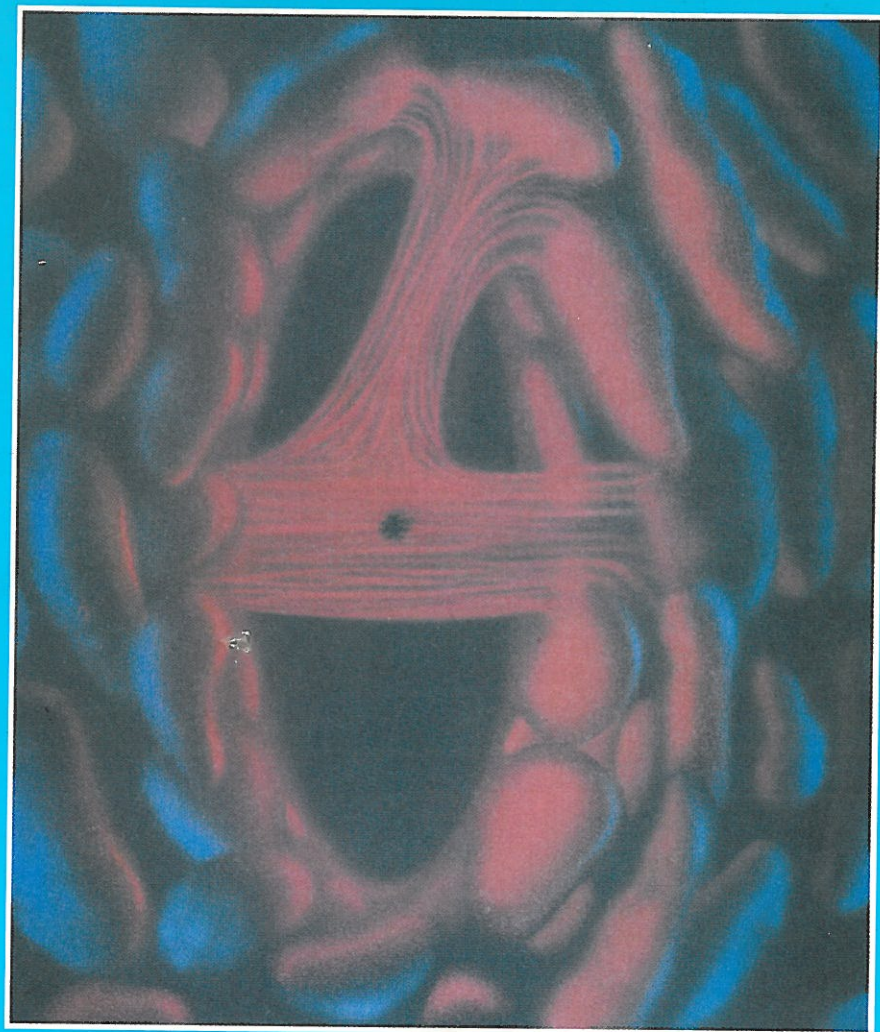


SÉRIE UNIVERSITÁRIA



SISTEMAS ENTROPIA, COESÃO

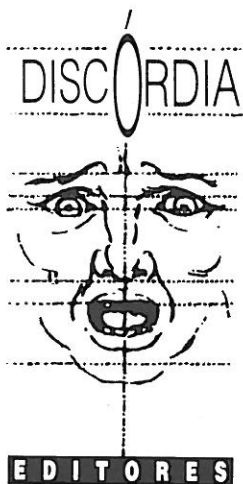
J. PINTO PEIXOTO

*Faculdade de Ciências da Universidade de Lisboa
Instituto Geofísica Infante D. Luís*

F. CARVALHO RODRIGUES

*Instituto Superior Técnico
Laboratório Nacional de Engenharia
e Tecnologia Industrial*





EDITORES

DISCÓRDIA, imagem e exemplo do contraditório, assume o **elogio do conflito** e a recusa do consenso. Porque no mundo do complexo e do excesso de informação, a comunicação pode ser, em milhares de bits, uma só frase, uma só palavra, urge procurar em cada coisa o seu oposto, gerir a mudança, resolver contradições sem conciliar os contrários, viver os saberes que aspiram à sabedoria. Caminho para a Discórdia.

Publicado por
Discórdia Editores, Lda.
Av. de Roma, 93, r/c Esq.
1700 Lisboa

Edição número 3/91

Autores

J. Pinto Peixoto
F. Carvalho Rodrigues

Título

Sistemas, Entropia, Coesão

Capa

J.M.M.

Composição e montagem

Estúdios MFCR, Lda.

Impressão

Eme Silva

Distribuição exclusiva em Portugal

Cultivar, edição, cultura e lazer, Lda

Acabado de imprimir aos 20 de Setembro de 1991

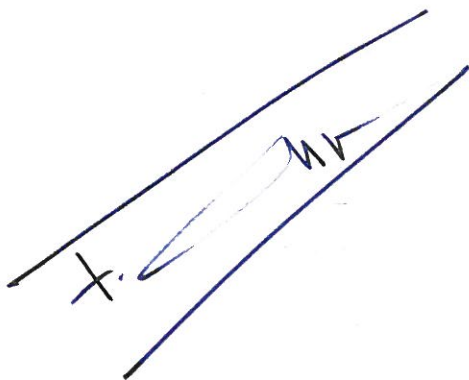
Direitos reservados por Discórdia Editores, Lda.

Esta publicação não pode ser, no todo ou em parte, reproduzida ou transmitida por qualquer forma ou processo, electrónico, mecânico ou fotográfico, incluindo fotocópia, xerocópia ou gravação, sem que tal tenha sido previamente, e por escrito, autorizado pelos editores. Exceptua-se naturalmente a transcrição de pequenos excertos a utilizar na apresentação, recensão ou crítica do livro. Tal não deverá no entanto ser tomado como extensivo à transcrição de textos em recolhas antológicas ou similares, que possam vir a reduzir o interesse pelo livro. Os editores reservam-se o direito de proceder judicialmente contra os transgressores.

Depósito legal nº 27619/91

J. PINTO PEIXOTO
F. CARVALHO RODRIGUES

SISTEMAS ENTROPIA, COESÃO



LISBOA
1991

PREFÁCIO	7
INTRODUÇÃO	11
I · NATUREZA ESTATÍSTICA DA ENTROPIA	15
1.1. Evolução do conceito de entropia	17
1.2. A termodinâmica, a mecânica e a estatística	21
2. Probabilidade termodinâmica	25
2.1. Microestados e macroestados dum sistema	25
2.2. Um exemplo	29
2.3. Colectividade estatística	33
3. Interpretação mecanicista do segundo princípio fundamental da termodinâmica	35
3.1. Fórmula de Boltzmann	35
3.2. O macroestado de probabilidade máxima e o segundo princípio fundamental da termodinâmica	41
3.3. O estado de probabilidade máxima. Lei de distribuição de Boltzmann	45
4. Conexão da mecânica estatística com a termodinâmica clássica	53
4.1. A função de partição e a determinação dos valores médios	53
5. Desordem e entropia	57
5.1. Desordem e o número de estados acessíveis	57
5.2. Carácter aleatório do segundo princípio fundamental da termodinâmica	61
5.3. O zero absoluto e o terceiro princípio da termodinâmica	67
5.4. O primeiro princípio da termodinâmica. Interpretação estatística e mecânica de calor e de trabalho	69
6. Epílogo: A natureza estatística da entropia	73
II ENTROPIA E INFORMAÇÃO	75
1. Noções elementares sobre a teoria da informação	77
1.1. Definição de informação	77
1.2. Incerteza, entropia, informação e probabilidade	79
2. Medida de informação	83
2.1. Incerteza, falta de informação e probabilidade	83
2.2. A fórmula fundamental da teoria da informação	89
3. Surpresa, ignorância e entropia	93
3.1. Surpresa e probabilidade	93
3.2. Ignorância e incerteza	95
3.3. Incerteza e informação	97
3.4. O valor da constante de Boltzmann	99
4. O sentido físico da dualidade - Entropia e informação	101
5. A conexão entre a teoria da informação e a termodinâmica	105
5.1. Formalismo de Jaynes. Fundamento do método da entropia máxima	105
5.2. Relação entre grandezas $\Omega(\lambda_1, \lambda_2 \dots \lambda_n)$ e $(S, r_1, r_2 \dots r_n)$ Dualidade Formal	111
5.3. Significado físico dos parâmetros de Lagrange, Ω e λ_i	115
6. O demónio de Maxwell e a teoria da informação	119
6.1. O demónio da Maxwell e o segundo princípio	119
6.2. A exorcização do demónio de Maxwell	123
7. Balanço energético da computação. Limite computacional	127
III A LARGUEZA DO CONCEITO DE ENTROPIA	131
1. Uma visão nova da entropia	133
2. Entropia numa rede hidrológica	135
3. A entropia e a ecologia: um exemplo	139

4.	Análise e entropia numa imagem	141
4.1.	Estatística numa imagem	141
4.2.	Indicadores de continuidade e de estrutura numa imagem	145
5.	O método de entropia máxima (MEM)	151
5.1.	Formalismo da entropia máxima	151
5.2.	A entropia numa distribuição normal de probabilidade	155
5.3.	Reconstituição de imagens	159
5.4.	Análise espectral de séries temporais	161
5.4.1.	O espectro de densidade espectral e o método da entropia máxima	161
5.4.2.	Fundamento e aplicação do método	163
6.	A entropia na linguagem	167
6.1.	Caracterização estatística da língua	167
6.2.	A aproximação de primeira ordem	169
6.3.	Matrizes de correlação	173
6.4.	A entropia de várias ordens na linguagem	177
6.5.	Alguns comentários	181
7.	Entropia como discriminante	183
7.1.	Medida da entropia da informação no conceito de Shannon	185
7.2.	Medida da entropia da informação	187
7.3.	Medida da entropia da informação no espectro visível	189
7.4.	Aplicações da entropia da informação como discriminante	191
7.4.1.	Sector têxtil	193
7.4.1.1.	Penteação	193
7.4.1.2.	Medidas de pilosidade de fios têxteis	195
7.4.1.3.	Caracterização de defeitos em tecidos	197
7.4.2.	Sector do papel	199
7.4.3.	Sector dos curtumes	201
7.4.4.	Cartografia	205
7.5.	Conclusões	209
IV	COMPLEXIDADE, INFORMAÇÃO, COESÃO	211
1.	Complexidade e complexidade de comportamento	213
2.	Função de coesão de estruturas ligadas por informação	217
2.1.	Da informação à coesão	217
2.2.	Definição da função de coesão	219
2.3.	Quantidade de informação observada em duas estruturas alternativas de um sistema	221
3.	Comportamento da função de coesão para o caso de duas estruturas alternativas	225
APÊNDICE - I		229
	Números extremamente grandes e extremamente pequenos e a sua representação em potência de 10	231
APÊNDICE - II		235
	Medida de informação	237
1.	Informação condicionada. Alguns exemplos	237
2.	Unidades de informação	241
APÊNDICE - III		245
	Redundância	247
APÊNDICE - IV		249
	A evolução da língua portuguesa (trabalho para a cadeira de termodinâmica generalizada em 1987 por Nuno Barradas	251
BIBLIOGRAFIA		267
AGRADECIMENTOS		271

PREFÁCIO

Para a eterna batalha entre $\chi\omicron\sigma\mu\omicron\varsigma$ (cosmos = ordem) e $\chi\alpha\omicron\varsigma$ (caos) tanto faz que o Universo tenha sido criado ex-nihil num big-bang ou esteja em contínua e cíclica criação. Sabe-se, há mais de três mil anos*, que o estabelecimento de estruturas coesas com maior complexidade requer fluxo de energia ao mesmo tempo que obriga à manutenção de uma grande simplicidade de informação.

De facto, Mummu, deus da trindade Babilónica, só foi impedido de levar o Universo dos deuses à total desordem porque Ea, deus da criatividade, o controlou. Mummu é de há cento e trinta anos a esta parte reconhecido pela designação de entropia e Ea, há pouco mais de meio século, chama-se informação.

Este livro é em parte o texto de aulas que um de nós proferiu, há cerca de trinta anos e outro de nós ouviu e alargou. Depois, junta a energia e a informação e com elas constroi as condições para obter estruturas coesas evolutivamente mais complexas. Finalmente contém neste périplo, que vai desde os princípios básicos até resultados de ponta da investigação, no domínio da informação um conjunto de aplicações industriais, metereológicas, ecológicas e sociais que mostram a largueza do conceito de entropia como caracterizador de sistemas.

* Poema da Criação da Babilónia

"... The law that entropy always increases - The second Law of Thermodynamics holds, I think, the supreme position among the laws of Nature..."

Arthur A. S. Eddington, 1928

"... Shannon replied: 'My greatest concern was what to call it. I thought of calling it information, but the word was overly used, so I decided to call it uncertainty. When I discussed it with John von Neumann he had a better idea. Von Neumann told me, you should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage...'"

Myron Tribus, 1971

"... Eis, salvo erro, uma versão possível de problemas em torno da 'Teoria da Informação' - palavra que até há pouco significava notícia ou mexerico, e que agora também quer dizer, mais finamente, trabalho fornecido, energia transformada, e daí também, por consequência, aumento da tal 'entropia' - e até a sua reconversão ou negação: a negantropia (contos largos ...)."

*Vitorino Nemésio
Diário Popular, 14-3-1972*

INTRODUÇÃO

Este livro resulta de um conjunto de lições e de temas de investigação que se verificou constituírem uma teoria para a aplicação do conceito e das várias medidas de entropia à caracterização de sistemas.

Sistema é aqui entendido como uma abstracção da realidade e é definido por um conjunto de variáveis escolhidas, caso a caso, para modelar em cada instante o comportamento dominante do objecto do nosso interesse.

Assim sendo, a entropia do sistema é uma propriedade que é independente do modelo atribuível aos seus componentes individuais. Como tal, tanto a física estatística como a física quântica, são susceptíveis de serem aplicadas para a previsão das medidas resultantes da observação da entropia dos sistemas. No entanto, não é a partir delas que pode resultar uma teoria unificadora de propriedades dos sistemas.

Neste livro o caminho seguido diz respeito à caracterização dos sistemas nos seus aspectos energéticos (Parte I) e nos seus aspectos da informação através de medidas entrópicas (Parte II). Na Parte III é explorada a relação existente entre energia e informação de um sistema. Alguns exemplos ilustram na Parte III a aplicabilidade do método da observação da entropia de sistemas, quer no domínio da energia, quer no domínio da informação, à previsão do seu comportamento. Finalmente a Parte IV aponta alguns caminhos possíveis para o estabelecimento de uma teoria unificadora de sistemas onde se relacionam entropia, complexidade e coesão.

Pretende-se assim ter construído um livro que é ao mesmo tempo um livro de texto para a tentativa de caracterização global de sistemas e um guia exploratório de áreas de investigação na previsão do seu comportamento.

Esperamos que o critério para a aceitação de novas ideias seja a verificação experimental dos resultados que porventura essas ideias possam levar a prever. Qualquer esclarecimento conceptual ou filosófico que uma nova ideia possa trazer, considera-se apenas como um benefício de importância secundária. A sua aceitação ou a sua rejeição é apenas uma questão do gosto pessoal.

NATUREZA ESTATÍSTICA DA ENTROPIA



1.1. EVOLUÇÃO DO CONCEITO DE ENTROPIA

O conceito de entropia, que surgiu em Termodinâmica Clássica Fenomenológica com Clausius, em 1865, tem-se revelado duma tal fecundidade, que hoje se estende da Física, à Química; da Engenharia às Ciências da Vida; da Economia, à Sociologia; da Matemática, à Linguística; da Arte, à Filosofia. O conceito de entropia, que durante tanto tempo, andou de mãos dadas com o de energia, universalizou-se e invadiu muitos ramos do saber e muitas manifestações de cultura.

Este conceito, tão vasto e tão profundo, apareceu em Termodinâmica, que é a Ciência que trata da energia e das suas relações com a matéria. Como qualquer disciplina de Física, é uma Ciência que assenta na observação e na experiência.

A estrutura da Termodinâmica é, na essência, muito simples e baseia-se num número reduzido de conceitos e de postulados. Esta estrutura lógica, que hoje achamos quase natural, é ao mesmo tempo, simples e completa, como reconhecia Einstein. Foi o resultado da reflexão e do trabalho de grandes homens de Ciência e não se obteve de forma linear e imediata. A sua edificação levou tempo a consolidar-se para hoje poder dar frutos, que se estendem a tantos domínios da Ciência e da Tecnologia.

Historicamente, a Termodinâmica desenvolveu-se a partir da necessidade de explicar o funcionamento das máquinas térmicas, que absorvem calor duma fonte quente e cedem parte a uma fonte fria, para produzir trabalho. Era, por isso, que se definia "Termodinâmica" como "Ciência que trata das relações entre calor e trabalho". Como vão longe esses tempos! A Termodinâmica estende-se hoje aos mais variados domínios das Ciências e da Tecnologia, envolvendo todas as formas de energia e a sua interacção com a matéria.

A via fenomenológica, macroscópica, da Termodinâmica não requer qualquer hipótese sobre a constituição e a estrutura da matéria e, por isso, as leis da Termodinâmica são universais. No entanto, não podia ficar indiferente aos progressos enormes, que as Teorias Cinética e Molecular levaram aos vários domínios da Física, da Química, da Engenharia e da Biologia. Ainda que a Termodinâmica fenomenológica, clássica, permita prever muitas das relações entre as propriedades da matéria e da energia, não pode interpretar, nem explicar, muitas vezes, a razão da forma dessas relações. Por isso, tem que se recorrer a uma via microscópica, aceitando um modelo atómico, ou molecular, e saber aplicar as leis que regem a Física das partículas (desde a Mecânica de Newton até à Mecânica Quântica, etc.).

