que atribui uma probabilidade fixa para estes pontos e uma probabilidade nula para todos os outros; o que mudaria em nossas previsões macroscópicas? Absolutamente nada! A energia, obviamente continua a mesma, já que estamos restritos a um valor fixo para a mesma. A distribuição de Maxwell-Boltzmann, por exemplo, teria sua variância modificada em, no máximo  $m k_B T (10^{10^{23}})^{-1}$  sendo que os outros momentos estatísticos seguem relações semelhantes. Moral da história: a física estatística é bastante robusta quanto à escolha da probabilidade  $P(x)$ .

mas a má notícia é que não há garantias que as modificações seriam pequenas. Estas mesmas observações também valem para a segunda maneira de quebrar as simetrias e consiste na situação bastante comum em que trocamos um modelo de interação realístico por um idealizado (i.e.: sistemas ideais). Ainda que seja possível mostrar que a energia dos dois é parecida, os vínculos dinâmicos, as simetrias internas, a resposta a influências externas e outras informações potencialmente importantes não são as mesmas. O resultado pode ser que, ao fazer a substituição do sistema real pelo ideal, alguma espécie de compensação artificial seja necessária; esta compensação pode ser representada pela escolha apropriada de  $P(x)$ .

Em sistemas de alta dimensionalidade, geralmente pequenas complicações tendem a ser eliminadas à irrelevância estatística (é preciso distorcer as probabilidades de uma maneira muito substancial para que os resultados apareçam após tirar uma média sobre todos os graus de liberdade). Existem situações, no entanto, em que elas se manifestam, mas não seria correto tratar estes casos sob suspeita: cada violação da equiprobabilidade é uma oportunidade para descobrir uma influência física muito relevante (tão relevante que aparece após o processo de média macroscópica); pode ser um novo vínculo, novos graus de liberdade ou qualquer outra coisa. Mesmo a física estatística se apresentando equiprovável no caso ideal, formulá-la segundo esta hipótese é tanto uma violência ao seu domínio de aplicabilidade, mas principalmente uma exigência indefensável sobre o tipo de informação que dispomos em situações reais. Neste espírito, seguimos uma visão mais ampla em que as simetria 1 a 3 podem ser violadas, mesmo encarando a situação ideal como o restabelecimento das mesmas. Na realidade, a questão se um modelo equiprovável é superior ou não a um não-equiprovável sequer pode ser colocada se a teoria trata apenas da primeira situação.

### 2.1.4 Teoria ergódica

Esta talvez seja uma história muito contada em física estatística<sup>11</sup>. As medidas macroscópicas levam um certo tempo para serem processadas, um tempo que é consideravelmente

---

<sup>11</sup>Ainda que o formalismo ergódico é normalmente identificado como a “teoria ortodoxa” de física estatística, uma consulta aos livros didáticos certamente não daria esta impressão. Dos livros consultados pelo autor, as referências [?, ?] se posicionam explicitamente contra a relevância da hipótese ergódica, enquanto [?] também dá a entender que além de ser sabidamente falsa, não há porque insistir no argumento, mas apenas em [?] vemos a “história padrão”. Pode se argumentar que Gibbs [?] também tomaria uma posição contrária à ortodoxa. Aliás, a não ser pelo formalismo que é comum, a estrutura lógica das argumentações, a eleição dos conceitos chave, as posições defendidas e atacadas pelos diversos autores raramente são as mesmas; de tal modo, o uso do termo *ortodoxia* referindo à física estatística talvez seja um tanto quanto vago.

longo com relação aos movimentos microscópicos. Deste modo, uma grandeza macroscópica expressaria nada mais que a média de uma função  $\mathcal{F}(x)$  ao longo da trajetória percorrida pelo sistema e, justificados pela observação anterior, seria legítimo passar o limite de integração para um tempo infinito. No que continua o argumento, existiria o *teorema ergódico* que garante que estas médias temporais convergem assintoticamente às médias de ensemble, basta que o intervalo de integração tenda ao infinito. Sem antes notar que existem usos legítimos para teorias ergódicas, e que a própria demonstração de ergodicidade exibe um interesse matemático por si só, expomos duas linhas de ataque ao seu uso como fundamento para a física estatística: uma porque ela não é necessária e a outra pelo fato que se baseia em expectativas absurdas [?]. Começamos pela segunda que é a mais simples.

A idéia por trás do teorema ergódico é que após um período de tempo muito grande, a folha de energia associada a um sistema no seu espaço de fase seria densamente preenchida e, portanto, todos pontos se veriam representados no ensemble. Para que isso aconteça, no entanto, é necessário que o sistema passe pelo menos uma vez por cada ponto, sob o risco de deixar sub-representado algum estado importante. O tempo necessário para que isso aconteça, naturalmente, é o tempo de recorrência de Poincaré. K. Huang elegantemente fecha a questão observando que “isto não tem nada a ver com física”<sup>12</sup>. Na percepção do autor, ir além deste ponto não seria sequer necessário.

Antes de proceder com a próxima linha de argumentação é interessante passar por um segundo ponto. É bem sabido que um sistema integrável não pode ser ergódico (no sentido expresso acima). A existência de invariantes adicionais quebraria a indecomposibilidade métrica do espaço de fase, impedindo que toda a superfície de energia pudesse ser visitada. Argumentos para contornar este fato geralmente se baseiam na suposta propriedade que, à medida que o número de partículas tende a infinito, a contribuição efetiva destes invariantes se associaria a um subespaço de medida nula. Verdade ou não, existem invariantes perfeitamente controláveis e mensuráveis que nunca se tornarão

---

<sup>12</sup>Para se ter uma noção de quão longo deveria se esperar considere um gás uniformemente distribuído pelo volume de um compartimento arbitrário. O tempo necessário para o sistema percorrer todos estados uniformemente depende, em grande maneira, da resolução utilizada para contar ocorrências repetidas do mesmo estado. Desta forma, fazendo o esquadramento grosseiro que distingue um estado do outro somente por uma das moléculas estar numa metade do compartimento diferente do outro estado, a afirmação “percorrer uniformemente todos os estados” ganha o caráter de “realizar todas as permutações dois a dois com  $10^{23}$  moléculas”. Existiriam, então,  $2^{10^{23}}$  estados a serem percorridos. Considere que a frequência de transição de um estado para outro se relaciona com o tempo  $T$  que cada molécula leva para atravessar o compartimento pela fórmula  $\nu = 10^{23}T^{-1}$ . O tempo necessário para percorrer pelo menos uma vez todos os estados seria maior que  $T \times 2^{10^{23}} \times 10^{-23}$ . Para todos efeitos práticos e todas ordens de valores aceitáveis para  $T$ , este tempo é uma eternidade!



irrelevantes por qualquer processo de limite  $N \rightarrow \infty$ : são os momentos lineares e angulares. O fato que podemos considerá-los ou não em nossos ensembles parece uma constatação ainda mais forte que a ergodicidade está sendo direcionada para um problema errado.

O apelo físico da teoria ergódica é que ela prescreve os ensembles de uma maneira que, efetivamente, não faz uso do conceito de probabilidade. O autor não poderia, nem sequer gostaria de fundamentar melhor esta idéia já que é curioso verificar este tipo de expectativa num ramo da ciência que se denomina *física estatística*. A primeira questão que se coloca é o que fazer com estes ensembles. Se uma resposta tentativa for *estatística de não-equilíbrio*, imediatamente descartaríamos a prescrição ergódica já que ela sequer permite formular a noção de tempo no ensemble. Se a motivação se situa exclusivamente no equilíbrio, avaliamos até que ponto a prescrição ergódica de fato elimina o uso de probabilidades.

Considere um sistema completamente integrável descrito em suas variáveis de ângulo e ação. A Hamiltoniana, escrita como  $H(J)$ , implica não só na invariância das variáveis  $J_i$  como numa relação linear entre os  $\theta_i \in [0 \dots 2\pi]$  com o tempo,  $\theta_i = \omega_i(t - t_0)$ ;  $\omega_i = \frac{\partial H}{\partial J_i}$ . Podemos especificar as coordenadas  $\theta_i$  pela simples especificação de  $t$ . Neste sentido, perguntar “qual a probabilidade de observar um certo valor de  $\theta_i$ ” se traduz imediatamente em “qual a probabilidade que a observação se dê em  $\frac{\theta_i}{\omega_i} + t_0$ ”. E assim é totalmente lícito substituir *média estatística* por *média temporal*.

É claro que a situação anterior pressupõe o conhecimento das variáveis  $J_i$  e  $t_0$ , o que claramente é inviável em sistemas de alta dimensionalidade. Para contornar esta dificuldade, postula-se a hipótese ergódica para a qual “quase todas” trajetórias percorrem “quase uniformemente” a folha de energia de um sistema — o que claramente não se aplica a sistemas completamente integráveis. Esta hipótese claramente é desnecessária para calcular  $P(\theta_i)$ , já que a questão da ignorância do estado macroscópico é sempre colocada ou em  $P(x)$  ou em  $P(t)$ , mas permite utilizar o raciocínio anterior que traduz ignorância em  $\theta_i$  por ignorância com relação ao tempo decorrido desde o início da série temporal. É claro que este esquema falha miseravelmente quando a pergunta feita é sobre “a probabilidade de  $\theta_i$  no instante  $t$ ”.

Isto nos coloca em uma situação peculiar. Neste trabalho tentamos fazer a menor distinção possível entre “termodinâmica de equilíbrio” e “termodinâmica de não-equilíbrio”, sendo a primeira um caso especial da segunda. Podemos abdicar completamente desta pretensão e validar os ensembles fornecidos pela prescrição ergódica; a motivação,

que devo insistir é bastante escusa, é eliminar as probabilidades da física estatística. Ergodicidade talvez seja o suficiente para prescrever os ensembles de equilíbrio, e como já foi mencionado anteriormente, é possível conceber ensembles bastante esdrúxulos que geram os resultados de equilíbrio de maneira igualmente satisfatória. A teoria de não-equilíbrio seria um grande mistério já que a principal ferramenta disponível atualmente (a equação de Liouville-von Neumann) lida com a evolução de *probabilidades*. De maneira geral, deveríamos nos sentir profundamente constrangidos se a física estatística necessitasse da noção de ergodicidade como fundamento. Felizmente este não é o caso.

### 2.1.5 Aproximação mecânica

Nesta seção nos voltamos à questão de determinar a dinâmica associada às variáveis macroscópicas. As regras de manipulação de probabilidades permitem decompor a probabilidade associada a um caminho nas probabilidades de transições entre sub-caminhos menores. Isto é feito em

$$P(F_f \dots F_1 | F_0) = P(F_1 | F_0) P(F_2 | F_1 F_0) \dots P(F_f | F_{f-1} \dots F_0). \quad (2.19)$$

Daí é fácil ver a relevância das transições condicionais  $P(F_{i+1} | F_i \dots F_0)$ . Calculamos estas probabilidades que, após alguma álgebra, resultam em

$$P(F_{i+1} | F_i \dots F_1 F_0) = \sum_{x_0} P(x_0) \frac{P(F_0 | x_0) P(F_1 | x_0) \dots P(F_i | x_0) P(F_{i+1} | x_0)}{P(F_1 \dots F_{i+1} | F_0) P(F_0)}. \quad (2.20)$$

Para a maioria dos casos, estas probabilidades são tão difíceis de se calcular quanto (2.9), mas em algumas situações é possível fazer a aproximação que o sistema evolui de maneira Markoviana, ou seja, que

$$P(F_f \dots F_1 | F_0) \simeq P(F_1 | F_0) P(F_2 | F_1) \dots P(F_f | F_{f-1}). \quad (2.21)$$

Uma condição para que isto seja válido, isto é, que  $P(F_{i+1} | F_i \dots F_0) = P(F_{i+1} | F_i)$ , pode ser expressa pela exigência que para “quase todas” trajetórias microscópicas associadas à uma determinada evolução macroscópica  $F_0, F_1, \dots, F_i$  que começam em  $x_0$ , valeria a relação

$$P(F_1 | x_0) \dots P(F_i | x_0) \simeq P(F_1 \dots F_{i+1} | F_0). \quad (2.22)$$

A parte associada à  $P(F_1 | x_0) \dots P(F_{i-1} | x_0)$  simplesmente seleciona a trajetória macroscópica compatível com o estado microscópico  $x_0$ . Sendo assim, elas são avaliadas em 0 se for  $x_0$  for incompatível com tal evolução e em 1, caso contrário. O termo do lado direito da

igualdade consiste na probabilidade que a evolução ocorre por esta trajetória parcial ao começar em  $F_0$ . A noção por trás da condição (2.22) é que existe uma trajetória macroscópica altamente provável  $P(\bar{F}_f \dots \bar{F}_1 | F_0) \simeq 1$ , de forma a manter a consistência entre o lado direito e lado esquerdo da equação (que só pode ser 0 ou 1). Este resultado está por trás de vários desenvolvimentos que faremos de agora em diante.

Na situação em que o sistema macroscópico obedece à aproximação (2.22) e uma vez de acordo que a propriedade  $P(\bar{F}_f \dots \bar{F}_1 | F_0) \simeq 1$  elimina as incertezas do problema, podemos dizer que, em um sentido específico, a condição inicial  $F_0$  *determina* a evolução macroscópica posterior. Em analogia ao que se passa com a física fundamental (equações diferenciais para a evolução dos estados), nos referimos à aproximação feita em (2.22) como *aproximação mecânica* e os sistemas para os quais a aproximação funciona são chamados *sistemas simples*.

Todo este raciocínio resolve uma parte do problema da atribuição de probabilidades aos estados macroscópicos. Para que a atribuição  $P(F)$  seja inteiramente especificada, é necessário fornecer  $P(x)$  explicitamente e este, talvez, seja o ponto de divergência conceitual mais radical entre a abordagem aqui exposta e a abordagem tradicional via teoria ergódica como foi discutido nas seções anteriores.

## 2.2 Irreversibilidade

No speculation, no body of knowledge ever claimed the equivalence between doing and undoing, between a plant that grows, has flowers and dies, and a plant that resuscitates, becomes younger and goes back to its primitive seed, between a man who learns and become mature and a man who becomes progressively a child, then an embryo, and finally a cell. Yet, since its origins, dynamics, the physical theory that identifies itself with the triumph of science, implied this radical negation of time.

(ILYA PRIGOGINE)

Irreversibilidade macroscópica é um tema capaz de acirrar os ânimos em certos círculos especializados. A disputa reside no fato que as leis da dinâmica microscópica são simétricas por inversão temporal, enquanto o comportamento de um conjunto muito grande de partículas demonstra várias características de evolução irreversível. Seres humanos nascem, vivem e morrem, mas nunca se observou o movimento oposto. Uma vez que as leis fundamentais são simétricas por inversão temporal, cada uma das duas formas de evolução é possível se e somente se a outra também o for, mas ainda

assim, somente uma delas é observada. Se as leis fundamentais não exibem preferência de direção numa linha temporal, porque a assimetria entre passado e futuro é observada em praticamente todos os processos de larga escala?

Considere o exemplo da assimetria entre uma evolução natural, nascimento  $\triangleright$  vida  $\triangleright$  morte, e o processo inverso, dificilmente observado. Seres humanos, assim como todos seres vivos, são configurações de átomos tão específicas que, sem oferecer indicações adicionais, dificilmente poderiam ser consideradas sequer razoavelmente plausíveis. Existe uma quantidade muito maior de modos de se distribuir as moléculas que formam uma pessoa em um sopa amorfa de elementos químicos que em pessoas de carne e osso, basta notar que a seqüência de fatos que levou ao nascimento de qualquer ser humano, inclusive o leitor, consiste em um encadeamento de eventos improváveis que, se não fossem considerados contextualmente, seriam o suficiente para se dizer que um ser humano é tão improvável que, racionalmente, jamais se esperaria que seres humanos existam!

A simetria de inversão temporal das equações fundamentais implica, de fato, que para cada configuração microscópica associada, por exemplo, ao movimento de nascimento  $\triangleright$  vida  $\triangleright$  morte, existe uma configuração que realiza precisamente o movimento inverso, basta que se isole completamente o sistema tratado<sup>13</sup>. Uma conclusão errônea, no entanto, é que a probabilidade de verificar a evolução inversa ou a evolução direta sejam *incondicionalmente* as mesmas (poderia se pensar isto invocando as probabilidades (2.4) da seção anterior). Este tipo de conclusão ignora o papel das condições iniciais na atribuição de probabilidades e seria como perguntar “Em qualquer época e independente da situação geopolítica do planeta, qual é a probabilidade que aconteça Hiroshima?” e esperar obter uma resposta conclusiva.

O fato que seres humanos ( $H$ ) representam configurações de moléculas muito específicas é expresso numericamente por  $P(H) = \sum_x P(x)P(H|x) \ll 1$ . A probabilidade que um ser humano nasça, viva, e morra, estaria associada ao número de trajetórias microscópicas que implementam um estado de “ser humano” em todos estes instantes. É claro que este número, em comparação a todas trajetórias possíveis é, na melhor das hipóteses, da ordem de  $P(H)$  — ou seja, incrivelmente pequeno. Do mesmo modo concluiríamos sobre a probabilidade do processo inverso, não-natural e certamente macabro em que

---

<sup>13</sup>É lógico que dificilmente seres vivos podem ser considerados um sistema isolado (pelo menos não por muito tempo). Para ser um pouco mais honestos seria necessário considerar que o ambiente que sustenta a vida do ser humano em questão seja contemplado em sua descrição microscópica. Isto, no entanto, não altera significativamente as conclusões.

se ressuscita, rejuvenesce, finalmente regride a um feto e desaparece. Ao colocar o problema nestes termos, a surpresa não consiste em que nunca se observou o processo funesto que leva da morte ao nascimento — esta configuração é, de fato, incrivelmente improvável — a maior surpresa no que diz respeito ao bom senso (ou as probabilidades) é o fato que algum ser humano jamais tenha existido, quem diria os atuais *6bi* deles!

É possível argumentar que a comparação não foi inteiramente justa. Quando se fala no processo nascimento, vida e morte, dificilmente alguém se referiria ao processo que cria um feto humano da matéria bruta e providencie os recursos necessários para o seu posterior amadurecimento, vida e morte, tudo isso como num lapso espontâneo da matéria. Este é justamente o ponto aonde se queria chegar! Contextualizando cada forma de evolução, ou seja, fazendo suposições a respeito das condições iniciais, as probabilidades podem se alterar significativamente. O que num contexto pareceria impossível (o aparecimento de seres humanos a partir da matéria bruta), em outro contexto seria perfeitamente plausível, e até mesmo provável (o aparecimento de um novo rebento de um casal de jovens apaixonados).

Existe um detalhe, no entanto, que é essencial para explicar a assimetria temporal associada a processos macroscópicos. Nem sempre (aliás, quase nunca) o conhecimento sobre estado inicial seleciona os mesmos estados microscópicos associados ao conhecimento sobre o estado final. Esta observação é importante porque a probabilidade destes estados, como se mostrou anteriormente, corresponde à fração dos microestados da condição inicial que atingem uma certa configuração final no instante  $t_f$ . De maneira semelhante, trocando a ordem temporal dos estados, muda-se o conjunto de estados pré-selecionados pela nova condição inicial (que é a condição final do primeiro problema) e a fração dos estados que atingem os estado final (correspondente ao estado inicial do primeiro problema) pode ser substancialmente diferente. Esta explicação fica mais clara em termos matemáticos; considere um processo arbitrário descrito pelo estado macroscópicos inicial  $I$  em  $t_0$  e pelo final  $F$  em  $t_1$ . Queremos mostrar que as probabilidades da evolução inversa e direta não são necessariamente as mesmas, ou seja

$$P(F_1|I_0) \neq P(I_1|F_0). \quad (2.23)$$

A avaliação explícita destas probabilidades ajuda a determinar a direção mais natural

para a transição

$$P(F_1|I_0) = \frac{1}{P(I_0)} \sum_{x_0} P(x_0)P(I_0|x_0)P(F_1|x_0) \quad (2.24)$$

$$= \frac{1}{P(I)} \sum_{x_0} P(x_0)\delta(I - \mathcal{F}(x_0))\delta(F - \mathcal{F}(\mathcal{U}(t_1) \cdot x_0)), \quad (2.25)$$

$$P(I_1|F_0) = \frac{1}{P(F_0)} \sum_{x_0} P(x_0)P(I_1|x_0)P(F_0|x_0) \quad (2.26)$$

$$= \frac{1}{P(F)} \sum_{x_0} P(x_0)\delta(F - \mathcal{F}(x_0))\delta(I - \mathcal{F}(\mathcal{U}(t_1) \cdot x_0)). \quad (2.27)$$

O fator com o somatório em ambas atribuições é o mesmo já que, devido à reversibilidade das equações microscópicas, para cada trajetória contabilizada na primeira avaliação (trajetórias que partem de  $I$  e chegam à  $F$ ), deve existir outra contabilizada na segunda (trajetória que parte de  $F$  e chega a  $I$ ). Desta forma é possível relacionar a probabilidade reversa com a direta pela equação

$$P(F_1|I_0) = \frac{P(F)}{P(I)}P(I_1|F_0). \quad (2.28)$$

Se o número de estados microscópicos associados ao estado final for muito maior que o número de estados associados ao estado inicial, e excluindo a possibilidade que a transição seja impossível, a primeira forma de evolução seria muito mais provável que a segunda de sorte que *conquanto se baseie numa descrição puramente macroscópica existe, de fato, uma seta do tempo* [?].

## 2.3 Entropia e termodinâmica

In contrast to the specificity of mechanics and electromagnetism, the hallmark of thermodynamics is generality. Generality first in the sense that thermodynamics applies to all types of systems in macroscopic aggregation, and second in the sense that thermodynamics does not predict specific numerical values for observables quantities. Instead, thermodynamics sets limits (*inequalities*) on permissible physical processes, and it establishes relationships among apparently unrelated properties.

(H. CALLEN - Thermodynamics and an introduction to thermostatics)

Chacoalhe a água de uma piscina; o movimento se propaga em ondas que, eventualmente, desaparecem sem deixar qualquer resquício da perturbação inicial. De maneira similar, coloque dois objetos com temperaturas diferentes em contato e ambos termalizarão; misture sal na água e ele se dissolverá; arremesse uma bola e eventualmente o atrito

cessará seu movimento. A natureza é cheia de exemplos de situações em que sistemas físicos em condições iniciais diversas evoluem para estados finais estacionários e bastante previsíveis. A estes estados macroscópicos associamos a condição de equilíbrio termodinâmico.

Tal constatação está em contraste patente com o que se conhece da dinâmica microscópica. O conhecimento sobre estado microscópico em qualquer instante determina tanto o passado como o futuro, sendo que nenhum estado é privilegiado no sentido que as outras condições convergem para ele. Já numa perspectiva macroscópica, a irreversibilidade do movimento é bastante fortuita; uma vez identificados, os estados de equilíbrio fornecem uma boa dica sobre o destino dinâmico dos processos físicos: *sistemas macroscópicos evoluem para o equilíbrio*.

Infelizmente esta constatação fenomenológica não é uma consequência geral das leis de movimento, já que nem sempre os termos em que  $P(F_1|F_0) \simeq 1$  convergem para um estado fixo quando  $t_1 \gg t_0$ . Uma análise detalhada, assim como qualquer cálculo microscópico em sistemas de alta dimensionalidade, envolve dificuldades técnicas proibitivas. O que fazemos, então, é inverter o argumento: assumindo que uma classe de sistemas macroscópicos evoluem para o equilíbrio, qual seria a maneira mais simples de caracterizá-lo? como se explicaria tal evolução? O método mais simples para se determinar estes estados, mas que de maneira alguma pode clamar por universalidade, é mediado pelo conceito de *entropia termodinâmica*. Voltamos a atenção para sistemas para os quais seja pertinente formular a seguinte questão: Uma vez que se impõe um certo conjunto de restrições macroscópicas, como se determinaria o estado de equilíbrio para o qual o sistema físico invariavelmente evolui?

Sob o rótulo de *sistemas termodinâmicos*, colocamos os fenômenos que permitam implementar um algoritmo que resolva a questão anterior inequivocamente. Mais precisamente, procura-se uma prescrição para calcular

$$\text{restrições macroscópicas} \Rightarrow \text{estado de equilíbrio.}$$

A formulação do problema nestes termos tão logo exclui a possibilidade que o estado de equilíbrio dependa significativamente de características específicas da trajetória que levou a ele. Isto, é claro, restringe o escopo da teoria às situações em que o equilíbrio pode ser caracterizado de maneira fácil, utilizando-se apenas da informação macroscópica disponível no instante em que é atingido (a história pregressa não possui qualquer papel). Alguns fenômenos não admitem uma abordagem tão simplista, para citar apenas um caso, citemos os sistemas biológicos: fora a morte, como diria o pessimista,

não há um estado de equilíbrio para o qual seres vivos “invariavelmente evoluem”. Ciclos de histerese, processamento algorítmico, sistemas biológicos, sociedades complexas, todos são fenômenos que desafiam a abordagem simplista implementada pela termodinâmica (ainda que em alguns casos ela possa fornecer indicações parciais). A caracterização do equilíbrio nestes casos *não* é (necessariamente) mediada pelo conceito de entropia de tal maneira que o formalismo termodinâmico é, se não errôneo, mas insuficiente.

Falamos em *termodinâmica*, mas nenhuma menção se fez aos conceitos de calor, trabalho e temperatura que, historicamente, foram determinantes da especificação desta área do conhecimento. Aqui se entende termodinâmica como uma certa metodologia de inferência estatística sobre o comportamento macroscópico de sistemas simples tomados em tempos distantes<sup>14</sup>. E assim como diz a citação no início da seção, a termodinâmica estabelece restrições e desigualdades, além disso seleciona, dentre os vários processos aparentemente permitidos pela mecânica, apenas aqueles que se situam num patamar de plausibilidade.

### 2.3.1 Seta do tempo e entropia

A primeira indicação significativa sobre a direção natural dos processos físicos vem da expressão (2.28) que seleciona, de um par de estados, aquele com maior chance de constituir um estado final ou inicial. Tomar isto como prescrição para encontrar estados de equilíbrio, no entanto, é bastante inadequado já que a seta temporal não diz se um processo é de fato provável, mas apenas compara a probabilidade de ele aconteça com o processo temporalmente revertido; na maioria das situações, tanto a evolução direta quanto a inversa podem ser altamente improváveis.

O caráter assimétrico da evolução temporal implica que, se o equilíbrio for atingido, estes estados, que denotamos por  $E$ , estariam associados a probabilidades de transição muito altas partindo de uma classe específica de condições iniciais  $\{I\}$  — em números, teríamos  $P(E_T|I_0) \simeq 1$ . Ao mesmo tempo exige-se que  $P(I_T|E_0) \simeq 0$ , caso contrário seria possível voltar à condição inicial a partir da condição de equilíbrio, o que representaria um regime cíclico. Estas duas exigências, junto com (2.28), implicam que os estados de equilíbrio estão associados a uma representatividade  $P(E)$  muito maior que os estados iniciais  $P(I)$ .