Um sistema macroscópico constituído por muitas partículas, não é apenas um sistema puramente mecânico e a *energia interna*, U , não é o factor exclusivo e determinante do seu comportamento. A energia interna dum sistema é a soma das energias dos elementos individuais que o constituem, que inclui as energias cinética e potencial, que resulta das interações de cada uma das partículas nos outros elementos. Cada um dos modos de energia que cada elemento pode armazenar define um grau de liberdade. O sistema é regido por leis da Termodinâmica e da Mecânica Estatística, em que outras grandezas, como *temperatura* T e a *entropia* S , que é uma medida da desordem, têm um papel decisivo.

Um sistema mecânico, simples, como uma bola que rola por uma encosta abaixo, atinge o estado de equilíbrio no fundo do vale em que a energia potencial é mínima. Por outro lado, um sistema termodinâmico, a temperatura constante, atinge o seu estado de equilíbrio quando a *energia livre*, dada por $F = U - TS$, atinge o valor mínimo e não quando a sua energia é mínima.

Da expressão da energia livre, $F = U - TS$, podem deduzir-se algumas das características mais salientes dum sistema. Para temperaturas baixas, é a energia das moléculas que dá a contribuição maior para a energia livre e, por isso, o estado favorecido, que minimiza esta energia, é o estado sólido. Para temperaturas mais elevadas, a entropia do sistema passa a ter a influência dominante. Por isso, as fases fluídas são aquelas que são mais estáveis para temperaturas elevadas, ainda que constituam configurações de energias mais elevadas do que as correspondentes ao estado sólido.

Os conceitos de entropia e de energia livre podem apreender-se facilmente quando se considera a relação entre as descrições microscópicas e macroscópicas dum sistema termodinâmico. O estado macroscópico do sistema é caracterizado por propriedades médias colectivas de todas as partículas, tais como a pressão, a temperatura e a densidade. O microestado de um sistema é definido, em cada instante, por exemplo, pela configuração do edifício molecular, desi-

gnadamente pela posição em cada instante das diferentes moléculas. Um ponto essencial é o de que um macroestado pode ser realizado por muitos estados microscópicos diferentes. Por outras palavras, há muitas configurações microscópicas que conduzem aos mesmos valores da temperatura, da densidade, da entropia e de outras grandezas fenomenológicas, que caracterizam o macroestado.

A incorporação da teoria molecular da matéria no estudo da Termodinâmica levou à estruturação duma nova disciplina, que é a Mecânica Estatística. Esta é fundamentalmente uma disciplina analítica, que assenta na aplicação da Teoria da Probabilidade às leis do movimento das partículas.

A Mecânica Estatística teria sido irrelevante se não permitisse determinar as propriedades estatísticas médias de muitas partículas e se estas propriedades médias não coincidissem com as propriedades termodinâmicas obtidas por via fenomenológica (observáveis). Uma das grandes vitórias da Mecânica Estatística foi dar a interpretação, por esta via microscópica, do conceito de entropia. A sua importância, como disciplina autónoma, consolidou-se a partir do momento em que surgiu a Mecânica Quântica. De facto, foi a incapacidade da Mecânica Estatística, (como então era concebida), para explicar as discrepâncias (aparentes) da radiação térmica, que levou Plank (1901) a propôr a "teoria quântica da radiação".

Esta, depois, foi alargada para constituir hoje a Mecânica Quântica. Hoje a Mecânica Estatística assenta numa concepção mais geral e unificadora em que a dinâmica de um sistema físico é representada por um conjunto de estados quânticos. E o desenvolvimento da Mecânica Estatística faz-se com base no formalismo quântico.

∴

Como se depreende das "citações" o conceito de entropia sofreu uma generalização inesperada, num contexto completamente novo, quando Shannon (1948) desenvolveu a Teoria da Informação, que mais tarde seria consolidada por Shannon e Warren Weaver (1949) com a publicação da Teoria Matemática da Comunicação.

Foi com surpresa que verificaram que a expressão da quantidade de informação coincidia com a da entropia. Expressaram essa descoberta surpreendente numa afirmação que ficou célebre:

"Thus, when one meets the concept of entropy in Communication Theory, he has a right to be rather excited - a right to suspect that one has hold of something that may turn out to be basic and important".

("Portanto, quando alguém encontra o conceito de entropia na teoria da comunicação, tem razão para ficar bastante excitado - uma razão para suspeitar que dispõe de qualquer coisa que pode ser fundamental e importante".)

∴

A Teoria da Informação conduziu a resultados que coincidem com vários conceitos da Termodinâmica. Como a entropia é uma medida da desordem, é fácil aceitar que pode estar relacionada com a ordem e desordem de mensagens.

Deve notar-se que a medida da quantidade de informação obtida por Shannon, foi estabelcida, independentemente da Física, a partir da Teoria das Probabilidades e da Lógica. A Teoria da Informação fornece um significado e uma interpretação da entropia, independentemente da Termodinâmica. Usando este significado (incerteza) é possível derivar os resultados das Termodinâmicas Clássica e Estatística duma forma directa e simples.

Quando comparada com a fórmula de Boltzmann, verifica-se que a fórmula de Shannon é muito mais universal. De facto, constitui uma ferramenta lógica, que é aplicável em muitos domínios da Ciência e da Tecnologia, muito para além da Termodinâmica:

Foi com base na Teoria de Shannon que se estabeleceu um novo método de inferência estatística, hoje tão importante na análise das séries cronológicas, e na análise de sinais, que é designado por "Método da Entropia Máxima" (E. T. Jaynes, 1958). Este assenta, fundamentalmente, num princípio que diz que *"a lei de distribuição de probabilidade que corresponde à distribuição menos viciada possível e mais isenta é a que corresponde ao máximo da incerteza (entropia)"*.

Este novo método de inferência estatística tem tido uma aplicação muito ampla em vários domínios da Geofísica e da Meteorologia. De repente, transformou-se numa ferramenta das Ciências Aplicadas, extremamente útil, na Climatologia para a análise de distribuição de redes meteorológicas e hidrológicas e para a análise espectral das séries climatológicas temporais e outras.

Em 1957, E. T. Jaynes avançou ainda, quando mostrou que as ideias e os conceitos da Teoria da Informação se podem considerar mais fundamentais do que os da Termodinâmica, porque todas as fórmulas de Mecânica Estatística se podem obter a partir daquelas. Uma vez obtidas as fórmulas de Mecânica Estatística podem derivar-se os próprios princípios da Termodinâmica.

Aceitando a concepção de Jaynes os princípios da Termodinâmica seriam apenas teoremas... É claro que esta atitude "tão polémica", tem originado "discussões" interessantes (Peixoto, 1972).

1.2. A TERMODINÂMICA, A MECÂNICA E A ESTATÍSTICA

A aplicação de alguns aspectos elementares da Teoria Cinética permite interpretar várias propriedades dos gases, partindo dum modelo molecular simples a que se aplicam as leis da Mecânica. Também é possível estudar algumas propriedades termodinâmicas dos sólidos e dos líquidos incluindo as transições de fase (Peixoto 1986).

Agora, utiliza-se um método mais geral baseado, essencialmente, na aplicação simultânea das leis da Mecânica e do Cálculo das Probabilidades a sistemas constituídos por grande número de partículas (moléculas, átomos, etc), e que constitui a Mecânica Estatística.

De facto, para estudar o comportamento mecânico dum sistema, será preciso resolver as equações do movimento de cada uma das partículas, tendo em consideração as condições iniciais, as acções mútuas entre as partículas e as que resultam de grupos estranhos. Ora uma tal tarefa para um sistema com um número muito grande de graus de liberdade, é humanamente impossível. No entanto, a mesma complexidade das evoluções a nível microscópico, permite abordar o problema duma forma diferente.

O que se conhece dum sistema é dado pelos resultados de medições, obtidos com instrumentos de medida. Estes não são afectados pelo movimento de cada partícula em separado, mas sim por um grande número de partículas, e se estas se puderem considerar membros duma população estatística, homogénea, tais grandezas que definem o estado *observável* do sistema, devido à regra da \sqrt{N} , raramente apresentarão grandes discrepâncias (flutuações) em relação aos seus valores medidos. Assim, os métodos estatísticos permitem deduzir leis que se cumprem com um grande grau de aproximação e até à ordem de grandeza, a prever para as suas variações, através de momentos estatísticos de ordem mais elevada.

Nesta concepção as previsões estatísticas, enquanto se aplicam a grandes populações são deterministas, apesar de se basearem no Cálculo das Probabilidades.

A Mecânica Estatística é um ramo da Física Teórica que tem por objectivo interpretar e prever as propriedades macroscópicas dos sistemas (observáveis), em função do comportamento dinâmico dos elementos microscópicos que os constituem. Assim, é possível, a partir de parâmetros atômicos e moleculares, determinar propriedades que só se obtinham, ou inferiam, por via experimental (v. g. capacidades caloríficas, entropia, etc.), os elementos e as "partículas" podem ser tomadas em sentido muito amplo, como estrelas no firmamento, bolas de bilhar, cartas de um baralho, gotículas das nuvens, moléculas, átomos, iões, electrões, fotões, etc. O conhecimento do comportamento dinâmico das partículas é suficiente para predizer a conduta estatística do sistema global, isto é, para prever as propriedades macroscópicas dos sistemas (observáveis). Estas são dadas pelos valores médios de determinadas grandezas microscópicas. Surge assim a necessidade de recorrer a métodos estatísticos na descrição das propriedades macroscópicas dos sistemas.

Poderia parecer, à primeira vista, que, quando o número de elementos dum sistema atinge valores extremamente elevados, a sua complexidade aumentaria e que o seu tratamento se tornaria impossível. Numa concepção determinista clássica seria esta a ideia dominante. Mas, estes sistemas, constituídos por um grande número de partículas, apresentam regularidades estatísticas bem definidas. De facto, estão sujeitos à "tirania estatística" da lei dos grandes números e aos métodos estatísticos.

Estas propriedades estatísticas, que surgem como consequência da existência de grande número de elementos, não podem justificar-se com base na Mecânica, mas também não faz sentido estendê-las a sistemas formados por um número pequeno de partículas. Além disso, os resultados mais imediatos da Mecânica Estatística são os que correspondem a configurações de equilíbrio, como acontece com a Termodinâmica Fenomenológica Clássica. Já com a Teoria Cinética, se podia fazer a sua aplicação fora do equilíbrio, como acontecia, por exemplo, com os fenómenos de transporte. (Equação de Boltzmann). A Teoria Cinética constitui provavelmente o caso mais conhecido em que vários conceitos da termodinâmica podem ser estabelecidos a partir dos princípios da Mecânica, aplicando-os à dinâmica molecular.

Deve acentuar-se que a Mecânica Estatística permite deduzir os princípios fundamentais da Termodinâmica que, por isso, passariam a ser, pura e simplesmente, teoremas. Assim, a Mecânica Estatística estabelece a conexão entre as propriedades a nível microscópico e as propriedades macroscópicas dum

sistema, surgindo assim como uma "explicação" ou uma "interpretação" da Termodinâmica.

Em Mecânica Estatística Clássica admite-se que as partículas, cujas energias podem assumir qualquer valor, são discerníveis, podendo distinguir-se umas das outras, e que são localizáveis. É o que se passa com os átomos de um cristal que podem distinguir-se uns dos outros pelas posições bem definidas, que ocupam na rede cristalina. Estas partículas seguem a estatística de Maxwell - Boltzmann (M-B).

Há partículas que se dizem quânticas; que só podem adquirir certos valores permitidos da energia. Estes distribuem-se por níveis, ou estados energéticos, que se especificam por meio duma série de níveis quânticos. As partículas quânticas são indiscerníveis, isto é, parecem iguais entre si, sem possibilidade de se distinguirem e de se localizarem. Para o estudo destas partículas há duas estatísticas quânticas diferentes:

- a estatística de Bose-Einstein (B-E) para as partículas de spin zero e inteiro, como as partículas α , fótons, etc;
- e a estatística de Fermi-Dirac (F-M) de spin semi-inteiro, como os electrões, os prótons, os neutrões, etc., e que satisfazem o princípio da exclusão de Pauli (em cada nível energético só pode haver uma, ou nenhuma partícula).

Numa concepção mais geral e unificadora a Mecânica Estatística engloba de facto, todos os casos (clássico, quântico, B-E, F-D, etc.) e tem feito grandes progressos nas últimas décadas.

A Mecânica Estatística, clássica, que é aquela a que, temos necessidade de recorrer nesta exposição pode considerar-se como um processo limite da Mecânica Estatística Quântica. De facto, para sistemas macroscópicos, no limite termodinâmico, os três formalismos são equivalentes.

2. PROBABILIDADE TERMODINÂMICA

2.1. MICROESTADOS E MACROESTADOS DUM SISTEMA

Começemos por considerar um sistema constituído por um grande número, N , de elementos microscópicos, por exemplo, moléculas ou átomos.

Imaginemos um observador ultra perspicaz, que conseguisse ver as moléculas, distingui-las umas das outras e que pudesse determinar as posições e o seu estado de movimento.

Podia, assim, descrever-se o estado microscópico dum sistema, ou o seu microestado, especificando, em cada instante, as coordenadas e as quantidades de movimento de todas as partículas que constituem o sistema. Cada uma das diferentes distribuições de N partículas, mesmo que só difiram entre si pelas coordenadas duma molécula, constitui já um microestado diferente do sistema.

Um microestado dum sistema termodinâmico é, portanto, a descrição completa e especificada até ao último detalhe em termos das propriedades, da natureza e do movimento de cada uma das partículas que constituem um sistema. Um estado macroscópico de um sistema ou um macroestado tal como é observado por um experimentador é um estado do sistema expresso em termos de propriedades médias globais de toda a colecção de partículas que o constituem. Um macroestado é realizado por um grupo de microestados com propriedades médias comuns. Num macroestado, a descrição do estado do sistema é feita em termos das propriedades integradas, tais como o volume, a massa, o número de moles, a energia e o movimento global do conjunto de partículas, etc.

Um sistema isolado atinge o estado de equilíbrio, quando, no decurso do tempo não ocorrem transformações a nível macroscópico ou, como veremos, quando as probabilidades dos vários microestados acessíveis não variarem com

o tempo. Quando se perturba o estado macrosópico de um sistema basta deixar decorrer um certo tempo, chamado tempo de relaxação, para que o sistema volte à situação de equilíbrio; nesta situação as propriedades extensivas e intensivas diferem tão pouco dos seus valores médios, que dificilmente se poderá medir essa diferença.

A hierarquização das descrições de microestado, macroestado e de estado termodinâmico, envolve um decréscimo de informação, mas é compensada por um aumento de simplicidade. O conceito de macroestado é muito útil para compreender como é que a estrutura microscópica do sistema afecta as suas propriedades globais; o estado termodinâmico é a descrição que corresponde à dimensão laboratorial do sistema.

Para elucidação das ideias expostas vamos recorrer a uma analogia. Consideremos uma cidade de 1 000 000 habitantes de que se pretendem representar as características. A informação "mais grosseira" será a especificação do próprio número de habitantes, que vive na cidade, o que corresponde ao "estado termodinâmico" do sistema. A especificação da distribuição da população pelos diferentes bairros, ou zonas, definiria "macroestados" da população. O censo completo da população, realizado em certo instante, indicando cada um dos habitantes com nomes, idades, moradas, profissões, localização, definiria um microestado da população. Com a informação contida no microestado poderiam definir-se outros "macroestados", organizando indivíduos por "idades" (quantos estão em idade escolar primária; quantos reformados); por bairros; por "actividades" (quantos carpinteiros, quantos estudantes, quantos contabilistas); por "rendimento económico" (quantos indigentes, quantos milionários). A cada um destes agrupamentos corresponderia um macroestado.

A probabilidade é a medida da nossa expectativa para a realização dum acontecimento futuro. Recordemos que a probabilidade matemática da realização dum acontecimento contingente é dada pela relação entre o número de casos favoráveis e o número de casos possíveis.

Vejamos um exemplo. A probabilidade de tirar à sorte um ás de um baralho de cartas é quatro vezes maior do que tirar o ás de copas. A probabilidade de extrair ao acaso uma carta de copas é treze vezes maior do que a de tirar o ás de copas. Quanto mais detalhada for a especificação dum acontecimento menor é a sua probabilidade, ou seja, menos frequente ele é. A especificação duma certa carta do baralho corresponde a um microestado e a especificação duma classe de cartas, como as cartas do mesmo naipe, as senas, ou as cartas distribuídas a um jogador, correspondem a outros tantos macroestados.

Como se vê, um macroestado de um sistema (estado definido à maneira clássica) pode ser realizado por um grande número de microestados diferentes.

No caso do modelo molecular ainda se torna mais evidente. Basta pensar, por exemplo, que, devido à agitação molecular, qualquer molécula no instante seguinte, está numa posição diferente e com outra velocidade. Daí a imensa variedade de microestados por que passa um sistema (tanto faz falar de moléculas, como de átomos, de electrões, de fotões, etc.). Aos microestados também se chamam *complexões* do sistema (Planck).

Da miríade de microestados possíveis, chamam-se microestados acessíveis àqueles que são compatíveis com as ligações, ou com os constrangimentos do sistema. Por exemplo, se se trata de um gás confinado a um dado volume, V , esta condição limita as posições possíveis das moléculas. Se a energia total de um sistema for E , esta condição condiciona as energias possíveis das partículas individuais E_i que o constituem.