---

<sup>14</sup>Estes intervalos de tempo são longos em comparação ao chamado tempo de relaxamento. De acordo com o sistema ou o fenômeno estudado, o tempo de relaxamento pode ser curto ou longo comparado aos tempos característicos da paciência humana.



Pelo menos em teoria, a existência de memória de longo prazo pode alterar estas conclusões. Considere que, para um dado estado inicial  $I$  e um estado final  $F$ , tenhamos  $P(F_T|I_0) \simeq 1$  e  $P(I_T|F_0) \simeq 0$ , indicando um processo de aproximação do equilíbrio. Se a memória joga uma papel fundamental, a probabilidade relativa ao conhecimento de uma situação pregressa a  $I_0$  ( $P_{-1}$ ) pode alterar significativamente a avaliação de forma que,  $P(F_T|I_0P_{-1}) \neq P(F_T|I_0)$ . Neste caso, uma simples análise das probabilidades relativas aos estados final e inicial não é totalmente conclusiva já que informação pregressa sempre pode alterar estas conclusões. De agora em diante nos restringimos, por uma questão pragmática, aos sistemas sem memória em que funcione a aproximação mecânica (seção 2.1.5).

Motivado pelo fato que os estados de equilíbrio estão associados à representatividades muito maiores que as condições iniciais, avaliemos como se dá a evolução da representatividade nos sistemas simples. Identificamos a trajetória estatisticamente predominante com a trajetória física denotada por  $F(t)$ . A representatividade dos estados nesta trajetória aumenta monotonicamente no tempo e a demonstração deste resultado é simples. Considere dois estados, o anterior  $A$  e o posterior  $B$ , tirados arbitrariamente de  $F(t)$ . Dado que a transição entre quaisquer dois estados em  $F(t)$  é extremamente provável temos que

$$P(B_f|A_0) \simeq 1 = \frac{P(B)}{P(A)} P(A_f|B_0) \leq \frac{P(B)}{P(A)}. \quad (2.29)$$

Numa escala logarítmica, a expressão resulta em

$$\ln P(B) \geq \ln P(A) \Rightarrow P(B) \geq P(A) \quad (2.30)$$

indicando que na trajetória física, a representatividade é monotonicamente não-decrescente. A mesma análise pode ser feita para o crescimento da representatividade sobre uma seqüência de estados arbitrária. Utilizamos a notação  $F^{(t)} = F(t)$  para indicar cada um destes estados tomados em tempos distintos. Desta maneira, temos que

$$P\left(F_f^{(f)} \dots F_1^{(1)} | F_0^{(0)}\right) \simeq 1 = P\left(F_f^{(f)} | F_{f-1}^{(f-1)}\right) \dots P\left(F_2^{(2)} | F_1^{(1)}\right) P\left(F_1^{(1)} | F_0^{(0)}\right) \quad (2.31)$$

$$= \frac{P\left(F^{(f)}\right)}{P\left(F^{(f)}\right)} P\left(F_f^{(f-1)} | F_{f-1}^{(f)}\right) \dots \frac{P\left(F^{(1)}\right)}{P\left(F^{(0)}\right)} P\left(F_1^{(0)} | F_0^{(1)}\right) \quad (2.32)$$

$$\leq \frac{P\left(F^{(f)}\right)}{P\left(F^{(f)}\right)} \dots \frac{P\left(F^{(2)}\right)}{P\left(F^{(1)}\right)} \frac{P\left(F^{(1)}\right)}{P\left(F^{(0)}\right)} = \frac{P\left(F^{(f)}\right)}{P\left(F^{(0)}\right)}. \quad (2.33)$$

O uso do logaritmo como escala para a representatividade é bastante conveniente para analisar como esta probabilidade se relaciona com as probabilidades de transição

de cada sub-caminho; em fórmulas temos

$$\ln P(F^{(f)}) - \ln P(F^{(0)}) = \sum_{i=1}^f [\ln P(F^{(i)}) - \ln P(F^{(i-1)})] \geq 0. \quad (2.34)$$

Neste ponto é conveniente introduzir uma função especial para representar o logaritmo da representatividade

$$S(F) \equiv k_B \ln P(F). \quad (2.35)$$

A esta grandeza damos o nome de *entropia termodinâmica*, e equivale à fórmula gravada no túmulo de Boltzmann em Zentralfriedhof, Viena (em contraste com a entropia de Gibbs que não é necessariamente igual a este valor). O uso do adjetivo *termodinâmico* serve para contrastar com a entropia informacional definida anteriormente; num quadro geral não existe conexão lógica nem formal entre os dois conceitos<sup>15</sup>. Enfatizamos que a termodinâmica se baseia no raciocínio plausível e o uso de uma designação especial para o logaritmo da probabilidade, e todas as relações que virão, não reduz em maneira nenhuma o seu conteúdo estatístico.

Usando a notação de entropia, é possível escrever a fórmula (2.34) simplesmente como

$$S_f - S_0 = \sum_{i=1}^f \delta S_i \geq 0, \quad (2.36)$$

onde se define  $S_i \equiv S(F^{(i)})$ . Isto indica que nos processos irreversíveis aqui considerados, a entropia sempre cresce. O uso de uma escala de representatividade logarítmica (a qual chamamos entropia) permite lidar com as multiplicações associadas à composição de probabilidades como simples somas. Afora a facilidade de cálculo, é possível generalizar o conceito para uma evolução contínua da trajetória  $F(t)$  já que é possível trocar a soma em (2.36) por uma integral. Veremos posteriormente, que variações de entropia podem ser determinadas a partir de experimentos macroscópicos simples, enquanto que tendemos ver a representatividade mais associada à descrição microscópica disponível. Neste sentido, a lei de aumento da entropia, que historicamente foi atribuída um caráter independente da mecânica, com Boltzmann adquire uma interpretação bastante razoável: isto é, estando de acordo com o modelo mecânico simplificado, a diminuição da representatividade microscópica (entropia) é simplesmente muito improvável.

Antes de se convencer totalmente disto, é necessário responder a algumas questões

---

<sup>15</sup>A entropia informacional é formalmente idêntica à entropia de Gibbs. Como mostraremos posteriormente, a entropia de Gibbs normalmente é *semelhante* à entropia de Boltzmann como foi aqui definida, especialmente para sistemas de muitas partículas fora dos regimes críticos — em especial, elas coincidem quando se trata de um sistema de partículas livres.

fundamentais. A primeira, que discutimos na seção 3.1, é se existem, de fato, tais sistemas simples — colocando melhor a questão, queremos saber se existe algum conjunto de variáveis macroscópicas que, para alguns sistemas, forneça uma descrição do tipo mecânica. A segunda, a ser discutida imediatamente, é se podemos determinar os estados de equilíbrio (máxima entropia) sem conhecer os detalhes da dinâmica macroscópica.

### 2.3.2 Determinação dos estados de equilíbrio

Uma vez detectado o estado de máxima representatividade,  $\bar{F}$ , é necessário determinar que razões fariam com que uma condição inicial,  $I$ , não evoluísse para o mesmo. Se, por exemplo, todos estados microscópicos nesta condição inicial possuírem um valor de energia diferente dos estados associados a  $\bar{F}$ , a probabilidade que a evolução se desse de  $I_0$  para  $\bar{F}_\infty$  seria rigorosamente *nula*, já que implicaria na violação de um invariante de movimento. As restrições ao movimento macroscópico são, portanto, outro elemento importante para caracterizar o equilíbrio.

Neste ponto é importante distinguir entre invariantes de movimento microscópicos e invariantes da dinâmica macroscópica. Ainda que existam situações em que os dois conceitos coincidam — por exemplo, a energia é um invariante microscópico assim como macroscópico —, os únicos invariantes microscópicos relevantes na caracterização do equilíbrio são aqueles que podem ser escritos diretamente *a partir* das variáveis macroscópicas. Como invariante macroscópico entendemos qualquer função  $\Gamma(F)$  para a qual a *todas* probabilidades de transições que impliquem na variação de  $\Gamma$  sejam nulas ou muito próximas de zero. Note que mesmo conhecendo os invariantes microscópicos, pode ser muito difícil determinar os invariantes macroscópicos. Geralmente apela-se para procedimentos heurísticos e o mais simples deles é considerar que a energia consiste no único invariante macroscopicamente relevante.

Supondo que seja possível levar em conta toda esta informação, o procedimento para encontrar os estados de equilíbrio se baseia em procurar o estado de maior probabilidade dentro de uma classe de invariantes macroscópicos associados à condição inicial  $I_0$ . Escrevendo em uma notação mais adequada, isto implica na busca pelo estado  $E$  tal que

$$P(E_T|I_0) = P(E_T|\Gamma(I_0)I_0) \simeq 1, \quad (2.37)$$

onde  $\Gamma(F)$  denota o valor dos invariantes macroscópicos associados a  $I_0$ .

A motivação inicial para desenvolver o formalismo termodinâmico era, dado um

conjunto de vínculos, determinar o estado de equilíbrio para o qual o sistema físico invariavelmente evolui. Isto pode ser expresso na notação acima como se a probabilidade  $P(E_T|I_0)$  dependesse de  $I_0$  apenas pelo valor dos invariantes macroscópicos. Sendo assim, o estado de equilíbrio para um problema de valor inicial é o que garante que

$$P(E_T|\Gamma) \simeq 1. \quad (2.38)$$

A prescrição para determinar  $E$  é a maximização de  $P(E_T|\Gamma)$  ou, alternativamente, a maximização de  $\ln P(E)$  restrita pelos valores dos vínculos. Dada a relação entre a entropia e a representatividade dos estados, esta prescrição equivale imediatamente à prescrição mais familiar de maximização da entropia.

## 3 *Estrutura dinâmica*

### 3.1 Distribuição número de partículas

Com o intuito de formular a dinâmica de  $F_t$  devemos procurar por um conjunto de variáveis macroscópicas para as quais se possa utilizar a aproximação mecânica. Neste espírito, analisamos a representação de distribuição dos números de partículas (DNP) que especifica o número de partículas em cada estado para uma dada configuração microscópica. Considere o movimento de uma partícula clássica denotada por  $i$  imersa num material constituído por partículas do mesmo tipo. As suas leis de movimento são determinadas pelas equações

$$\frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial q_i} = -\frac{\partial}{\partial q_i} \sum_{j=1}^N V(q_i, q_j) \quad (3.1)$$

$$\frac{\partial q_i}{\partial t} = \frac{\partial H}{\partial p_i} = \frac{1}{m} p_i. \quad (3.2)$$

Outras partículas que, neste sistema, possuam a mesma posição e momento obedecem a equações de movimento virtualmente idênticas.

Permutação entre as partículas em um mesmo estado não altera as suas respectivas equações de movimento. Assim, do ponto de vista dinâmico, não interessa que partícula possui qual designação (um *label*  $i, j$ , etc); mas interessa o número delas presentes em cada estado. Neste sentido, a evolução temporal do vetor de ocupação dos estados, depende apenas do valor desta distribuição num instante inicial e não da permutação específica que ele representa. Isto fica óbvio quando, nas equações acima, realiza-se a substituição da soma  $\sum_{j=1}^N V(q_i, q_j)$  pela integral  $\int dp dq \rho(q, p) \times V(q, q_j)$ . Neste caso,

valem as equações de movimento

$$\frac{\partial \rho}{\partial t} = \frac{\partial \rho}{\partial p} \frac{\partial p}{\partial t} + \frac{\partial \rho}{\partial q} \frac{\partial q}{\partial t} \quad (3.3)$$

$$= \frac{p}{m} \frac{\partial \rho}{\partial q} - \frac{\partial \rho}{\partial p} \frac{\partial}{\partial q} \int dp' dq' \rho(q', p'; t) \times V(q', q), \quad (3.4)$$

onde, obviamente escolheu-se representar a densidade de partículas  $\rho(q, p)$  por

$$\rho(q, p; t) = \sum_{i=1}^N \delta(q - q_i) \times \delta(p - p_i). \quad (3.5)$$

Quando tentamos interpretar a DNP acima como variável macroscópica, surge um dilema: a forma (3.5) pressupõe a disponibilidade de informação microscópica, enquanto que em laboratório normalmente se obtêm DNP's com baixíssima resolução. Comparativamente ao movimento realizado pelas moléculas, os instrumentos de medição macroscópicos são grosseiros tanto no tempo quanto no espaço e, portanto, é preciso desenvolver uma teoria que generalize o resultado anterior para estas situações. Como se escreveria a equação de movimento que leva em conta esta imprecisão?

A solução é considerar que existe uma escala natural associada ao processo de medição que, na medida em que se consegue resoluções mais finas, o resultado acima é atingido e na medida que se considera resoluções mais grosseiras, outros termos poderiam se somar à equação anterior. Resoluções muito grosseiras implicam no aparecimento de um comportamento probabilístico já que a especificação de qualquer ponto no espaço de estados corresponderia a um conjunto muito abrangente de trajetórias potencialmente divergentes. Resoluções muito finas, no entanto, podem implicar na especificação de estados que, do ponto de vista prático, são inacessíveis. Em determinadas situações, é razoável supor a existência de uma escala o suficientemente grosseira para que os estados  $\rho(q, p)$  possuam sentido laboratorial, mas fina o suficiente para que o raciocínio probabilístico não seja necessário<sup>1</sup>. Estes sistemas são os que obedecem à aproximação mecânica.

Entre os sistemas em que se espera que tal aproximação seja válida, estão todos sistemas tipicamente termodinâmicos como gases simples, sistemas químicos de poucos reagentes/produtos, sólidos cristalinos, etc. Já os sistemas em que a estrutura específica da organização molecular é relevante, se suporia que tal aproximação seja inválida. Um exemplo, os sistemas biológicos, evoluem de acordo com reações químicas extremamente

---

<sup>1</sup>Em terminologia usual isto é descrito como se os elementos de volume  $dq, dp$  fossem grandes em escala microscópica, mas pequenos numa escala macroscópica.

elaboradas que acontecem no interior das organelas celulares. Imagine que numa escala “suficientemente grosseira para a determinação laboratorial” certo ponto represente um volume no espaço de estados que possui a seguinte especificação: “existem  $C$  átomos de carbono,  $H$  de hidrogênio,  $O$  de oxigênio, ...”. Note que dentro deste volume “suficientemente grosseiro” a mesma configuração macroscópica pode representar um tipo de proteína, outro tipo de proteína, uma sopa grosseira os mesmos elementos químicos etc. É claro que cada uma destas situações produz processos biológicos totalmente diferentes. Com o intuito de contrastar com sistemas deste tipo, escolheu-se a denominação de *sistemas simples* àqueles que vale a aproximação mecânica.

Considere um sistema com  $N$  moléculas ou outros graus de liberdade, os quais se configuram em estados “grosseiros” enumerados de 1 a  $m$ . Este sistema tanto pode ser representado por uma lista com todos estados moleculares,

$$x = (x_1, x_2, \dots, x_N); \quad x_i \in \{1, \dots, m\} \quad (3.6)$$

quanto pela a representação *distribuição do número de partículas*, relacionada com o estado microscópico por um processo de contagem e apresentada como o vetor de ocupação

$$\mathbf{n}(x) = (n_1(x), n_2(x), \dots, n_m(x)). \quad (3.7)$$

Na medida que consideramos resoluções mais grosseiras, torna-se razoável representar as coordenadas de  $\mathbf{n}$  por valores fracionários. Note que muitos fenômenos característicos da descrição macroscópica são consequência do uso de resoluções grosseiras — ou lentes embaçadas — e entre estes se inclui o aparecimento da irreversibilidade.

## 3.2 Uso de integrais funcionais

Acreditava em infinitas séries de tempos, numa rede crescente e vertiginosa de tempos divergentes, convergentes e paralelos. Essa trama de tempos que se aproximam, se bifurcam, se cortam ou que secularmente se ignoram, *abrange* todas as possibilidades.

(J. L. BORGES — O Jardim de Veredas que se Bifurcam)

Nesta seção esboçamos uma metodologia para lidar com a evolução temporal de  $P(\mathbf{n}_t|\mathbf{n}_0)$  se o sistema em questão obedece à aproximação mecânica. O formalismo é levemente inspirado em [?, ?] e na formulação de Feynmann da mecânica quântica. Analisemos as probabilidades de que a evolução temporal se dê por cada uma das possíveis trajetórias que partem de um estado inicial, a DNP  $\mathbf{n}_0$ . Para isto utilizamos

as probabilidades

$$P(\mathbf{n}_f(t_f) \dots \mathbf{n}_2(t_2) \mathbf{n}_1(t_1) | \mathbf{n}_0(t_0)), \quad (3.8)$$

que representam a chance de  $\mathbf{n}$  evoluir através dos estados  $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_f$  nos tempos  $t_1, t_2, \dots, t_f$  uma vez fixado o estado inicial  $\mathbf{n}_0$ . Sistemas simples possuem uma evolução Markoviana de tal forma que as probabilidades acima são, simplesmente

$$P(\mathbf{n}_f \dots \mathbf{n}_2 \mathbf{n}_1 | \mathbf{n}_0) = P(\mathbf{n}_f | \mathbf{n}_n) \dots P(\mathbf{n}_2 | \mathbf{n}_1) P(\mathbf{n}_1 | \mathbf{n}_0). \quad (3.9)$$

A probabilidade de transição de  $\mathbf{n}_0$  para  $\mathbf{n}_f$  é dada pela marginalização desta função sobre todos os estados intermediários. Seja  $P(\mathbf{n}' | \mathbf{n}) = \int f(\mathbf{n}' | \mathbf{n}) d\mathbf{n}'$

$$P(\mathbf{n}_f | \mathbf{n}_0) = \int d\mathbf{n}_1 \dots d\mathbf{n}_n f(\mathbf{n}_f | \mathbf{n}_n) \dots f(\mathbf{n}_1 | \mathbf{n}_0), \quad (3.10)$$

que fazendo seu número ir a infinito, e introduzindo a notação  $\mathcal{D}\mathbf{n} = d\mathbf{n}_1 d\mathbf{n}_2 \dots$ , se escreve como

$$P(\mathbf{n}_f | \mathbf{n}_0) = \int \mathcal{D}\mathbf{n} f(\mathbf{n}_f | \mathbf{n}_k) \dots f(\mathbf{n}_1 | \mathbf{n}_0). \quad (3.11)$$

Uma análise do processo de limite induz a certas considerações simplificadoras sobre o comportamento da transição infinitesimal  $P(\mathbf{n}_{i+1} | \mathbf{n}_i)$  para tempos  $t_{i+1}$  e  $t_i$  infinitamente próximos. A probabilidade  $P(\mathbf{n}_{i+1} | \mathbf{n}_i)$  é escrita como função dos parâmetros  $\mathbf{n}_i, \mathbf{n}_{i+1}, t_i$ , e  $t_{i+1}$ . Uma mudança de coordenadas para  $\mathbf{n}, \dot{\mathbf{n}}, t_i$  e  $\delta t$  onde  $\mathbf{n} \equiv \frac{\mathbf{n}_{i+1} + \mathbf{n}_i}{2}$ ,  $\dot{\mathbf{n}} \equiv \frac{\mathbf{n}_{i+1} - \mathbf{n}_i}{\delta t}$  e  $\delta t \equiv t_{i+1} - t_i$  facilita os cálculos e, portanto, será extensivamente utilizada.

Por uma questão de coerência, é necessário que a função que descreve a probabilidade de transição  $P(\mathbf{n}_{i+1} | \mathbf{n}_i) \equiv \int f(\mathbf{n}, \dot{\mathbf{n}}, t_i, \delta t) d\mathbf{n}$  obedeça certas propriedades. Uma delas é que o limite  $\delta t \rightarrow 0$  implique que  $P(\mathbf{n}'_{i+1} | \mathbf{n}_i) \rightarrow \delta(\mathbf{n}_{i+1} - \mathbf{n}_i) g(\mathbf{n}, \dot{\mathbf{n}}; t, \delta t)$  onde  $g(\mathbf{n}, \dot{\mathbf{n}}; t, \delta t)$  é uma função qualquer "bem comportada" e a delta garante que a transição de  $\mathbf{n}_i$  para  $\mathbf{n}_{i+1}$  só seja permitida se  $\mathbf{n}_i \rightarrow \mathbf{n}_{i+1}$  quando  $\delta t \rightarrow 0$ . Uma representação útil da função delta é dada pelo limite da Gaussiana

$$\delta(\mathbf{n}_{i+1} - \mathbf{n}_i) = \lim_{\delta t \rightarrow 0} \frac{\kappa}{2\sqrt{\pi\delta t}} e^{-\frac{\kappa^2}{4\delta t} (\mathbf{n}_{i+1} - \mathbf{n}_i)^2}, \quad (3.12)$$

onde  $\kappa$  é uma constante que mantém o argumento da exponencial adimensional. Pela substituição de  $\dot{\mathbf{n}} \equiv \frac{\mathbf{n}_{i+1} - \mathbf{n}_i}{\delta t}$  ficamos com

$$\delta(\mathbf{n}_{i+1} - \mathbf{n}_i) = \lim_{\delta t \rightarrow 0} \frac{\kappa}{2\sqrt{\pi\delta t}} e^{-\frac{1}{4}\kappa^2 \delta t \dot{\mathbf{n}}^2}, \quad (3.13)$$

de forma que devemos esperar que a probabilidade  $P(\mathbf{n}_f | \mathbf{n}_0)$  contenha um termo multiplicativo semelhante a este.



Escrevemos, por conveniência, a função  $g(\mathbf{n}, \dot{\mathbf{n}}; t, \delta t)$  como  $\exp[\delta t \alpha^{-1} V(\mathbf{n}, \dot{\mathbf{n}}; t, \delta t)]$  de sorte que a probabilidade de transição é dada por

$$P(\mathbf{n}_{i+1}|\mathbf{n}_i) \propto e^{-\delta t \alpha^{-1} (K \dot{\mathbf{n}}^2 - V(\mathbf{n}, \dot{\mathbf{n}}, t, \delta t))} d\mathbf{n}_{i+1}, \quad (3.14)$$

onde a constante temporal  $\alpha$  mantém o argumento da exponencial adimensional e tentativamente identificamos com o tempo de relaxação do sistema. O termo  $K$  determina a relevância estatística relativa do termo cinético e, finalmente, o termo de interação  $V(\mathbf{n}, \dot{\mathbf{n}}, t, \delta t)$  caracteriza detalhadamente o processo estatístico. Note que no limite  $\delta t \rightarrow 0$  esperamos que o termo de interação convirja para uma função específica de  $\mathbf{n}, \dot{\mathbf{n}}$  e  $t$  a qual nos referimos como  $V(\mathbf{n}, \dot{\mathbf{n}}; t)$ . Desta maneira, podemos escrever a "Lagrangiana" do processo,  $L(\mathbf{n}, \dot{\mathbf{n}}; t) \equiv K \dot{\mathbf{n}}^2 - V(\mathbf{n}, \dot{\mathbf{n}}, t)$ , que é uma função arbitrária e adimensional que especifica completamente a natureza das transições entre estados. Ao substituir este valor em (3.11) temos

$$P(\mathbf{n}_f|\mathbf{n}_0) \propto \int \mathcal{D}\mathbf{n} e^{-\alpha^{-1} \mathcal{A}(\mathbf{n}_f|\mathbf{n}_0)}, \quad (3.15)$$

de sorte que a "integral de ação"  $\mathcal{A}(\mathbf{n}_f|\mathbf{n}_0) \equiv \int_{t_0}^{t_f} dt L(\mathbf{n}(t), \dot{\mathbf{n}}(t); t)$  sobre os pontos iniciais e finais fornece a probabilidade atribuída a cada trajetória que liga os mesmos.

A escolha das designações "Lagrangiana" e "ação" não é acidental. Ao tomar o limite  $\alpha \rightarrow 0$ , a probabilidade associada à trajetória de mínima ação torna-se estatisticamente dominante sobre as outras — recuperando uma dinâmica determinista. Deste modo, interpretamos  $\alpha$  como uma medida do grau de "aleatoriedade" do sistema. Isto não implica, é claro, que a mecânica seja estocástica, só que para  $\alpha$  muito grande as probabilidades de transição  $P(\mathbf{n}_f|\mathbf{n}_i)$  se tornam cada vez mais dispersas, dificultando fazer qualquer inferência mais conclusiva sobre o estado final.

Para avaliar melhor este limite, considere a troca de variáveis  $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_f \rightarrow S, \omega_1, \dots$  onde  $S$  representa a ação definida pela trajetória  $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_f$  e os parâmetros  $\omega_i$  consistem numa parametrização que implemente tal mudança de variáveis. Assim, definindo o Jacobiano da transformação por  $J(S, \omega)$  ficamos com uma probabilidade de transição

$$P(\mathbf{n}_f|\mathbf{n}_0) \propto \int d\omega_1 \dots d\omega_f \times dS J(S, \omega) \times e^{-\alpha^{-1} S}. \quad (3.16)$$

Pela posterior integração dos  $\omega_i$ 's, e realizando uma integração por partes em  $S$  ficamos com

$$P(\mathbf{n}_f|\mathbf{n}_0) \propto \int_{S_0}^{\infty} dS J(S) \times e^{-\alpha^{-1} S} = \alpha J(S) e^{-\alpha^{-1} S} \Big|_{S_0}^{\infty} + \alpha \int_{S_0}^{\infty} dS \frac{\partial J(S)}{\partial S} e^{-\alpha^{-1} S}. \quad (3.17)$$

Supondo que  $J(S \rightarrow \infty)$  não tenha ordem exponencial em  $S$ , é possível eliminar todos termos do tipo  $\lim_{S \rightarrow \infty} \frac{\partial^n J(S)}{\partial S^n} e^{-\alpha^{-1}S}$ . A motivação é que queremos que as trajetória com ação infinita sejam avaliadas com probabilidade nula. Integrando por partes recursivamente, ficamos com a série

$$P(\mathbf{n}_f|\mathbf{n}_0) = A\alpha \left( \sum_{n=0}^{\infty} \alpha^n \frac{\partial^n J(S_0)}{\partial S_0^n} e^{-\alpha^{-1}S_0} \right), \quad (3.18)$$

onde  $A$  é uma constante de normalização. Desta forma, ao passar o limite  $\alpha \rightarrow 0$ , apenas a trajetória de mínima ação que liga  $\mathbf{n}_0$  à  $\mathbf{n}_f$  possui relevância estatística. Na realidade, podemos comparar trajetórias com pontos finais distintos, já que  $S_0$  é, na realidade, uma função de  $\mathbf{n}_f$ . Deste modo, seja  $\bar{\mathbf{n}}$  o estado final associado ao mínimo global de  $S_0(\mathbf{n}_f)$ , de forma que  $\bar{S} = S_0(\mathbf{n}_f)$ . Desta forma é possível comparar a probabilidade de um estado  $\bar{\mathbf{n}} + \delta\mathbf{n}$  com  $\bar{\mathbf{n}}$ . Expandindo em segunda ordem em torno da menor ação, e lembrando que  $\frac{\partial \bar{S}}{\partial \mathbf{n}} = 0$ , temos

$$P(\bar{\mathbf{n}} + \delta\mathbf{n}|\mathbf{n}_0) = J(S(\bar{\mathbf{n}} + \delta\mathbf{n})) \exp \left[ -\alpha^{-1}\bar{S} + \alpha^{-1} \frac{\partial \bar{S}}{\partial \mathbf{n}} \delta\mathbf{n} - \frac{\alpha^{-1}}{2!} \frac{\partial^2 \bar{S}}{\partial \mathbf{n}^2} (\delta\mathbf{n})^2 \right], \quad (3.19)$$

$$= J(S(\bar{\mathbf{n}} + \delta\mathbf{n})) e^{-\alpha^{-1}\bar{S}} \times \exp \left[ -\frac{\alpha^{-1}}{2!} \frac{\partial^2 \bar{S}}{\partial \mathbf{n}^2} (\delta\mathbf{n})^2 \right]. \quad (3.20)$$

Agora comparamos com a trajetória máxima de sorte que

$$\frac{P(\bar{\mathbf{n}} + \delta\mathbf{n}|\mathbf{n}_0)}{P(\bar{\mathbf{n}}|\mathbf{n}_0)} = \frac{J(S(\bar{\mathbf{n}} + \delta\mathbf{n}))}{J(\bar{S})} e^{-\frac{\alpha^{-1}}{2!} \frac{\partial^2 \bar{S}}{\partial \mathbf{n}^2} (\delta\mathbf{n})^2}. \quad (3.21)$$

A condição de mínimo para  $\bar{S}$  garante que a derivada segunda é positiva, de maneira que, no limite  $\alpha \rightarrow 0$ , de todos pontos  $\mathbf{n}_f$ , apenas aquele  $\bar{\mathbf{n}}$  com a menor ação estaria estatisticamente representado.

Uma distinção importante com relação ao papel que a Lagrangiana possui na mecânica é que o acréscimo de termos com derivada total na Lagrangiana *altera* as probabilidades relativas a cada transição — normalmente somos levados à crer que estes termos são irrelevantes. Isto se dá porque a ação é acrescida de termos como  $G(\mathbf{n}_f) - G(\mathbf{n}_0)$  que, no que diz respeito ao estado inicial fixo, apenas acrescenta uma constante global  $\exp(-\alpha^{-1}G(\mathbf{n}_0))$ , mas ao considerar diferentes estados finais,  $\exp(-\alpha^{-1}G(\mathbf{n}_f))$  modifica o peso estatístico relativo de cada um deles. Considere o caso extremo em que  $\alpha \rightarrow 0$ , ou seja, a dinâmica é determinista. É fácil ver que a adição de termos como o anterior modifica o estado avaliado com ação nula e, conseqüentemente, probabilidade 1. Dada a experiência prévia em mecânica, sabemos que um problema de contorno para  $\mathbf{n}_0$  e  $\mathbf{n}_f$  equivale a outro equivalente especificado pelas condições iniciais  $\mathbf{n}_0$  e  $\dot{\mathbf{n}}_0$ . Desta forma,

as divergências totais são entendidas como um termo que incorpora informação sobre a derivada na origem, ou seja  $\dot{\mathbf{n}}(t_0)$ .

### 3.2.1 Incorporando informação

Na seção (2.3) que caracteriza os sistemas de interesse termodinâmico vimos que uma prescrição adequada para determinar os estados de equilíbrio depende, em grande parte, da especificação correta dos invariantes de movimento macroscópico. Neste sentido, estes estados correspondem aos que maximizam a entropia sujeitos aos vínculos macroscópicos relevantes. Incorporaremos explicitamente este tipo de informação na formulação por integrais funcionais para garantir que, a não ser que se queira, o sistema físico não viole nenhum invariante de movimento.

Utiliza-se a conservação de energia (expressa como  $U(\mathbf{n})$ ) como exemplo típico do raciocínio empregado para considerar a informação sobre invariantes dinâmicos. A chave para incorporar este tipo de informação é pela introdução dos termos apropriados às probabilidades de transição infinitesimais  $P(\mathbf{n}_{i+1}|\mathbf{n}_i)$ . Uma maneira de proceder seria multiplicar  $P(\mathbf{n}_{i+1}|\mathbf{n}_i)$  por um termo que tenda à  $\delta(U(\mathbf{n}) - U)$  no limite  $\delta t \rightarrow 0$ . Infelizmente este procedimento é inviável ou, pelo menos, exige um grau de engenho maior que a imaginação do autor.

Outro procedimento que possui um apelo físico bastante interessante é o método da máxima entropia. Queremos especificar a probabilidade  $P(\mathbf{n}_{i+1}|\mathbf{n}_i U_i)$  em função de  $P(\mathbf{n}_{i+1}|\mathbf{n}_i)$  ao aceitar a consideração  $U_i \equiv$  "sistema possui uma energia  $U(\mathbf{n})$  com um valor  $U$  no instante  $t_i$ ". Consideramos  $P(\mathbf{n}_{i+1}|\mathbf{n}_i)$  como a probabilidade anterior do nosso problema para a qual a informação  $U_i$  é incorporada maximizando a entropia relativa

$$\int d\mathbf{n} P(\mathbf{n}_{i+1}|\mathbf{n}_i U_i) \ln \frac{P(\mathbf{n}_{i+1}|\mathbf{n}_i U_i)}{P(\mathbf{n}_{i+1}|\mathbf{n}_i)} \quad (3.22)$$

sujeita a um vínculo em  $U_i = \int d\mathbf{n} U(\mathbf{n}) P(\mathbf{n}_{i+1}|\mathbf{n}_i U_i)$ . O resultado deste procedimento é

$$P(\mathbf{n}_{i+1}|\mathbf{n}_i U_i) \propto e^{-\beta_i U(\mathbf{n}) - \delta t \alpha^{-1} L(\mathbf{n}, \dot{\mathbf{n}}, t_i)}. \quad (3.23)$$

Redefinimos o multiplicador de Lagrange  $\beta_i \mapsto \frac{\delta t}{\alpha}$ , e assim a probabilidade de uma trajetória que leve em conta informação sobre a invariância de  $U(\mathbf{n})$  é acrescida do termo multiplicativo  $\exp\left(-\alpha^{-1} \int dt \beta'(t) U(\mathbf{n}(t))\right)$ . Desta maneira, o método da máxima entropia pode ser entendido como uma prescrição para acrescentar informação que

efetivamente corresponde a acrescentar os termos de vínculo na Lagrangiana

$$L(\mathbf{n}, \dot{\mathbf{n}}; t) \mapsto L(\mathbf{n}, \dot{\mathbf{n}}; t|UN) = L(\mathbf{n}, \dot{\mathbf{n}}; t) + \beta(t)U(\mathbf{n}) + \lambda(t)N(\mathbf{n}). \quad (3.24)$$

Existe uma dificuldade técnica relativa à determinação da função  $\beta(t)$ . Note que a prescrição mencionada, além de implementar leis de conservação, também é capaz de especificar a variação temporal de qualquer função do estado do sistema. Desta forma, o par  $\beta(t), U(t)$  opera como grandezas termodinâmicas conjugadas: ao escolher uma certa evolução  $\beta(t)$ , a função  $U(t)$  é automaticamente determinada. A variável  $\beta(t)$  realiza o papel (do inverso) da temperatura, o que é extremamente conveniente já que a temperatura é um parâmetro mais fácil de se controlar que a energia. O problema de inverter a relação, ou seja: a partir da evolução  $U(t)$ , determinar a escolha da função  $\beta(t)$  apropriada apresenta uma dificuldade matemática desanimadora. Uma perspectiva futura deste trabalho é simular alguns processos numericamente e avaliar como se dá a dinâmica das coordenadas térmicas  $\beta(t), \mu(t) = \beta(t)^{-1}\lambda(t)$ , etc no processo de aproximação ao equilíbrio.

-

### 3.2.2 Tempos assintóticos

Sistemas simples evoluem irreversivelmente para o equilíbrio; isto significa que, partindo de uma classe específica de condições iniciais, e após um longo tempo, um determinado estado  $\bar{\mathbf{n}}_f$  será atingido. De maneira semelhante, para esta mesma classe de condições iniciais, o algoritmo de maximização da entropia encontra  $\bar{\mathbf{n}}$  a partir do conhecimento sobre os invariantes de movimento macroscópico. Estes dois procedimentos, por uma questão de consistência, devem fornecer os mesmos resultados.

A aproximação mecânica, como vimos anteriormente, determina que existe uma trajetória  $\bar{\mathbf{n}}(t)$  muito mais provável que as outras. Dado que o parâmetro  $\alpha$  regula o grau de “aleatoriedade” do processo, ele deve ser próximo de zero em sistemas mecânicos. Neste caso, ao invés de somar sobre todas as trajetórias possíveis, podemos aproximar as atribuições de probabilidade por

$$P(\mathbf{n}|\mathbf{n}_0UN) \propto e^{-\alpha^{-1}\mathcal{A}(\mathbf{n},\mathbf{n}_0)}, \quad (3.25)$$

em que o trajeto  $\bar{\mathbf{n}}(t)$  que vai de  $\mathbf{n}_0$  a  $\mathbf{n}$  minimiza a ação.

A irreversibilidade temporal se traduz no fato que o sistema evolui assintoticamente

para o estado estacionário  $\bar{\mathbf{n}} \equiv \mathbf{n}(t_f \rightarrow \infty)$ . Desejando consistência com esta observação, a condição de invariância sobre as variáveis  $U$  e  $N$  se traduz em multiplicadores de Lagrange  $\beta(t)$  e  $\lambda(t)$  que também convergem para um valor estacionário. Ao separar a parte de equilíbrio da parte transiente, a integral da ação sobre os tempos  $t_0$  até  $t_f$  torna-se arbitrariamente próxima de<sup>2</sup>

$$\lim_{t_f \rightarrow \infty} \frac{1}{t_f} (\mathcal{A}_{eq} + \mathcal{A}_{trans}) \simeq \lim_{t_f \rightarrow \infty} \frac{1}{t_f} \mathcal{A}_{eq} = \beta U(\bar{\mathbf{n}}) + \lambda \bar{\mathbf{n}} + L_A(\bar{\mathbf{n}}, \dot{\mathbf{n}} \rightarrow 0). \quad (3.26)$$

que, substituindo em  $P(\mathbf{n}|\mathbf{n}_0UN)$ , resulta em

$$P(\mathbf{n}|\mathbf{n}_0UN) \propto e^{-k(L_A(\mathbf{n},0) + \beta U(\mathbf{n}) + \lambda N(\mathbf{n}))}, \quad (3.27)$$

para o qual definiu-se  $k \equiv \frac{t_f}{\alpha}$ .

Em um sistema fechado,  $U(t)$  e  $N(t)$  necessariamente expressam leis de conservação. Uma expectativa razoável sobre a evolução macroscópica mencionada em (2.3.2) é que, na medida que se considera intervalos de evolução progressivamente longos, a informação sobre a condição inicial se resumiria aos invariantes de movimento associados a este estado, pouco importaria outras especificidades do mesmo. Esta condição, verdadeira ou não, parece ser necessária para fundamentar o método da maximização da entropia que determina os estados de equilíbrio e, portanto, será assumida. Isto nos permite identificar a probabilidade  $P(\mathbf{n}|\mathbf{n}_0UN)$  com  $P(\mathbf{n}|UN)$  já que a informação sobre  $\mathbf{n}_0$  é redundante. Nesta expectativa, calculamos a representatividade  $P(\mathbf{n})$  necessária para determinar a entropia do estado de equilíbrio.

Utilizando o método MaxEnt para incorporar informação sobre  $U$  e  $N$  na representatividade  $P(\mathbf{n})$ , temos que

$$P(\mathbf{n}|UN) = \frac{1}{Z(\beta, \lambda)} P(\mathbf{n}) e^{-\beta' U(\mathbf{n}) - \lambda' N(\mathbf{n})} \quad (3.28)$$

e, pela eliminação dos termos apropriados em (3.27), calcula-se a representatividade dos estados como

$$P(\mathbf{n}) \propto e^{-k L(\mathbf{n},0) + \text{cte}}. \quad (3.29)$$

Comparando este resultado com a fórmula de Boltzmann para a entropia, vemos que a

---

<sup>2</sup>Utilizamos a designação  $L_A$  para se referir a Lagrangiana em tempos muito longos. Note que, em geral, não se espera que a Lagrangiana dependa explicitamente do tempo devido às considerações de simetria por translação temporal. É lógico que a especificação da energia do sistema, como foi discutido anteriormente, quebra esta simetria, o que é expresso pela existência do termo dependente do tempo  $\beta(t)$  na Lagrangiana  $L(\mathbf{n}, \dot{\mathbf{n}}|NU)$ .

entropia de um estado pode ser identificada com os termos de interação da Lagrangiana<sup>3</sup>

$$H(\mathbf{n}) \propto -L(\mathbf{n}, 0) + \text{cte.} \quad (3.30)$$

O limite assintótico da teoria de não-equilíbrio constitui a peça fundamental do formalismo que se segue. As considerações acima mostram que o comportamento no equilíbrio é elegantemente encapsulado na teoria do não-equilíbrio, tudo isso utilizando uma linguagem unificada. Nas duas situações utiliza-se a prescrição de minimizar a ação para determinar a trajetória mais provável. No formalismo fora do equilíbrio, isto reflete numa equação diferencial. Já no outro, devido à eliminação dos termos transientes, o resultado é uma equação algébrica — esta é a condição de máxima entropia.

### 3.3 Formalismo de equilíbrio

Uma vez de posse da representatividade dos estados  $P(\mathbf{n})$ , seja pelo limite assintótico da teoria de não-equilíbrio ou por considerações puramente estatísticas, os estados de equilíbrio são determinados pela maximização da entropia  $H(\mathbf{n}) = k_B \ln P(\mathbf{n})$  utilizando vínculos apropriados. A representatividade de cada DNP pode ser determinada pela somatória sobre todas as possibilidades condizentes com uma configuração macroscópica,

$$P(\mathbf{n}) = \sum_{[x]} P(x); \quad \mathbf{n}(x) = \mathbf{n}. \quad (3.31)$$

Esta expressão pode ser simplificada pela suposição que as probabilidades dependem do estado microscópico  $x$  pelos números  $\mathbf{n}(x)$  de partículas em cada estado molecular. Utilizando uma notação em que esta probabilidade (a quebra da equiprobabilidade) é expressa como  $\gamma(\mathbf{n}) \equiv P(x) = P(\mathbf{n}(x))$ , é possível reescrever a representatividade como

$$P(\mathbf{n}) = \frac{N!}{n_1!n_2! \times \cdots \times n_m!} \gamma(\mathbf{n}), \quad (3.32)$$

onde o fator de multiplicidade  $\frac{N!}{n_1!n_2! \times \cdots \times n_m!}$  determina o número de seqüências diferentes associadas à DNP  $\mathbf{n} = n_1, n_2, \dots, n_m$ .

Seguindo a argumentação em (2.1.3), é interessante permitir que  $\gamma(\mathbf{n})$  assumia valores

---

<sup>3</sup>Note que assumiu-se implicitamente que a utilização do método da máxima entropia fornece os mesmos resultados que aqueles encontrados na seção (2.1.1). Esta equivalência não é rigorosamente válida, mas dedicamos o capítulo (5) para mostrar que na maioria das situações de interesse ela consiste numa boa aproximação.

diferentes de uma constante (associada à equiprobabilidade em  $P(x)$ ). Além de dar um caráter mais robusto e mais amplo às afirmações da termodinâmica, veremos que a não-equiprobabilidade se trata de uma ferramenta heurística poderosa para lidar com as situações em que modelos com  $P(x) = \text{cte}$  falham. Ainda que dificilmente exista uma justificativa geral que fundamente quebras de equiprobabilidade universais (como Tsallis), a escolha por uma metodologia mais abrangente pode ser importante para descobrir o motivo por trás de uma determinada estatística generalizada. Veremos, posteriormente, que um mapa com as propriedades térmicas de um elemento, junto com o valor de ocupação relacionado a cada nível de energia é tudo que se precisa para determinar a probabilidade  $\gamma(\mathbf{n})$  das DNPs.

A entropia de Boltzmann associada a uma DNP é facilmente calculável utilizando a aproximação de Stirling, (fazendo  $k_B \equiv 1$ )

$$H(\mathbf{n}) = - \sum_{i=1}^m n_i \ln \frac{n_i}{N} + \ln \gamma(\mathbf{n}). \quad (3.33)$$

O caso em que  $\gamma(\mathbf{n}) = \text{cte}$  resulta trivialmente na entropia de Boltzmann dada por

$$H_1 \equiv - \sum_{i=1}^m n_i \ln \frac{n_i}{N} + \text{cte}. \quad (3.34)$$

Se os vínculos funcionam de forma a distorcer o peso estatístico dos estados a 1-partícula, mas sem acrescentar correlações, atribuímos as probabilidades  $\omega_1, \omega_2, \dots, \omega_m$  a cada estado de sorte que a probabilidade de uma seqüência é dada por  $\gamma(\mathbf{n}) = \omega_1^{n_1} \omega_2^{n_2} \dots \omega_m^{n_m}$ . Isto conduz à *entropia não-equiprovável*

$$H = - \sum_{i=1}^m n_i \ln \frac{n_i}{N\omega_i}. \quad (3.35)$$

Se as restrições dinâmicas introduzem correlações entre partículas, mas sem correlacionar estados diferentes entre si, a probabilidade atribuída a uma seqüência escreve-se como um produto de funções  $h_i(n_i)$ . A forma funcional específica dos  $h_i(n_i)$  determina o tipo de correlação e generaliza o caso de partículas independentes onde  $f_i(n) = \omega_i^{n_i}$ . O apelo físico desta condição é que as interações de curto alcance cessam fora da região associada à cada estado de grão grosseiro.

A entropia relacionada a este tipo de estatística pode ser escrita como

$$H = - \sum_i n_i \ln \frac{n_i}{N} + \sum_i \ln h_i(n), \quad (3.36)$$

ou, já que as funções  $h_i$ 's são arbitrárias,

$$H = \sum_i f_i(n_i; i). \quad (3.37)$$

Atribuições de probabilidade onde a informação sobre um subconjunto de estados não afeta as probabilidades atribuídas a outro sub-conjunto são escritas como um exemplo especial da equação (3.37) acima. Estas atribuições definem o que se chama *entropias locais*, já que toda estatística que viola esta expressão indica a existência de uma propriedade que correlaciona estados diferentes entre si. Consideramos que estas entropias são o ponto de partida mais geral para uma análise heurística do problema com vínculos indeterminados — para nos referir a elas utilizamos a designação de *entropias efetivas* ou *entropias generalizadas*. Exigências razoáveis sobre o comportamento destas classes de entropia serão feitas na próxima seção.

Um uso legítimo para as entropias efetivas é evitar a proliferação de variáveis termodinâmicas. Cada vínculo utilizado na maximização da entropia acrescenta um ou mais parâmetros ao estado termodinâmico. Normalmente estas variáveis não estão disponíveis no laboratório, de forma que, na impossibilidade de medi-las ou controlá-las, é mais adequado escolher valores típicos para as mesmas e eliminá-las do formalismo. Representando estes vínculos por  $\Gamma_k(\mathbf{n})$  e os valores típicos para os multiplicadores de Lagrange denotados por  $\lambda_k^*$ , podemos utilizar qualquer entropia que se assemelhe à forma

$$H(\mathbf{n}) = H_1(\mathbf{n}) + \sum_{k=1}^M \lambda_k^* \Gamma_k(\mathbf{n}). \quad (3.38)$$

Definindo  $\gamma(\mathbf{n}) \equiv \sum_{k=1}^M \lambda_k^* \Gamma_k(\mathbf{n})$ , é possível se basear neste raciocínio para fundamentar o uso de entropias generalizadas que, restringindo-se apenas ao vínculo na energia, sabidamente descrevem alguns processos de maneira mais adequada que a entropia de Boltzmann.

Relacionado a isto temos o reconhecimento que alguns problemas aparentemente são melhores descritos pelas chamadas  $q$ -entropias de Rényi ou Tsallis [?] que a entropia



de Boltzmann. Elas são dadas por

$$H_q^{(R)} = \frac{N}{1-q} \ln \left( \sum_{i=1}^m p_i^q \right); \quad H_q^{(T)} = \frac{N}{1-q} \left( \sum_{i=1}^m p_i^q - 1 \right), \quad (3.39)$$

onde  $p_i = \frac{n_i}{N}$  e o parâmetro adicional  $q$  deve ser especificado em cada modelo, sendo que se  $q = 1$ , ambas se equivalem e recuperam a fórmula de Boltzmann.

Ainda que nenhum argumento convincente tenha sido dado para justificar esta quebra específica de probabilidades, podemos pensar nas entropias acima como uma metodologia aproximativa para um problema dado pela equação (3.38). Isto é bem discutido em [?], onde mostrou-se que ao considerar um invariante adicional no procedimento de maximização da entropia de Boltzmann, é possível determinar adequadamente a estatística de um sistema que só admitia uma descrição a partir de uma destas entropias. A eficiência da estatística de Rényi ou Tsallis pode ser explicada pelo fato que o ajuste de  $q$  corresponde ao ajuste de um momento estatístico adicional na distribuição  $n_i$  resultante do processo de maximização. Isto basta, em muitas situações, para fornecer aproximações bastante adequadas para fenômenos reais.

### 3.4 Entropias fisicamente plausíveis

A prescrição de máxima entropia, aplicada a entropias generalizadas, vale para uma classe tão grande de problemas que parece improvável que sistemas físicos reais compreendam todas estas infinitas possibilidades. Nesta seção analisamos o caso mais simples: um sistema quântico ideal. Posteriormente generalizamos os argumentos para outras situações mais realistas. Suponha que este sistema de teste não exiba entrelaçamento e os níveis de energia não-degenerados são aproximadamente uniformemente espaçados. Deste sistema, exigimos que o comportamento termodinâmico seja o mais simples possível: não há transições de fase, não há instabilidades termodinâmicas etc. É lógico que tal sistema não descreve toda riqueza associada à fenomenologia da termodinâmica — seu comportamento é trivial —, mas eventualmente algumas propriedades serão quebradas ao acrescentar interações entre partículas ou outras complicações. A motivação para proceder desta maneira é a expectativa que regimes críticos sejam induzidos por agentes dinâmicos bem definidos (como a interação entre partículas) e não decorrem simplesmente da estatística associada.

O sistema de teste possui uma energia  $U = \sum_i n_i \epsilon_i$  que determina a condição de

máximo

$$\frac{\partial H}{\partial n_i} = \beta \epsilon_i + \lambda. \quad (3.40)$$

Nos restringimos inicialmente a entropias locais escritas como  $H(\mathbf{n}) = \sum_{i=1}^m f_i(n_i)$ , onde exige-se a condição de equiprobabilidade: a maximização sem vínculos não seleciona nenhum estado em detrimento dos outros, levando à DNP  $n_i = n_j = \frac{1}{m}$ . Tal condição implica em

$$\frac{\partial f_i(n_i)}{\partial n_i} \equiv g_i(n_i) = g_i(m^{-1}) = \lambda, \quad (3.41)$$

onde, dada a arbitrariedade do número de estados  $m$ , se conclui que

$$g_i(n) = g_j(n) \Rightarrow f_i(n) = f_j(n) + \text{cte}. \quad (3.42)$$

Esta propriedade significa que as funções  $f_i(n_i)$  são essencialmente as mesmas a menos de uma constante aditiva imaterial. Colecionando-as numa constante global,  $H(\mathbf{n})$  fica escrita como

$$H(\mathbf{n}) = \sum_i f(n_i) + \text{cte}, \quad (3.43)$$

que corresponde à uma probabilidade fatorável nos estados dada por

$$P(\mathbf{n}) = \prod_{i=1}^m e^{f(n_i)} \times \text{cte}. \quad (3.44)$$

Aderimos à convenção que  $f(0) = 0$ , o que, além de simplificar algumas derivações, possui o apelo físico que níveis vazios não contribuem para a entropia.

A expressão (3.43) pode implicar em propriedades termodinâmicas extremamente inadequadas pela escolha de  $f(x)$ . Exigimos que a maximização de  $H(\mathbf{n})$  esteja associada a um único máximo local — ou seja, só existe um estado mais representativo que os vizinhos no subespaço determinado por uma escolha de invariantes de movimento. Desta forma, é necessário que  $g(n) \equiv \frac{\partial f}{\partial n}$  possua inversa para que a equação (3.40) tenha uma única solução. O comportamento monotônico de  $g(n)$  implica na existência de um sinal bem definido para a sua derivada, mas o único caso de interesse é  $g'(n) < 0$ , que garante que  $f(n)$  seja côncava. A outra possibilidade implicaria que a extremização de  $H(\mathbf{n})$  fornece um mínimo, situação que obviamente não interessa.

Uma segunda exigência tem a ver com o fato que em sistemas com infinitos níveis de energia, os níveis mais altos (com energia indo à infinito) devem, eventualmente, se tornar estatisticamente irrelevantes. Isto é o mesmo que dizer que quando  $\epsilon_i$ , e portanto  $\beta \epsilon_i + \lambda$ , tendem ao infinito  $n$  deve se anular. Para que isso aconteça é necessário que a

função  $g(n) = \beta\epsilon + \lambda$  seja tal que<sup>4</sup>

$$g(n \rightarrow 0) = \infty. \quad (3.45)$$

Assim os níveis de altíssima energia são avaliados corretamente com uma probabilidade assintoticamente nula.