Chama-se probabilidade termodinâmica, ou peso estatístico, W , ao número de microestados, ou de complexões, que realizam um mesmo macroestado de um sistema. O número W (*) serve para nos dar uma medida da probabilidade de ocorrência do estado termodinâmico. É sempre um número inteiro, superior à unidade, ao contrário da probabilidade matemática, que varia entre 0 e 1.

A probabilidade matemática, p , para que o sistema global se encontre num dos W microestados acessíveis, é dada por:

$$p = 1 / W$$

Se for W_0 o número total de microestados possíveis, a probabilidade matemática, p , de que o estado seja realizado por qualquer dos microestados acessíveis é dada por:

$$p = W / W_0 \equiv \frac{\text{número de microestados acessíveis}}{\text{número de microestados possíveis}}$$

É fácil verificar que os pesos estatísticos e as probabilidades matemáticas, referentes a dois estados, i e j , são proporcionais:

$$p_i / p_j = W_i / W_j$$

Um sistema isolado (caracterizado por E, V, N), diz-se que está em equilíbrio, quando as probabilidades dos vários estados acessíveis não variam com o tempo.

* Wahrscheinlichkeit = probabilidade

Para os vários microestados, aceita-se o princípio fundamental da equiprobabilidade *a priori*:

Num sistema isolado em equilíbrio todos os microestados acessíveis são igualmente prováveis, ou ainda: um sistema isolado pode encontrar-se em qualquer dos microestados acessíveis. Não há, portanto, microestados privilegiados: todos são igualmente prováveis.

2.2. UM EXEMPLO

Vejamos, por meio dum exemplo, a diferença entre microestado e macroestado e entre peso estatístico, ou probabilidade termodinâmica, e probabilidade matemática.

Consideremos 5 moedas numeradas (discerníveis), que lançamos ao ar, e analisemos as várias possibilidades de obter caras ou cunhos. Em cada lançamento há duas possibilidades para cada moeda e a probabilidade de sair cara ou cunho é $1/2$.

Registemos o número de caras e cunhos que vão aparecendo nos sucessivos lançamentos, especificando a que moeda pertencem. Qualquer combinação de caras (C) ou cunhos (c) que imaginemos é igualmente provável. De quantas maneiras, na sequência do lançamento das 5 moedas é que se pode obter só caras ou só cunhos. É evidente que só há uma combinação possível em cada caso, isto é, obter-se as combinações CCCCC e ccccc. E de obter quatro caras e um cunho? Neste caso o número de maneiras aumenta, visto que as combinações possíveis são cinco CCCCC; CCCcC; CCcCC; CcCCC; cCCCC. Fazendo o mesmo exercício para o caso de saírem 3 caras e 2 cunhos, chegaríamos à conclusão que haveria 10 hipóteses possíveis.

Assim, com o lançamento das cinco moedas, podem dar-se várias possibilidades, que passamos a enumerar:

1 possibilidade de obter 5 caras	$W_1 = 1$
5 possibilidades de obter 4 caras e 1 cunho	$W_2 = 5$
10 possibilidades de obter 3 caras e 2 cunhos	$W_3 = 10$
10 possibilidades de obter 2 caras e 3 cunhos	$W_4 = 10$
5 possibilidades de obter 1 cara e 4 cunhos	$W_5 = 5$
1 possibilidade de obter 5 cunhos	$W_6 = 1$
	<hr/>
	$\Sigma W = 32$

A especificação do número de caras e de cunhos define o macroestado e cada uma das suas distribuições possíveis define um microestado. O número dessas distribuições (microestados) define o peso estatístico, W .

Ao todo, há $2^5 = 32$ casos possíveis, isto é, cada microestado tem a mesma probabilidade de $1/32$. As possibilidades de que apareça um macroestado formado por 4 caras e 1 cunho, são $W_2 = 5$. Logo, a probabilidade matemática para que se verifique este macroestado é $p_2 = 5/32$. O macroestado constituído por 3 caras e 2 cunhos tem um peso estatístico $W_3 = 10$ e uma probabilidade matemática de $p_3 = 10/32$, o que significa que em 320 lançamentos das cinco moedas apareceriam, em média, cem vezes 3 caras e 2 cunhos..

É fácil constatar que os pesos estatísticos, W , são sempre números positivos e inteiros e que não são todos iguais. Quanto mais repartidos e melhor divididos estiverem as caras e os cunhos das moedas, maior é o peso estatístico e maior é o número de microestados que realizam um dado macroestado. Assim mais difícil será "saber" qual é o microestado num dado instante do sistema, ou seja, maior é a desordem. As probabilidades matemáticas, p , são também números positivos que variam entre 0 e 1. Como se pode verificar os valores dos "pp" e dos "WW" são proporcionais:

$$p_2 / p_3 = W_2 / W_3$$

Como vemos, para definir o peso estatístico, não interessa saber quais, mas quantas são. E o problema geral pode pôr-se nestes termos: Suponhamos N objectos que se podem distribuir por g caixas (células), todas iguais. Qual será a probabilidade de que haja N_1 objectos na 1ª caixa, N_2 na segunda, etc., e sabendo que $N = \sum N_i$. O número de maneiras para que haja N_1 objectos na 1ª caixa é dado pelo número de combinações do N objectos, N_1 a N_1 : $C_{N_1}^N$. Ficam agora $N - N_1$ objectos de entre os quais se vão tirar N_2 , para a caixa 2 o que se pode dar por $C_{N_2}^{N-N_1}$, maneiras diferentes. Restam $N - N_1 - N_2$ dos quais se vão tirar N_3 , o que acontece de $C_{N_3}^{N-N_1-N_2}$ modos diferentes e assim sucessivamente.

O número W_3 de maneiras diferentes de realização a distribuição desejada:

$$W = C_{N_1}^N \times C_{N_2}^{N-N_1} \times C_{N_3}^{N-N_1-N_2} \times \dots, \text{ ou seja, notando que } C_j^n = \frac{n!}{(n-j)! j!}$$

$$W = N! / N_1! N_2! N_3! \dots N_j!$$

Assim, W representa o número de microestados que corresponde a haver N_1 partículas na célula 1, N_2 na célula 2, ..., N_j na célula j . De facto, a probabilidade

termodinâmica para o caso das N partículas se distribuírem por j células poderia também deduzir-se notando que:

a) o número de permutações de N objectos é $N!$;

b) que se devem excluir as que resultam de permutar entre si as partículas de cada recinto. Para isso tem que se dividir sucessivamente, pelos valores de $N_1!$, de $N_2!$... de $N_j!$, que correspondem às permutações "ineficazes" em cada célula.

∴

No nosso exemplo é $N = 5$ (moedas numeradas) e há 2 células (caras ou cunho). Logo as probabilidades termodinâmicas dos seis macroestados, atrás referidos, são:

$$\begin{array}{lll} W_1 = 5!/5!0! = 1 & W_2 = 5!/4!1! = 5 & W_3 = 5!/3!2! = 10 \\ W_4 = 5!/2!3! = 10 & W_5 = 5!/1!4! = 5 & W_6 = 5!/0!5! = 1 \end{array}$$

valores que já tínhamos obtido anteriormente, por via intuitiva.

2.3. COLECTIVIDADE ESTATÍSTICA

Consideremos agora o caso geral de um sistema em contacto com um reservatório com o qual pode ou não interactivar. Nós podemos imaginar que preparamos um grande número de sistemas independentes entre si, mas idênticos ao sistema que pretendemos estudar e que todos eles interactivam com o mesmo reservatório. Então o número relativo desses sistemas em cada um dos estados possíveis reflecte a probabilidade do sistema de referência se encontrar nesses estados. Este conjunto, ou colecção de um número elevado de sistemas independentes, identicamente preparados e sujeitos às mesmas condições exteriores constitui uma *Colectividade estatística* («ensemble»).

A *colectividade* é uma colecção fictícia, conceptual, de um grande número de sistemas, que ocupam volumes iguais, têm o mesmo número de partículas com estruturas análogas e que estão submetidas às mesmas condições fronteira. Ora, os valores médios das propriedades do sistema de referência no decurso do tempo podem substituir-se pelos valores médios de todos os membros da colectividade («cópias» do sistema) num instante determinado (*hipótese quasi-ergódica*).

Por exemplo, suponhamos que um sistema S pode estar num dos três estados S_1 , S_2 , S_3 . Suponhamos que «preparamos» uma colectividade de 1000 sistemas idênticos («cópias» de S) e que os membros da colectividade interactivam com o mesmo reservatório. Num certo instante verificamos que 90 estão no estado S_1 , 650 no estado S_2 e 260 no estado S_3 . Então as probabilidades de que o sistema S esteja num desses estados no decurso do tempo serão respectivamente $p_1=0,09$; $p_2=0,65$ e $p_3=0,26$.

É claro que no caso real os sistemas têm muito mais do que três estados possíveis e as Colectividades terão muitíssimos mais do que 1000 membros.

As Colectividades estatísticas podem ser hierarquisadas de acordo com o tipo de interacção que os seus membros têm com o reservatório. Ora, os tipos de interacção entre sistemas incluem trocas de calor, de trabalho e de partículas. Os membros de uma Colectividade podem não interactuar por nenhum dos processos sem o seu reservatório ou interactuarem pelos três processos simultaneamente, ou por qualquer combinação destes.

Dos vários casos possíveis, os mais usuais são os três tipos de interacção que, por isso têm nomes especiais. Assim se os membros da Colectividade não interactuam com a sua vizinhança, a Colectividade diz-se *microcanónica*. Se permuta trabalho e calor com o reservatório a Colectividade diz-se *canónica*, se permuta calor, trabalho e partículas a Colectividade diz-se *grande canónica*.

Os membros duma Colectividade *microcanónica* estão limitados por paredes restritivas à massa e à energia (paredes rígidas, impermeáveis e *adiatérmicas*). São sistemas fechados e isolados. As grandezas E , V , N permanecem constantes. Na *Colectividade canónica* as paredes são rígidas, impermeáveis à massa e *diatérmicas* (não restritivas para a «energia em trânsito»). As grandezas T , V e N de cada membro da Colectividade, mantêm-se constante, enquanto que E varia. Por último, na *Colectividade grande canónica* os membros são limitados por paredes rígidas, *permeáveis* e *diatérmicas*. As grandezas T , V e μ (potencial químico) permanecem constantes enquanto que V e N são variáveis. Os sistemas da Colectividade permutam energia e massa.

A «Colectividade estatística» não tem realidade física. Foi uma estrutura conceptual que Gibbs introduziu para analisar o comportamento de um sistema e estudar as suas propriedades e analisar as flutuações.

3. INTERPRETAÇÃO MECANICISTA DO SEGUNDO PRINCÍPIO FUNDAMENTAL DA TERMODINÂMICA

3.1. FÓRMULA DE BOLTZMANN

A primeira tentativa de dar uma interpretação mecanicista do Segundo Princípio Fundamental da Termodinâmica foi apresentado por Ludwig Boltzmann, em 1866. A sua preocupação foi a de dar uma prova, puramente analítica e completamente geral, do Segundo Princípio e, ao mesmo tempo, de descobrir um teorema da Mecânica, que lhe correspondesse. Era, então, preocupação dominante reduzir as leis da fenomenologia da Física às leis da Mecânica.

Se o Primeiro Princípio da Termodinâmica se podia interpretar, à luz da teoria Cinética, de forma clara, introduzindo as energias potencial e cinética das moléculas, não seria também possível encontrar uma interpretação mecanicista do Segundo Princípio?

Esta era a grande questão que Boltzmann pôs a si próprio e que procurou resolver.

Boltzmann não foi o único físico que tentou reduzir o Segundo Princípio da Termodinâmica a um teorema da Mecânica. Clausius reconhecia que, à escala molecular, o Segundo Princípio era muito mais difícil de apreender do que o Primeiro. Introduziu o conceito de "desagregação", como uma medida do grau de dispersão das moléculas. No entanto, foi Boltzmann que descobriu um teorema - o *teorema H*, que relaciona a entropia com a probabilidade termodinâmica.

O grande sucesso de Boltzmann e da Mecânica Estatística, foi conseguir dar uma interpretação da entropia, ao demonstrar que é uma grandeza que mede o grau de probabilidade de um sistema.

Com o teorema *H*, Boltzmann demonstrou que a *entropia é uma função crescente da probabilidade termodinâmica, isto é, do número W de complexões.*

Mais tarde, Boltzmann enunciou um princípio que alarga, consideravelmente, o alcance do teorema *H*. Pode enunciar-se assim:

A entropia S , dum macroestado dum sistema é uma função universal, exclusiva, da probabilidade termodinâmica W : $S = f(W)$.

Vamos ver, como é que, com um enunciado tão geral como o deste princípio, se pode estabelecer a forma da função f , que dá aquela dependência.

Comecemos por considerar um sistema A , fechado, cuja probabilidade termodinâmica é W , subdividido em dois sistemas: A_1 e A_2 : $A = A_1 \cup A_2$.

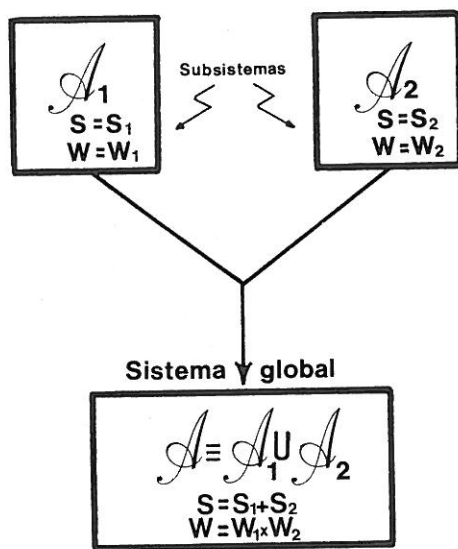


Fig. 1 - A entropia é aditiva e o número de microestados é multiplicativo.

Sejam W_1 e W_2 as suas probabilidades termodinâmicas respectivas.

É evidente que a probabilidade da realização de A é a que resulta das probabilidades W_1 e W_2 , da realização de A_1 e de A_2 , simultaneamente.

Cada forma de realização (microestado) de um dos subsistemas do sistema, pode combinar-se com todos os microestados do outro subsistema, para dar novos microestados do sistema global. Logo, quando todos os estados W_1 , de

A_1 , se combinam com os W_2 estados de A_2 , obtêm-se a probabilidade termodinâmica do sistema total, W , que é igual ao produto das probabilidades dos componentes:

$$W = W_1 \times W_2$$

Portanto, pelo princípio de Boltzmann, a entropia S do sistema A , satisfaz a expressão:

$$S = f(W) = f(W_1 \times W_2)$$

Como, por outro lado, a entropia é uma propriedade aditiva, verifica-se ainda, para o sistema A , a condição:

$$S = S_1 + S_2 = f(W_1) + f(W_2)$$

ou seja:

$$f(W) = f(W_1 \times W_2) = f(W_1) + f(W_2)$$

A partir desta relação funcional pode determinar-se a forma da função f , que satisfaz a esta igualdade. De facto, derivando ambos os membros em ordem a W_1 , (mantendo W_2 constante) e em ordem a W_2 (mantendo W_1 constante) resulta.

$$\partial f(W) / \partial W_1 = \partial f(W) / \partial W \cdot \partial W / \partial W_1 = \partial f(W) / \partial W \cdot W_2 = \partial f(W_1) / \partial W_1$$

e analogamente, para $\partial f(W) / \partial W_2$:

$$\partial f(W) / \partial (W \cdot W_1) = \partial f(W_2) / \partial W_2$$

E, assim pode obter-se o eliminante:

$$\partial f(W) / \partial W = 1/W_2 \cdot \partial f(W_1) / \partial W_1 = 1/W_1 \cdot \partial f(W_2) / \partial W_2$$

Multiplicando estas equações por $W = W_1 \times W_2$ obtêm-se expressões que são funções exclusivas duma mesma variável.

$$W \partial f(W) / \partial W = W_1 \cdot \partial f(W_1) / \partial W_1 = W_2 \cdot \partial f(W_2) / \partial W_2 = \text{const.} = k$$

Como cada membro das duas equações é uma função exclusiva de uma só das variáveis, e como W_1 , ou W_2 , são independentes entre si, aquelas igualdades só se podem verificar, se cada membro assumir um mesmo valor constante, k .

As equações anteriores são da forma:

$$df(x) = c \cdot dx / x$$

que por integração, conduz a:

$$f(x) = c \log x + \text{const.}$$

Portanto, no nosso caso, tem-se:

$$f(W_1) = k \log W_1 + C_1$$

e

$$f(W_2) = k \log W_2 + C_2$$

Como vimos, pelo princípio de Boltzmann é: $S = f(W)$. Por outro lado, devido à aditividade da entropia S , tem-se:

$$f(W) = f(W_1) + f(W_2),$$

e simultaneamente:

$$S = f(W) = f(W_1 \cdot W_2) + \text{const.}$$

expressão equivalente a:

$$S = k \log W + \text{const.}$$

Para dois estados, a e b , do sistema, esta expressão permite escrever a variação de entropia:

$$S_b - S_a = k \log W_b / W_a$$

ou, obter a razão das complexões na forma exponencial, isto é:

$$W_b / W_a = \exp (\Delta S / k)$$

em que $S_b - S_a = \Delta S$, representa a variação da entropia dum sistema, quando passa do estado inicial, a , ao estado final, b , em que as probabilidades termodinâmicas correspondentes são W_a e W_b , respectivamente.

As expressões anteriores constituem versões da famosa equação de Boltzmann, no formalismo microcanónico, que relaciona, de forma explícita, o conceito termodinâmico de entropia S com a probabilidade termodinâmica W de um sistema caracterizado por E , V e N . A constante k é a constante universal de Boltzmann, que, como veremos, é dada pelo quociente da constante dos gases ideais, R , pelo número de Avogadro, N_A :

$$k = \frac{R}{N_A} = \frac{1,987 \times 4,18 \text{ Joule/graude mole}}{6,022 \times 10^{23} \text{ mole}} =$$

$$= 1,38 \times 10^{-23} \text{ J/K}$$

A fórmula anterior mostra que a constante de Boltzmann é a constante dos gases ideais referida a uma molécula.

3.2. O MACROESTADO DE PROBABILIDADE MÁXIMA E O SEGUNDO PRINCÍPIO FUNDAMENTAL DA TERMODINÂMICA

Consideraremos, apenas, sistemas constituídos por partículas independentes. Ainda que o sistema esteja em equilíbrio termodinâmico, sabe-se que cada partícula está a variar constantemente de posição e de velocidade e que o sistema passa, muito rapidamente, por vários microestados.

Num sistema isolado, em equilíbrio termodinâmico, o macroestado, assim como o seu estado termodinâmico, não variam com o tempo.

Ainda que possa haver vários macroestados possíveis dum sistema, em equilíbrio, com as mesmas propriedades globais E , V , N , há um macroestado, que se observa "sempre", ou um que "se observa" mais frequentemente. É o macroestado de probabilidade máxima, o qual, por sua vez, pode assumir uma miríade de microestados.

Nestas condições, é de presumir que as propriedades fenomenológicas do sistema sejam as que se obtêm do macroestado de maior probabilidade. Mas, porque é que no estado de equilíbrio se observa "sempre" um mesmo macroestado. Ora, tal só se pode atribuir ao facto do número de microestados, que realizam aquele macroestado, ser muito superior ao número restante de todos os outros microestados combinados. Assim se compreende que o macroestado de probabilidade máxima corresponda, portanto, ao estado de equilíbrio do sistema que seria o estado compatível com o maior número de microestados acessíveis..

Boltzmann reconheceu que o Segundo Princípio está intimamente ligado com a teoria das probabilidades e que o aumento da entropia pode ser expresso exclusivamente em termos de probabilidade. A entropia surge, assim, como um aferidor da probabilidade. A entropia aparece, de forma mais natural e não tem

nada de mágico. É apenas uma medida conveniente do número de microestados acessíveis a um sistema num dado estado de equilíbrio. Mas ainda mais: a entropia, sendo proporcional ao logaritmo daquele número de estados, é uma medida mais adequada para representar a população de microestados W , porque os números resultantes são menores e muito mais manuseáveis e, o que é mais importante, porque as entropias dos vários subsistemas que constituem um sistema são aditivas, ao passo que os números de microestados são multiplicativos.

Algebricamente, o logaritmo de um produto, sendo aditivo, tornam a entropia mais próxima das variáveis de estado mais comuns do sistema, tais como a energia E , o volume V , o número de partículas N , que são também aditivas:

$$E = E_1 + E_2 \qquad V = V_1 + V_2 \qquad N = N_1 + N_2$$

e, analogamente:

$$S = S_1 + S_2$$

Um exemplo ajudará a aclarar estas considerações.

Consideremos um sistema A formado por dois subsistemas A_1 e A_2 , cujas probabilidades termodinâmicas são, respectivamente, $W_1 = 10^{20}$ e $W_2 = 2 \times 10^{20}$. Qual será o número de estados acessíveis do sistema global

$$A = A_1 \cup A_2 \qquad \text{E qual é a sua entropia?}$$

Vejam os:

$$W = W_1 \times W_2 = 2 \times 10^{40}$$

$$S_1 = k \log W_1 = k \log (10^{20}) = k \log (e^{2,30})^{20} = 46,0 \, k$$

$$S_2 = k \log W_2 = k \log (2) + k \log (e^{2,30})^{20} = 46,7 \, k$$

Logo:

$$S_A = S_1 + S_2 = 92,7 \, k$$

Os valores da entropia vêm expressos em "unidades k ". Como se vê, os valores de S são números muito menores e mais manuseáveis do que os de W .

Pelo seu alcance, a fórmula de Boltzmann é uma das fórmulas fundamentais da Física. A entropia e a probabilidade são grandezas muito diferentes, uma de natureza física e outra de natureza matemática.

A constante k assegura a homogeneidade física da fórmula de Boltzmann. A interpretação estatística da entropia consubstancia conceitos profundos da Física e foi decisiva para estabelecer os fundamentos da Mecânica Quântica, como o próprio Planck fez notar.

Muitas vezes, a entropia, à Boltzmann, é designada por entropia configuracional, que como se pode mostrar tem as mesmas propriedades que a entropia termodinâmica, à Clausius.

O macroestado mais provável é de extrema importância e as suas propriedades médias são quase idênticas às propriedades fenomenológicas. Boltzmann admitiu que o estado de um sistema, que corresponde ao estado de equilíbrio estatístico, é aquele cuja probabilidade é máxima. De facto, com a entropia, S , é o logaritmo do número de estados acessíveis, W , á medida que W aumenta, a entropia S também aumenta e quando W é máximo, S também o é.

Logo, o Segundo Princípio pode enunciar-se, em termos de entropia, ou do número de estados acessíveis, assim: *Um sistema isolado tende a evoluir no sentido em que a sua entropia aumenta até atingir o valor máximo, no estado de equilíbrio, ou ainda: o estado de equilíbrio dum sistema é o estado de probabilidade máxima.*

3.3. O ESTADO DE PROBABILIDADE MÁXIMA. LEI DE DISTRIBUIÇÃO DE BOLTZMANN

Consideremos um sistema formado por um grande número N de partículas idênticas, mas discerníveis (átomos, moléculas, etc.), em equilíbrio térmico à temperatura T , repartidas por n células (estados de energia). Seja N_i a população duma das células, contidas num recinto isolado. Qual é o estado de probabilidade termodinâmica máxima do sistema?

Para isso, basta calcular o máximo da função W , compatível com as condições do sistema. Ora, W é dada por:

$$W = N! / N_1! N_2! \dots N_n! = N! / \prod N_i!$$

Como os números $N_i!$ são muito grandes o cálculo directo dos factoriais é praticamente impossível e temos que recorrer à conhecida fórmula de Stirling ($\log N! = N \log N - N$). Vamos, por isso, aplicar logaritmos daquela expressão e, aplicando a fórmula de Stirling, vem:

$$\begin{aligned} \log W &= \log N! - \sum \log N_i! = N \log N - N - \sum N_i \log N_i + \sum N_i \\ &= N \log N - \sum N_i \log N_i \end{aligned}$$

em que o somatório se estende a todas as células n .

A fórmula anterior pode ainda escrever-se, com alguns arranjos, sob a forma:

$$\log W = -N \sum N_i / N \cdot \log N_i / N + N \log N = -N \sum N_i / N \cdot \log N_i / N$$

e se nela fizermos:

$$F_i = N_i/N$$

vem:

$$\text{Log}W = - N \sum F_i \text{Log}F_i$$

que é uma versão da fórmula de Boltzmann, precursora da fórmula de Shannon, como veremos oportunamente, e que justifica a afirmação de Von Newmann, referida na citação de abertura.

O estado de probabilidade máxima é aquele para o qual W é máximo, isto é, para o qual se anula a variação δW , ou seja, para a qual, $\delta \text{Log}W = 0$. Mas:

$$\delta \text{Log}W = \sum N_i \delta \text{Log}N_i + \sum \text{Log}N_i \delta N_i = 0$$

O segundo termo $\sum N_i \delta \text{Log}N_i$, é nulo, visto que:

$$\sum N_i \delta \text{Log}N_i = \sum N_i/N_i$$

$$\sum \delta N_i = \delta \sum N_i = \delta N = 0$$

Logo, a condição de máximo de W fica reduzida apenas, a:

$$\sum \text{Log}N_i \delta N_i = 0$$

sujeita à condição de ser $\sum \delta N_i = 0$ para todas as variações possíveis de N_i .

∴

Trata-se de resolver um problema de máximos condicionados, o que se pode fazer utilizando, por exemplo, o método dos multiplicadores indeterminados de Lagrange. Multiplicando a última expressão por $-\lambda$ e somando-a à anterior, tem-se:

$$\sum (\text{Log}N_i - \lambda) \delta N_i = 0$$

Como δN_i pode tomar qualquer valor, a condição anterior impõe que seja nulo o factor binómio, isto é, que se verifique a igualdade:

$$\text{Log}N_i = \lambda$$

Como λ é constante conclui-se que:

$$N_1 = N_2 = \dots = e^\lambda$$

Logo podemos dizer que o estado de probabilidade máxima corresponde aos microestados em que as partículas se encontram igualmente repartidas pelas n células. O valor de λ fica determinado notando que

$$N_1 + N_2 + \dots + N_n = n e^\lambda = N$$

Portanto como $e^\lambda = N/n$, vem:

$$N_1 = N_2 = \dots = N/n$$

Esta conclusão da equipartição do número de partículas pelas diferentes células é um caso especial e muito simples da estatística de Boltzmann, em que a única restrição que se impôs foi a da invariância do número total de partículas (massa constante).

Vamos considerar um caso mais geral e mais próximo da realidade, que é o da energia total do sistema ser também constante. De facto, o equilíbrio térmico exige que as partículas permutem energia entre si de modo que, se uma ganha energia, outra perde-a. Logo, a condição $\sum N_i = N = \text{const.}$, deve ser completada com a condição de ser também:

$$E_{\text{total}} = \sum N_i \epsilon_i = \text{const.}$$

em que ϵ_i representa o nível energético da célula i .

Temos, portanto, que resolver um novo problema de extremos condicionados em que as equações de condição são agora:

$$\sum \text{Log} N_i \delta N_i = 0$$

e

$$\sum \delta N_i = 0$$

e mais a da energia ser constante:

$$\sum \epsilon_i \delta N_i = 0$$

Aplicando novamente o método de *Lagrange*, multiplicando a segunda equação pelo coeficiente indeterminado α e a última por β e somando à primeira equação (multiplicada por 1), tem-se esquematicamente:

1	$\sum \text{Log } N_i \delta N_i = 0$
α	$\sum \delta N_i = 0$
β	$\sum \epsilon_i \delta N_i = 0$
$\text{Log } N_i + \alpha + \beta \epsilon_i = 0$	

em que α é um número puro e β tem as dimensões do inverso da energia.

Desta equação conclui-se que, W será máximo, quando a "população", N_i , de cada célula for dada pela expressão:

$$N_i = 1/e^{\alpha + \beta \epsilon_i}$$

A repartição estatística que corresponde a um estado macroscópico de probabilidade máxima é dada por uma lei exponencial que constitui a *lei da distribuição de Boltzmann*.

Se fizermos $e^{-\alpha} = A$, resulta para N_i uma nova expressão:

$$N_i = A e^{-\beta \epsilon_i}$$

em que A se pode determinar aplicando a condição $\sum N_i = N$, isto é:

$$A \sum e^{-\beta \epsilon_i} = N$$

O somatório que figura nesta expressão

$$Z = \sum e^{-\beta \epsilon_i}$$

estendido a todos os níveis energéticos, ϵ_i , tem um papel muito importante na Mecânica Estatística e chama-se *função de partição*, ou *soma de estados*

(*Zustandssumme, Planck*), porque depende da forma como se repartem as partículas pelos vários níveis energéticos.

A função Z é fundamental, porque, a partir dela, se podem deduzir as propriedades termodinâmicas do sistema. A função Z é uma propriedade termodinâmica porque depende exclusivamente do estado do sistema. É função de T através de β e de E . Como se verá, é através da função Z , que se estabelece a conexão da Mecânica Estatística Clássica com a Termodinâmica. Mas voltemos à *lei de distribuição de Boltzmann*. Das expressões anteriores conclui-se que $A = N / Z$ e, portanto, $\alpha = \log Z / N$.

A *lei de distribuição de Boltzmann* toma então forma mais usual:

$$N_i = (N / Z) e^{-\beta \epsilon_i}$$

É claro que

$$p_i = N_i / N = (e^{-\beta \epsilon_i}) / (\sum e^{\beta \epsilon_i})$$

representa a probabilidade de que existam N_i partículas no nível energético ϵ_i . Se notarmos que $\sum p_i = 1$, vê-se porque é que, muitas vezes, se diz que a função Z é um factor de normalização.

Vamos agora dar o significado de β . Para isso, escreve-se a equação de Boltzmann da entropia máxima : $S = k \log W_{\max}$ na forma:

$$\begin{aligned} S &= k N \log N - k \sum N_i \log N_i = k N \log N - k \sum N_i (\log N - \log Z - \beta \epsilon_i) = \\ &= k (N \log Z + \beta U) \end{aligned}$$

em que fazemos $\sum N_i \epsilon_i = U$

Recordando que

$$(\partial S / \partial U)_{V,N} = 1/T$$

se derivarmos expressão anterior de S , depois de algumas operações matemáticas, conclui-se que:

$$1/T = k \beta \quad \text{ou} \quad \beta = 1/k T$$

Logo, β é, em unidades k , o inverso da temperatura absoluta T (função de Carnot).

Nestas condições a lei da distribuição de *Boltzmann*, que dá a distribuição mais provável das partículas pelos diversos estados de energia, para um sistema em que N e U são constantes (colectividade canónica) é:

$$N_i = (N/Z) e^{-\epsilon_i / kT}$$

Na Figura 2 dá-se a representação gráfica da lei de *Boltzmann*.

Note-se que a população relativa N_i/N_j , corresponde a dois níveis de energia distintos, seria:

$$N_i/N_j = e^{-(\epsilon_i - \epsilon_j)/kT} = p_i/p_j$$

Se todos as células tivessem igual energia, a distribuição mais provável seria a homogénea: $N_1 = N_2 = \dots = N/n$, como tínhamos verificado de início.

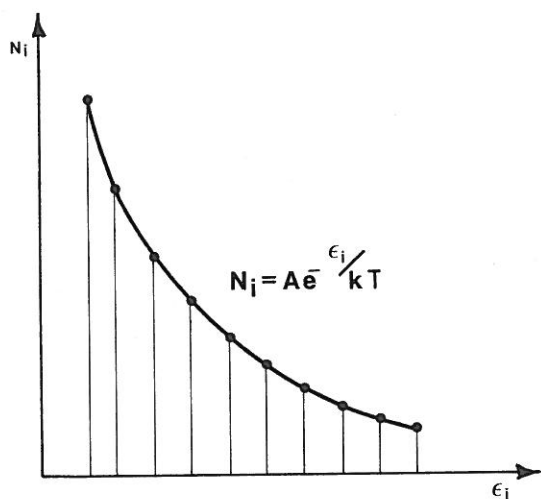


Fig. 2 - Distribuição de Boltzmann.

A expressão $e^{-\epsilon_i/kT}$ costuma designar-se por *factor de Boltzmann*. O efeito da temperatura traduz-se num declive diferente das curvas das distribuições (Fig. 2).

Da expressão anterior conclui-se que a probabilidade dum estado energético, ϵ_i , é proporcional ao factor de *Boltzmann*. Em condições normais, os estados de energia mais elevados são sempre ocupados por um número menor de partículas do que os estados de energia mais baixa.

O valor numérico de Z depende, evidentemente, da origem de contagem das energias ϵ_i . Em geral, toma-se para zero a energia do estado fundamental (o de valor mais baixo possível).

Nestas condições, o primeiro termo do somatório da função de partição é igual à unidade. A função Z reduz-se a:

$$Z = 1 + e^{-\epsilon_1/kT} + e^{-\epsilon_2/kT} + \dots$$

A série dos termos da função Z é rapidamente convergente. A convergência é tanto mais acentuada, quanto menor for a temperatura, T , e mais distanciados forem os níveis energéticos.

Quando $\epsilon_i/kT \gg 1$ o factor de *Boltzmann* torna-se desprezável.

Para calcular a função Z convém substituir o somatório por um integral, o que só é legítimo, se houver uma distribuição contínua de estados energéticos.

Estados degenerados

Em muitos casos os sistemas podem apresentar para um mesmo nível energético, vários estados moleculares distintos. É o caso de um átomo isolado que vibra harmonicamente com uma certa amplitude a que corresponde uma energia ϵ_i , mas em que falta o eixo de vibração pode ter g_i direcções espaciais distintas e discerníveis.

Estes estados dizem-se degenerados e o número g_i chama-se factor degenerescência, ou peso estatístico do nível energético ϵ_i . Em Mecânica Estatística Clássica este caso corresponde à existência de partículas com posições e quantidades de movimento diferentes mas com a mesma energia. Em Mecânica Quântica significa que há estados com séries de números quânticos diferentes, mas com igual energia.

4. CONEXÃO DA MECÂNICA ESTATÍSTICA COM A TERMODINÂMICA CLÁSSICA

4.1. A FUNÇÃO DE PARTIÇÃO E A DETERMINAÇÃO DOS VALORES MÉDIOS

A distribuição das probabilidades P_s dum sistema num dado estado s , com a energia E_s , é dada, como vimos, por:

$$p_s = e^{-\beta E_s} / Z$$

em que Z é a função de partição. A partir desta distribuição de probabilidades, podem determinar-se os valores médios de qualquer propriedade do sistema. Por exemplo, para a propriedade F_s do sistema s , o valor médio, \bar{F} , é dado por:

$$\bar{F} = \sum p_s F_s = \sum Z^{-1} e^{-\beta E_s} F_s$$

Mas, este processo de cálculo, dado o número elevadíssimo de estados acessíveis, s , é moroso e, por vezes, difícil. Ora, os valores médios podem obter-se duma forma muito mais simples, recorrendo à função de partição, cujo cálculo envolve também uma "soma sobre todos os estados", mas muito mais fácil de executar. Além disso, uma vez conhecida a função de partição, Z , é possível determinar, directamente a partir desta, muitas propriedades do sistema, sem ter que repetir um somatório, de cada vez que se estuda uma propriedade diferente.