A concavidade de  $f(n)$  é um resultado muito importante porque implica no critério de estabilidade termodinâmica, ou seja, que a entropia de equilíbrio é uma função côncava da energia. Para verificar isto, dividimos o sistema em duas partições  $\lambda_1$  e  $\lambda_2$  isoladamente em equilíbrio, onde  $\lambda_1 + \lambda_2 = 1$ , de forma que uma partição recebe um incremento de energia  $\delta U$  e outra é subtraída da mesma quantidade. Dizemos que um sistema é extensivo se segue a igualdade

$$\mathbf{n}(U) = \lambda_1 \mathbf{n}(U - \delta U) + \lambda_2 \mathbf{n}(U + \delta U). \quad (3.46)$$

Isto será sempre verdade se a energia for linear como mostra a relação

$$n_i(U \pm \delta U) = g^{-1}((\beta \mp \delta\beta)\epsilon_i + \lambda \pm \delta\lambda) \quad (3.47)$$

$$= n_i(U) \mp (\delta\beta\epsilon_i + \delta\lambda) \frac{\partial n_i(U)}{\partial \lambda} + O(\delta\beta^2) + O(\delta\lambda^2). \quad (3.48)$$

De forma que substituindo em (3.46), temos

$$n_i(U) = n_i(U) + \lambda_1 (\delta\beta\epsilon_i + \delta\lambda) \frac{\partial n_i(U)}{\partial \lambda} - \lambda_2 (\delta\beta\epsilon_i + \delta\lambda) \frac{\partial n_i(U)}{\partial \lambda} \quad (3.49)$$

$$= \lambda_1 n_i(U + \delta U) + \lambda_2 n_i(U - \delta U). \quad (3.50)$$

Uma função é côncava se e somente se  $f(\lambda_1 x_1 + \lambda_2 x_2) \geq \lambda_1 f(x_1) + \lambda_2 f(x_2)$ , onde as constantes positivas  $\lambda_1$  e  $\lambda_2$  são escolhidas de forma que  $\lambda_1 + \lambda_2 = 1$ . Utilizando esta expressão em  $H(\mathbf{n}) = \sum_i f(n_i)$ , dada a concavidade de  $f(n_i)$ , é trivial mostrar que

$$S(U) \geq \lambda_1 S(U - \delta U) + \lambda_2 S(U + \delta U). \quad (3.51)$$

Este resultado é interessante; mostra que para que um sistema ideal seja termodinamicamente estável, basta que exista um único equilíbrio termodinâmico por valor de energia. Como é bem conhecido, a estabilidade termodinâmica implica que tais sistemas se comportem de maneira trivial: os estados de equilíbrio tendem a ser espacialmente uniformes, as grandezas intensivas de sistemas fracamente interagentes tendem a se

---

<sup>4</sup>Aqui assumimos que  $\beta > 0$ . Caso fosse o contrário, concluiríamos que  $g(n \rightarrow 0) = -\infty$ , o que é incompatível com a idéia que  $g(n)$  é uma função monotônica decrescente.

igualar e não há transições de fase.

Uma vez de acordo com esta exigência, procuramos por outras classes de entropias que implicam na estabilidade termodinâmica, mas não sejam escritas como (3.43). A primeira generalização é considerar entropias não-equiprováveis. Supomos que a distribuição de equilíbrio seja modificada pela transformação  $n_i \mapsto n'_i = \omega_i n_i$  devido à existência de um peso  $\omega_i$  que altere a relevância estatística de cada estado. Esta transformação pode ser facilmente implementada substituindo a entropia do tipo (3.43) pela forma mais geral

$$H(\mathbf{n}) = \sum_i \omega_i f\left(\frac{n_i}{\omega_i}\right) + \text{cte.} \quad (3.52)$$

Esta expressão, como a anterior, é termodinamicamente estável. A demonstração deste resultado segue em paralelo à do resultado anterior e portanto não será explicitada. A segunda generalização é dada pela expressão

$$H(\mathbf{n}) = \eta\left(\sum_i \omega_i f\left(\frac{n_i}{\omega_i}\right)\right). \quad (3.53)$$

Esta é, até onde foi possível verificar, a expressão mais geral com possível relevância física. Querendo recapitular os resultados anteriores, exigimos que as funções  $f(x)$  e  $\eta(x)$  sejam côncavas e  $\eta(x)$  monotonicamente crescentes, além disso a derivada de  $f(x)$  deve divergir positivamente no limite  $x \rightarrow 0$ .

A introdução da função  $\eta(x)$  não altera a estatística subjacente à cada entropia, isto é bem conhecido pelo exemplo das entropias de Tsallis e Renyi. No entanto, ela altera a resposta térmica do sistema já que modifica a relação funcional entre entropia e energia. Este resultado é claramente demonstrado em [?]. A motivação para que se utilize o esquema de entropias generalizadas (3.53) é que esta expressão fornece um controle fino para inferir as entropias efetivas a partir dos parâmetros experimentais. O primeiro destes parâmetros, a estatística de ocupação dada pelos números  $n_i$ , determina  $f(x)$  e as probabilidades  $\omega_i$ . O segundo, que é a relação de  $S$  com  $U$ , pode ser ajustado por uma escolha apropriada da função  $\eta(x)$ .

### 3.5 Entropias para bósons e férmions

O argumento que utilizamos para derivar a equação (3.32), rigorosamente não é válido para sistemas quânticos já que utilizou-se um método de contagem “clássico”

para determinar a multiplicidade dos estados. Em estatística quântica, tanto os bósons quanto os férmions não admitem este tipo de contagem já que as restrições impostas na ocupação dos níveis limitam severamente o modo que podemos distribuir as partículas pelos estados. Na realidade, a estatística quântica difere da estatística clássica no que a noção de partícula é substituída pela noção de estado [?]; não se pergunta de quantas maneiras é possível arrumar um certo conjunto de partículas segundo uma distribuição  $\mathbf{n}$ , mas sim em quantas maneiras é possível ocupar cada estado  $i$  com respectivamente com  $n_i$ 's partículas. Isto talvez seja melhor justificado pela noção que as partículas são entendidas como uma excitação de um campo térmico, do que pela idéia de indistinguibilidade. Antes de prosseguir, no entanto, é interessante rever alguns aspectos da estatística de Boltzmann usual.

Considere um sistema quântico com  $N$  partículas idênticas não interagentes com níveis de energia  $\epsilon_1, \epsilon_2, \dots, \epsilon_m$ . A configuração energética deste sistema consiste em  $m$  níveis degenerados por um fator  $g_i N$ , onde  $g_i$  é a degenerescência do  $i$ -ésimo nível a 1-partícula. Nesta situação, mas somente nela, é possível realizar o processo de contagem exposto anteriormente sem quaisquer modificações. Ao permitir que as partículas interajam, a degenerescência dos níveis usualmente é quebrada e as linhas energéticas degeneradas em  $g_i N$  dão lugar a *bandas de energia*, com  $g_i N$  níveis distintos, e usualmente com uma dispersão energética considerável.

O procedimento que nos livra desta ambigüidade é agrupar níveis energéticos com energia similar, efetivamente redefinindo os estados quânticos. Desta maneira, o número de estados relevantes em nossa estatística cresce de  $m$  para um novo valor  $m'$  tal que  $mN > m' > m$  determinado pelo tipo de partição realizada. Suponha que o  $i$ -ésimo destes novos estados seja constituído por  $G_i$  estados originais, os quais, num primeiro momento, podemos preencher sem qualquer restrição. A probabilidade associada à esta nova configuração de partículas é simplesmente

$$P(\mathbf{n}) = \frac{N!}{n_1! n_2! \dots n_{m'}!} \left(\frac{G_1}{N}\right)^{n_1} \left(\frac{G_2}{N}\right)^{n_2} \dots \left(\frac{G_{m'}}{N}\right)^{n_{m'}} \quad (3.54)$$

já que cada estado é ponderado por um fator de degenerescência  $G_i$ . A entropia de Boltzmann para este sistema é dada simplesmente por

$$H = - \sum_{i=1}^{m'} n_i \ln \frac{n_i}{G_i} + \text{cte}, \quad (3.55)$$

onde podemos recuperar a expressão original simplesmente tomando partições  $G_i$

iguais, de forma que o resultado apresentado é equivalente a uma média de “grão grosseiro” de (3.34).

Nesta perspectiva, podemos escrever as entropias para férmions e bósons segundo os princípios adequados para lidar com a contagem de possíveis arranjos de estados. Mostramos como chegar à entropia bosônica de maneira similar ao raciocínio anterior enquanto, logo em seguida, ilustramos a metodologia proposta na seção anterior para chegar à forma para a entropia de férmions. Desta maneira, definimos o fator  $\Omega_i(n_i)$  como o número de arranjos com  $n_i$  “excitações” diferentes no  $i$ -ésimo conjunto com degenerescência  $G_i$  do tipo “grão grosseiro” considerado anteriormente. Em se tratando de bósons, obviamente não há qualquer restrição nos números de ocupação dos estados de  $G_i$  de tal maneira que podemos representar uma certa ocupação da seguinte maneira

$$| - - | - | - - - | - - - | - | - - - |. \quad (3.56)$$

Cada hífen representa uma partícula e, separando-as, os traços verticais  $|$  delimitam um estado de  $G_i$ . Deste modo, os símbolos acima prescrevem um vetor de ocupação  $(2, 1, 3, 3, 1, 3)$  com  $n_i = 13$  e  $G_i = 6$ . O número de maneiras distribuir estas  $n_i$  partículas nos  $G_i$  estados consiste simplesmente no número de permutações distintas em que podemos dispor os símbolos acima. Dado que os  $|$  das extremidades são fixos, existem  $n_i + G_i - 1$  símbolos de interesse, de forma que o número de permutações é dado por  $(n_i + G_i - 1)!$  descontado pelo número de permutações de hífen, ou seja,  $n_i!$  e o número de permutações dos traços verticais, que é  $(G_i - 1)!$ . Deste modo, a multiplicidade para cada estado bosônico é dada por

$$\Omega(n_i) = \frac{(n_i + G_i - 1)!}{n_i!(G_i - 1)!}. \quad (3.57)$$

Este termo permite escrever a probabilidade

$$P(\mathbf{n}) = \text{cte} \times \prod_{i=1}^{m'} \Omega(n_i), \quad (3.58)$$

de forma que, utilizando a aproximação de Stirling e organizando os termos obtemos a entropia para bósons

$$H_B(\mathbf{n}) = \sum_i G_i [(1 + \bar{n}_i) \ln(1 + \bar{n}_i) - \bar{n}_i \ln \bar{n}_i], \quad (3.59)$$

onde definimos  $\bar{n}_i \equiv \frac{n_i}{G_i}$  como a fração dos estados de  $G_i$  ocupados.

O resultado da maximização desta entropia com um vínculo na energia fornece os

números de ocupação

$$n_i = \frac{G_i}{e^{-\beta(\epsilon_i - \mu)} - 1}, \quad (3.60)$$

onde  $G_i$  simplesmente conta a degenerescência do estado com energia  $\epsilon_i$ , de maneira que este resultado é igual ao esperado.

Utilizaremos agora um raciocínio completamente diferente para derivar a forma para a entropia de férmions. Vale ressaltar que a diferença reside exclusivamente na forma de apresentação, sendo que tanto o argumento anterior quanto o próximo podem ser utilizados para derivar as entropias para bósons e férmions. O intuito é ilustrar os resultados da seção anterior, mostrando como se daria um raciocínio do tipo *a posteriori*. Ou seja, imaginamos que um certo físico experimentalista, o Sr. X, se depare com um sistema fermiônico, mas ainda sem conhecer os princípios por trás da estatística de férmions (suponha que Pauli não existiu), consiga apenas determinar experimentalmente os números de ocupação pela fórmula

$$n_i = \frac{G_i}{e^{\beta(\epsilon_i + \mu)} + 1}, \quad (3.61)$$

onde novamente  $G_i$  representa a degenerescência do  $i$ -ésimo nível de energia. Conhecendo a fórmula (3.53), e tentando entender porque não encontrou os fatores de Boltzmann em baixas temperaturas, nosso físico experimentalista certamente poderia encontrar alguma luz na estatística. Assuma inicialmente uma entropia local equiprovável, para a qual se determina a condição de máximo pela expressão

$$\frac{\partial f}{\partial n_i} = \beta\epsilon_i + \lambda. \quad (3.62)$$

Definindo  $a_i \equiv \beta\epsilon_i + \lambda = \beta(\epsilon_i + \mu)$  e  $g(n_i) \equiv \frac{\partial f(n_i)}{\partial n_i}$ , é possível expressar a condição acima como  $g(n_i) = a_i$ . Ao relacionar estas grandezas com o número de ocupação conhecido, ficamos com

$$n_i = g^{-1}(a_i) = \frac{G_i}{e^{a_i} + 1} \Rightarrow g(n_i) = \ln \left( \frac{G_i - n_i}{n_i} \right). \quad (3.63)$$

Isto implica na equação diferencial

$$\frac{\partial}{\partial n} f(n) = \ln \left( \frac{G - n}{n} \right) \Rightarrow f(n) = -n \ln(n) - (G - n) \ln(G - n) + G \ln G, \quad (3.64)$$

de forma que, facilmente encontramos a entropia de férmions

$$H_F(\mathbf{n}) = - \sum_{i=1}^m [(G_i - n_i) \ln (G_i - n_i) + n_i \ln n_i - G_i \ln G_i], \quad (3.65)$$

$$= - \sum_{i=1}^m G_i [(1 - \bar{n}_i) \ln (1 - \bar{n}_i) + \bar{n}_i \ln \bar{n}_i]. \quad (3.66)$$

Note que este resultado especifica apenas os números de ocupação de equilíbrio, mas não determina a curva térmica do sistema que é inteiramente especificada por  $\eta [H_B(\mathbf{n})]$ . Suponha que o nosso experimentalista, o Sr. X, mediu a resposta térmica do sistema e os resultados foram compatíveis com a entropia anterior. Desta forma, não é necessário modificar nada e justo fazer  $\eta(x) = x + \text{cte}$ . Uma vez de posse da entropia de férmions, Sr. X naturalmente se pergunta sobre que tipo de estatística estaria associada a uma forma tão peculiar de entropia. Nesta linha de raciocínio, utilizamos a relação de Boltzmann,  $P(\mathbf{n}) = e^{H(\mathbf{n})}$  para obter

$$P_B(\mathbf{n}) \propto \prod_{i=1}^m [n_i^{-n_i} (G_i - n_i)^{G_i - n_i} G_i^{G_i}]. \quad (3.67)$$

Agora notamos que é possível utilizar a aproximação de Stirling para calcular os termos do tipo  $x^x$ , ou seja

$$n^n = e^{n \ln n - n + n} = e^{\ln(n!) + n} = n! e^n. \quad (3.68)$$

Daí, ficamos com

$$P_F(\mathbf{n}) \propto \prod_{i=1}^m \frac{G_i!}{n_i! (G_i - n_i)!} = \prod_{i=1}^m \binom{G_i}{n_i}. \quad (3.69)$$

Os fatores binomiais na probabilidade acima descrevem o número de possibilidades de preencher  $n_i$  estados de um total de  $G_i$  disponíveis. Ou seja,  $G_i$  corresponde a uma urna em que se pode tirar no máximo  $G_i$  partículas, onde  $G_i$  é o número de estados encerrados nesta urna. Mas esta é justamente a restrição esperada para as ocupações de férmions! Que maravilhosa seria a descoberta do Sr. X.

### 3.6 Aproximação de campo médio e as distintas fases

Ainda que a maioria dos sistemas físicos reais não sejam do tipo simplificado considerado em na seção 3.4, a estabilidade termodinâmica é a regra e não a exceção nos processos observados na natureza. O motivo para isto é que os sistemas que



exibem interações podem ser razoavelmente bem representados por um sistema ideal dentro de regiões específicas do espaço de variáveis termodinâmicas — a cada uma destas regiões nos referimos como uma fase da matéria. Cada fase é estável e apenas fronteiras reduzidas do espaço termodinâmico exibem o comportamento não-ideal característico da transição de fase.

O termo que implementa a interação entre partículas é a forma quadrática  $\sum_{ij} n_i n_j V_{ij}$ . Se a energia de interação for relevante, a presença deste termo dificulta consideravelmente a determinação da entropia máxima devido ao acoplamento da ocupação de cada nível energético com os demais. Há situações em que este acoplamento implica na proliferação de máximos locais para um mesmo problema de vínculo [?]. Consideremos situações deste tipo.

Uma vez de acordo com o estado de equilíbrio associado a uma certa configuração termodinâmica, o que certamente se dará em um dos vários máximos locais disponíveis, é possível simplificar a análise ao perceber que a prescrição de maximização da entropia fornece o mesmo resultado se substituíssemos o sistema com interação pelo sistema ideal em que os níveis de energia livre  $\epsilon_i$  são deslocados por um valor específico

$$\epsilon_i \rightarrow \bar{\epsilon}_i = \epsilon_i + \sum_j V_{ij} n_j^*(N, U, V, \dots). \quad (3.70)$$

É claro que a determinação de  $\delta\epsilon_i(N, U, V, \dots) = \sum_j V_{ij} n_j^*(N, U, V, \dots)$  pode ser extremamente trabalhosa não só pela dificuldade em inverter a dependência dos multiplicadores de Lagrange com  $U$  e  $N$ , mas também pelo fato que ela seria diferente para máximos locais distintos. A cada um destes máximos, associamos uma fase termodinâmica; a diferença nas possíveis configurações energéticas reflete o fato que cada fase possui propriedades diferentes das outras. Desta maneira, esperamos que o panorama completo para a descrição do equilíbrio seja dado simplesmente por uma teoria para sistemas ideais complementada por um critério para identificar a fase a ser prescrita aos sistemas reais.

Assim somos levados a uma questão crucial: das diversas fases em que a matéria se manifesta, que fase seria observada nos experimentos reais? Ainda que não seja possível fornecer uma resposta inteiramente conclusiva — ela depende, em última instância, da preparação experimental —, a interpretação da entropia como representatividade do estado macroscópico indica que, com maior probabilidade, a fase observada seria aquela associada ao máximo global da entropia. A medida que se aumenta o número de partículas, a diferença entre as probabilidades associadas a DNPs distintas tende a se tornar cada vez mais acentuada, de forma que a probabilidade associada ao máximo

global tenderia a eliminar as outras opções à irrelevância estatística.

Um caso típico em que esta prescrição não é contemplada consiste numa evolução associada a baixas probabilidade de transição entre duas fases distintas. Uma razão para isto é que o sistema eventualmente necessita passar por regiões de baixíssima representatividade para percorrer qualquer trajetória que ligue uma fase à outra. Isto é verificado cotidianamente no fenómeno dos líquidos super-resfriados, muito observado no verão. Uma certa bebida possui uma pequena probabilidade de transição entre as fases *líquido*→*sólido* e permanece na fase líquida, a despeito das condições de temperatura reduzida em que ela eventualmente encontre. Pequenas perturbações externas, no entanto, modificam a dinâmica do sistema e podem implicar no congelamento imediato deste líquido.

O palco para os fenómenos críticos são as regiões em que existem pelo menos duas fases muito representativas avaliadas com probabilidades/entropias semelhantes. Neste caso, o sistema real pode ser encontrado com igual justiça em duas configurações distintas e uma fenomenologia diferente pode ser identificada de acordo com a topologia da interface que une as curvas de entropia associadas a cada fase. Transições de fase do *primeiro tipo* estão associados a curvas que se cortam, ou seja, existem fases tal que, em uma certa região a primeira é estatisticamente preponderante sobre a segunda enquanto em outra região ocorre o contrário. Separando as duas, está a fronteira de equiprobabilidade que determina a região crítica. Este tipo de transição caracteriza, por exemplo, as mudanças de fase líquido/sólido.

As transições do *segundo tipo* estão associadas a bifurcações. Neste caso, existem fases competitivas que, a partir de uma certa fronteira, convergem para uma fase única. Deste modo, a região crítica associada a transições do segundo tipo se posiciona na borda de uma região crítica do primeiro tipo. Ela possui uma dimensionalidade 2 unidades menor que o espaço de coordenadas termodinâmicas, sendo que, usualmente, correspondem a um simples ponto. Devido ao fato que no ponto crítico as fases coincidem, transições do segundo tipo estão associadas à ausência de descontinuidades. Exemplos típicos são a transição vapor/líquido a partir do ponto crítico ou à magnetização/desmagnetização pela mudança na temperatura de materiais ferromagnéticos.

## 4 Termodinâmica

Existem quatro leis. A terceira delas, a Segunda Lei, foi reconhecida primeiro; a primeira, a Lei Zero, foi formulada por último; a Primeira Lei foi a segunda; a Terceira Lei talvez nem seja uma lei no mesmo sentido das outras

(P.W. ATKINS)

Neste capítulo escrevemos as leis da termodinâmica como *teoremas* de uma teoria estatístico-mecânica. A motivação é esclarecer os fundamentos da termodinâmica; mais especificamente, mostrar que estas leis estão implicitamente contidas na mecânica microscópica como formas de raciocínio plausível. A partir daí, a termodinâmica se torna uma teoria independente e em praticamente todas as situações o formalismo tradicional segue inalterado. Para tornar mais explícita a relação entre o esquema mecânico-estatístico e a termodinâmica tradicional, tomamos como base uma apresentação padrão sobre o assunto. Daí se deriva, uma a uma, as propriedades assumidas no esquema axiomático. Um possível ganho em tornar esta relação explícita é escrever os potenciais termodinâmicos a partir de termos conhecidos com origem microscópica.

### 4.1 Postulados termodinâmicos

Na esperança de evitar o tortuoso desenvolvimento histórico que conduziu às diversas apresentações das 0 + 3 leis da termodinâmica, adotamos o ponto de vista desenvolvido por Callen, Tiza e outros em [?]. Esta apresentação é fornecida pelo conjunto de postulados que se segue.

Postulado I: *Os estados de equilíbrio podem ser inteiramente caracterizados por variáveis macroscópicas tais como a energia interna  $U$ , volume  $V$ , o número de partículas  $N_i$ , etc.. Nós os denotamos pelo vetor de estado  $F = (U, V, N_1, \dots)$ .*

Durante o raciocínio desenvolvido até agora, assumiu-se a validade da aproximação mecânica em vários pontos, o que ainda segue válido nesta discussão. Aqui consideramos

que a caracterização do estado macroscópico é dada pela DNP que evolui de acordo com a teoria apresentada no 3º capítulo. À partir da DNP, é possível calcular praticamente todas as quantidades de interesse, tanto as já mencionadas coordenadas termodinâmicas  $U, V, N_1, \dots$  como grandezas associadas a processos fora do equilíbrio. Ainda que a DNP caracterize estados macroscópicos arbitrários, as variáveis termodinâmicas podem ser entendidas como o conjunto mais simples de coordenadas que especifica um estado no *equilíbrio*.

Esta informação inclui não só o valor dos vínculos utilizados na maximização da entropia, mas também os parâmetros que apareçam indiretamente. A introdução das coordenadas do primeiro tipo é essencial, mas a escolha dos outros parâmetros depende do que é acessível ao experimento ou não. Num patamar menos abstrato, uma componente associada à energia dificilmente poderia ser desprezada em  $X$ , mas os parâmetros que aparecem nos níveis de energia tais como o volume, campo magnético, massa, constante de Planck, etc. podem constituir variáveis termodinâmicas adequadas ou não conforme a nossa habilidade de manipulá-los.

*Postulado II: Os estados são preparados fixando os valores de algumas (mas não necessariamente todas) variáveis extensivas  $F$ . Existe uma função  $S(F)$  que assume o valor máximo nos estados do equilíbrio. Isto define o equilíbrio termodinâmico.*

Este postulado não deve parecer nada surpreendente. A função  $S(F)$  mencionada corresponde à entropia de equilíbrio determinada pela maximização de  $H(\mathbf{n})$  sujeita ao vínculo na energia (utilizamos a notação usual de se referir à entropia maximizada pela letra  $S$ ). O motivo para isso já foi discutido anteriormente mas se resume essencialmente ao fato que o estado de equilíbrio é o mais *provável* numa classe de invariantes.

*Postulado III: A entropia de um sistema composto é a soma das entropias de cada sub-sistema. A entropia é uma função contínua, diferenciável e monotônica crescente da energia.*

A aditividade da entropia é conseqüência imediata da ausência de correlações estatísticas entre subsistemas [?] onde a probabilidade de cada seqüência definida em (3.32) é dada por  $\gamma(\mathbf{n}, \mathbf{n}') = \gamma(\mathbf{n})\gamma'(\mathbf{n}')$ . Apenas sistemas não correlacionados entre si podem ser separados de maneira inequívoca, caso contrário, seria necessário eliminar o agente que cria tais correlações, o que só pode ser feito às custas da integridade de um ou outro sistema. Neste sentido, podemos dizer que apenas subsistemas não-correlacionados e, portanto, com entropia aditiva, fazem sentido separados.

A seção 3.4 mostra que a entropia relacionada à sistemas simples é uma função

côncava e diferenciável da entropia. A monotonicidade é fácil de demonstrar. Uma vez de acordo que  $\beta = \frac{\partial S}{\partial U}$ , o que será demonstrado em breve, o caráter monotônico crescente equivale à  $\beta$  ser positivo. A justificativa para isto pode ser encontrada na discussão em torno da equação (3.45) e, resumidamente vem na necessidade que os estados de alta energia sejam estatisticamente irrelevantes, pois caso não fossem, a energia do sistema seria infinita. Estas propriedades garantem que a função  $S(U)$  possui inversa e, em vista disto, que  $\frac{1}{\beta} = \frac{\partial U}{\partial S}$ . Verificaremos posteriormente que o parâmetro  $\beta^{-1}$  é o que entendemos normalmente como temperatura.

Postulado IV: *A entropia se anula quando temperatura se aproxima do 0:*

$$\left( \frac{\partial U}{\partial S} \right)_{V, N_i, \dots} = 0 \Rightarrow S = 0. \quad (4.1)$$

Este postulado é compreendido como se o termo aditivo indeterminado fosse tal que a entropia seja zero à temperatura nula. Para que este limite esteja bem definido a entropia não pode divergir quando a temperatura vai a zero e também é necessário que a variação dos outros parâmetros termodinâmicos não afete este valor. Este parece ser o único ponto em aberto com relação à compatibilidade com a termodinâmica. A seção 4.5 verifica que, em geral, o Postulado IV, ou Terceira Lei, é esperado para todas classes de entropias plausíveis.