Suponhamos que se pretende determinar a energia interna média, \bar{U} , dum sistema, de que se conhece a função de partição Z .

O valor médio da energia interna é dado por:

$$\bar{U} = \sum p_s E_s = Z^{-1} \sum e^{-\beta E_s} E_s$$

ou, tendo em vista a forma de função Z e da sua derivada $\frac{\partial Z}{\partial \beta}$:

$$\begin{aligned}\bar{U} &= -Z^{-1} \partial / \partial \beta (\sum e^{-\beta E_s}) = -Z^{-1} \partial Z / \partial \beta \\ &= -\partial \text{Log} Z / \partial \beta\end{aligned}$$

De forma análoga, temos para o valor médio quadrático da energia interna, \bar{U}^2 , a seguinte expressão:

$$\begin{aligned}\bar{U}^2 &= \sum p_s E_s^2 = Z^{-1} \sum e^{-\beta E_s} E_s^2 = Z^{-1} \cdot \partial^2 / \partial \beta^2 \cdot (\sum e^{-\beta E_s}) \\ &= Z^{-1} \cdot \partial^2 Z / \partial \beta^2\end{aligned}$$

O desvio médio quadrático, σ , como é fácil verificar, pode ser expresso em termos da segunda derivada do logaritmo da função Z, como se segue:

$$\begin{aligned}\sigma^2 &= (\Delta \bar{U})^2 = \bar{U}^2 - \bar{U}^2 \\ &= \partial^2 \text{Log} Z / \partial \beta^2\end{aligned}$$

Vejamos como é que a entropia e os potenciais termodinâmicos estão relacionados com a função de partição. Já vimos que a entropia S é dada por:

$$S = k N \text{Log} Z + U/T$$

Esta fórmula dá a entropia sem constantes arbitrárias, o que só se conseguia em Termodinâmica Clássica introduzindo o Terceiro Princípio.

Se agruparmos os termos da expressão anterior, tem-se:

$$U - T S \equiv F \equiv -k T N \text{Log} Z$$

Como se sabe, $F = U - TS$ designa a energia livre do sistema.

Esta expressão permite obter o significado físico de $\text{Log} Z$, ou, ainda, de Z.

De facto, a função de partição Z pode escrever-se agora, em termos de energia livre, F , como uma simples função exponencial, em vez dum somatório de funções exponenciais, em termos da energia. E assim a função de partição Z é dada por:

$$Z = e^{-F/NkT}$$

Vejamos mais algumas consequências importantes. Ora, a entropia, S , a pressão p , e o potencial químico μ podem ser expressos em termos de derivadas parciais da função F . Logo:

$$S = -(\partial F / \partial T)_{V,N} = kN \log Z + KTN \cdot \partial \log Z / \partial T$$

$$p = -(\partial F / \partial V)_{T,V} = -kN \cdot \partial \log Z / \partial V$$

$$\mu = -(\partial F / \partial N)_{T,V} = -kT \cdot \partial \log Z / \partial N$$

A expressão da entalpia, $H = U + pV$, em termos da função de partição, Z , é então dada por:

$$H = U + pV = NkT^2 \cdot \partial \log Z / \partial T + NTV \cdot \partial \log Z / \partial V$$

Finalmente, para o potencial Gibbs, G , tem-se:

$$G = F + pV = -kNT \log Z + kNTV \cdot \partial \log Z / \partial V$$

∴

Vejamos ainda a relação que existe entre a função de partição Z e a probabilidade termodinâmica W .

Notando que a função termodinâmica, F , por partícula (molécula) é $F = kT \log Z$, vem para U :

$$U = F + TS = -kT \log Z + kT \log W$$

ou seja:

$$U = kT \log (W/Z)$$

e portanto:

$$Z = W e^{-U/kT}$$

o que permite afirmar que a *função Z mede o número de microestados do sistema "amortecido" pela exponencial da energia expressa em "unidades" kT.*

∴

Consideremos agora um sistema constituído por dois sistemas *A* e *B*, cujas energias individuais, nos estados *a* e *b*, são respectivamente $\epsilon_a + \epsilon_b$. A função de partição do sistema global, $A \cup B$, cuja energia no estado *i* é $\epsilon_a + \epsilon_b$, vem dada por:

$$Z = \sum e^{-\beta \epsilon_i} = \sum e^{-\beta (\epsilon_a + \epsilon_b)} = \sum e^{-\beta \epsilon_a} \sum e^{-\beta \epsilon_b} = Z_A Z_B$$

o que mostra que a *função de partição do sistema é igual ao produto das funções de partição dos subsistemas que o constituem.*

5. DESORDEM E ENTROPIA

5.1. DESORDEM E O NÚMERO DE ESTADOS ACESSÍVEIS

A fórmula de Boltzmann revela o carácter probabilístico da entropia. De facto, aquela fórmula mostra que o aumento da entropia de um sistema isolado é consequência da tendência natural de um sistema passar de um estado de menor probabilidade, para outro de maior probabilidade.

Mas, esta probabilidade é tanto maior, quanto mais desordenada for a distribuição das partículas e diversificadas as suas velocidades. Os estados são, ainda, tanto mais desordenados, quanto maior for o número de arranjos moleculares, de modo que W e $\log W$ sejam números grandes, como acontece com os gases. Os estados ordenados têm apenas um pequeno número possível de arranjos (caso de um cristal à temperatura $^{\circ}\text{K}$, v.g., $W = 1$).

A entropia mede o grau de desordem interna, a nível molecular e constitui uma medida de incerteza do estado microscópico, que nesse instante realiza o estado macroscópico do sistema. O estado microscópico de um gás está a mudar constantemente, apresentando uma hipervariada de complexões. Não podemos saber, nunca, onde se encontra uma dada molécula e, ainda menos saber, o microestado que, em cada instante, realiza um dado estado macroscópico (estado clássico). A entropia reflecte, afinal, a incerteza do estado microscópico que ocorre em cada instante.

A entropia é aditiva, como exige a nossa intuição: desordem mais desordem, confusão mais confusão, caos mais caos, incerteza mais incerteza só podem conduzir a valores muito maiores da desordem, da confusão, do caos e da incerteza.

A entropia está relacionada com a desordem microscópica e com a tendência para a equalização. Por ordem entende-se a qualidade caracterizada pela diferenciação das partes diversificadas de um sistema e da sua separação e organização por categorias.

As configurações instantâneas são, em geral, desordenadas e variam no decurso do tempo. Mesmo que se começasse por um estado ordenado, com o decorrer do tempo, a ordem deteriorava-se e a desordem torna-se cada vez maior, até atingir o valor máximo, a que corresponde o maior valor possível da probabilidade termodinâmica.

É o que se passa quando se baralham as cartas de jogar, inicialmente ordenadas por naipes e por valores. Depois de se baralharem, é muito pouco provável que se obtenha, devido apenas ao acaso, uma distribuição pré-estabelecida das cartas, ou que, por se baralharem sucessivamente se volte a recuperar o ordenamento inicial. A desordem, a aleatoriedade e a dispersão vão sempre aumentando e andam de mãos dadas com a confusão e com a entropia.

Consideremos um sistema constituído por dois subsistemas, cada um com um gás diferente, separados por uma partição com igual número de moléculas mas de cada lado da partição. Ora, este sistema tem um certo ordenamento, visto que moléculas da mesma espécie estão em subsistemas diferentes.

Suponhamos que se retira a partição. Os gases difundem-se um no outro e ficam uniformemente distribuídos por todo o volume do sistema.

O ordenamento inicial desapareceu e a desordem aumentou, porque agora cada partícula pode estar num volume duas vezes maior. É claro, que na fase inicial, o número de microestados era muito menor do que é depois de remover a partição.

∴

Vamos mostrar que a desordem e, portanto, o número de microestados acessíveis varia rapidamente, quando se verifica uma interacção térmica, mecânica, ou mássica de um sistema com o seu universo. De facto, o número de microestados acessíveis muda rapidamente, quando se fornece energia (calor, trabalho) a um sistema, ou quando se altera o seu volume, ou ainda, quando se juntam ou subtraem partículas ao sistema. Esta conclusão resulta da constante de Boltzmann ter um valor extremamente pequeno. Consideremos um sistema

à temperatura T , à pressão p , com um potencial químico μ . A sua variação da entropia ds é dada por:

$$ds = 1/T \cdot dU + p/T \cdot dV - \mu/T \cdot dN$$

em que dU , dV e dN designam as variações da energia interna, do volume e do número de moles. Desta expressão, recorrendo à fórmula de Boltzman, vem:

$$a) 1/T = (\partial S/\partial U)_{V,N} = \partial k \text{ Log } W/\partial U$$

ou seja:

$$\partial \text{ Log } W/\partial U = 1/k T$$

Quando o sistema passa do estado a ao estado b a integração da equação anterior conduz a:

$$W_b/W_a = e^{\Delta U/k T}$$

b) Analogamente, como

$$p/T = (\partial S/\partial V)_{U,N} = \partial k \text{ Log } W/\partial V$$

vem depois da integração:

$$W_b/W_a = e^{p \Delta V/k T}$$

c) Finalmente, por ser:

$$\mu/T = (\partial S/\partial N)_{U,V} = \partial k \text{ Log } W/\partial N$$

tem-se

$$W_b/W_a = e^{-\mu \Delta N/k T}$$

O número de microestados acessíveis ao sistema é uma função exponencial rapidamente crescente. De facto, se a energia interna de um sistema a 27°C variar de 1 e V ($1,6 \times 10^{-12}$ erg), vê-se que o número de microestados acessíveis varia na razão de

$$W_b/W_a = e^{\Delta U/k T} = e^{39} = 5,5 \times 10^{16}$$

que é um número muitíssimo grande.

O mesmo se poderia concluir quando se verifica uma variação do volume e do número de partículas.

Por exemplo, se 20 m^3 de ar, à temperatura de 300°K e à pressão duma atmosfera, sofrem uma expansão adiabática, o número de estados acessíveis aumenta de

$$W_b/W_a = e^{11,8 \times 10^{21}} = 10^{2,1 \times 10^{21}}$$

que é um número muitíssimo grande.

Ainda outro exemplo: Vamos calcular o aumento do número de microestados acessíveis, quando se junta uma partícula a um sistema supondo que o potencial químico é $\mu = -3,0$ e a temperatura é $T = 300^\circ \text{ K}$.

O resultado final é:

$$W_b/W_a = e^{116} = 2,4 \times 10^{50}$$

o que é um número extraordinariamente grande.

∴

Claro que, se a entropia varia de Δs , o número de microestados acessíveis varia de

$$W_b/W_a = e^{\Delta S/T}$$

que é um número inconcebivelmente grande. Por exemplo, se

$$\Delta s = 10 \text{ cal}/^\circ\text{K}$$

vem

$$W_b/W_a = e^{\Delta S/k} = 10^{\Delta S/2,303k} = 10^{1,32 \times 10^{24}}$$

Como dissemos anteriormente, estes números tão elevados, resultam do valor da constante de *Boltzmann* ser extremamente pequeno e figurar em denominador na expressão de β ($\beta = 1/k T$).

5.2. CARÁCTER ALEATÓRIO DO SEGUNDO PRINCÍPIO FUNDAMENTAL DA TERMODINÂMICA

De acordo com a Termodinâmica Clássica, num sistema isolado, a entropia nunca pode diminuir, isto é, ou aumenta, ou se mantém constante. A diminuição da entropia não seria nunca permitida.

Ora, em face da nova concepção estatística (*à Boltzmann*) da entropia, temos que aceitar uma posição menos dogmática. Assim, se considerarmos um sistema num estado de probabilidade máxima (ou de entropia máxima), não significa que, ocasionalmente, o sistema não passe por microestados de menor probabilidade e, portanto, de menor entropia. Estes estados, não sendo impossíveis, são, no entanto, muitíssimo pouco prováveis e extremamente raros.

Notaremos que o Segundo Princípio Fundamental da Termodinâmica assenta no conceito de probabilidade, ao contrário do Primeiro Princípio.

Mas, enquanto este se pode aplicar a sistemas pequenos, aquele só se aplica a sistemas com um número elevadíssimo de partículas. No entanto, mesmo para sistemas com muitas partículas, há sempre uma probabilidade ínfima de violação do Segundo Princípio. É extremamente provável que um bloco de gelo funda

naturalmente quando se mergulha em água a uma temperatura superior a 0°C . Mas é muitíssimo pouco provável que se forme espontaneamente um bloco de gelo nessa mesma água (Fig. 3).

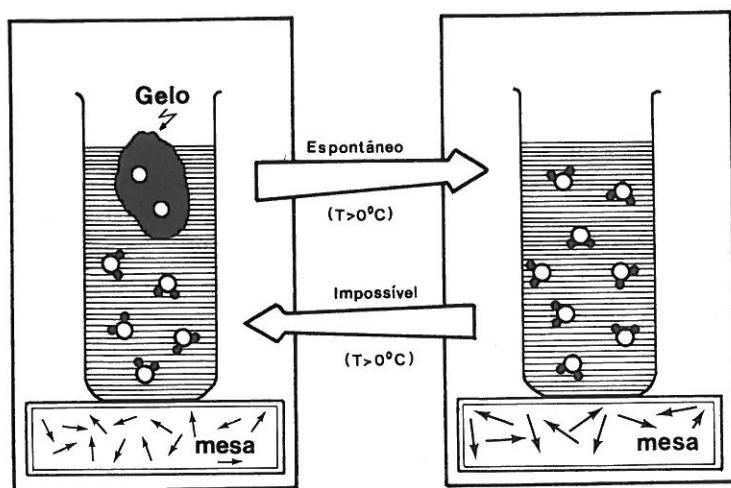


Fig.3 - O espontâneo e o impossível na fusão, ou na solidificação do gelo não violam o Primeiro Princípio, mas podem desafiar o Segundo Princípio.

Por exemplo, há sempre uma probabilidade de que o ar de uma sala se concentre todo num dos cantos. Há sempre uma certa probabilidade de que os rios corram da foz para a nascente, ou de que os objectos duma secretária saltem espontaneamente, mas estes acontecimentos são extremamente raros!

Mesmo com meios de observação extremamente rigorosos, capazes de medir com um rigor de 1 parte por bilião e se a máquina repetisse a medição 10^6 vezes por segundo, teríamos que aguardar $10^{360 \times 10^{12}}$ vezes a idade do universo para assistir a um acontecimento desses.

A probabilidade da violação do Segundo Princípio é tão ínfima, que podemos estar descansados: não veremos, na História da Humanidade, qualquer violação deste Princípio.

Vamos apresentar alguns exemplos que, embora demonstrem o carácter aleatório do Segundo Princípio, mostram também, que as probabilidades de violação daquele Princípio são extremamente pequenas.

Exemplo 1:

Suponhamos que, devido a uma flutuação, 2g de água, inicialmente em equilíbrio, a 20°C , dão lugar, espontaneamente, a 1g a 19°C mais 1g a 21°C .

Pergunta-se: qual será a relação entre a probabilidade para que se verifique esta separação e a probabilidade de manter o estado de equilíbrio normal?

Solução:

Ora vejamos: a variação de entropia, Δs , verificada neste processo, em que a pressão constante, é dada por:

$$\Delta s = c (\log T_1/T_0 + \log T_2/T_0)$$

em que c é o calor específico da água.

Ou seja:

$$\Delta s = c \log T_1 T_2 / T_0^2 = 4,18 \log (293 - 1) (293 + 1) / 293^2 = - 4,86 \times 10^{-5} \text{ J/K}$$

Portanto, a relação entre a probabilidade W_2 para que ocorra este processo, e a probabilidade W_1 do estado normal de equilíbrio será:

$$W_1/W_2 = \exp. (- 4,86 \times 10^{-5} / 1,38 \times 10^{-23}) = 10^{-3,52 \times 10^{18}}$$

$$= 10^{-3,52/2,30 \times 10^{18}} = 10^{-18}$$

Esta relação é um número extremamente pequeno. Logo, existe uma certa probabilidade para que ocorra a separação, mas é tão pequena, tão pequena, tão pequena, que temos a "certeza prática" que tal fenómeno nunca se dará, espontaneamente.

Exemplo 2:

Suponhamos que dois corpos, às temperaturas de 27° C e 28° C, são postos em contacto e que há entre eles um fluxo de calor de 1 erg no "sentido normal".

- Qual é a variação do número de estados acessíveis devido a esta transferência de energia?
- Qual a probabilidade de que o fluxo se dê em sentido contrário?
- E se o fluxo for muito menor, por exemplo, de $1,2 \times 10^{-9}$ erg?

Solução:

A entropia total do sistema formado pelos dois corpos, antes e depois do contacto, sofreu uma variação dada por:

$$\Delta s = (1/301) = 1/9 \times 10^{-4} \text{ erg/K}$$

a) De acordo com a fórmula de *Boltzmann* sabe-se que:

$$\Delta s = k \text{ Log } (W_f / W_i)$$

em que W_i e W_f são as probabilidades inicial e final, do sistema, ou seja:

$$W_f = W_i e^{\Delta s/k} = W_i \exp(10^{-4}/9 \times 1,38 \times 10^{-16}) = W_i \exp(10^{12}/12) = W_i e^{10^{11}}$$

Como se vê, o número de complexões acessíveis no estado final excede o número inicial por um factor extremamente grande. Neste caso o calor flui do corpo quente para o corpo frio.

b) Se o fluxo se tivesse verificado do corpo frio para o corpo quente, seria:

$$\Delta s = - 1/9 \times 10^{-4} \text{ erg/k}$$

e, portanto, agora, tinha-se:

$$W_f = W_i \exp(-10^{12}/12) = W_i e^{-10^{11}}$$

Logo, a probabilidade para que se verificasse um fluxo de calor do corpo frio para o corpo quente seria extremamente pequena. De facto, em $W_f/W_i = 10 \times 10^{10}$ casos só num caso se verificaria o fluxo "contra natura". Na prática, este resultado corresponde, afinal, a uma confirmação dos resultados da Termodinâmica Clássica, em que tal fluxo nunca se poderia verificar.

c) Vejamos agora o caso do fluxo ser muito inferior a um erg e igual a $1,2 \times 10^{-9}$ erg;

O resultado seria agora:

$$W_2/W_1 = e = 2,7$$

O fluxo ainda se verificará com maior frequência no sentido imposto pela Termodinâmica Clássica, mas a frequência do fluxo "contra natura" verifica-se agora 1 vez em cada três casos.

Exemplo 3:

Consideremos uma bola de 10 gramas em repouso sobre uma mesa, à temperatura de 27° C. Qual seria a probabilidade de que a bola saltasse espontaneamente, até atingir uma altura de 1cm, à custa da energia da mesa?

Solução:

A energia potencial da bola aumentaria de $m \cdot g \cdot \Delta h$ e a energia fornecida pela secretária seria $-\Delta Q$, de tal modo que $-\Delta Q + m \cdot g \cdot \Delta h = 0$. Mas, o fornecimento da energia ΔQ , corresponde uma variação de entropia $\Delta s = \Delta Q/T = -m \cdot g \cdot \Delta h/T$. Portanto, recorrendo à fórmula de *Boltzmann*, a probabilidade de que tal acontecesse, W_2 , em relação à probabilidade do que normalmente se passa, W_1 , seria:

$$W_2/W_1 = e^{\Delta s/k} = e^{-9800/300k} = 10^{-1,32 \times 10^{17}} = 1/10^{1,32 \times 10^{17}}$$

que é um número extremamente pequeno. Em $10^{1,32 \times 10^{17}}$ realizações potenciais ("tentativas") só 1 teria sucesso! Este resultado mostra que, para todos os efeitos práticos, a bola continuará em repouso sobre a secretária e, talvez nunca venha a saltar. (Fig. 4).

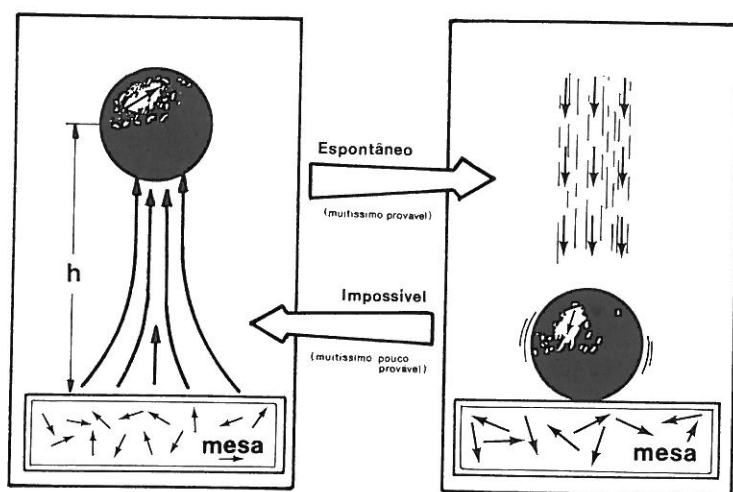


Fig. 4 - O Segundo Princípio diz que "espontâneo" significa ser muitíssimo provável e o "impossível" é equivalente a ser muitíssimo pouco provável.

5.3. O ZERO ABSOLUTO E O TERCEIRO PRINCÍPIO DA TERMODINÂMICA

Vejam os o que se passa, quando se remove energia dum sistema. O número de estados acessíveis é uma função que decresce, rapidamente, com a energia interna e, portanto, o número de estados acessíveis sofre um declínio acentuado. Suponhamos que se continua a remover energia, até que o número de estados acessíveis se reduza a um. Neste estado, o sistema estaria no estado de energia mínima, porque não poderia continuar a ceder mais energia, visto que não há nenhum outro estado possível com menos estados acessíveis do que 1. A energia atinge o nível zero de referência e a sua temperatura será, portanto, o "zero absoluto".

Como o número de estados acessíveis é exactamente 1, a sua entropia é zero:

$$S = k \log W = k \log 1 = 0$$

Devemos salientar que este resultado não depende das dimensões, da pressão, ou da intensidade de qualquer campo. Quaisquer que sejam as condições, ou os constrangimentos a que o sistema esteja submetido, só existe um estado de energia mínima e, neste estado, a entropia é nula.

Podemos, pois, enunciar assim o Terceiro Princípio da Termodinâmica:

"A entropia dum sistema tende para zero, à medida que a temperatura tende também para zero, independentemente dos valores dos parâmetros exteriores (pressão, volume, número de partículas, etc.)."

Esta formulação assenta na hipótese de que não há degenerescência. Se o nível de energia mínima (nível de referência) tiver uma degenerescência 5, por exemplo, então no zero absoluto haveria 5 estados acessíveis e a entropia seria:

$$S_0 = k \log 5$$

Por isso, muitas vezes, no enunciado do Terceiro Princípio se diz que "*a entropia tende para uma constante*" em vez de "*a entropia tende para zero*". Entre outras coisas, o Terceiro Princípio requer que os calores específicos devam também tender para zero no zero absoluto.

5.4. O PRIMEIRO PRINCÍPIO DA TERMODINÂMICA. INTERPRETAÇÃO ESTATÍSTICA E MECÂNICA DE CALOR E DE TRABALHO

Como já referimos, a energia interna U duma colecção de N partículas independentes é $U = N \sum p_i \epsilon_i = \sum N_i \epsilon_i$, em que ϵ_i são os níveis de energia, cuja distribuição de probabilidade é p_i , a que corresponde uma distribuição da população N_i . Nós sabemos da termodinâmica clássica que a energia interna dum sistema pode variar devido à quantidade de calor absorvido, dQ , e ao trabalho realizado, dW :

$$dU = dQ + dW = dQ + p dV$$

Por outro lado, da expressão de U tem-se:

$$dU = \sum \epsilon_i dN_i + \sum N_i d\epsilon_i$$

e as duas expressões devem ser idênticas.

Vamos mostrar que os dois termos da última expressão correspondem, respectivamente, às contribuições do calor e do trabalho para a variação da energia interna (Fig. 5).

De facto, aquecendo o sistema a volume V , constante, os níveis da energia não variam, porque $\epsilon_i = f(V)$ e $d\epsilon_i = (\partial \epsilon_i / \partial V) dV$. Como $dV = 0$ não há realização de trabalho e a variação da energia é devida apenas ao calor absorvido, ou $dU = dQ_{\text{rev}}$. Se fizermos $d\epsilon_i = 0$ na expressão anterior vem:

$$dU_v = dQ_{\text{rev}} = \sum \epsilon_i dN_i$$

Logo, a adição de calor produz variações nas populações dos vários níveis de energia, mas não os altera.

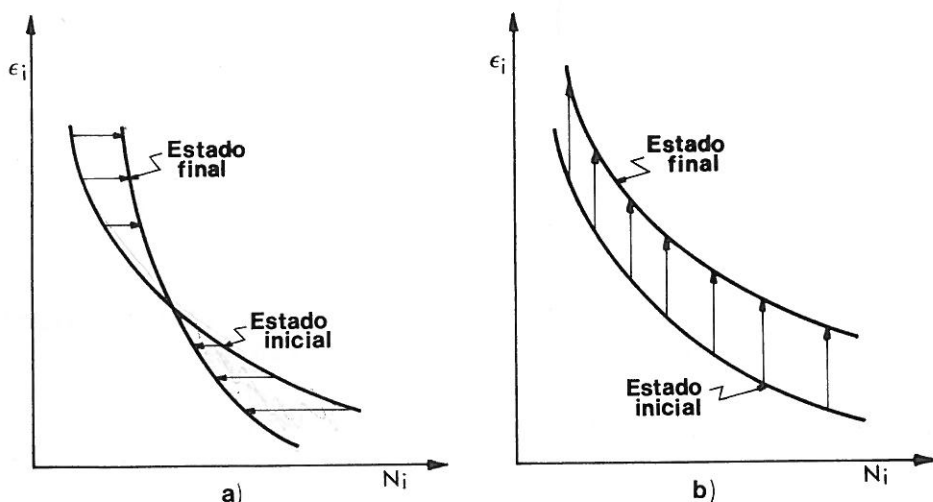


Fig. 5 - Interpretação microscópica do calor (a) e do trabalho (b). A adição de calor produz variações nas distribuições dos vários níveis; o trabalho recebido aumenta a energia dos diferentes níveis.

O termo do trabalho deve então ser atribuído à outra parcela. Com efeito:

$$(dU)_Q = dW = \sum N_i d\epsilon_i = \sum N_i \left(\frac{\partial \epsilon_i}{\partial V} \right) \cdot dV$$

Portanto, o trabalho implica uma variação da energia interna associada à variação dos níveis energéticos, sem alteração das populações de cada nível.

Em síntese, um incremento de calor pode considerar-se como se correspondesse a uma mudança nas populações dos diferentes níveis de energia, enquanto o trabalho se reflecte nos próprios níveis de energia.

Uma analogia simples ajuda a compreender melhor as considerações anteriores. Suponhamos um escadote com uma certa distribuição de bolas nos degraus. Os degraus correspondem aos níveis de energia e as bolas em cada degrau à distribuição das partículas. Se houver bolas que passem dum degrau para o outro, a energia potencial do conjunto de todas as bolas varia, mesmo sem actuar no escadote, mantendo os degraus fixos. Mas, a energia potencial também pode variar, sem se alterar o número de bolas por degrau; basta abrir ou fechar mais o escadote, porque então o nível dos degraus varia. No primeiro caso, deu-se uma transferência interna de bolas, de forma mais ou menos

desordenada, mas sem ter quaisquer repercursões "visíveis" no movimento global do escadote: corresponde à transferência de calor. No segundo caso, o aumento de energia potencial é "visível", através do movimento ordenado do escadote; corresponde ao trabalho realizado pelo sistema.

6. EPÍLOGO: A NATUREZA ESTATÍSTICA DA ENTROPIA

Boltzmann reconhecendo "como o segundo princípio está tão intimamente ligado com a teoria das probabilidades" concluía que o aumento da entropia podia ser aceite apenas em termos de probabilidade. O aumento de entropia seria, portanto, altamente provável, mas não *certo*. Foi uma conclusão a que chegou quando formulou o conceito da entropia de um macroestado termodinâmico em termos do número de complexões, ou de configurações microscópicas, compatíveis com aquele estado. A entropia constitui assim, uma medida de probabilidade termodinâmica.

Como foi referido, a fórmula de *Boltzmann*, $S = k \log W$, constitui pela sua elegância, simplicidade e concisão e, sobretudo, pelo seu alcance, uma das fórmulas mais notáveis da Ciência. A entropia e a probabilidade são grandezas muito diferentes; uma, S , é de natureza física, outra, W , é de natureza matemática. A constante k assegura a homogeneidade física da fórmula de *Boltzmann*. É, todavia, a relação entre duas grandezas tão heterogêneas, que *a priori* não pareciam ter qualquer relação entre elas, que constitui a originalidade e um dos motivos de interesse desta fórmula. Mas a grande contribuição de *Boltzmann* foi a de ter trazido à Física uma nova dimensão: a coexistência, lado a lado, de duas concepções determinista e probabilística. A natureza estatística da entropia abriu novos horizontes na Ciência e, em particular, na Física. Como diz *Planck*, sem as suas reflexões sobre o conceito de entropia, não se teria abalançado a propôr uma teoria dos quântos, que depois levaria à Mecânica Quântica.

Por isso, a fórmula de *Boltzmann*, mais do que uma confirmação brilhante da Mecânica Estatística, constituiu o rastilho que desencadearia o progresso da Física, revelado a partir das décadas de vinte e trinta.

A Física Clássica era determinista. O sucesso das leis de *Newton*, ao permitirem prevêr os movimentos dos Planetas, teve um efeito decisivo nesta

atitude científica. O facto dos astrónomos poderem determinar com antecedência de séculos as posições dos Planetas, ou dos cometas, e poderem prevêêr a existência de outros astros desconhecidos, constitui como o que o padrão *non plus ultra* da perfeição mecanicista de *Newton*. Era este padrão que se procurava generalizar e alargar a outros campos da Ciência.

Laplace sugeriu, mesmo, a possibilidade de vir a existir um supergênio matemático - o feiticeiro de *Laplace* - que, a partir das condições iniciais das posições e das velocidades de todas e quaisquer partículas do Universo, poderia prevêêr a sua evolução futura.

Hoje podemos negar a existência dum feiticeiro de *Laplace*, a nível atómico ou a nível molecular, mesmo de concepção electrónica, ou de qualquer outra, por várias razões. Uma, é devida ao princípio da incerteza de *Heisenberg* que nos impede, à partida, de conhecer simultaneamente as posições e as velocidades de partículas quânticas. Logo, o feiticeiro de *Laplace* não poderia integrar as equações de qualquer das partículas, por não se conhecerem completamente as condições iniciais. Todos sabemos a influência que as pequenas variações das condições iniciais têm nas trajectórias (basta saber jogar bilhar ...).

Mas, mesmo que não houvesse esta incerteza essencial, o feiticeiro de *Laplace*, não disporia de capacidade para armazenar os valores numéricos das várias grandezas que descrevem os processos que ocorrem em sistemas constituídos por grande número de elementos, como acontece, por exemplo, com os processos em escala atómica ou molecular. Para registar um número, é necessário reter uma quantidade adequada de átomos, ou de electrões e dispô-los numa configuração específica, estável, e que possa ser detectada e localizada quando, mais tarde, for necessário utilizar o número em cálculos (sinais de tinta no papel, dipolos magnéticos, fluorescência dum écran, etc.). Ora, acontece que não há sequer átomos suficientes no Universo para registar todos os dados referentes às moléculas de ar existente na atmosfera.

Somos, portanto, levados a abandonar a esperança de poder seguir, por via determinista, o que ocorre em escala atómica, numa dada amostra de matéria. Temos que nos contentar com certo conhecimento limitado e tentar obter, por inferência estatística, várias conclusões que são correctas, em média. Surge, assim, a Mecânica Estatística, que é uma teoria matemática, que diz como se hão-de combinar, estatisticamente, as propriedades das partículas individuais, para poder prever, como dizia *Gibbs*, "aquelas propriedades acessíveis aos nossos sentidos e susceptíveis de medição". Depois, surgirá a Mecânica Quântica, que nos vai abrir as grandes vias da Física para o estudo do mundo microscópico e penetrar, mais profundamente, na estrutura da matéria.

ENTROPIA E INFORMAÇÃO

II

1 - NOÇÕES ELEMENTARES SOBRE A TEORIA DA INFORMAÇÃO

1.1. DEFINIÇÃO DE INFORMAÇÃO

A Teoria da Informação teve origem no estudo do problema da transmissão de mensagens e muitos dos seus conceitos e designações enquadram-se na Teoria da Comunicação.

Informação é a entidade que mede a diferença entre saber e não saber; entre conhecer e não conhecer; ou ainda, a diferença que há entre encarar várias possibilidades sobre a realização dum acontecimento futuro e saber qual delas se realiza.

A nossa expectativa sobre a realização dum acontecimento futuro está associada a um certo grau de incerteza, ou a uma falta de informação. Desde o momento em que compramos um bilhete de lotaria, mantemos uma certa expectativa, associada a um certo grau de incerteza. A incerteza seria menor, se houvesse "mais prémios" e "menos números". Nós, sem querer, começamos a pensar, em termos da probabilidade, que temos de nos sair a "sorte grande", antes de andar "à roda". É, portanto, uma probabilidade *a priori*, que traduz sempre um certo grau de esperança.

Segundo Claude Shannon, fundador da Teoria da Informação, o que é importante numa mensagem que recebemos, é a alteração da incerteza que existia antes de receber a mensagem. O efeito da mensagem depende, portanto, do nosso estado de conhecimento em relação a um dado acontecimento. Logo, o conteúdo de informação recebido é, portanto, dado pelo decréscimo da falta de informação, isto é, pela diferença entre a incerteza que existia, antes de receber a mensagem, e a incerteza que resta, depois de receber a mensagem. Se, depois de receber a mensagem, não resta qualquer incerteza, o conteúdo da

informação é exactamente igual à informação completa, que desfaz toda a incerteza! É evidente, que para se obter informação, a mensagem que recebemos tem que trazer "coisas de novo", que não esperássemos, ou qualquer coisa que ainda não conhecéssemos.

1.2. INCERTEZA, ENTROPIA, INFORMAÇÃO E PROBABILIDADE

Suponhamos agora um sistema, cujo estado pode ser realizado por um grande número de complexões. Diz-se, então, que esse estado é um estado de grande probabilidade. Há uma grande indeterminação sobre o microestado, que num dado instante realiza o macroestado. Fica-se sempre com uma grande incerteza em face de um sistema que se apresenta com tanta desordem nos microestados, isto é, com tanta entropia. Mas, um sistema com muita desordem, ou com muita entropia, é um sistema de que se tem muito pouca informação, ou uma grande falta de informação. A um aumento de informação corresponde, portanto, uma diminuição de entropia, ou da desordem.