## 4.2 O conceito de temperatura

A temperatura quantifica as noções intuitivas de quente e frio ao regular o sentido natural de transferência de energia entre sub-sistemas em interação. Considere um processo próximo do equilíbrio em que há uma transferência espontânea de energia do sistema (1) para o sistema (2) ocasionada por uma disparidade entre suas respectivas temperaturas. Para o sistema composto, a entropia é dada por  $S = S_1 + S_2$  e a energia do sistema (1) se relaciona com a o outro sistema por  $U_1(t) = U - U_2(t)$ . Define-se o fluxo de energia  $W$  como  $W(t) \equiv \frac{\partial U_2}{\partial t} = -\frac{\partial U_1}{\partial t} \geq 0$ , de forma que

$$\frac{\partial S}{\partial t} = \frac{\partial S}{\partial U_1} \frac{\partial U_1}{\partial t} + \frac{\partial S}{\partial U_2} \frac{\partial U_2}{\partial t} = W \left( \frac{\partial S}{\partial U_2} - \frac{\partial S}{\partial U_1} \right). \quad (4.2)$$

Os processos físicos plausíveis estão associados a incrementos positivos de entropia. Avaliando o crescimento infinitesimal da entropia total,  $\delta H = W \left( \frac{\partial S}{\partial U_2} - \frac{\partial S}{\partial U_1} \right) \delta t$ , concluímos que  $\frac{\partial S}{\partial U_2} \geq \frac{\partial S}{\partial U_1}$ . Podemos relacionar estes parâmetros com as temperaturas de cada

sistema já que exercem o mesmo papel de regular o fluxo energético espontâneo e irreversível associado à interação térmica. Mais ainda, quando o equilíbrio é atingido, é necessário que  $\frac{\partial S}{\partial U_2} = \frac{\partial S}{\partial U_1}$ , para que cesse o aumento da entropia. Uma vez mostrada a semelhança qualitativa entre os dois conceitos, basta verificar a consistência entre as magnitudes dadas por  $T_1 > T_2$  e  $\frac{\partial S}{\partial U_1} < \frac{\partial S}{\partial U_2}$ , e definir uma escala de temperatura pela razão

$$\frac{\partial S}{\partial U_1} \equiv \frac{1}{T}. \quad (4.3)$$

Este valor pode ser calculado explicitamente quando consideramos a condição de máximo

$$\frac{\partial}{\partial n_i} S(\mathbf{n}) = \beta \frac{\partial}{\partial n_i} U(\mathbf{n}) - \lambda \quad (4.4)$$

para entropia escrita como função da DNP e resulta em

$$\frac{\partial S}{\partial U} = \sum_{i=1}^m \lambda \frac{\partial n_i}{\partial U} + \beta \sum_{i=1}^m \frac{\partial U}{\partial n_i} \frac{\partial n_i}{\partial U}. \quad (4.5)$$

$$= \lambda \frac{\partial N}{\partial U} + \beta \frac{\partial U}{\partial U}, \quad (4.6)$$

o que resulta imediatamente que

$$\frac{1}{T} \equiv \frac{\partial S}{\partial U} = \beta. \quad (4.7)$$

Isto implica numa interpretação física muito fortuita para o multiplicador de Lagrange. Uma vez que  $\beta$  normalmente é acessível experimentalmente, é possível lidar com este parâmetro como variável independente, ao invés de exigir que o determine a partir da relação implícita entre  $\beta$  e  $U$ .

### 4.3 Transformada de Legendre

O fato que as derivadas  $\frac{\partial S}{\partial U}$  e  $\frac{\partial U}{\partial S}$  possuem um apelo físico tão evidente indica que a exploração de representações em termo das coordenadas  $\beta$  ou  $T$  possa fornecer resultados úteis. A mudança de representação da entropia, que é função explícita da energia e outras coordenadas, para outra função equivalente que dependa de  $\beta$  é conduzida por uma transformada de Legendre de  $S$ . Considere, como um exemplo típico que implementa tal transformação, o potencial de Massieu

$$\Phi(\beta) = S - \beta U. \quad (4.8)$$

Desta definição, segue que

$$\frac{\partial \Phi}{\partial \beta} = \frac{\partial S}{\partial U} \frac{\partial U}{\partial \beta} - U - \beta \frac{\partial U}{\partial \beta} = -U, \quad (4.9)$$

onde também podemos mostrar facilmente que  $\Phi$  depende de  $\beta$  e independe de  $U$ . Chamamos variáveis relacionadas por transformações de Legendre da entropia ou da energia (ou outro potencial associado a um princípio de extremo) de conjugados termodinâmicos.

Outros pares conjugados podem ser definidos derivando a entropia ou energia a partir de outros parâmetros que, por exemplo, estejam presentes nos níveis energéticos que caracterizam o sistema. Coordenadas como pressão, potencial químico etc são definidas desta maneira

$$p \equiv -\frac{\partial U}{\partial V}; \quad \mu \equiv \frac{\partial U}{\partial N}. \quad (4.10)$$

A cada uma delas está associado um potencial termodinâmico que executa um papel análogo ao potencial de Massieu  $\Phi(\beta)$  mostrado anteriormente. Desta forma se define a função de entalpia e a entalpia- $\mu$  (mais utilizada em química)

$$H(p) \equiv E(V) + pV; \quad H_{\mu_i}(p, \mu_i) = H(p, N_i) - \mu_i N_i \quad (4.11)$$

Onde, naturalmente, valem as relações

$$\frac{\partial H}{\partial p} = V; \quad \frac{\partial H_{\mu_i}}{\partial \mu_i} = -N_i \quad (4.12)$$

(não confundir  $H$  e  $H_{\mu_i}$  com entropia).

As relações acima, na maioria das situações, envolvem cálculos complicados já que raramente é possível obter uma fórmula fechada da entropia com relação às outras coordenadas intensivas. Muitas vezes, no entanto, é possível calcular estes parâmetros diretamente a partir do valor dos níveis microscópicos. Considere, por exemplo, o cálculo da pressão onde a dependência no volume é introduzida através dos níveis de energia do sistema. Seja a energia com interação quadrática dada por

$$U = \sum_i \epsilon_i n_i + \sum_{ij} \epsilon_{ij} n_i n_j, \quad (4.13)$$

que pode ser escrita compactamente como

$$U = \langle \epsilon_i \rangle + \langle \epsilon_{ij} \rangle. \quad (4.14)$$

Apesar do cálculo ser aparentemente simples, basta derivar  $\epsilon_i$  e  $\epsilon_{ij}$  por  $V$ , existe

uma sutileza a ser considerada. Suponha um espaço termodinâmico caracterizado somente pelas três coordenadas  $U, V, N$ . Toda vez que nos referimos à entropia se assume que ela seja dada como  $S = S(U, V, N)$ . De maneira similar, ao referir à energia assumimos a relação funcional  $U = U(S, V, N)$ . Estas coordenadas implícitas na definição de cada potencial termodinâmico são as chamadas coordenadas naturais. As relações do tipo  $p = -\frac{\partial U}{\partial V}$  são válidas se expressas em coordenadas naturais, mas em muitos casos estas coordenadas são diferentes daquelas obtidas mais facilmente a partir dos modelos mecânicos.

De uma maneira geral, podemos considerar que um potencial termodinâmico, tal como a energia, é uma função a 6 coordenadas,  $U = U(S, V, N, \beta, p, \mu)$ , mas apenas 3 delas são, de fato, independentes<sup>1</sup>. Muitas vezes pode ser necessário especificar as coordenadas independentes ao derivar  $U$  por qualquer um destes parâmetros. Desta forma utiliza-se a notação de parênteses

$$p = \left( \frac{\partial U}{\partial V} \right)_{S, N} = \frac{\partial}{\partial V} U(S, V, N), \quad (4.15)$$

que é considerada mandatária apenas se o potencial termodinâmico não for escrito em suas coordenadas naturais.

Voltando ao problema de calcular a pressão de um sistema em termos das variáveis microscópicas, fica claro que devemos reescrever  $U$  nas suas coordenadas naturais. Note que a expressão para a energia que se obtêm de um problema de maximização típico é expressa nas variáveis  $\beta, V, N$  através da dependência dos níveis de energia com  $V$  e das ocupações com  $\beta\epsilon_i + \lambda(N)$ . Desta maneira, é necessário fazer a derivação implícita

$$\frac{\partial U}{\partial V} \equiv \frac{\partial}{\partial V} U(S(\beta, V, N), V, N) = \frac{\partial U}{\partial S} \frac{\partial S}{\partial V} + \frac{\partial}{\partial V} U(\beta, V, N), \quad (4.16)$$

ou ainda

$$\frac{\partial}{\partial V} U(S(\mathbf{n}), V, N(\mathbf{n})) = \frac{\partial U}{\partial S} \sum_i \frac{\partial S}{\partial n_i} \frac{\partial n_i}{\partial V} + \frac{\partial}{\partial V} U(\mathbf{n}, \epsilon). \quad (4.17)$$

Lembrando da relação de máximo para a entropia (4.4) e substituindo  $\frac{\partial U}{\partial S}$  pela temperatura, calculamos cada termo

$$\frac{\partial}{\partial V} U(\mathbf{n}, \epsilon) = \sum_i \frac{\partial U}{\partial n_i} \frac{\partial n_i}{\partial V} + \sum_i \frac{\partial U}{\partial \epsilon_i} \frac{\partial \epsilon_i}{\partial V} + \sum_{ij} \frac{\partial U}{\partial \epsilon_{ij}} \frac{\partial \epsilon_{ij}}{\partial V}, \quad (4.18)$$

---

<sup>1</sup>Sistemas extensivos, pela exigência que para  $\lambda$  arbitrário  $S(\lambda U, \lambda V, \lambda N) = \lambda S(U, V, N)$ , teriam um grau de liberdade a menos.



e ainda

$$\sum_i \frac{\partial S}{\partial n_i} \frac{\partial n_i}{\partial V} = \beta \sum_i \frac{\partial U}{\partial n_i} \frac{\partial n_i}{\partial V} + \lambda \sum_i \frac{\partial N}{\partial n_i} \frac{\partial n_i}{\partial V} = \beta \sum_i \frac{\partial U}{\partial n_i} \frac{\partial n_i}{\partial V}. \quad (4.19)$$

Substituindo estes termos e utilizando a notação anterior ficamos simplesmente com

$$p = - \left\langle \frac{\partial \epsilon_i}{\partial V} \right\rangle - \left\langle \frac{\partial \epsilon_{ij}}{\partial V} \right\rangle. \quad (4.20)$$

Existem várias argumentos que permitem escrever expressões deste tipo que relacionam os parâmetros microscópicos com as diversas variáveis termodinâmicas. Isto trás à luz alguns pontos válidos sobre a interpretação destes parâmetros. À medida que se comprime um sistema, a tendência é que os níveis energéticos se tornem mais afastados. Se as partículas permanecessem em seus estados originais seria necessário suprir esta energia adicional. A quantidade de energia a ser suprida para alimentar este processo é  $\langle \delta \epsilon_i \rangle + \langle \delta \epsilon_{ij} \rangle = p \delta V$  de sorte que a pressão determina a dificuldade em função do custo energético para que se altere o volume de um determinado sistema.

Poderia se pensar que, ao estabelecer relações como (4.20), todas grandezas termodinâmicas teriam uma interpretação microscópica convincente como é o caso da pressão. Isto não é verificado. Ainda que a temperatura eventualmente seja interpretada como “o grau de agitação das moléculas”, não é possível elaborar uma expressão tal qual a anterior que embase esta conclusão. Na realidade existem sistemas em que a temperatura *não* pode ser interpretada desta maneira, pois resulta em conclusões absurdas.

## 4.4 Calor e trabalho

Um dos resultados mais importantes da mecânica é que a energia, manifestada de diferentes maneiras, sempre se conserva. E um dos resultados mais importantes da termodinâmica é que estas diferentes manifestações não são livremente conversíveis entre si. Exemplos disto são a energia de movimento que se perde por atrito, a energia elétrica desperdiçada por efeito Joule, ondulações atenuadas devido à viscosidade e assim por diante. Energia que se transforma desta maneira não pode ser reaproveitada para realizar outras tarefas, é energia inútil. Para capturar esta noção introduzimos os conceitos de calor e trabalho.

Calor é uma variação da energia associada a um processo irreversível e, portanto, representa energia irremediavelmente “desperdiçada”. Desta forma, definimos calor

como a variação de energia devido à mudança na entropia do sistema

$$\delta Q \equiv \frac{\partial U}{\partial S} \delta S = T \delta S. \quad (4.21)$$

Trabalho representa simplesmente as outras formas de variação de energia que garantem a conservação ao final das contas. O trabalho é definido simplesmente pela relação de conservação, ou primeira lei da termodinâmica,

$$\delta U = \delta Q + \delta W. \quad (4.22)$$

É possível relacionar o trabalho com os parâmetros microscópicos do sistema em um processo termodinâmico qualquer. Considere que tal processo resulte em mudanças devido tanto à variação nas ocupações  $n_i$  quanto por variações nos níveis de energia. Desta forma, utilizando as propriedades do potencial de Massieu em (4.8), temos que

$$\delta S = \beta \delta Q = \beta \overline{\delta U} + U \delta \beta + \delta \Phi(\beta), \quad (4.23)$$

onde

$$\delta \Phi(\beta) = \frac{\partial \Phi}{\partial \beta} \delta \beta = -U \delta \beta. \quad (4.24)$$

A quantidade  $\overline{\delta U}$  corresponde à variação total de energia perdida irremediavelmente no processo. Nisto se contabiliza tanto alterações devido à redefinição dos níveis de energia quanto em mudanças na energia interna (aquela que é sempre conservada em sistemas isolados) expressa em termos da DNP. Utilizando a expressão acima concluímos que  $\overline{\delta U}$  é igual ao calor

$$\delta Q = \overline{\delta U}. \quad (4.25)$$

Explicitamos os termos referentes à variação de energia interna e trabalho implícitos na Primeira Lei

$$\overline{\delta U} = \delta U(\mathbf{n}, \epsilon_i, \epsilon_{ij}) = \sum_{i=1}^m \frac{\partial U}{\partial n_i} \delta n_i + \sum_{i=1}^m n_i \delta \epsilon_i + \sum_{i,j=1}^m n_i n_j \delta \epsilon_{ij} \quad (4.26)$$

$$= \delta U + \langle \delta \epsilon_i \rangle + \langle \delta \epsilon_{ij} \rangle = \delta U + \delta W. \quad (4.27)$$

De sorte a variação na energia interna é dada por

$$\delta U(\mathbf{n}) = \sum_{i=1}^m \frac{\partial U}{\partial n_i} \delta n_i = \delta \langle \epsilon_i \rangle + \delta \langle \epsilon_{ij} \rangle \quad (4.28)$$

e o trabalho conseqüentemente é definido por

$$\delta W(\epsilon_i, \epsilon_{ij}) \equiv - \langle \delta \epsilon_i \rangle - \langle \delta \epsilon_{ij} \rangle, \quad (4.29)$$

que é simplesmente variação da energia ocasionada pelo reposicionamento dos níveis de energia. Esta, portanto, é a variação que diz respeito às mudanças nos parâmetros macroscópicos que regulam o valor dos níveis de energia; assim se refere ao volume, campos eletromagnéticos, o número de partículas e outros. Deste modo, é possível escrever o trabalho como função das variações infinitesimais nas coordenadas termodinâmicas

$$\delta W = \frac{\partial U}{\partial V} \delta V + \dots = p \delta V \quad (4.30)$$

$$= \left\langle \frac{\partial \epsilon_i}{\partial V} \right\rangle \delta V + \left\langle \frac{\partial \epsilon_{ij}}{\partial V} \right\rangle \delta V + \dots, \quad (4.31)$$

onde na última linha utilizou-se o resultado (4.20) demonstrado na seção anterior.

## 4.5 Entropia à temperatura zero

Concluimos este capítulo analisando o comportamento da entropia na medida que a temperatura se anula. Esta demonstração da 3ª Lei é a única peça que falta para concluir que as leis da termodinâmica são decorrentes de raciocínio plausível acerca da dinâmica microscópica. É bem conhecido que a entropia de Boltzmann obedece à Terceira Lei, pelo menos no que se refere aos sistemas com estados energéticos enumeráveis. Mostraremos esta propriedade para entropias generalizadas, completando o quadro geral que fundamenta a termodinâmica sem fazer qualquer apelo específico à equiprobabilidade implícita na entropia de Boltzmann. Isto é importante porque trás a discussão da equiprobabilidade para um nível microscópico (que simetrias de movimento são violadas?) já que macroscopicamente ela não se manifesta de nenhuma maneira especial.

Considere, inicialmente, entropias locais e equiprováveis do tipo

$$H(\mathbf{n}) = \sum_i f(n_i) + \text{cte.} \quad (4.32)$$

Estas entropias podem ser classificadas em “bosônicas” ou “fermiônicas” pela maneira que  $f(x)$  se comporta afastado da origem.

As entropias bosônicas, são caracterizadas pelo fato que a função  $g(N) = \left. \frac{\partial f(x)}{\partial x} \right|_{x=N}$  tem um valor finito. Este é o caso da entropia de Boltzmann, onde  $-\frac{\partial}{\partial x} x \ln \frac{x}{N} = -\ln \frac{x}{N} - 1$  não diverge para nenhum valor de  $x$ . A condição de entropia máxima para

a temperatura nula,

$$\frac{\epsilon_i - \mu}{g(n_i)} = \frac{1}{\beta} = T = 0, \quad (4.33)$$

é satisfeita em duas situações. Na primeira, é necessário que  $g(n_i) \rightarrow \infty$ , o que segundo as considerações em (3.4) implica que este estado possui ocupação nula. A segunda maneira de satisfazer a igualdade é que  $\epsilon_i = \mu$ , indicando que a temperatura nula seleciona o estado com energia igual ao potencial químico. O potencial químico  $\mu$  é especificado pela relação de vínculo  $\sum_i n_i(\mu, T) = N$ , e neste sentido, não é uma variável independente de  $T$ . De fato, é possível mostrar que o estado selecionado à temperatura nula corresponde ao estado de mais baixa energia. Para verificar este fato, lembramos que os estados menos energéticos possuem ocupações maiores que os mais energéticos. Esta propriedade é requerida por construção, com o intuito de eliminar os níveis de energia infinita que certamente não são observados na natureza e consiste numa escolha apropriada para o sinal de  $\beta$ . Uma vez que em temperatura nula a DNP bosônica necessariamente colapsa em para um estado único, este deve ser o de mais alta ocupação já que os outros permanecem vazios.

As entropias fermiônicas são caracterizadas pela existência de uma divergência negativa em  $g(x)$  que é verificada a partir do ponto  $\hat{x}$  ( $g(\hat{x}) = -\infty$ ). Isto introduz uma possibilidade adicional em satisfazer a condição (4.33) —  $g(n_i) \rightarrow \pm\infty$  indica que  $n_i$  pode ser tanto nulo quanto igual à  $\hat{x}$ , a ocupação determinada pelo ponto de divergência de  $g(x)$ . É o potencial químico que dita que níveis de energia terão uma ocupação nula ou uma ocupação  $\hat{x}$  uma vez que regula o sinal da divergência em  $g(x)$ . Assim, é fácil determinar o valor específico do potencial químico a partir da condição de vínculo: basta selecionar os  $N/\hat{x}$  estados menos energéticos. Isto, é claro, especifica integralmente a DNP fermiônica à temperatura zero.

A existência de entropias fermiônicas, normalmente está associado à existência de restrições de contagem. A presença de divergência em  $g(x)$  implica que existe um limite intransponível na ocupação máxima de cada estado dado por  $\hat{x}$ . Isto generaliza esta propriedade bem conhecida da estatística fermiônica para o qual apenas 1 partícula pode se encontrar em cada estado no mesmo instante de tempo. Neste sentido, a ocupação associada à temperatura nula representa a menor energia que um sistema deste tipo pode obter. De maneira similar, o caso bosônico também verifica o fato que a temperatura nula implica que o sistema se encontra em sua configuração menos energética. Em ambos os casos, vale notar, existe apenas uma configuração microscópica associada ao estado fundamental.

Uma vez que à temperatura nula, independente das outras coordenadas termodinâmicas, o mesmo estado é sempre selecionado, a entropia  $S(\mathbf{n}_{T=0})$  tende a um valor fixo que pode ser ajustado por uma simples redefinição de constantes aditivas. Uma vez que as entropias generalizadas do tipo (1.45) são obtidas por transformações a partir das entropias locais aqui consideradas, elas possuiriam propriedades análogas em  $T = 0$ . Isto completa a demonstração da última propriedade necessária para que a prescrição da mecânica estatística esteja em total acordo com a termodinâmica usual.



## 5 *Estatística de Gibbs*

The ideal situation occurs when the things that we regard as beautiful are also regarded by other people as useful.

(DONALD KNUTH)

A metodologia tradicional utilizada na física estatística do equilíbrio se baseia numa prescrição muito simples. Uma vez de acordo com a representação microscópica utilizada para descrever o sistema, que pode ser clássica ou quântica, primeiramente se identifica as grandezas macroscópicas com a média sobre a distribuição de probabilidades de equilíbrio

$$U = \sum_x H(x)P(x|U) \quad (5.1)$$

e obtêm a distribuição  $P(x|U)$  pela maximização da entropia

$$H(P) = - \sum_x P(x|U) \ln P(x|U) \quad (5.2)$$

com um vínculo na energia (daí se inclui a informação condicional em  $U$ ). Apesar de quase ninguém discordar sobre a validade deste algoritmo, nem todos estão de acordo sobre *porque* ele funciona.

A justificativa mais simples para eficiência da física estatística provavelmente foi fornecida por E. T. Jaynes ao interpretar o processo de maximização da entropia como uma metodologia de inferência estatística. Neste sentido, o algoritmo especificado acima simplesmente considera em nossas atribuições de probabilidade uma das constatações mais gerais sobre a dinâmica dos sistemas físicos, que é a conservação da energia. É claro que não se espera que somente esta lei de conservação forneça um panorama geral sobre a dinâmica de sistemas macroscópicos: a distribuição de posições das partículas, suas velocidades, os fluxos de matéria, tudo isso certamente faria parte de um quadro mais geral. É de certo modo surpreendente que para caracterizar os estados de equilíbrio, uma especificação tão insignificante — a energia do sistema —

parece ser o suficiente.

A maximização da entropia utilizando este vínculo fornece o conhecido resultado

$$P(x|U) = \frac{1}{Z(\beta)} e^{-\beta H(x)}; \quad Z(\beta) = \sum_x e^{-\beta H(x)}, \quad (5.3)$$

onde, pelos mesmos motivos mostrados anteriormente,  $\beta$  faz o papel da temperatura e a constante de normalização  $Z(\beta)$  determina o potencial de Massieu pela transformada de Legendre

$$S(U, V, N) = \beta U + \ln Z(\beta, V, N). \quad (5.4)$$

Note que todas as propriedades com relevância para a termodinâmica de entropias generalizadas valem neste caso particular — é fácil ver isto dada a semelhança formal entre as duas entropias. O esquema de Jaynes/Gibbs, à semelhança do que foi mostrado no capítulo anterior, é compatível com a termodinâmica. Ele difere da metodologia desenvolvida ao longo deste trabalho no que enfatiza a especificação do micro-estado enquanto até então se utilizou a *distribuição de densidades* sobre os estados associados a *partículas*. É uma questão de gosto pessoal, mas talvez a segunda abordagem ofereça um panorama mais convincente e mais fácil visualização dos argumentos. Fora isto, ela explicita um ponto crucial para entender vários dos dilemas e falsos dilemas que se apresentam à física estatística pois leva em conta a distinção qualitativa entre os níveis microscópicos e macroscópicos. Em todo caso, estando os dois esquemas em concordância qualitativa com a termodinâmica, será que ambos também fornecem as mesmas respostas numéricas?

## 5.1 Equivalência entre os formalismos

Para verificar mais detalhadamente a equivalência entre os dois formalismos, considere um sistema simples formado por um único tipo de partícula com a Hamiltoniana dada por

$$U(x) = \sum_{i=1}^N T(i) + \sum_{i,j=1}^N V(i, j), \quad (5.5)$$

onde  $x$  representa o estado microscópico do sistema e as coordenadas  $i, j, \dots$  representam o estado de cada partícula.

Esta Hamiltoniana é claramente simétrica por permutação de partículas. Isto nos permite modificar a representação do sistema de  $x$ , que especifica que partícula está



em cada estado, para a distribuição  $\mathbf{n}$  que determina quantas partículas  $n_i$  estão em cada estado  $i$ . Assumindo que cada molécula pode estar em um estado enumerado de 1 à  $m$ , a Hamiltoniana se escreveria como

$$U(\mathbf{n}) = \sum_{k=1}^m T(i)n_i + \sum_{i,j=1}^m V(i,j)n_i n_j \quad (5.6)$$

reduzindo a dependência com respeito ao estado à dependência nos números de ocupação ou DNPs em (3.7). Deste modo, podemos utilizar o método da máxima entropia para derivar a probabilidade de ocorrência de cada ocupação, tornando mais fácil a comparação entre os resultados de Gibbs e os mostrados anteriormente. Uma sutileza, no entanto, deve ser levada em conta porque esta representação, ao contrário da representação no espaço de fase, não é equiprovável. Sendo o número de diferentes configurações microscópicas associadas à mesma configuração macroscópica é dada pelo fator de multiplicidade, é necessário utilizar a probabilidade anterior

$$P(\mathbf{n}) = \frac{N!}{n_1! n_2! \dots n_m!} \frac{1}{N^m} \propto e^{H_1(\mathbf{n})} \quad (5.7)$$

onde  $H_1(\mathbf{n})$  é a entropia de Boltzmann definida em (3.34). Deste modo, a prescrição de Gibbs consiste em maximizar a entropia

$$H(P) = \sum_{\mathbf{n}} P(\mathbf{n}) \ln \frac{P(\mathbf{n})}{e^{H_1(\mathbf{n})}}, \quad (5.8)$$

sujeita a vínculos, sendo que a sua extensão para entropias generalizadas é imediata, basta trocar  $H_1(\mathbf{n})$  pela entropia efetiva considerada. Esta fórmula faz a ligação entre o formalismo de Gibbs com o formalismo anterior associado à contagem e atribuição de probabilidades às DNPs.

Desta forma, a probabilidade de cada configuração é dada por

$$P(\mathbf{n}|U) = \frac{1}{Z(\beta)} e^{-\beta U(\mathbf{n}) + H(\mathbf{n})}; \quad Z(\beta) = \sum_{\mathbf{n}} e^{-\beta U(\mathbf{n}) + H(\mathbf{n})}. \quad (5.9)$$

Se o estado mais provável  $\bar{\mathbf{n}}$  for estatisticamente preponderante sobre os outros,  $P(\bar{\mathbf{n}}|U) \simeq 1$ , é possível substituir as médias sobre a distribuição pelo valor da grandeza considerada em  $\bar{\mathbf{n}}$ . Neste caso,

$$\langle Y \rangle = \sum_{\mathbf{n}} P(\mathbf{n}|U) Y(\mathbf{n}) \simeq Y(\bar{\mathbf{n}}). \quad (5.10)$$

Para calcular a ocupação mais provável, basta maximizar  $P(\mathbf{n}|U)$ , isto é, encontrar

o máximo de  $H(\mathbf{n}) - \beta U(\mathbf{n})$ . Note que esta é exatamente a mesma prescrição para encontrar os estados de equilíbrio fornecida pelo formalismo anterior. Desta forma, basta que as distribuições de probabilidade resultantes sejam muito concentradas em torno de um valor, o formalismo de Gibbs para o equilíbrio e o formalismo apresentado anteriormente coincidem.

### 5.1.1 Método do ponto de sela

Uma maneira muito interessante de fundamentar a estatística de Gibbs é dada pelo método do ponto de sela, *Sattelpunktmethode*, elaborado por Darwin e Fowler. Este método foi elaborado para calcular as médias associadas a grandezas macroscópicas,

$$\langle \Psi \rangle = \sum_{[\mathbf{n}]} P(\mathbf{n}) \Psi(\mathbf{n}), \quad (5.11)$$

restrita ao vínculo da energia. A solução do problema envolve uma certa dose de considerações engenhosas e virtuosismo matemático que, à despeito do tecnicismo, resulta numa demonstração muito elegante. Mostramos que, na medida que aumenta o número de partículas, é possível substituir a média microcanônica pelas médias canônicas dadas por

$$\langle \Psi \rangle = \sum_{[\mathbf{n}]_{NU}} P(\mathbf{n}) Y(\mathbf{n}) = \sum_{[\mathbf{n}]_N} \frac{1}{Z(\beta)} e^{-\beta U(\mathbf{n})} \Psi(\mathbf{n}), \quad (5.12)$$

onde a notação  $[\mathbf{n}]_{NU}$  especifica que o somatório está restrito a valores para o número de partículas e energia.

Define-se  $\Psi_U$  como o valor médio da grandeza  $\Psi(\mathbf{n})$  calculado sobre uma distribuição de probabilidades que possua a forma geral (3.32) restrita a um valor específico de energia, ou seja

$$\Psi_U \equiv \sum_{[\mathbf{n}]_{N,U}} \frac{N!}{n_1! \dots n_m!} \gamma(\mathbf{n}) \psi(\mathbf{n}). \quad (5.13)$$

Admita que existe uma expansão em série de potências para a função  $\chi(\mathbf{n}) \equiv \gamma(\mathbf{n}) \psi(\mathbf{n})$ . Define-se o operador  $\chi(\partial_\eta)$  pela troca da dependência nas coordenadas do vetor de ocupação pelas derivadas em  $\eta_i$ , isto é

$$\chi(\partial_\eta) = \chi_0 + \sum_{i=1}^m \chi_i \partial_{\eta_i} + \sum_{i,j=1}^m \chi_{ij} \partial_{\eta_i} \partial_{\eta_j} + \dots \quad (5.14)$$

Então é fácil mostrar que a função  $\chi(\mathbf{n})$  também se escreve como a aplicação

$$\chi(\mathbf{n}) = \chi(\partial_\eta) \prod_i e^{\eta_i n_i} \Big|_{\eta=0}, \quad (5.15)$$

de sorte que é possível passar a dependência na função  $\chi(x)$  em (5.13) para fora do somatório. Para mostrar este resultado, basta definir a função

$$G_{NU} \equiv \sum_{[\mathbf{n}]_{N,U}} \frac{N!}{n_1! \dots n_m!} \prod e^{\eta_i n_i}, \quad (5.16)$$

onde fica claro que a média  $\Psi_U$  pode ser escrita simplesmente como

$$\Psi_U = \chi(\partial_\eta) G_{NU} \Big|_{\eta=0}. \quad (5.17)$$

A função  $G_{NU}$  pode ser avaliada aproximadamente nos sistemas onde  $U$  obedece a uma dependência linear com o número das partículas. Isto restringe a aplicabilidade do método a sistemas ideais ou aproximadamente ideais, mas as considerações feitas em 3.6 indicam que esta exigência pode não ser tão restritiva quanto parece. Definimos a função

$$G_N(z) \equiv \sum_{[\mathbf{n}]_N} \frac{N!}{n_1! \dots n_m!} \prod_{i=1}^m e^{\eta_i n_i} z^{U(\mathbf{n})-U-1}; \quad z \in \mathbb{Z}, \quad (5.18)$$

que *não* está restrita por um vínculo no valor da energia. Admitindo uma escala de energias onde os números  $U(\mathbf{n})$  e  $U$  possam ser considerados inteiros, a expressão anterior consiste na expansão em série de Laurent da função  $G_N(z)$  em torno da origem. Ainda que a notação não torne isto explícito, os coeficientes da expansão podem ser calculados simplesmente colecionando todos termos com um valor determinado para  $U(\mathbf{n}) - U - 1$ . Cada um destes coeficientes corresponde aos valores de  $G_{NU}$  associados a diferentes folhas de energia —  $G_N(z)$  é, em suma,

$$G_N(z) = G_{NU_0} z^{U_0-U-1} + G_{N(U_0+1)} z^{U_0-U} + \dots + G_{NU} z^{-1} + \dots \quad (5.19)$$

A nós interessa calcular o termo onde  $U = U(\mathbf{n})$  ou, de maneira análoga, o termo de ordem  $-1$  em  $z$  na expansão acima. Este termo corresponde ao resíduo de  $G_N(z)$  que é dado pela integral de caminho

$$\frac{1}{2\pi i} \oint G_N(z) dz = \text{res} G_N(z) = G_{NU}. \quad (5.20)$$

Uma avaliação precisa desta integral pode ser complicada, mas é possível calculá-

la com boa aproximação no limite em que  $N \gg 1$ . Considere que a energia  $U(\mathbf{n})$  seja dada por  $\sum_i n_i \epsilon_i$ , de modo que

$$G_N(z) = \frac{1}{z^{U+1}} \sum_{[\mathbf{n}]_N} \frac{N!}{n_1! \dots n_m!} \prod_i (e^{\eta_i} z^{\epsilon_i})^{n_i} \quad (5.21)$$

onde nesta passagem usamos a fórmula para expansão multinomial  $(\sum_{i=1}^m a_i)^N = \sum_{[\mathbf{n}]_N} \frac{N!}{n_1! \dots n_m!} a_1^{n_1} \dots a_m^{n_m}$ .

Uma vez que todos fatores  $e^{\eta_i}$ 's são positivos,  $\phi(z) \equiv \sum_i e^{\eta_i} z^{\epsilon_i}$  e  $\frac{1}{z^{u+1}}$  devem ser funções respectivamente monotônicas crescentes e decrescentes de  $z$  sobre o eixo real positivo. A multiplicação destas funções — que fornece essencialmente  $G_N(z)$  — possui somente um mínimo neste domínio, o qual denotamos por  $z_0$ . Este número pode ser calculado pela extremização de  $G(z)$ , ou do seu logaritmo,

$$\frac{\partial}{\partial z_0} \ln G_N(z_0) \equiv g'(z_0) = 0. \quad (5.22)$$

Seja um trajeto circular em torno da origem; na escala de energias mencionada em que todos  $\epsilon_i$ 's são os inteiros, a função  $\phi(r e^{i\theta})$  é nada mais que a representação em série de Fourier de uma função periódica em  $\theta$ . Deste modo,  $\phi(\theta)$  possui ao menos um máximo acentuado em  $\theta = 0$ , onde a soma dos fatores positivos  $e^{\eta_i}$  contribui construtivamente, e eventualmente existiriam outros máximos se o período associado à variável  $\theta$  for menor que  $2\pi$ . Dado um  $N$  suficientemente grande, a maior parte da integral de caminho corresponderia a pontos vizinhos destes máximos quando  $\phi(\theta)$  é levado à  $N$ -ésima potência.

Calcularemos a integral sobre um círculo do raio  $z_0$  por um motivo que ficará óbvio logo em seguida. Usando a denominação de  $\hat{G}_{NU}$  para a integral sobre um pequeno arco nas proximidades de  $z_0 + i0$  para distinguir da integral realizada sobre todo círculo e, ao considerar pequenos deslocamentos imaginários  $z_0 \rightarrow z_0 + iy$ , o integrando se torna

$$G_N(z_0 + iy) = e^{g(z_0) + ig'(z_0)y - \frac{1}{2}g''(z_0)y^2 + O(y^3)}. \quad (5.23)$$

Recordando da condição mínimo, somos levados à uma integral gaussiana pura

$$\hat{G}_{NU} = \frac{1}{2\pi i} \frac{\phi(z_0)^N}{z_0^{B+1}} \int_{-\delta y}^{\delta y} e^{-\frac{1}{2}g''(z_0)y^2} i dy, \quad (5.24)$$

onde mostra-se que o fator  $g''(z_0)$ , o inverso da dispersão associada à esta gaussiana,

se escreve explicitamente como

$$g''(z_0) = N \left[ \frac{\phi''(z_0)}{\phi(z_0)} - \left( \frac{\phi'(z_0)}{\phi(z_0)} \right)^2 \right] + \frac{U+1}{z_0^2}. \quad (5.25)$$

Uma vez reconhecido que  $g''(z_0)$  é uma função extensiva de  $N$ , o que é facilmente verificado pelo resultado anterior, deve existir um  $N$  suficientemente grande para que seja possível trocar os limites de integração em (5.24) de  $y \in [\mp \delta y]$  para  $y \in [\mp \infty]$ , na medida que a Gaussiana se aproxima de uma delta de Dirac. Assim  $\hat{G}_{NU}$  se escreveria segundo a forma fechada

$$\hat{G}_{NU} = \frac{\phi(z_0)^N}{z_0^{U+1}} \sqrt{\frac{1}{2\pi g''(z_0)}}. \quad (5.26)$$

Se o período de  $\phi(\theta)$  for um múltiplo de  $2\pi$ , a integral exibiria diversos picos que podem ser avaliados de uma maneira essencialmente igual ao resultado acima. Podemos eliminar estes picos por uma escolha apropriada de escala energética. Lembrando que assumiu-se que os valores de  $\epsilon_i$  eram números inteiros, o fato que  $\phi(\theta)$  realiza  $C$  ciclos no período de variação  $2\pi$  indica simplesmente que a escala escolhida possui um divisor comum  $C$ . Redefinimos a escala para que isto não aconteça de forma que a integral completa em torno do círculo de raio  $z_0$  é dada simplesmente por

$$G_{NU} = \text{cte} \times \frac{\phi(z_0)^N}{z_0^{U+1} \sqrt{g''(z_0)}}. \quad (5.27)$$

É possível determinar esta constante a partir da condição de normalização das probabilidades, simplesmente exigindo que

$$\gamma(\partial_\eta) G_{NU} |_{\eta=0} = 1. \quad (5.28)$$

Isto completa a demonstração do método de Darwin e Fowler para somatórias sujeitas a vínculos.

### 5.1.2 Entropia de Gibbs

Uma vez de acordo com a metodologia empregada para realizar as somas sujeitas a vínculos, podemos utilizar o método anterior para calcular a representatividade dos estados macroscópicos e, a partir daí, definir a entropia dos mesmos. Considere atribuições

de probabilidade

$$P(U) = \sum_{[\mathbf{n}]_{N,U}} P(\mathbf{n}) = \sum_{[\mathbf{n}]_{N,U}} \frac{N!}{n_1! \dots n_m!} \gamma(\mathbf{n}). \quad (5.29)$$

Utilizando o método do ponto de sela, calculamos o resultado anterior que é, simplesmente,

$$\sum_{[\mathbf{n}]_{N,U}} P(\mathbf{n}) \propto \frac{1}{z^{U+1}} \gamma(\partial_\eta) \left[ \phi(z_0) g''(z_0)^{-\frac{1}{N}} \right]^N. \quad (5.30)$$

Para um valor suficientemente grande de  $N$ , o termo  $g''(z_0)^{\frac{1}{N}}$  pode ser desprezado no cálculo de  $\ln \sum_{[\mathbf{n}]_{nu}} p(\mathbf{n})$ . Empregando esta aproximação e definindo a constante  $z_0 = e^{-\beta}$  que expressa mais adequadamente o fato de  $z_0$  ser positivo, podemos escrever o termo da esquerda como

$$\gamma(\partial_\eta) \phi(z_0)^N = \sum_{[\mathbf{n}]_N} \frac{N!}{n_1! \dots n_m!} \gamma(\mathbf{n}) e^{-\beta U(\mathbf{n})} \quad (5.31)$$

$$= \sum_{[\mathbf{n}]_N} e^{H(\mathbf{n}) - \beta U(\mathbf{n})}, \quad (5.32)$$

onde  $H(\mathbf{n})$  é a entropia generalizada associada à DNP  $\mathbf{n}$ . Definindo  $Z = \sum_{[\mathbf{n}]_N} e^{H(\mathbf{n}) - \beta U(\mathbf{n})}$ , temos finalmente que, fora uma constante irrelevante,

$$S(U, V, N) \equiv \ln P(U, V, N) = \beta U + \ln Z(\beta, V, N). \quad (5.33)$$

Este resultado é, obviamente, idêntico ao processo de maximização da entropia de Gibbs mostrado anteriormente. O raciocínio mostra que a entropia de Gibbs *maximizada* pode ser corretamente interpretada como o logaritmo da representatividade de um estado macroscópico (aquele utilizado com vínculo) e, portanto, realiza exatamente o mesmo papel da entropia de Boltzmann definida anteriormente como

$$S(U, V, N) \equiv k_B \ln P(U, V, N). \quad (5.34)$$

Um ponto em que devemos ser cuidadosos é que, ainda que este resultado seja correto em um domínio estrito, algumas aproximações foram utilizadas extensivamente para obter o resultado. O teorema só é verdadeiro se  $N \gg 1$ , mas o que é mais dramático, somente se a energia for uma função linear da ocupação de cada estado, i.e. o resultado está restrito às situações em que a energia é livre. Ainda que não conseguiu-se generalizar a demonstração para as situações em que há interação entre partículas, o inverso também não parece óbvio: que a presença de interação necessariamente

invalidaria o resultado.

Uma solução heurística dada pelo paralelo entre o panorama de Gibbs e o panorama de Boltzmann expresso pela fórmula da entropia (5.34). A presença de termos quadráticos na energia pode implicar na proliferação de vários máximos locais no problema de maximização. Isto, geralmente, não representa um problema. A diferença de probabilidades se torna cada vez mais pronunciada na medida em que aumenta  $N$  de sorte que, para um  $N$  o suficientemente grande, apenas o máximo global seria relevante. Explorando a possibilidade que no método Darwin e Fowler o mesmo ocorreria na integral de caminho em volta do círculo de raio  $z_0$ , de forma que  $N \gg 1$  justificaria considerar apenas o máximo associado à  $z \simeq z_0 + i0$ . A exceção está, obviamente, nos pontos relacionados à transição de fase em que existiria pelo menos dois máximos de entropia de valores equivalentes.

Existem basicamente dois meios por onde a inclusão de termos não-lineares na energia poderia violar o método de Darwin e Fowler. O primeiro deles seria se estes termos implicassem que o valor de  $g''(z_0)$ , ao contrário do que aparece na fórmula (5.25), não cresça o suficientemente rápido para que seja possível fazer a integral gaussiana em (5.24). Este cenário não parece plausível, principalmente pela presença de um termo de energia  $\frac{U+1}{z_0^2}$  que, mesmo para as dependências mais esdrúxulas de  $\phi(z_0)$  com  $N$  ainda poderia salvar o dia. Isto leva ao segundo motivo em que a demonstração poderia falhar.

Este caso correspondente à situação em que existem vários máximos equivalentes. Isto pode vir desde a existência de máximos comparáveis no círculo de integração em torno de  $z_0$ , à existência de vários pontos de sela no eixo real. O segundo caso é irrelevante já que todos estes caminhos de integração devem necessariamente fornecer o mesmo valor uma vez que contornam o mesmo resíduo. A primeira situação, no entanto, é potencialmente problemática. Sabemos que existe um máximo local em  $z = z_0 + i0$ , mas a não linearidade de  $\phi(\theta)$  torna a procura pelos outros máximos bastante problemática. Interpretamos esta proliferação de máximos como reflexo do que acontece com a maximização da entropia de Boltzmann, mas ainda sem poder demonstrá-lo. Não está claro também se o resultado final, supondo que possa ser obtido, seria igual à entropia de Gibbs. Consideramos estas especulações como motivo de análise para um trabalho futuro.

## 5.2 Teoria fora do equilíbrio

Uma aparente dificuldade reside em se utilizar o esquema de Gibbs em situações fora do equilíbrio. O motivo para isto não está relacionado à inexistência de uma equação dinâmica para a probabilidade  $P(x|U)$  — esta equação existe —, mas decorre do fato que os resultados obtidos parecem, a uma primeira vista, absurdos. A evolução para as probabilidades é dada pela equação de Liouville-von Neumann, que é derivada a partir da dinâmica microscópica de maneira provavelmente muito mais elegante que todo o formalismo exposto no capítulo 3. O problema é que, de acordo com esta equação, a entropia de Gibbs é um invariante de movimento e portanto nunca cresce. A irreversibilidade da dinâmica macroscópica, que é a pedra fundamental para a eficácia do conceito de entropia, não é observada na equação de Liouville-von Neumann — assim todo arcabouço conceitual que justifica a termodinâmica é posto imediatamente em xeque.

Há como se livrar destas objeções?

A solução para o dilema é razoavelmente simples. O fato da entropia de Gibbs ser estática, ao invés de ser uma violação do senso comum, é o comportamento esperado. A entropia não só mede a nossa capacidade de obter informação sobre o sistema, mas também a quantidade de informação necessária para prepará-lo em um determinado estado (especificado por um conjunto de vínculos). Neste espírito, a evolução irreversível não é necessariamente representada por um  $H(P)$  que cresce no tempo; uma vez preparada uma configuração macroscópica inicial, o fato que a evolução Hamiltoniana conserva volume implica que o conhecimento sobre o sistema físico não é modificado substancialmente — o número de configurações microscópicas permitidas não muda e junte isso ao fato que a evolução é reversível, o conhecimento sobre a condição inicial sequer se deteriora.

No entanto isto não implica que a nossa capacidade de *inferir* o estado microscópico permaneça a mesma. Ainda que, em tese, poderíamos determinar a região do espaço de fase associada à evolução de uma configuração macroscópica, esta informação não constitui informação mensurável no laboratório. A equação de Liouville determina a evolução do volume de fase e conseqüentemente das médias macroscópicas  $F(t) = \langle F(x) \rangle_t$ . Isto pode dar a ilusão que a equação de Liouville também determina a evolução das grandezas macroscópicas — isto é, dado o valor inicial de  $F(0)$ , determinaria a trajetória  $F(t)$ . Isto não poderia ser mais falso. À cada especificação  $\langle F(x) \rangle =$



$F(0)$ , existem infinitas distribuições de probabilidade que, mesmo satisfazendo este critério na condição inicial, fornecem valores distintos para  $F(t)$ . A análise correta deste problema é feita no capítulo 2.

Do ponto de vista da entropia de Gibbs, também é relativamente fácil escapar deste problema. Primeiro temos que reconhecer que um volume de fase arbitrário não representa uma informação macroscópica muito útil. Não há laboratório com resolução suficiente para escrutinar todos meandros do espaço de fase e determinar este volume experimentalmente<sup>1</sup>. Isto leva ao segundo ponto do argumento: informação útil é *somente* informação macroscópica. Para definir uma entropia com utilidade operacional, seria necessário avaliar que configurações macroscópicas corresponderiam aos pontos encerrados neste volume. Não é possível preparar experimentalmente um volume de fase arbitrário, mas é razoável que se possa reproduzir as configurações macroscópicas a ele associadas e posteriormente determinar o volume de fase que elas correspondem..Note que o segundo volume de fase pode ser bem maior que o primeiro, mas nunca aconteceria o contrário. Uma vez que a nossa capacidade de medição se refere ao segundo caso, e não ao primeiro, a entropia encontrada em laboratório seria necessariamente maior ou igual à entropia calculada a partir da equação de Liouville.

Assim como bem enfatiza Jaynes, a questão de qual destas entropias seria a entropia “física” ou entropia “ontológica” é completamente imaterial. Uma vez que se aceita o caráter probabilístico do conceito, a especificação da entropia possui um caráter subjetivo (ou antropomórfico, contextual, como se queira): ela especifica a informação microscópica associada a um certo conhecimento macroscópico. E se, como bons cientistas, quisermos confrontar as teorias com os dados experimentais é melhor esquecer a evolução da entropia dada pela equação de Liouville. Neste sentido, definições úteis de entropia, irreversibilidade macroscópica e, em última instância a termodinâmica dependem impossibilidade em determinar o estado microscópico experimentalmente. Caso isso fosse possível, não só a imagem de mundo seria completamente diferente, mas também os conceitos “macroscópicos” seriam desnecessários e fundamentalmente misteriosos. Este também é o motivo que qualquer programa de “primeiros princípios” para fundamentar a termodinâmica a partir das equações mecânicas, esquecendo esta limitação incontornável, está fadado a um fracasso retumbante.

---

<sup>1</sup>Uma metáfora bastante repetida é que os efeitos de incerteza na mecânica quântica aparecem porque tentamos medir bolas de bilhar com outras bolas de bilhar. A situação em que se encontra a física estatística é muito mais dramática. Uma metáfora mais apropriada para ilustrar a tentativa de determinar o estado microscópico utilizando instrumentos macroscópicos seria como medir a posição desta mesma bolha de bilhar arremessando-lhe o sistema solar!

Um programa para a termodinâmica de não-equilíbrio que se baseie na teoria de Gibbs deve incorporar elementos muito mais sutis que, por exemplo, a dinâmica da equação de Liouville-von Neumann. Esta equação não possui pontos fixos e, mais ainda, é trivial mostrar que a distribuição canônica  $Z(\beta)^{-1} \exp(-\beta H)$  é uma constante de movimento. Isto, é claro, somente expressa o fato que esta equação não assume limites na nossa capacidade de apreender informação sobre o mundo. Considerar estes limites, no entanto, não é uma tarefa simples.

### 5.2.1 Médias de grão grosseiro

Um procedimento canônico para incorporar a irreversibilidade na equação de Liouville-von Neumann é conhecido por médias de grão grosseiro. Consiste em discretizar o espaço de fase em células de tamanho finito e, a partir daí, expressar a evolução das probabilidades na forma de uma equação mestra. A surpresa está em que, em algumas versões discretizadas, a equação de Liouville-von Neumann torna-se irreversível. A justificativa para proceder desta maneira é a já mencionada imprecisão das medidas realizadas com instrumentos macroscópicos.

Alguns pontos devem ser esclarecidos para tornar o argumento convincente. O primeiro é que o procedimento, em princípio, não contradiz nada conhecido sobre a dinâmica microscópica. Esta irreversibilidade está associada à nossa capacidade de predição do estado do sistema, não à existência de um novo fator dinâmico. Neste sentido, não há problema que a dinâmica dependa, por exemplo, de fatores arbitrários como o tamanho dos “grãos grosseiros”. Isto apenas expressa o fato que arranjos experimentais com resoluções distintas permitem fazer inferências mais ou menos detalhadas. Um estado de conhecimento de grão muito grosseiro se deteriora rapidamente na medida em que as trajetórias compatíveis com uma certa medição inicial diverjam entre si. No outro extremo, os estados de grão muito fino, selecionam uma única trajetória. Por isso, nunca se deterioram já que é possível acompanhar esta trajetória indefinidamente. A irreversibilidade já não se colocaria pois o sistema permaneceria numa exploração contínua do espaço de fase.

Este argumento não parece suficiente para justificar a irreversibilidade macroscópica (que é muito melhor exposta, por exemplo, em 2.2). Além disso ele dá a falsa impressão de que é necessário modificar as equações dinâmicas para dar conta da irreversibilidade macroscópica, isto não é verdade. Entendendo os grãos grosseiros como um simples reflexo da dificuldade de acesso ao estado macroscópico, ele chega ao ponto crucial

da questão: irreversibilidade não está na natureza, mas sim em nossas limitações ao observá-la.

## 5.2.2 Formalismo MaxEnt fora do equilíbrio

Ainda que o método da máxima entropia de Gibbs seja amplamente utilizado para descrever o equilíbrio, a extensão para situações de não-equilíbrio de acordo com esquema conceitual “informacional” proposto por Jaynes, é imediata. Para generalizar a teoria para estes regimes, é necessário incorporar informação dinâmica à distribuição de probabilidades — como, por exemplo, informação do tipo  $\langle F(x) \rangle = F(t)$ . Isto é feito da mesma maneira que o caso anterior pela prescrição

$$P(x|F(t)) = \frac{1}{Z(\sigma(t))} e^{-\int dt \sigma(t)F(x)}; \quad Z(\sigma(t)) \equiv \sum_x e^{-\int dt \sigma(t)F(x)}. \quad (5.35)$$

À partir da informação sobre a evolução de uma determinada grandeza,  $\langle F(x) \rangle = F(t)$ , o método da máxima entropia permite fazer inferências sobre o desenvolvimento temporal de outras grandezas. Note que a evolução da probabilidade no tempo é fornecida pela designação da função  $F(t)$  e não é consequência da equação de Liouville-von Neumann. A irreversibilidade pode estar contemplada, ou não, de acordo com o processo associado a  $F(t)$ .

Esta extensão elegante do formalismo de equilíbrio foi utilizada, de fato, para prever o comportamento irreversível de alguns processos como é mostrado em [?]. Com destaque está uma derivação particularmente elegante das equações hidrodinâmicas. Neste caso, a prescrição de maximizar a entropia sujeita a um vínculo na evolução da distribuição de energia e outro no valor conhecido para a evolução das densidades de partículas implica no aparecimento dos multiplicadores de Lagrange com interpretação física de temperatura e termos de fluxo de corrente. Daí, uma matemática elaborada leva às leis usuais da teoria fenomenológica.

Vários paralelos podem ser feitos entre a teoria desenvolvida no capítulo 3 com a teoria de Jaynes/Grandy. Do ponto de vista conceitual, a ênfase de ambos é interpretar a dinâmica macroscópica como um esquema de inferência a partir da pouca informação microscópica disponível. Desta maneira, é necessário sistematizar o procedimento pelo qual se incorpora informação e elaborar modelos que utilizem esta informação de maneira construtiva. Nenhuma destas teorias se propõe a ser *a teoria definitiva*, elas são modelos plausíveis que fornecem um esquema geral para elaborar descrições

macroscópicas tão refinadas quanto a informação disponível possibilita — e *somente* o quanto ela possibilita. Uma investigação promissora prorrogada para um trabalho futuro é analisar até que ponto a teoria de Jaynes se reflete no esquema esboçado no capítulo 3. Não parece existir nenhum impedimento para que as duas visões eventualmente sejam unificadas.

# Conclusão

Yeah! uh! I am a scientist  
 Yeah! uh! I am a scientist  
 Yeah! uh! I am a scientist  
 We've gotta live on science alone

(The Dandy Warhols - I am a Scientist)

## 5.1 Uma pequena digressão

Em uma das citações mais impressionantes de toda tradição científica, Galileu define a atitude do físico, em oposição ao eclesiástico, como:

A filosofia está escrita neste enorme livro que continuamente está aberto diante de nossos olhos (me refiro ao o universo), mas que não se pode compreender sem que antes se aprenda a língua e conheça os caracteres no qual está escrito. Ele é escrito em linguagem matemática, e os caracteres são triângulos, círculos, e outras figuras geométricas, sem as quais é impossível entender uma só palavra; sem as quais caminha por um labirinto escuro.

A visão muito difundida que a matemática consiste na principal, se não na única, ferramenta para compreender a linguagem da Natureza possivelmente teve um certo impacto negativo na física do século *XX*. Não se trata de qualquer limitação inerente às matemáticas, nem mesmo da ingenuidade geral do ponto de vista filosófico implícito nesta expectativa — estas são opiniões que, sinceramente, o autor não compartilha. É possível que tais pretensões em reescrever o Grande Livro do próprio punho resultaram no obscurecimento de um problema um pouco mais mundano, mas igualmente importante, que é a maneira como se dá a aquisição de informação sobre a Natureza e como expressá-la em nossas teorias. A resposta de Galileu, que o “livro [da natureza] é interpretado e lido apenas pela experiência”, consiste fundamentalmente no tipo de crença mais difundida nos dias de hoje.