A desordem constitui uma medida da nossa ignorância, ou da falta de informação sobre a complexão que realiza, naquele instante, o estado do sistema. Desordem é, essencialmente, o mesmo que ignorância, ou falta de informação. A informação e a entropia não são, de facto, grandezas idênticas, porque o sentido da variação da indeterminação é oposto ao sentido da variação da informação. É o valor da indeterminação, expressa pela entropia, que se toma como a medida da quantidade de informação necessária para eliminar a indeterminação respectiva. Quanto maior é a indeterminação, maior será a quantidade de informação que se obtém, quando se conhecer o resultado.

Podemos afirmar que a incerteza e a informação estão relacionadas por uma lei de conservação. Assim, a soma da incerteza, H , com a informação I é constante e igual à informação máxima, ou à entropia máxima, sob dadas condições.

Usando a linguagem matemática, tem-se:

$$H + I = \text{const.} = H_{\text{max}} = I_{\text{max}}$$

em que H representa a incerteza, ou a falta de informação; I o conteúdo da informação e H_{\max} e I_{\max} os valores máximos possíveis, correspondentes. Logo:

$$\Delta H + \Delta I = 0$$

e um ganho de informação é compensado por uma perda igual de incerteza.

Embora se possa utilizar a relação de Boltzmann para o cálculo do número de distribuições para uma dada entropia do sistema, pode interpretar-se esta relação de forma diferente em termos da ignorância, ou falta de informação sobre o sistema. Se o sistema se pode apresentar num do grande número de microestados, o estado, que inclui estes microestados, é altamente provável, com um caos microscópico, ou uma desordem muito pronunciados.

Todavia, se um sistema existe com um número reduzido de microestados, tem muito menos desordem e muito menos entropia. Pode dizer-se que a ignorância, ou a incerteza, sobre a distribuição é muito maior para um estado de entropia mais elevada (muitas distribuições possíveis). Logo, um aumento de entropia corresponde a uma diminuição de informação sobre a distribuição dos microestados dum sistema.

Sabemos que um sistema em desordem é um sistema de que não temos muita "informação". Mas, com o crescimento da desordem, aumenta a entropia; logo podemos "sentir" que a um aumento de entropia corresponde uma "perda" de informação.

Mas, será possível medir a informação e exprimi-la por uma fórmula, isto é, será possível quantificar a informação?

A "informação" provém de "mensagens" que "recebemos". É evidente que para obter informação, a "mensagem" que se recebe tem que trazer alguma coisa "de novo", isto é, qualquer coisa que não se esperasse. Caso contrário, a mensagem não passa de um papel em branco... Por exemplo, não adianta nada receber uma mensagem (notícia) que é já completamente esperada, ou conhecida.

Suponhamos que se podem receber N mensagens. Quanto menos provável for uma mensagem, isto é, quanto mais inesperada ela for, maior será a informação obtida, se a mensagem for recebida.

O que é um segredo?

É uma mensagem, cuja probabilidade de nos chegar ao conhecimento é muito pequena, mas se chegar aumenta, e de que maneira, a informação.

Somos, assim, levados a pressentir uma certa relação entre informação e probabilidade.

Como se poderá definir a probabilidade de ocorrência de um acontecimento? A probabilidade é dada pelo quociente entre o número de casos favoráveis e o

número de casos possíveis da ocorrência do acontecimento. Esta probabilidade varia entre 0 e 1. O valor 0 corresponde à impossibilidade e o seu valor, máximo, de 1, corresponde à certeza. Na I PARTE considerou-se contudo, mais conveniente tomar simplesmente o número de casos favoráveis, como a medida da probabilidade termodinâmica.

Um macroestado pode ser realizado por um grande número de diferentes microestados. Com efeito, as moléculas estão em agitação contínua donde resulta uma imensa variedade de microestados. O número de microdistribuições é, pois, a probabilidade termodinâmica. Quanto maior for o número de complexões, maior é a probabilidade e maior é a desordem, mas menor é a informação que se tem sobre qual o microestado em que, num dado instante, se encontra o sistema.

2. MEDIDA DA INFORMAÇÃO

2.1. INCERTEZA, FALTA DE INFORMAÇÃO E PROBABILIDADE

A quantidade de informação pode avaliar-se pela variação do grau de incerteza. A quantidade de informação recebida por uma mensagem é dada pela diferença entre o grau de incerteza *a priori* e o grau de incerteza *a posteriori*.

Vamos abordar o problema da matematização da incerteza.

Começemos por considerar um exemplo: Suponhamos um objecto que se encontra dentro de uma das n gavetas dum contador, todas iguais e, como é evidente, mutuamente exclusivas. Além disso, todas as gavetas são igualmente prováveis, isto é, não há razões de preferência para a escolha duma qualquer sobre as outras. Por outro lado, o objecto não pode estar em mais do que uma gaveta. Logo, há n possibilidades para saber em que gaveta está o objecto e as possibilidades excluem-se, também, mutuamente.

A incerteza em se saber em que gaveta se encontra o objecto traduz, de facto, uma falta de informação. Ora, a incerteza, ou a falta de informação, é tanto maior quanto maior for o número n de gavetas (possibilidades). Se designarmos por H a falta de informação, ou a incerteza, vemos que H é uma função de n :

$$H \equiv H(n)$$

e mais, que H é uma função crescente de n , isto é:

$$H(n) > H(r)$$

se $n > r$

Se recebermos uma "mensagem" com a indicação da gaveta em que o objecto se encontra, então, toda a incerteza desaparece. Só há uma escolha possível ($n=1$) e não há já qualquer falta de informação, isto é:

$$H(1) = 0$$

Esta expressão traduz a informação completa. Deve acentuar-se que, quanto maior é o número de possibilidades, n , menor é a probabilidade de acertar na gaveta onde está o objecto. A probabilidade p vale "um contra n ", isto é, $p = 1/n$.

Assim, no caso dos acontecimentos terem todos a mesma probabilidade, a expressão da incerteza é uma função da probabilidade:

$$H \equiv H(n) = H(1/p) = H^*(p)$$

Agora, quanto maior for n , menor é a probabilidade p ; portanto, a falta de informação, ou a incerteza, H , é uma função decrescente da probabilidade da realização dum acontecimento.

$$H(p_1) > H(p_2)$$

$$\text{se } p_1 < p_2$$

Consideremos, ainda, o problema de uma escolha que, por sua vez, resulta de duas escolhas independentes, isto é, que se excluem mutuamente. Continuemos com o exemplo do contador. Este caso corresponde a cada uma das n gavetas estar dividida em m compartimentos. Para localizar o objecto tem que se identificar não só a gaveta, mas também o compartimento. O número de possibilidades é dado agora, pelo número total de compartimentos: $n \times m$.

A falta de informação, ou a incerteza, H , é, neste caso, uma função de $n \times m$, isto é:

$$H \equiv H(m \times n)$$

No entanto, esta incerteza pode ser superada por duas operações, ou fases distintas: numa primeira, procura identificar-se a gaveta, a que corresponde uma incerteza $H(n)$; na segunda, há que determinar o compartimento da gaveta em que o objecto se encontra, a que corresponde a incerteza $H(m)$. Estas duas

operações sucessivas, conduzem, exactamente, ao mesmo resultado que se obteria, se se identificassem, duma só vez, a gaveta e o compartimento, em que a incerteza inicial era $H(m \times n)$. Como as duas operações sucessivas são equivalentes à operação global inicial, conclui-se que a incerteza é uma propriedade aditiva: junta-se a incerteza sobre a gaveta à incerteza sobre o compartimento. Em linguagem matemática, tem-se:

$$H(m \times n) = H(m) + H(n)$$

em que as funções H são funções crescentes dos argumentos.

Esta relação funcional, tão geral, vai permitir determinar a forma da função H (o problema é semelhante ao que tivemos que resolver para a equação de Boltzmann). Para isso, por ser mais cómodo, procede-se a uma mudança de variável, fazendo $m \times n = z$.

Começemos por derivar ambos os membros daquela relação, sucessivamente, em ordem a m , com n constante, e em ordem a n , mantendo m constante. Para facilitar as operações convém aplicar o teorema da derivação das funções compostas, às derivadas do primeiro membro:

$$\partial H(z)/\partial m = \partial H(z)/\partial z \cdot \partial z/\partial m$$

e

$$\partial H(z)/\partial n = \partial H(z)/\partial z \cdot \partial z/\partial n$$

Posto isto, podemos escrever as expressões das derivadas em ordem a m e em ordem a n da relação funcional como se segue:

$$\partial H(z)/\partial z \cdot \partial z/\partial m = \partial H(m)/\partial m$$

$$\partial H(z)/\partial z \cdot \partial z/\partial n = \partial H(n)/\partial n$$

porque $H(n)$ não é função de m , nem $H(m)$ é função de n ; além disso, m e n são independentes e $\partial m/\partial n = 0$. Mas como $z = m \times n$, vem:

$$\partial z/\partial m = n \quad \text{e} \quad \partial z/\partial n = m$$

Logo, as expressões das derivadas podem escrever-se:

$$\partial H(z)/\partial z \cdot n = \partial H(m)/\partial m$$

$$\partial H(z)/\partial z \cdot m = \partial H(n)/\partial n$$

donde se deduz que:

$$\partial H(z)/\partial z = 1/n \cdot \partial H(m)/\partial m = 1/m \cdot \partial H(n)/\partial n$$

Estas igualdades podem escrever-se, multiplicando-as por $z = m \times n$, sob a forma:

$$z \cdot \partial H(z)/\partial z = m \cdot \partial H(m)/\partial m = n \cdot \partial H(n)/\partial n = \text{const.} = k$$

Como cada membro das equações só é função duma única variável, para serem iguais os seus valores têm que ser constantes. Por isso, a forma da função H obtém-se resolvendo as equações:

$$\partial H(z)/\partial z = k/z; \quad \partial H(m)/\partial m = K/m; \quad \partial H(n)/\partial n = K/n$$

Estas equações são do tipo $df(x) = k dx/x$, cuja integração, permite determinar a forma da função $f(x)$, que neste caso é evidentemente:

$$f(x) = k \text{Log} x$$

Logo pode escrever-se para H :

$$H(m) = k \text{Log } m + \text{const.}$$

$$H(n) = k \text{Log } n + \text{const.}$$

Para $H(m \times n)$ tem-se uma função da mesma forma:

$$H(mn) = k \text{Log } (mn) + \text{const.}$$

Notando que o logaritmo dum produto é igual à soma dos logaritmos dos factores, vem:

$$H(mn) = H(m) + H(n) + \text{const.}$$

o que confirma a aditividade da expressão funcional geral.

Vê-se que $H(n)$ é uma função logarítmica e, à partida, para n possibilidades a incerteza, será:

$$H(n) = k \log n + \text{const.}$$

e o valor numérico da incerteza não depende da maneira como o problema é posto. Quando for $n = 1$ não há qualquer incerteza (certeza), visto que:

$$H(1) = k \log 1 + \text{const.}; \quad 0 = 0 + \text{const.}$$

e, portanto, pode fazer-se a parcela constante igual a zero.

Retomemos a expressão anterior no caso dos n acontecimentos serem equiprováveis. Notando que $p = 1/n$, a expressão da incerteza pode escrever-se explicitamente:

$$H(n) = k \log n$$

ou

$$H(p) = k \log 1/p = -k \log p$$

Esta expressão é muito interessante, porque exprime a incerteza, H , em função da probabilidade p e confirma que H é uma função decrescente da probabilidade.

2.2. A FÓRMULA FUNDAMENTAL DA TEORIA DA INFORMAÇÃO

Vamos ver, agora, qual é a expressão da incerteza, quando os acontecimentos não são igualmente prováveis. Consideremos um ensaio de n acontecimentos, todos com a mesma probabilidade, $1/n$, mas distribuídos por grupos diferentes: o superior com n_1 elementos; o médio com n_2 elementos e o inferior com n_3 elementos, isto é, (Fig.6) tais que:

$$n = n_1 + n_2 + n_3$$

Como os n acontecimentos são igualmente prováveis, a probabilidade da ocorrência de um dos n acontecimentos, qualquer que seja o grupo a que ele

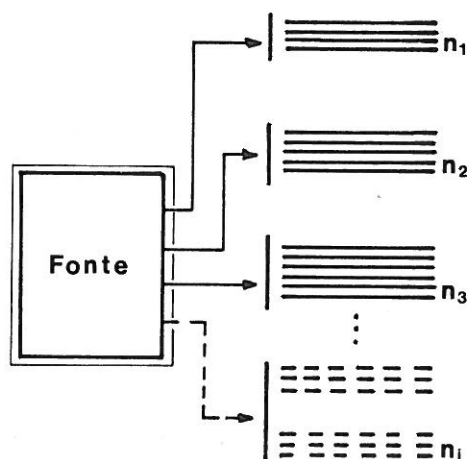


Fig. 6 - Uma fonte geradora de acontecimentos com probabilidades diferentes.

pertença, é $1/n$. No entanto, se se quiser saber qual o grupo em que ocorre o acontecimento, há que considerar três possibilidades, com probabilidades diferentes. Para que o acontecimento se verifique no grupo superior com n_1 elementos, (Fig.6), a probabilidade p_1 é dada por:

$$p_1 = n_1/(n_1 + n_2 + n_3) = n_1/n$$

Para que o acontecimento se verifique no grupo médio com n_2 casos prováveis, a probabilidade p_2 será:

$$p_2 = n_2/(n_1 + n_2 + n_3) = n_2/n$$

Para que se verifique o acontecimento no grupo inferior, a probabilidade p_3 é dada por:

$$p_3 = n_3/(n_1 + n_2 + n_3) = n_3/n$$

É evidente que as probabilidades p_1 , p_2 e p_3 são diferentes entre si, porque $n_1 \neq n_2 \neq n_3$.

Vejamos agora, como se determina a função que traduz a incerteza, em que os acontecimentos têm probabilidades diferentes.

Seja A o acontecimento parcial, que consiste na escolha de um dos grupos, n_1 , n_2 e n_3 e seja B o acontecimento seguinte, que consiste na escolha de um dos elementos que pertença a um dos grupos seleccionados. Logo, o acontecimento final C é o produto da realização dos dois acontecimentos: a realização de A e a realização de B, isto é, $C = A \times B$. Mas, por outro lado, o acontecimento C consiste, afinal, na realização de um dos n acontecimentos, igualmente prováveis.

Como H é uma função aditiva, tem-se:

$$H(C) = H(A) + H(B)$$

Logo, para se saber a incerteza resultante, basta subtrair à incerteza associada aos n acontecimentos equiprováveis, dada pela expressão $H(n)$, as incertezas referentes à realização do acontecimento em cada um dos três subgrupos. Notando que em cada um dos subgrupos, os acontecimentos continuam a ser equiprováveis, a função $H(n_i)$, por exemplo, é dada por:

$$H(n_i) = K \log n_i$$

Mas, a falta de informação referente à realização de um só acontecimento deste subgrupo é uma fracção da falta de informação total do subgrupo, isto é, neste caso, será, apenas, $H(n_1)$. O mesmo se verifica com $H(n_2)$ e com $H(n_3)$.

Logo, para o acontecimento A, a incerteza $H(A)$ é dada por:

$$H(A) = K \log n - (n_1/n \cdot K \log n_1 + n_2/n \cdot K \log n_2 + n_3/n \cdot K \log n_3)$$

ou, como $p_i = n_i/n$:

$$\begin{aligned} H(p_1, p_2, p_3) &= K \log n - \sum p_i K \log n_i \\ &= K \log n - \sum p_i K \log p_i - K \log n \\ &= -K \sum p_i \log p_i \end{aligned}$$

Se em vez dos três grupos, que considerámos, generalizarmos para N grupos, teremos:

$$H(p_1, p_2, \dots, p_N) = -K \sum p_i \log p_i$$

que é a famosa fórmula de Shannon, fundamental na Teoria da Informação. Esta é uma generalização da fórmula que já se tinha obtido anteriormente (§ 3.3 I parte), quando os acontecimentos eram igualmente prováveis. De facto, se for:

$$p_1 = p_2 = \dots = p_N = 1/N, \text{ a expressão anterior reduz-se a}$$

$$H = -K \sum 1/N \log (1/N) = K \log N$$

De forma análoga se pode definir a entropia duma distribuição contínua, com a introdução de uma densidade de probabilidade $p(x)$, pela expressão:

$$H = - \int p(x) \log p(x) dx$$

3. SURPRESA, IGNORÂNCIA E ENTROPIA

3.1. SURPRESA E PROBABILIDADE

Vamos agora ver o que se entende por *função surpresa*, $\phi(p_i)$.

Suponhamos que associamos a cada estado de conhecimento uma certa probabilidade. Quanto maior for o valor da probabilidade, p_i , maior é o conhecimento que temos sobre um acontecimento. A probabilidade p_i é a medida da nossa expectativa baseada num dado estado de conhecimento, ou, se quisermos, de ignorância. Se a probabilidade, p_i , da realização de um certo conhecimento, é muito pequena, mas se, mesmo assim, se realiza, tem que se concordar que constitui uma grande "surpresa". A "surpresa", referente à realização dum acontecimento, traduz a estupefação e a admiração pela sua realização. Que "surpresa", tão agradável, quando sai a "sorte grande" no único vigésimo que se comprou ...

Vejamos, com outro exemplo, como a *surpresa* depende da probabilidade. Suponhamos que o Benfica joga no Estádio da Luz com o Vilar Formoso F. C. e que aquele perde o jogo! *A priori*, atribuímos uma grande probabilidade à vitória do Benfica, e com que surpresa ficámos quando o jogo acabou! A probabilidade do Vilar Formoso F. C. ganhar era ... "ínfima". Quem havia de dizer! Foi um "escândalo" para os benfiquistas e uma surpresa geral, tão inacreditável, mesmo para os apoiantes do Vilar Formoso, porque estes aceitavam de antemão, que a probabilidade que tinham de ganhar era pequeníssima.