Certamente não há como questionar a essência da posição de Galileu, mas antes alguns fatos básicos devem ser reconhecidos. O primeiro é que existem limites práticos

à nossa capacidade de apreensão do mundo. Isto será sempre verdade a despeito de todo progresso técnico da humanidade. O segundo é que as conclusões úteis de uma teoria só podem ser formuladas a partir de parâmetros que podemos observar, apreender e, idealmente, manipular. Muitas das questões mais pertinentes para a nossa vida não são expressas da maneira mais natural, ou de maneira alguma, no Grande Livro. Na verdade, a maior parte da ciência está condicionada por fatores acidentais tanto de origem histórica (como biologia, geologia, etc) quanto devido a certas limitações de nossos experimentos (termodinâmica, química, etc) além das ciências em que o elemento humano é o centro do discurso (economia, medicina, etc). Em nenhum destes casos uma Lagrangiana a vários parâmetros, mesmo que expresse o universo como solução particular, permitiria sequer formular as perguntas mais interessantes.

É importante ressaltar isto porque, particularmente em termoestatística, existe uma expectativa muito forte em derivações de “primeiros princípios”, onde se esquece facilmente de algum destes aspectos para que se chegue a resultados pré-definidos. Neste sentido, o presente trabalho reitera a linha de pensamento associada às figuras de Laplace, Maxwell, Boltzmann, Gibbs, Planck e, finalmente, Jaynes, entre outros. A tese central é que a termodinâmica decorre, quase que integralmente, da simples identificação entre entropia e a probabilidade (representatividade) relacionada ao estado macroscópico, sendo que a justificativa para isto é dada, em última instância, pelo reconhecimento formal das limitações expostas acima. Acreditamos que, uma vez que o modelo estatístico esteja bem posto, extensões recentes sob os nomes de “não-extensividade” [?] ou uma “irreversibilidade mecânica irreduzível” [?] parecem um tanto quanto vazias.

Laplace provavelmente foi o fundador do tipo de teoria de probabilidades necessária para tratar de algumas das limitações expostas acima. Boltzmann foi um pioneiro não só pela defesa da hipótese atômica em física, mas também por mostrar o papel da probabilidade em termodinâmica em sua fórmula emblemática. Questões que incomodavam os físicos de sua época, como a aparente incompatibilidade entre o crescimento monotônico da entropia e a reversibilidade das equações fundamentais foram logo esclarecidas. Reforço a deferência a este grande mestre com as palavras de Jaynes, “Cem anos no futuro, a sua equação de transporte será um detalhe quase esquecido da história da ciência; mas mil anos adiante, a relação  $S = k \ln W$  ainda será a pedra fundamental desta matéria. Um epitáfio mais adequado dificilmente poderia ser imaginado”. Esta fórmula é o ponto de partida de tudo que foi feito neste trabalho. Nas próximas seções discutimos em maiores detalhes os principais pontos abordados.

## 5.2 Probabilidades

Existem várias maneiras de interpretar o conceito de probabilidade, sendo que algumas versões se adequam melhor ao modelo da física estatística que outras. A identificação estrita entre probabilidade e frequência leva a sérios problemas interpretativos quando se tenta dar sentido físico aos ensembles já que, por motivos óbvios, não existe experimento aleatório bem definido para determinar as frequências estatísticas dos estados microscópicos. Uma segunda expectativa, em parte relacionada à primeira, é que a probabilidade se define pelo tempo de permanência em cada estado microscópico. É o raciocínio por trás da hipótese ergódica que leva a problemas relacionados aos tempos de Poincaré ou e à qualquer tentativa de formulação dinâmica para a evolução das probabilidades e entropia.

Há uma discussão vigorosa em certos círculos de matemática e estatística sobre como interpretar o cálculo de probabilidades. Ainda que seja adequado se manter informado sobre estes assuntos, a postura do físico se baseia fundamentalmente em uma escolha pragmática sobre que papel as probabilidades devem exercer nas suas próprias teorias e não precisa ser condicionada pelo papel que elas exercem em geral. O problema colocado para a física estatística, em certa medida se relaciona àquele encontrado em outras áreas da física como a mecânica quântica, diz respeito sobre como codificar um certo conhecimento a priori sobre as propriedades dos objetos que se estuda, mas que é necessariamente incompleto. Esta é a interpretação que iremos adotar<sup>1</sup>.

Ainda que seja fácil reconhecer que a questão do conhecimento incompleto existe, em física há uma certa resistência em incorporá-la explicitamente nas teorias. O aparecimento da física estatística no final do século *XIX* e da física quântica no início do *XX* poderiam ter colocado estas questões mais ao centro das atenções, mas certamente não o fizeram. As duas disciplinas seguem, respectivamente, de constatações muito simples sobre a natureza das nossas observações e do tipo de conhecimento que podemos adquirir da Natureza

**Física Estatística:** Não é possível observar nem controlar o estado microscópico de um sistema de muitas partículas com precisão absoluta.

**Física Quântica:** Não é possível observar nem controlar simultaneamente o momento

---

<sup>1</sup>Vale lembrar que os teoremas limite, sobre a convergência das frequências relativas em probabilidades, permitem testar se estas atribuições, de fato, fazem sentido.

e a posição de uma partícula com precisão absoluta.

Respostas muito convincentes foram elaboradas por Gibbs, Maxwell e Boltzmann com relação à primeira e em certa medida por Bohr com relação à segunda. Infelizmente, cada oportunidade de trazer a questão da aquisição de informação mais para o centro da prática comum em física, foi deliberadamente perdida. Muitas vezes beirando o limite da incredulidade, ou até mesmo do ridículo<sup>2</sup>, qualquer solução que perpetuasse a separação paradigmática entre “teoria da natureza” e “conhecimento obtido sobre a natureza” foi sistematicamente adotada. Jaynes ilustra de maneira dramática o dano causado por estas intervenções em mecânica estatística na introdução do excelente artigo “The Second Law as physical fact and as human inference” [?].

In spite of the spectacular advances in experimental techniques all about us, we live in an age of inexplicable decadence where theory is concerned. A wild variety of different views about entropy and reversibility, their place in fundamental physics, and the role of information for science in general, is being expressed. But important facts that were well understood and clearly explained by Maxwell and Gibbs over 100 years ago, and which played a crucial role in the work of Planck and Einstein 80 years ago, have been lost and are no longer comprehended at all by some who try to work in this area.

Expressamos este ponto de vista por acreditar que qualquer desenvolvimento substancial na área de física estatística é impossível sem reconhecer que em sistemas macroscópicos a ignorância sobre o microestado sempre existirá e não pode ser removida. Muitos conceitos chave da disciplina são meras expressões desta ignorância e, mais ainda, o tratamento da ignorância nas teorias físicas pode ser facilmente formalizado a partir do conceito de probabilidades. Alguns teóricos tratam com desconfiança afirmações de caráter epistemológico como “segundo dita o bom senso, a entropia de um sistema não decresce”. Respondemos a este tipo de ceticismo com a máxima de Jaynes que a necessidade injustificada em fundamentar objetivamente as probabilidades que aparecem nas teorias fazem dos físicos “os mais ingênuos dos cientistas”<sup>3</sup>.

<sup>2</sup>Uma solução que possivelmente pode ser classificada como ridícula foi proposta por J. von Neumann para explicar a redução do pacote de ondas em mecânica quântica. Para von Neumann, a consciência, pela sua capacidade singular de introspecção, seria o agente responsável por reduzir o pacote de ondas no processo de medição. Esta redução pode ser representada por uma interação física: um ser humano consciente é capaz de modificar o estado da matéria inanimada pela simples contemplação do próprio umbigo. Felizmente o número de pessoas que atualmente se oporiam à solução de von Neumann é crescente. Talvez às custas da crescente popularidade da interpretação de muitos mundos, no entanto.

<sup>3</sup>Mude a afirmação para “segundo dita o bom senso, ninguém ganha 2 vezes seguidas na loteria comprando apenas 2 bilhetes”. Esta é aproximadamente estimativa superior da ordem de grandeza associada à



## 5.3 Entropia

A próxima questão é estabelecer uma metodologia geral de inferência. Recapitulando novamente a fórmula de Boltzmann, ou seja,

$$S = k_B \ln W, \quad (5.1)$$

várias propriedades da entropia podem ser melhor entendidas, dada a sua relação com a probabilidade ( $W(F) \propto P(F)$ ). O critério de máxima entropia (Segunda Lei da Termodinâmica) é traduzido imediatamente em escolher o estado mais provável segundo a informação macroscópica disponível. É claro que, vista desta maneira, a Segunda Lei não expressa uma verdade ontológica, mas é apenas a expectativa razoável a ser feita com relação ao estado macroscópico que se espera observar. Deste modo, não há como ter certeza que a informação disponível é suficiente para fazer inferências corretas sobre todas propriedades interesse. Antes que não se saiba que a informação é insuficiente, e na ausência da informação adicional necessária, o melhor a se fazer ainda é apostar nas previsões do método da máxima entropia. Como é bem ilustrado na seção 3.5, as situações em que o método falha podem abrir portas para novo e importante conhecimento.

Neste momento, velhos vícios impedem alguns autores a aceitar o tipo de raciocínio aqui exposto. É claro que teríamos maior confiança na termodinâmica se a Segunda Lei expressasse uma lei imutável da Natureza, consequência direta da dinâmica microscópica e incondicionalmente válida. Mas quando levamos em conta a reversibilidade temporal da mecânica microscópica, a possibilidade que a Segunda Lei seja violada, junto ao fato que ela é *realmente* violada em situações controladas (i.e.: observações mesoscópicas como em [?]) consiste numa indicação muito forte que o raciocínio empregado está correto.

Neste sentido, tentamos mostrar que boa parte do formalismo tradicional de física estatística do equilíbrio pode ser entendido simplesmente como a prescrição de encontrar os estados mais prováveis. Isto fica óbvio quando se utiliza a entropia de Boltzmann definida para uma representação em função das DNPs. Também tivemos a preocupação de mostrar que, ao menos em situações especiais, a maximização da entropia de Gibbs pode ser entendida de maneira similar. A partir desta motivação simples — encontrar

---

probabilidade para que ocorra flutuações da ordem de 1 microcaloria em um sistema macroscópico típico. Estas flutuações estão no limite detectável pelas técnicas atuais e correspondem a uma probabilidade de cerca de  $10^{-15}$ .

os estados macroscópicos mais prováveis —, praticamente toda a termodinâmica pode ser obtida; as únicas suposições adicionais dizem respeito à forma da probabilidade  $\gamma(\mathbf{n})$  que por questões de consistência.

## 5.4 Perspectivas futuras

Nenhuma linha de investigação pode ser considerada muito promissora se já é apresentada como um produto acabado. Um aspecto importante do trabalho foi a tentativa de formular a física estatística de uma maneira bastante econômica, com esforço para reduzi-la a um simples modelo estatístico com o mínimo de hipóteses. Desta maneira acreditamos que alguns falsos problemas podem ser facilmente reconhecidos, além do que a maior clareza de exposição prepara o terreno para investigações futuras.

Um aspecto talvez pouco usual na forma de apresentação é a ênfase em utilizar a DNP como representação do estado macroscópico, enquanto normalmente a estatística de equilíbrio é feita no espaço de fase (ou de Hilbert). Partindo do ponto de vista que ambas estatísticas podem ser tratadas no mesmo pé de igualdade, é possível que a exposição nos dê maior confiança para utilizar alguns resultados relacionados à literatura de entropias generalizadas para, por exemplo, tratar das entropias fermiônicas e bosônicas. Um objetivo posterior é tornar viáveis as simulações utilizando-as e obter resultados numéricos. Além de dar um maior sentido de completude à teoria, é possível que o tratamento a partir da DNP facilite a inclusão de termos de interação ou vínculos adicionais.

Do ponto de vista puramente técnico, existem várias lacunas a ser preenchidas. No que se refere à literatura de entropias generalizadas, uma delas é esclarecer melhor o papel da interação na quebra do comportamento ideal associado a sistemas livres. Já está bem apontado na referência [?], que a inclusão de vínculos não-lineares pode resultar em propriedades qualitativamente novas no comportamento da entropia. Entre elas, está a presença de vários máximos. Acreditamos que esta propriedade pode ajudar a entender os mecanismos de transição de fase e uma perspectiva futura é tratar de modelos simples e, de preferência, analiticamente solúveis em que a não-linearidade da energia (ou outro vínculo) com relação à DNP desencadeie a existência de múltiplas fases. Em um certo ponto será necessário desenvolver simulações para avaliar detalhadamente o comportamento de sistemas mais complicados.

Ainda com relação ao tratamento de sistemas com energia de interação, um aspecto

importante que ainda não foi completamente resolvido na literatura é se a estatística de Gibbs se equivale, ou até que ponto ela equivale à estatística de Boltzmann. É possível mostrar pelo método do ponto de sela que sistemas livres satisfazem esta expectativa completamente, e nesta demonstração incluímos um meio de considerar probabilidades do tipo  $\gamma(\mathbf{n})$ . Infelizmente a generalização para incluir termos quadráticos na energia ainda parece bastante complicada, de maneira que a argumentação sobre como ambas se relacionam não está completamente assentada.

Para finalizar, uma parte do trabalho que ainda permanece praticamente inexplorada é o formalismo de não-equilíbrio. O fato que encontramos uma formulação por integrais funcionais é extremamente positivo dado a vasta literatura que existe a este respeito. Futuramente é necessário explicitar os procedimentos usuais de Kubo e Schwinger para o cálculo das funções de correlação e, mais ainda, seria interessante elaborar a contrapartida visão de DNPs como operadores. Uma formulação existente para lidar com processos estocásticos [?, ?], utiliza a linguagem de espaços de Fock de maneira muito convincente para a descrição de processos Markovianos e equação de Liouville, de sorte que já existe uma linha traçada para conectar ambos formalismos. Dado que cada escolha de Lagrangiana se traduz no formalismo apresentado como a instância de um processo Markoviano específico, é relativamente direto elaborar simulações para as séries temporais de DNPs. Os primeiros passos nesta direção estão sendo tomados, o que além de desenvolver maior familiaridade com o formalismo, permite extrair resultados numéricos a partir dos modelos mais simples.

Uma questão extremamente importante que foi deixada um pouco de lado nesta apresentação é sobre como podemos relacionar as médias sobre séries temporais da DNP com as médias de equilíbrio. Se existir uma propriedade de “ergodicidade” que identifique uma com a outra, seria possível determinar as probabilidades  $P(\mathbf{n}|U)$  (e após algum cálculo, a forma da entropia) a partir dos resultados de uma série temporal obtida por simulação ou experimento. Note que, em princípio, uma “teoria ergódica” para a DNP não sofreria de várias críticas relacionadas à teoria ergódica tradicional como o aparecimento de tempos de Poincaré (a dimensão do espaço para  $\mathbf{n}$  é muito menor), além do que  $\mathbf{n}$  representa uma grandeza macroscópica acessível experimentalmente. Esperamos que seja possível explorar mais adequadamente esta questão com um formalismo do tipo exposto acima.

Isto talvez resuma as expectativas mais imediatas para trabalhos futuros. Visto que se trata de uma dissertação para a conclusão de um curso de mestrado, pareceu

muito adequado explorar alguns problemas conceituais da física estatística e teoria de probabilidades que, vale a pena insistir, já foram resolvidos há vários anos ainda que se manifestem de tempos em tempos. Mais ainda, tentamos desenvolver, até onde o tempo nos permitiu, uma versão da física estatística fundamentalmente calcada na representação macroscópica da DNP. Isto a distingue da metodologia usual de estatística de equilíbrio de Gibbs, mas mostrou-se que em várias situações os resultados são os mesmos. Acreditamos que a representação dada pela DNP permite ver com muito mais clareza alguns aspectos importantes como a distinção entre descrição macroscópica e microscópica, o aparecimento da irreversibilidade e o significado da entropia além do papel da hipótese da equiprobabilidade *a priori* no esquema conceitual que fundamenta a termodinâmica.

Algumas posições adotadas recentemente na literatura (i.e.: a ênfase em subdinâmica e não-extensividade), parecem uma forma tenebrosa de conduzir a atividade científica em física estatística. Este trabalho também serve como uma provocação a estas tentativas. É claro que, sendo a ciência um empreendimento de longo prazo, qualquer linha de investigação que se situe em padrões mínimos de adequação é automaticamente legitimada. Ao explorar os caminhos que levam diretamente ao abismo, ainda que com um intuito de mera contemplação, aprendemos mais sobre como funciona a Natureza que aquilo que conhecíamos anteriormente. Tomando as palavras de Einstein, “a matemática é o sexto sentido dos homens e o sétimo das mulheres”, de sorte que tateando com paciência, andamos confortavelmente no escuro, longas distâncias através da confusão. Acreditamos que, em sua essência, este trabalho se inclui numa tradição saudável que tenta colocar claramente o papel das probabilidades em física estatística, e dentro desta, mais especificamente, exploramos sistematicamente o papel que a representação da DNP e sua relação com as entropias generalizadas.

## *APÊNDICE A – Regras de Cox*

Uma vez de acordo com Laplace que as probabilidades representam um dos “principais meios de alcançar a verdade” à partir do nosso conhecimento limitado sobre praticamente tudo, é razoável que o raciocínio probabilístico esteja por trás das decisões racionais sobre os diversos problemas da vida. As regras do cálculo de probabilidades são como uma espécie de processadores de informação — elas representam uma forma racional de conduzir um raciocínio à partir das coisas que sabemos com certeza para dizer algo sobre as muitas coisas sobre as quais pouco se sabe.

Todo curso sobre filosofia da ciência exhibe uma aula à respeito de uma característica muito ingênua no pensamento dos cientistas. O argumento consiste em que, apenas à partir da experiência — tome como exemplo a observação que o sol nasceu em todos os dias durante os milhares de anos de existência da humanidade — é impossível fazer generalizações sobre o comportamento da natureza. A razão é que a conclusão falaciosa que o sol nascerá amanhã é logicamente infundada. Neste sentido, a ciência apenas poderia descobrir quais *não são* as leis da natureza, o que se daria pela violação das mesmas em algum experimento, mas não diz absolutamente nada a respeito das leis “verdadeiras”. A teoria de probabilidades representa uma resposta formal à este tipo de crítica.

Ainda que não possamos ter certeza que o sol nascerá amanhã, mesmo esquecendo tudo que se sabe sobre gravitação, a observação acumulada durante todos esses anos indica que esta possibilidade é altamente plausível — somente um cínico ou um louco diriam que estas observações são irrelevantes. Neste sentido, mesmo que a indução não possa fornecer certezas sobre o mundo, a observação exaustiva permite chegar gradualmente à verdade. Aceitar a indução como estratégia de raciocínio válida justifica vários aspectos do pensamento dos cientistas de um ponto de vista racional; eles não são, afinal, desprovidos de razão. Se a lógica for entendida como a formalização de todo tipo de pensamento racional, a teoria das probabilidades, que é a lógica estendida para tratar da incerteza, representaria a verdadeira lógica da ciência enquanto aquela

estudada pelos filósofos e matemáticos consiste apenas numa manifestação parcial. É claro que ambas maneiras de pensar podem ser questionadas de acordo com a sanidade do interjeitor, mas em qualquer argumentação que se estabeleça num patamar estritamente racional, seria igualmente repreensivo ignorar a lógica ou as probabilidades.

E como se daria um programa de formalização do pensamento indutivo? Mesmo sem conhecer absolutamente nada sobre a tal Teoria das Probabilidades, existe pelo menos uma situação em que sabemos os resultados corretos: ela consiste na dedução lógica. Utilizamos a formalização feita no final do século XVII pelo matemático inglês George Boole no seu tratado intitulado “Leis do Pensamento”. Ainda que a ambição de Boole seja louvável, existe um problema à respeito da lógica que está, fundamentalmente, em nossas experiências. A maioria das coisas que realmente interessam, não sabemos com certeza, e desta forma, a lógica possui um domínio de aplicabilidade extremamente limitado, só pra dizer o mínimo. Seguindo uma tradição filosófica que começa no reverendo Thomas Bayes, passa por Laplace, Bernoulli e outros, o físico americano Richard T. Cox, formalizou todas estas intuições e mostrou explicitamente que as probabilidades, de fato, representam as leis adequadas para o pensamento indutivo.

## A.1 Cálculo de predicados

As regras para a dedução lógica naturalmente são regras que dizem respeito a *afirmações*. Tudo o que elas dizem é se, a partir de um certo conjunto de pressupostos, podemos concluir ou não sobre a veracidade ou falsidade de uma terceira afirmação. Suponha que as afirmações lógicas sejam representadas formalmente por letras,  $A$ ,  $B$ ,  $C$ , etc. Estas afirmações podem ser praticamente qualquer coisa, desde que se possa dizer que sejam *ou* falsas *ou* verdadeiras. À partir de um conjunto básico de afirmações, é possível criar outras afirmações mais complexas misturando-as entre si a partir de algumas regras de composição simples:

1. *Disjunção*:  $A + B = 1 \iff$  “ $A$  e/ou  $B$  são verdadeiros”
2. *Conjunção*:  $AB = 1 \iff$  “ $A$  e  $B$  são verdadeiros”
3. *Negação*:  $\bar{A} = 1 \iff$  “ $A$  é falso”
4. *Implicação*:  $A \Rightarrow B \iff$  “ $A$  implica  $B$ ”
5. *Bi-implicação*:  $A \Leftrightarrow B \iff$  “ $A$  implica  $B$  e  $B$  implica  $A$ ”

## 6. e outras...

Ainda que outras regras poderiam ser inventadas, este conjunto reduzido aparentemente captura as formas mais comuns de pensamento racional. Boole percebeu que, no que diz respeito à disjunção e à conjunção, as regras para a manipulação formal são muito parecidas com as regras de álgebra elementar. Assim, explorou esta similaridade de uma maneira que é muito conveniente, por exemplo, para demonstrar teoremas. Uma vez que seja possível traduzir um problema para o nível formal (como afirmações do tipo exposto acima), é possível determinar a veracidade de certas proposições por meio de cálculos algébricos os mais simples o possível.

Antes de prosseguir nesta direção, fornecemos a as regras de composição para estas operações que, além das regras ordinárias relativas às adições e multiplicações, são:

$$A = AA = A + A = \bar{\bar{A}} = A + 0 = 1A \quad (\text{A.1})$$

$$\overline{A + B} = \bar{A}\bar{B} \quad (\text{A.2})$$

$$\overline{AB} = \bar{A} + \bar{B} \quad (\text{A.3})$$

$$A\bar{A} = 0 \quad (\text{A.4})$$

$$A + \bar{A} = 1 \quad (\text{A.5})$$

Aqui usamos a convenção que 0 representa a falsidade e 1 a verdade. A partir destas regras podemos exprimir qualquer operação lógica, como por exemplo, a implicação:  $A \Rightarrow B = \bar{A} + B$ , e outras operações inventadas por motivos técnicos como o XOR:  $A \text{ XOR } B = (A + B)\overline{AB}$ , o NOR:  $A \text{ NOR } B = \overline{A + B}$ , o NAND:  $A \text{ NAND } B = \overline{AB}$  etc. Daí se traduz uma derivação lógica em cálculos simples como, por exemplo, partindo de  $A \Rightarrow B = \bar{A} + B$ , se obtêm facilmente o silogismo Aristotélico  $A \Rightarrow B = \bar{A} + B = \bar{\bar{B}} + \bar{A} = \bar{B} \Rightarrow \bar{A}$ .

Existem várias propriedades do cálculo Booleano que, a não ser que se trate de um lógico profissional, um programador ou um engenheiro de circuitos eletrônicos, não possuem realmente muitas aplicações. Uma delas é que todas operações lógicas podem ser reescritas a partir do encadeamento adequado de uma única operação, esta operação pode ser tanto a NAND quanto o NOR<sup>1</sup>. Uma questão legítima e correlata à

<sup>1</sup>Este fato é explorado na construção de circuitos eletrônicos. Um circuito eletrônico típico é formado por duas portas de entrada conectadas a uma de saída; se ambas recebem uma voltagem prefixada ou e ambas não recebem esta voltagem, a saída do circuito é 0V, representando a falsidade. Se uma das portas recebe voltagem e a outra não, a saída é equivalente à voltagem da primeira. Este tipo de resposta implementa uma função lógica do tipo NAND, sendo que a passagem de corrente representa a verdade

teoria de probabilidades é encontrar conjuntos de operações suficientes para reproduzir todas as outras operações que existem. Deste modo, ao exigir a consistência entre a manipulação de probabilidades e o cálculo Booleano, seria suficiente mostrar a consistência para as operações de qualquer um destes conjuntos — a adequação com relação às outras decorreria automaticamente. Uma resposta parcial se refere aos conjuntos  $\{NAND\}$  e  $\{NOR\}$ . Na realidade, é muito mais conveniente utilizar as operações de disjunção/conjunção/negação devido ao apelo intuitivo. Mostra-se que é possível escrever toda a álgebra de Boole escolhendo apenas duas das três operações citadas, elas seriam a negação acrescida ou da disjunção ou da conjunção. É fácil ver isto pois  $A + B = \overline{\overline{A}B}$  e ainda  $AB = \overline{\overline{A} + \overline{B}}$ .

## A.2 Raciocínio indutivo

Uma das constatações que o reverendo Thomas Bayes se deparou em seus estudos sobre as probabilidades é o caráter contextual sobre tudo aquilo que nós sabemos, ou julgamos saber. Por exemplo, uma questão pertinente para o reverendo, que é a veracidade da afirmação  $A \equiv$  "Deus existe" expressa pela probabilidade  $P(A) \equiv$  "Probabilidade que Deus exista", seria certamente avaliada de maneiras muito diferentes por Bayes que, por exemplo, por um filósofo como Nietzsche. Como, então, duas pessoas perfeitamente racionais em seus juízos poderiam discordar sobre algo tão importante? A resposta, obviamente, está em que as motivações de cada indivíduo envolvido são tão radicalmente diferentes que as coisas que uma pessoa aceita como verdade e a outra não, geram, no fim das contas, julgamentos muito distintos sobre os mais variados problemas da existência.

Desta maneira, a crença no quanto uma certa afirmação  $A$  é verdadeira só possui um significado operacional quando relacionada às hipóteses  $H = H_1 H_2 \dots H_N$  implícitas em tal julgamento. Define-se

$$(A|H) \equiv \text{"o quanto } A \text{ é plausível dada a hipótese } H". \quad (\text{A.6})$$

"Plausibilidades" deste tipo são o objeto formal básico da teoria apresentada. Uma vez de acordo que é possível capturar estas idéias formalmente, a próxima etapa é descobrir as possíveis regras de manipulação de *plausibilidades* consistentes com o cálculo Booleano. Isto é feito exigindo algumas propriedades simples como, por exemplo que  $(AB|H) = (BA|H)$ .

---

e a ausência, falsidade. Pela composição de NAND's é possível definir todas as outras operações de sorte que este desenho cria uma máquina capaz de computar qualquer coisa.



Desta forma, R. T. Cox mostrou que as regras para o cálculo de probabilidade seguem de três requisitos simples sobre o mapeamento  $(\mathcal{B}|\mathcal{B}') \mapsto \mathbb{R}$  que define a relação de plausibilidade. Estes requerimentos são

**Postulado I:** As plausibilidades são representadas por números reais.

Qualquer teoria de inferência que se preste à indução lógica, que é a nossa motivação fundamental, deve possuir alguma razão de ordenamento bem definida entre as plausibilidades de diferentes proposições  $A$ ,  $A'$ ,  $A''$  etc. Isso permite, por exemplo, eliminar as hipóteses implausíveis do discurso ou selecionar as mais prováveis — queremos escolher onde apostar as fichas. Na ausência de um ordenamento definido, a indução estaria condenada à todo tipo de pensamento circular onde, por exemplo,  $A$  é mais plausível que  $A'$  que é mais plausível que  $A''$  que é mais plausível que  $A$ .

Assim definimos a operação de comparação  $(A|H) \succ (B|H)$  que determina que  $A$  é mais plausível que  $B$  segundo a hipótese  $H$ . O significado operacional dos símbolos  $\succ$  e  $\prec$  varia de acordo com o objeto matemático utilizado para representar a plausibilidade e, no caso que sejam números reais, equivalem às relações tradicionais  $>$  e  $<$ . O Postulado I, acima, pode ser decomposto nos dois sub-postulados:

**Postulado I.I:** Transitividade nas afirmações. Seja um conjunto de hipóteses  $H$ , é necessário que se  $(A|H) \succ (B|H)$  e  $(B|H) \succ (C|H)$  então  $(A|H) \succ (C|H)$ .

**Postulado I.II:** Comparabilidade Universal. Dadas quaisquer proposições  $A$ ,  $B$  e  $H$  é necessário que uma e apenas uma destas relações seja válida:  $(A|H) \succ (B|H)$ ,  $(A|H) \prec (B|H)$  ou  $(A|H) = (B|H)$ .

As condições acima implicam que os objetos matemáticos que representam as plausibilidades sejam isomorfos a números reais.

**Postulado II:** Consistência.

Dada qualquer afirmação lógica  $A$  que possa ser representada igualmente pelas funções  $A = f(B, C, \dots)$  ou  $A = g(B, C, \dots)$ , é necessário que para todas hipóteses, vale que

$$(f(B, C, \dots)|H) = (g(B, C, \dots)|H).$$

Colocado de maneira mais precisa, considere uma operação tal como a negação. A plausibilidade  $(\bar{A}|H)$  deve ser escrita como uma função da plausibilidade  $(A|H)$ , ou seja,  $(\bar{A}|H) = S(A|H)$ . Neste caso, o requerimento de consistência é que se  $\bar{\bar{A}} = A$  então  $(A|H) = S(S(A|H))$ , garantindo que  $S(x)$  é tal que  $S(S(x)) = x$ .

**Postulado III:** Correspondência com a lógica.

No limite em que as plausibilidades se referem à certeza ( $V$ ) ou à impossibilidade ( $F$ ) representadas por números a serem especificados, a estrutura lógica deve ser integralmente recuperada. A lógica é apenas um caso especial do pensamento indutivo.

Estas regras simples são tudo necessário pra determinar, univocamente, as regras de cálculo de probabilidades. Existem extensões desta demonstração para números imaginários e elementos da álgebra de Clifford (como produtos vetoriais, variáveis de Grassmann e spinors de Dirac) [?, ?]. A relevância das mesmas para questões relativas à mecânica quântica é bastante óbvia, mas a relação que estas teorias de probabilidades exóticas possuem com campos distintos da física ainda é uma questão em aberto. As próximas seções se preocupam em estabelecer as regras relativas à manipulação destes elementos tanto para a operação  $(AB|H)$  quanto para  $(\bar{A}|H)$  e, à partir destas, para todas as outras.

### A.2.1 Plausibilidade da conjunção

Para que a relação de plausibilidade  $(AB|H)$  seja bem definida, é necessário obedecer a algumas regras básicas relativas à álgebra Booleana como, por exemplo, que  $(AB|H) = (BA|H)$ . Desta forma, supondo que existe uma relação funcional entre  $(AB|H)$  e as plausibilidades mais simples  $(A|H)$ ,  $(B|H)$ ,  $(A|BH)$  e  $(B|AH)$ , os questionamentos sobre a consistência das atribuições podem ser colocados numa forma matematicamente mais precisa que são as propriedades de invariância da função

$$(AB|H) \equiv F((A|H), (B|H), (A|BH), (B|AH)). \quad (\text{A.7})$$

$F(w, x, y, z)$  obviamente mapeia os números reais  $w = (A|H)$ ,  $x = (B|H)$ ,  $y = (A|BH)$  e  $z = (B|AH)$  em  $(AB|H)$ . Note que adotamos o ponto de vista mais geral possível sobre a dependência da função  $F$  com os seu argumentos. Intuitivamente é razoável

supor que a dependência não se dê efetivamente em todas estas variáveis.

Considere as duas afirmações  $V_D \equiv$  "fulano tem o olho direito verde" e  $C_E \equiv$  "fulano tem o olho esquerdo castanho". Desconhecendo fulano, é razoável supor que são razoavelmente plausíveis tanto  $(V_D|H)$  quanto  $(C_E|H)$ . Já a afirmação composta  $V_D C_E \equiv$  "fulano tem o olho direito verde e o olho esquerdo castanho" é bastante implausível, ainda que nem  $V_E V_D$  e  $C_E C_D$  sejam avaliadas desta maneira. O raciocínio que nos leva à esta conclusão é que, para que a afirmação composta  $V_D C_E$  seja plausível, tanto devem ser plausíveis  $V_D$  quanto deve ser  $C_E$  condicionado à verdade de  $V_D$ . Desta forma, é razoável supor que a dependência funcional da equação (A.7) seja restrita à

$$(AB|H) \equiv F((A|H), (B|AH)) \quad (\text{A.8})$$

$$= F((B|H), (A|BH)). \quad (\text{A.9})$$

Uma análise mais exaustiva de (A.7) mostra que estas são as únicas opções compatíveis com a lógica nos limites que  $A$  e  $B$  tendem à certeza ou impossibilidade (e  $A \Rightarrow B$ , ou  $A \Rightarrow \bar{B}$  etc). Portanto, invocamos o Postulado III para justificar o uso das relações acima.

Utilizando esta dependência simplificada, podemos calcular a plausibilidade de  $(ABC|H)$ , que deve ser compatível com as duas formas

$$(ABC|H) = (A(BC)|H) = F((A|H), F((B|AH), (C|ABH))) \quad (\text{A.10})$$

$$= ((AB)C|H) = F(F((A|H), (B|AH)), (C|ABH)). \quad (\text{A.11})$$

Definindo as variáveis  $x = (A|H)$ ,  $y = (B|AH)$  e  $z = (C|ABH)$ , ficamos com a relação funcional

$$F(x, F(y, z)) = F(F(x, y), z) \quad (\text{A.12})$$

que é conhecida como a equação de associatividade. Para resolvê-la, definimos as grandezas auxiliares

$$F_1(a, b) \equiv \frac{\partial F}{\partial a}; \quad F_2(a, b) \equiv \frac{\partial F}{\partial b} \quad (\text{A.13})$$

e ainda

$$u \equiv F(x, y); \quad v \equiv F(y, z). \quad (\text{A.14})$$

Desta forma, a equação funcional fica

$$F(x, v) = F(u, z). \quad (\text{A.15})$$

Diferenciando com relação a  $x$  e a  $y$  e utilizando a regra da cadeia obtemos respectivamente

$$F_1(x, v) = F_1(u, z)F_1(x, z) \quad (\text{A.16})$$

$$F_2(x, v)F_1(y, z) = F_1(u, z)F_2(x, y), \quad (\text{A.17})$$

que pela eliminação do termo comum  $F_1(u, z)$  resulta em

$$\frac{F_1(x, v)}{F_1(x, z)} = \frac{F_2(x, v)F_1(y, z)}{F_2(x, y)}. \quad (\text{A.18})$$

Definindo a função

$$G(x, y) \equiv \frac{F_2(x, y)}{F_1(x, y)}, \quad (\text{A.19})$$

e substituindo em (A.18), podemos escrever as duas expressões

$$G(x, v)F_1(y, z) = G(x, y) \quad (\text{A.20})$$

$$G(x, v)F_2(y, z) = G(x, y)G(y, z). \quad (\text{A.21})$$

Note que não existe dependência em  $z$  na expressão (A.20), de forma que

$$\frac{\partial}{\partial z} [G(x, v)F_1(y, z)] = \frac{\partial G(x, v)}{\partial v} F_2(y, z)F_1(y, z) + G(x, v) \frac{\partial^2}{\partial z \partial y} F(y, z) = 0. \quad (\text{A.22})$$

De maneira semelhante, derivando o lado esquerdo da equação (A.21) por  $y$  encontramos que

$$\frac{\partial}{\partial y} [G(x, v)F_2(y, z)] = \frac{\partial}{\partial z} [G(x, v)F_1(y, z)] = 0, \quad (\text{A.23})$$

o que nos leva à concluir que  $G(x, y)G(y, z)$  não exhibe dependência em  $y$ . A forma mais geral para  $G(x, y)$  que possui esta propriedade é dada por

$$G(x, y) = k \frac{H(x)}{H(y)}, \quad (\text{A.24})$$

onde  $k$  é uma constante arbitrária e  $H(x)$  é uma função indeterminada de  $x$ . Substituindo esta relação respectivamente em (A.20) e (A.21) ficamos com

$$F_1(y, z) = \frac{H(v)}{H(y)}; \quad F_2(y, z) = r \frac{H(v)}{H(z)}. \quad (\text{A.25})$$

Agora utilizando (A.14) para escrever

$$dv \equiv dF(y, z) = F_1(y, z)dy + F_2(y, z)dz,$$

e substituindo os valores de (A.25) nesta equação, após dividir a expressão por  $H(v)$  e

integrar ambos lados, obtemos a relação

$$\int^v \frac{dv'}{H(v')} = \int^y \frac{dy'}{H(y')} + r \int^z \frac{dz'}{H(z')}. \quad (\text{A.26})$$

Define-se a função

$$w(x) \equiv e^{\int^x \frac{dx'}{H(x')}}, \quad (\text{A.27})$$

de maneira que a expressão (A.26) se escreve simplesmente como

$$w(v) = w(F(y, z)) = w(y) [w(z)]^r. \quad (\text{A.28})$$

Substituindo a expressão acima em cada lado da equação funcional (A.15), obtemos a relação

$$w(x) [w(v)]^r = w(u) [w(z)]^r, \quad (\text{A.29})$$

que, substituindo a expressão (A.15) novamente para  $w(v)$  e  $w(u)$  resulta em

$$w(x) [w(y)]^r [w(z)]^{r^2} = w(x) [w(y)]^r [w(z)]^r. \quad (\text{A.30})$$

A única forma de evitar a solução trivial para a equação  $w(z)^{r^2-r} = 1$  é que  $r = 0$  ou  $r = 1$ . A primeira opção implica numa solução trivial para (A.24), o que obviamente não interessa. Desta forma somos restritos à escolha  $r = 1$  que resulta na seguinte regra para a conjunção

$$w(F(x, y)) = w(x)w(y). \quad (\text{A.31})$$

A função  $w(x)$  apenas define uma escala de plausibilidade conveniente a qual nos referimos como  $w(A|H)$ . Desta maneira, lembrando da definição das variáveis  $x$ ,  $y$  e  $z$ , as plausibilidades relativas a afirmações compostas, numa escala  $w(x)$ , obedecem necessariamente à relação

$$w(AB|H) = w(A|H)w(B|AH) \quad (\text{A.32})$$

$$= w(B|H)w(A|BH). \quad (\text{A.33})$$

## A.2.2 Valores de certeza e impossibilidade

De acordo com o Postulado III a respeito das exigências sobre a plausibilidade, é necessário que as plausibilidades relativas à quaisquer duas afirmações verdadeiras ou falsas sejam as mesmas. À partir da regra (A.32) para o cálculo de afirmações compostas é possível determinar o valor numérico específico a ser atribuído tanto à

impossibilidade quanto à certeza. Desta maneira, calculamos

$$w(A|H) = w(AA|H) = w(A|H)w(A|AH). \quad (\text{A.34})$$

Note que o termo  $w(A|H)$  se cancela para qualquer  $A$  de forma que ficamos com  $w(A|AH) = 1$ .  $w(A|AH)$  obviamente representa uma situação de certeza de sorte que mostrou-se que na escala definida por  $w(x)$ , se atribui uma plausibilidade 1 à certeza. O cálculo do valor numérico associado à impossibilidade é igualmente simples. Seja a expressão

$$w(A\bar{A}|H) = w(A|H)w(\bar{A}|AH) = w(\bar{A}|H)w(A|\bar{A}H). \quad (\text{A.35})$$

Segundo o Postulado III, os valores associados às atribuições impossíveis  $w(A\bar{A}|H)$ ,  $w(A|\bar{A}H)$  e  $w(\bar{A}|AH)$  devem ser os mesmos. A única possibilidade não trivial é que todos sejam avaliados em 0, caso contrário teríamos, segundo a expressão (A.35) que, para todas afirmações,  $w(A|H) = w(\bar{A}|H)$ . Pelo menos uma situação em que isso leva à inconsistências é quando  $A$  é certa em relação à  $H$ , deste modo,  $\bar{A}$  seria impossível e a expressão anterior implicaria que a certeza e impossibilidade estariam associadas ao mesmo valor, que é obviamente um contra-senso. Desta forma concluímos que:

Valor de verdade	Plausibilidade
V $(H \Rightarrow A)$	$w(A H) = 1$
F $(H \Rightarrow \bar{A})$	$w(\bar{A} H) = 0$

### A.2.3 Plausibilidade da negação

Como já foi exposto anteriormente, a relação de negação está associada à equação funcional

$$S(S(x)) = x \Rightarrow S(x) = S^{-1}(x). \quad (\text{A.36})$$

Note que restrições adicionais à forma de  $S(x)$  podem ser impostas pelo quesito de coerência com relação à operação de conjunção já considerada anteriormente. Esta restrição adicional pode ser obtida facilmente a partir de

$$w(AB|H) = w(A\bar{\bar{B}}|H) = w(A|H)w(\bar{\bar{B}}|AH) = w(A|H)S[w(\bar{B}|AH)], \quad (\text{A.37})$$

$$= w(\bar{\bar{A}}B|H) = w(B|H)w(\bar{\bar{A}}|BH) = w(B|H)S[w(\bar{A}|BH)]. \quad (\text{A.38})$$

Tomamos o termo à direita de ambas equações, para ser reescrito como

$$w(A|H)S \left[ \frac{w(A\bar{B}|H)}{w(A|H)} \right] = w(B|H)S \left[ \frac{w(\bar{A}B|H)}{w(B|H)} \right]. \quad (\text{A.39})$$

A relação acima pode ser escrita para quaisquer afirmações  $A$  e  $B$  arbitrárias, em especial, também deve ser verdadeira se  $\bar{B} = AC$ . Utilizando as regras do cálculo Booleano temos que  $\bar{B} = AC \Rightarrow A\bar{B} = AAC = AC = \bar{B}$  e também que  $A = AC + A\bar{C} = \bar{B} + A\bar{C}$ , de sorte que  $\bar{A} = \overline{\bar{B} + A\bar{C}} + B\bar{A}\bar{C} = B\bar{A}\bar{C} = \bar{A}B$ . Desta forma escrevemos

$$w(A\bar{B}|H) = w(AC|H) = w(\bar{B}|H), \quad (\text{A.40})$$

$$w(\bar{A}B|H) = w(\bar{A} + \bar{A}\bar{C}|H) = w(\bar{A}|H). \quad (\text{A.41})$$

Definindo  $x \equiv w(A|H)$  e  $y \equiv w(B|H)$ , é possível escrever a condição adicional sobre  $S(x)$  substituindo (A.40) e (A.41) em (A.39), para que

$$xS \left[ \frac{S(y)}{x} \right] = yS \left[ \frac{S(x)}{y} \right]. \quad (\text{A.42})$$

Definimos as variáveis

$$u \equiv \frac{S(y)}{x}; \quad v \equiv \frac{S(x)}{y}, \quad (\text{A.43})$$

de forma que é possível expressar a relação (A.42) e as suas derivadas com respeito à  $x$ ,  $y$  e em segunda para  $x$  e  $y$ , respectivamente como

$$xS(u) = yS(v), \quad (\text{A.44})$$

$$S'(v)S'(x) = S(u) - \frac{S(y)}{x}S'(u) = S(u) - uS'(u), \quad (\text{A.45})$$

$$S'(u)S'(y) = S(v) - \frac{S(x)}{y}S'(v) = S(v) - vS'(v), \quad (\text{A.46})$$

$$uS''(u)\frac{S'(x)}{y} = vS''(v)\frac{S'(y)}{x}. \quad (\text{A.47})$$

Multiplicando os termos correspondentes de (A.44) e de (A.47) resulta em

$$uS''(u)S(u)S'(x) = vS''(v)S(v)S'(y). \quad (\text{A.48})$$

Utilizando (A.45) e (A.46) para eliminar  $S'(x)$  e  $S'(y)$  em (A.48) resulta em

$$\frac{uS''(u)S(u)}{[uS'(u) - S(u)]S'(u)} = \frac{vS''(v)S(v)}{[vS'(v) - S(v)]S'(v)}. \quad (\text{A.49})$$

Os dois lados da equação correspondem à mesma função avaliadas em pontos independentes  $u$  e  $v$ . Desta maneira, ambos devem corresponder à uma constante que denominamos

$k$ . Daí se obtêm a equação diferencial não-linear

$$xS''(x)S(x) = kS'(x) [xS'(x) - S(x)]. \quad (\text{A.50})$$

que pode ser reescrita de maneira mais conveniente como

$$\frac{\partial}{\partial x} \ln S'(x) = k \frac{\partial}{\partial x} [\ln S(x) - x], \quad (\text{A.51})$$

que implica em

$$\ln S'(x) - k \ln S(x) + kx = \text{cte}. \quad (\text{A.52})$$

Deste modo ficamos com uma equação diferencial de primeira ordem

$$S'(x) = A \left[ \frac{S(x)}{x} \right]^k, \quad (\text{A.53})$$

que resolvendo por separação de variáveis,  $\int \frac{dS}{S^k} = A \int \frac{dx}{x^k}$ , tem a solução geral dada por

$$[S(x)]^{1-k} = Ax^{1-k} + B. \quad (\text{A.54})$$

Note que a solução geral não satisfaz os critérios de consistência para todas escolhas de  $A$  e  $B$ . Definindo  $c \equiv 1 - k$  e substituindo a expressão acima em (A.42), concluímos que

$$A^2y^c + Bx^c = A^2x^c + By^c, \quad (\text{A.55})$$

que só é satisfeito para todo  $x$  e  $y$  se  $B = A^2$ . Já pela substituição na primeira equação de consistência (A.36), ficamos com

$$A^2x^c + A^2 + A^3 = x^c \quad (\text{A.56})$$

de sorte que a única escolha de  $A$  que deixa esta equação válida para todo  $x$  é  $A = -1$ . Assim concluímos que

$$[S(x)]^c = 1 - x^c. \quad (\text{A.57})$$

Definindo a função  $P(x) \equiv [w(x)]^c$ , podemos reescrever a regra para a negação simplesmente como

$$P(\bar{A}|H) = 1 - P(A|H). \quad (\text{A.58})$$

## A.2.4 Plausibilidade da disjunção

Uma vez de posse das regras para o cálculo da plausibilidades para a conjunção e para a negação, é possível calcular a regra para a disjunção utilizando a relação



$A + B = \overline{\bar{A}\bar{B}}$ . Utilizando a escala  $P(x)$  introduzida em (A.58) é possível escrever a plausibilidade da conjunção de duas afirmações como

$$P(AB|H) = P(A|H)P(B|AH) = P(B|H)P(A|BH). \quad (\text{A.59})$$

Utilizando esta relação e a expressão para a disjunção temos que

$$P(A + B|H) = 1 - P(\bar{A}\bar{B}|H), \quad (\text{A.60})$$

onde o segundo termo do lado direito da equação pode ser escrito como

$$P(\bar{A}\bar{B}|H) = P(\bar{A}|H)P(\bar{B}|\bar{A}H) = [1 - P(A|H)] \times [1 - P(B|\bar{A})], \quad (\text{A.61})$$

$$= 1 - P(A|H) - P(B|\bar{A}H) [1 - P(\bar{A}|H)] + P(B|\bar{A}H), \quad (\text{A.62})$$

$$1 - P(A|H) + P(\bar{A}B|H). \quad (\text{A.63})$$

Substituindo esta expressão em (A.60), ficamos com

$$P(A + B|H) = P(A|H) + P(\bar{A}B|H), \quad (\text{A.64})$$

onde o segundo termo pode ser escrito como

$$P(\bar{A}B|H) = P(B|H)P(\bar{A}|BH) = P(B|H) - P(B|H)P(A|BH) \quad (\text{A.65})$$

$$= P(B|H) - P(AB|H). \quad (\text{A.66})$$

Desta forma completamos demonstração das regras para o cálculo de plausibilidades compostas

$$P(AB|H) = P(A|H)P(B|AH) = P(B|H)P(A|BH), \quad (\text{A.67})$$

$$P(A + B|H) = P(A|H) + P(B|H) - P(AB|H), \quad (\text{A.68})$$

$$P(\bar{A}|H) = 1 - P(A|H). \quad (\text{A.69})$$

À esta escala específica  $P(x) = [w(x)]^c$  chamamos de *probabilidade*. À partir de (A.67), (A.68) e (A.69) é fácil calcular a probabilidade/plausibilidade associada à qualquer

função Booleana como, por exemplo, as funções

$$P(A \Rightarrow B|H) = P(A\bar{B}|H) = 1 + P(A|H) - P(B|AH), \quad (\text{A.70})$$

$$P(A \Leftrightarrow B|H) = P(AB + \bar{A}\bar{B}|H) = 1 + 2P(AB|H) - P(A|H) - P(B|H), \quad (\text{A.71})$$

$$P(ANAND B|H) = P(\overline{AB}|H) = 1 - P(AB|H), \quad (\text{A.72})$$

$$P(ANORB|H) = P(\overline{A+B}|H) = 1 - P(A|H) - P(B|H) + P(AB|H), \quad (\text{A.73})$$

$$P(AXORB|H) = P((A+B)\overline{AB}|H) = P(A|H) + P(B|H) - 2P(AB|H), \quad (\text{A.74})$$

$$(\dots) \quad (\text{A.75})$$

Com estas propriedades definimos completamente o objeto matemático que representa a *plausibilidade*.

### A.2.5 Conjuntos exaustivos e mutualmente exclusivos (EME)

À partir das regras acima, é possível definir as regras de probabilidade associadas a conjuntos EME que são de grande valia na teoria de probabilidades. Um conjunto mutualmente exclusivo de afirmações  $\{A_I\}$  é tal que para todo par  $A_i A_j = F$  se  $i \neq j$ . Neste caso, calculamos

$$P(A_1 + A_2 + \dots + A_N|H) = P(A_1|H) + P(A_2 + \dots + A_N) - P(A_1 A_2 + \dots + A_1 A_N|H). \quad (\text{A.76})$$

Para o conjunto exaustivo, temos que  $A_1 + \dots + A_N = V$ , o que implica em  $P(A_1 + A_2 + \dots + A_N|H) = 1$ . O segundo termo à direita consiste numa disjunção de afirmações falsas já que  $A_1 A_{i \neq 1} = F$  e, portanto, também representa uma falsidade. Deste modo, ficamos com

$$1 = P(A_1|H) + P(A_2 + \dots + A_N|H) = P(A_1|H) + P(A_2|H) - P(A_2 A_3 + \dots + A_2 A_N|H), \quad (\text{A.77})$$

e, utilizando o mesmo argumento anterior, elimina-se o último termo à direita. Utilizando este raciocínio recursivamente, mostramos que para todo conjunto EME de afirmações  $\{A_i\}$ , vale que

$$\sum_i P(A_i|H) = 1. \quad (\text{A.78})$$

Daí também demonstra-se a regra de marginalização

$$\sum_i P(A_i B|H) = \sum_i P(B|H)P(A_i|BH) = P(B|H). \quad (\text{A.79})$$

Resumo: propriedades de uma probabilidade	
Álgebra Booleana	
Conjunção:	$P(AB H) = P(A H)P(B AH) = P(B H)P(A BH)$
Disjunção:	$P(A + B H) = P(A H) + P(B H) - P(AB H)$
Negação:	$P(\bar{A} H) = 1 - P(A H)$
Domínio de definição	
Limites de definição:	$0 \leq P(A H) \leq 1$
Certeza ( $H \Rightarrow A$ ):	$P(A H) = 1$
Impossibilidade ( $H \Rightarrow \bar{A}$ ):	$P(A H) = 0$
Conjunto de afirmações EME	
Normalização:	$\sum_i P(A_i H) = 1$
Marginalização:	$P(B H) = \sum_i P(A_i B H)$



# Índice Remissivo

- aproximação
  - campo médio, 96
  - mecânica, 66
- Bernoulli, teorema, 27
- calor, 105
- condição
  - de localidade, 40
  - de sub-localidade, 43
- coordenadas naturais, 104
- distância de Kullback-Leibler, 36
- distribuição número de partículas, 79
- distribuições de probabilidades
  - contínuo, 30
  - discreto, 28
- DNP
  - definição, 77
  - equilíbrio, 86
  - evolução, 79
- ensemble, 52
- entropia
  - bósons, 94
  - bosônica, 107
  - efetiva, 88
  - estatística, 32
  - férmions, 96
  - fermiônica, 108
  - generalizada, 88, 92
  - Gibbs, 111
  - inferencial, 39
  - local, 88
  - relativa, 36
  - Shannon, 33
  - termodinâmica, 70
- entropia de Boltzmann, 87
- entropia Local, 88
- equação
  - da seta do tempo, 70
  - Liouville-von Neumann, 120
- equilíbrio termodinâmico, 71
- equiprobabilidade, 62
- fase da matéria, 97
- irreversibilidade, 67
- Jaynes
  - máxima entropia, 112
- Legendre, transformada, 102
- método
  - do ponto de sela, 114
- mínimos quadrados, 38
- MaxEnt, 112
  - não-equilíbrio, 123
- plausibilidade, 26
- potencial termodinâmico, 102
- princípio
  - da máxima entropia, 28, 34
  - da razão insuficiente, 27
  - mínima ação, 81
- probabilidade, 23
- sistemas simples, 67, 79
- temperatura, 101
- teorema
  - de Bayes, 36
- teoria
  - ergódica, 63
- termodinâmica
  - leis da, 99
- trabalho, 106