A *função* que traduz a *surpresa*, $\phi(p_i)$, deve ser, portanto, uma função monotonicamente decrescente da probabilidade: quanto menos provável for a realização dum acontecimento, maior será a surpresa, se esse acontecimento se realizar.

Se se receberem duas mensagens independentes referentes a acontecimentos inesperados, a surpresa aumenta; este aumento é dado pela soma das surpresas contidas em cada uma das mensagens.

A *função surpresa* satisfaz à condição de aditividade: a surpresa que resulta da realização simultânea de vários acontecimentos, independentes, é a soma das surpresas que teríamos se cada um deles se realizasse, separadamente.

Por outro lado, se for p_j a probabilidade de receber a mensagem j e p_k a probabilidade de receber a mensagem k , a probabilidade de receber as duas, simultaneamente, será $p_j \times p_k$, como é evidente, pelo teorema das probabilidades compostas.

Como a surpresa é aditiva, tem-se:

$$\phi(p_j \times p_k) = \phi(p_j) + \phi(p_k)$$

e, portanto, esta condição impõe que a função de p , seja uma função logarítmica de p_j , como se demonstraria de forma análoga à que se usou para o estabelecimento da fórmula de Boltzmann, ou da fórmula de Shannon.

Logo, a forma da função $\phi(p_i)$ é:

$$\phi(p_i) = -C \log p_i$$

em que C é uma constante positiva. Como $0 \leq p_i \leq 1$, a função $\phi(p)$ é positiva e aumenta com $1/p_i$.

Esta expressão corresponde, à ideia que se tem sobre a surpresa. Assim, o que adiante receber uma mensagem (notícia) que é já completamente esperada? De facto, a probabilidade, *a priori*, é 1; logo: $\phi = -C \log 1 = 0$ e não trouxe nenhuma surpresa.

Quanto menos provável for uma notícia (isto é, quanto mais inesperada for), maior é a surpresa, ao receber a mensagem. Nestas circunstâncias, p_i é muito pequeno, e $\phi = -C \log p_i$ é muito grande.

3.2. IGNORÂNCIA E INCERTEZA

O valor expectável, ou valor médio duma grandeza g , cuja probabilidade de realização é p_a , é a "média pesada" de g definida por:

$$\langle g \rangle = \sum p_a g_a$$

No caso da probabilidade ser constante e igual a $1/n$, o valor médio é, simplesmente, a média aritmética. Ora, a probabilidade p_a implica uma "surpresa" $\phi(p_a)$, se se verificar o acontecimento E_a . Pergunta-se agora: qual será o valor expectável da surpresa para a realização de vários acontecimentos?

De acordo com a definição anterior, o valor médio será:

$$\langle \phi(p_i) \rangle = \sum p_i \phi(p_i) = -C \sum p_i \log p_i$$

Qual é, de facto, o significado do valor expectável da surpresa na vida corrente?

É evidente, que se pode considerar como uma medida da ignorância em relação aos resultados dos ensaios realizados sobre um conjunto de acontecimentos E_i que se representa por $\{E\}$. Logo:

$$\text{ign} \{E\} = - \sum p_i \log p_i$$

A ignorância refere-se à surpresa média sobre os acontecimentos de $\{E\}$ e não, apenas, à realização de um só deles, E_i .

Qual é o valor mínimo da ignorância? É evidente que a ignorância mínima se verifica quando um dos p_r é a unidade ($p_r = 1$) e todos os outros são zero, o que equivale a saber a resposta "certa". Portanto:

$$\{ \text{ign } E \}_{\min} = 0$$

E a ignorância máxima? Corresponde à situação de não haver quaisquer indicações preferenciais sobre cada um dos acontecimentos E_i . São todos igualmente prováveis.

A ignorância total sobre $\{ E \}$ traduz-se, portanto, por uma distribuição equitativa das probabilidades das várias respostas possíveis, o que equivale a ser sempre $p_i = 1/n$, qualquer que seja $i = 1, 2 \dots n$.

Suponhamos que, antes da realização dum ensaio, as probabilidades são p_i e que, depois da realização dum ensaio, são $p'_k = 1$ e $p'_j = 0$, para todos os $j \neq k$.

A ignorância depois da realização do ensaio, em que se fica com a resposta certa, é:

$$\text{ign } \{ E \} = 0$$

3.3. INCERTEZA E INFORMAÇÃO

A exposição anterior foi deliberadamente subjectiva, porque pretendemos usar as noções usuais e intuitivas de surpresa, de expectativa e de ignorância. Aceitando esta versão, pode, interpretar-se $\text{ign} \{ E \}$ como uma medida de incerteza, ou de indeterminação do estado de conhecimento.

Suponhamos, agora, que se está num estado de conhecimento a que corresponde uma distribuição de probabilidade p . Depois da realização dum ensaio sobre $\{ E \}$ o estado de conhecimento variou, assim como variou a lei de distribuição da probabilidade, p , que passou a ser p' .

Se a incerteza, a priori, era $\text{ign} \{ E \}$, a incerteza à posteriori será $\text{ign}' \{ E \}$. Em geral, quando a incerteza (ignorância) diminui, aumenta o "conhecimento" sobre $\{ E \}$.

Logo, a todo o decréscimo de ignorância, corresponde um aumento de informação sobre $\{ E \}$.

Define-se *informação* $I(p/p')$, entre os estados de conhecimento caracterizados pelas leis de distribuição de probabilidades p e p' , pelo decréscimo de ignorância:

$$\begin{aligned} I(p/p') &= \text{ign} \{ E \} - \text{ign}' \{ E \} \\ &= - \sum p_i \text{Log } p_i + \sum p'_i \text{Log } p'_i \end{aligned}$$

O máximo de informação obtém-se quando a ignorância final for nula (ficou-se a "saber tudo" ...) o que equivale a ser $p'_i = 1$ e $p'_j = 0$, para $j \neq i$:

$$I(p/1) = I_{\max}(p) \equiv I(p) = - \sum p_i \text{Log } p_i$$

que é a fórmula de Shannon, que já se tinha encontrado, por uma via diferente.

Note-se que pode surgir uma certa confusão, pois, embora os conceitos de ignorância e de informação sejam antagônicos, aparecem expressos pela mesma formulação analítica. Ora, esta incoerência aparente resultou de ser $\text{ign} \{ E \} = 0$.

Atendendo à interpretação, a grandeza $\text{ign} \{ E \}$ pode ser designada por *função entropia*, definida pela distribuição de probabilidade p_i sobre o espectro de acontecimentos, $\{ E_i \}$; e podemos representá-la por $S(p)$ ou $S(E)$, ou ainda $H(p)$ ou $H(E)$, conforme compararmos várias distribuições de p sobre o mesmo espectro, ou considerarmos as entropias de vários espectros E ao mesmo tempo:

$$H(p) \equiv H(E) \equiv S(p) \equiv S(E) = - \sum p_i \text{Log } p_i$$

Quando se usam logaritmos na base 2, $S(p)$ vem expressa em unidades chamadas "bit" (contração do **binary digit**). Um bit de informação corresponde a um par de alternativas: sim ou não; (+) ou (-), com igual probabilidade. Por ser

$$\lim_{\delta \rightarrow 0} \delta \text{Log } \delta = 0$$

tomaremos:

$$0 \text{Log } 0 = 0$$

A fórmula de Shannon, como é correntemente utilizada (expressa em bits), escreve-se, portanto, sob a forma:

$$H(p) = - \sum p_i \text{Log}_2 p_i, \text{ (em bits)}$$

Como está tacitamente implícito na expressão anterior, o emprego doutra base de logaritmos (maior que 1) implica apenas o uso de unidades diferentes para mesma grandeza, (ver Apêndice II).

O emprego de logaritmos naturais (base e) permite simplificar os cálculos que envolvem diferenciações e integrações. Em Física, por vezes, utiliza-se o factor $k \text{Log } 2$ vezes a entropia expressa em bits em que k é a constante de Boltzmann.

O problema das unidades é tratado, no Apêndice II.

3.4. O VALOR DA CONSTANTE DE BOLTZMANN

Suponhamos uma partícula confinada a um volume V . A densidade de probabilidade de presença da partícula em V é $p(x)$, de modo que:

$$p(x) = \begin{cases} 1/V & \text{- dentro do volume} \\ 0 & \text{- fora do volume} \end{cases}$$

e, tal que, $\int p(x) dx = 1$, por definição.

A incerteza será dada por:

$$H = -k \int 1/V \cdot \log 1/V \cdot dx$$

Como $V = \text{const.}$, vem:

$$H = -k \cdot 1/V \cdot \log 1/V \cdot V = k \log V$$

Suponhamos agora que o volume acessível à partícula aumenta para $V'(V' > V)$. A densidade de probabilidade passa a ser $p'(x) = 1/V'$ e a incerteza será:

$$H' = k \log V'$$

Como $H' > H$, perdemos a quantidade de informação, I :

$$I = H - H' = k \log V/V'$$

Analiseemos agora o problema da variação da entropia, à Clausius. Admitindo que a transformação é isotérmica, a variação da entropia para um mole de gás ideal é dada por:

$$\Delta S = R \log V/V'$$

Logo, a variação da entropia, por molécula será:

$$\delta S = R/N_A \cdot \log V/V'$$

em que N_A é o número de Avogadro.

Se as duas concepções são equivalentes, então $\delta S \equiv I$, e o valor da constante de Boltzmann, k , é dado por:

$$k = R/N_A$$

como indicámos anteriormente.

4. O SENTIDO FÍSICO DA DUALIDADE ENTROPIA E INFORMAÇÃO

No § 3.3. da I PARTE referiu-se, com alguma antecipação, que uma das versões da fórmula de Boltzmann coincidia com a fórmula de Shannon.

Por intuição e "ciência certa" Von Newmann, como vimos na citação de abertura, "esclareceu" Shannon, quando lhe apontou o nome de *entropia* para designar a *falta de informação*. Justifica-se, ainda, que a grandeza H se designe por entropia, visto que a fórmula de Shannon se reduz à fórmula célebre de Boltzmann.

De facto, se for $p_1 = p_2 = \dots = p_n = 1/W$, designando por W o peso estatístico, ou a probabilidade termodinâmica, de um dado macroestado (número total de microestados que realizou esse macroestado) vem:

$$H = - \sum 1/W. \log 1/W = C \log W = S(W)$$

Consideremos agora o problema inverso, isto é, vejamos como a aplicação dos conceitos da Teoria da Informação e do postulado da equiprobabilidade dos microestados, conduz à fórmula de Boltzmann.

Com efeito, vamos considerar o espectro $\{E\}$ dos W microestados, ou das W complexões, que realizam o estado macroscópico de equilíbrio do sistema termodinâmico. Não temos razões para aceitar que haja uns microestados "mais privilegiados" do que outros: todos os microestados são igualmente prováveis. Portanto a *ignorância* sobre o sistema é *máxima*.

Se se pretende saber o "microestado" que, num certo instante, realiza o estado do sistema, tem que se proceder a medições cujos resultados constituem mensagens, que contêm informação. Se as medições mostram que o sistema está

no microestado E_i , a informação parcelar, obtida com a medição, seria a diminuição da surpresa, i. e. $-c \log p(E_i)$ em que $p(E_i)$ seria a probabilidade do estado E_i .

O valor médio da informação parcelar, obtida em várias medições seria de acordo com a fórmula de Shannon:

$$\langle \phi \rangle = \sum p_i \phi(p_i) = -C \sum p_i \log p_i$$

Mas, $p_i = 1/W$, visto que o estado de equilíbrio se realiza de W maneiras diferentes. Substituindo na fórmula de Shannon, obtemos o máximo de falta de informação do sistema:

$$H = -C \sum 1/W \cdot \log 1/W = C \sum 1/W \cdot \log W = C \log W$$

Como W representa o número de microestados que realizam o macroestado, e como todos os microestados são igualmente prováveis é:

$$H = C/K \cdot S$$

Portanto, a *informação média* necessária para definir um sistema é proporcional à *entropia do sistema*.

Quanto menos soubermos sobre um sistema, ou quanto mais numerosos forem os microestados em que o estado do sistema se possa realizar, maior é a incerteza ou a entropia. As grandezas representadas pelos símbolos S e H , ainda que pareçam de natureza diferente, têm expressões formais análogas:

$$S = k \log W$$

e

$$H = C \log W$$

e, de facto, temos que admitir que as grandezas são, essencialmente, idênticas.

Se o estado dum sistema pode ser realizado por um número muito elevado de complexões, diremos que o estado tem uma grande probabilidade, com uma desorganização muito elevada (caos), enquanto que um sistema cujo estado é realizado apenas por poucas distribuições, está num estado de menor desorganização, menos caótico e de menor entropia. Logo, a nossa ignorância, ou a nossa falta de informação, sobre a distribuição que realiza o estado do sistema, é maior

para o estado de entropia mais elevada (há mais distribuições possíveis de microestados), do que para estados de entropia mais baixa (são poucas as complexões possíveis). A um aumento de entropia corresponde um decréscimo de informação sobre a distribuição microscópica de um sistema.

A entropia e a informação devem ser consideradas como dois aspectos diferentes da mesma entidade, ou na célebre expressão de Bohr, constituem "duas faces complementares da mesma realidade física". Assim, quando a entropia dum sistema aumenta, a informação sobre a sua estrutura microscópica diminui e inversamente. Por outro lado, ao mínimo da entropia ($S=0$) corresponde, pelo contrário, à informação mais completa possível.

A face "entropia" e a face "informação" excluem-se e completam-se mutuamente, como se se tratasse das faces duma mesma moeda. Pode, no entanto, ser mais cómodo ter duas grandezas que variem no mesmo sentido, do que duas grandezas a variar em sentido oposto. Basta inverter o sinal e uma delas. Foi o que sugeriu Brillouin (1956) introduzindo a entropia negativa $-S$ a que chamou "neguentropia". Nestas condições, num *processo natural a neguentropia decresce sempre*.

Vamos partir da fórmula de Boltzmann,

$$S = k \log W$$

Suponhamos que W passa dum valor inicial, W_0 , para outro valor menor, W_1 . Então a incerteza diminui. Se se definir a grandeza I , pela diferença $S_0 - S_1$, tem-se

$$I_1 = S_0 - S_1 = k \log (W_0/W_1)$$

Mas se o sistema for deixado a si próprio, a entropia numa transformação espontânea não pode nunca diminuir, $\Delta S_1 > 0$. Da definição de I , vem:

$$\Delta S_1 = (S_0 - S_1) \geq 0$$

Se se introduzir a neguentropia $N = -S$, aquela desigualdade pode escrever-se:

$$\Delta (I_1 + N_0) \leq 0$$

ou seja: *a soma da neguentropia com a informação, ou se mantém constante (transformação reversível), ou decresce.*

Mas, poderia também, introduzir-se o neologismo *neguinformação*. De facto, $f(p)$, designa a falta de informação, antes da resolução dum problema (ou da realização de um ensaio), ou, o que, é o mesmo, a informação que se obtém depois da resolução do problema, (ou da efectivação do ensaio). *Neguinformação*, coincidiria afinal com a *ignorância*, ou com a *falta de informação*. Num sistema isolado a *entropia e a neguinformação tendem, num processo natural, para um certo valor máximo*.

A dualidade entropia - informação não faz senão traduzir a mesma realidade física: a entropia é uma grandeza que define o estado macroscópico de um sistema na concepção fenomenológica, enquanto a informação é uma grandeza que define o estado de um sistema a partir do conhecimento da complexão microscópica, que realiza esse estado.

Desde que o estado de desordem molecular dum sistema aumente, a sua entropia cresce, enquanto a informação sobre o estado microscópico do sistema diminui. Se, pelo contrário, o sistema evolui para estados mais ordenados (o sistema não pode ser então isolado) a sua entropia diminui e a informação sobre o seu estado macroscópico do sistema aumenta. Esta dualidade para um mesmo sistema afigura-se, pois, incontestável e pode demonstrar-se que corresponde a uma dualidade formal matemática que permite estabelecer a conexão, em termos quantitativos, entre Teoria da Informação e a Termodinâmica.

5. A CONEXÃO ENTRE A TEORIA DA INFORMAÇÃO E A TERMODINÂMICA

5.1. FORMALISMO DE JAYNES. FUNDAMENTO DO MÉTODO DA ENTROPIA MÁXIMA

A solução do demónio de Maxwell, de que trataremos mais tarde, e a sua exorcização por Brillouin, recorrendo à Teoria da Informação, veio reforçar a ideia de que os problemas da Termodinâmica deveriam ser encarados sob uma nova perspectiva.

A dualidade entropia - informação, pode ser explorada no sentido de tomar a Teoria da Informação como a base lógica para a "construção" de toda uma teoria da Termodinâmica Clássica e da Termodinâmica Estatística. E.T. Jaynes em 1957, foi mesmo mais longe, ao partir do pressuposto de que as ideias da Teoria da Informação são mais fundamentais do que as da Termodinâmica, porque toda a Mecânica Estatística se pode enquadrar naquela teoria. Uma vez obtidas as expressões analíticas da Mecânica Estatística, podem derivar-se, como se sabe, as leis da Termodinâmica.

Os fundamentos do formalismo de Jaynes partem de certas afirmações de carácter intuitivo e, muito gerais, feitas sobre a natureza da distribuição das probabilidades que maximiza a informação (entropia) sobre o estado dum sistema. Assim dizia Jaynes em 1957:

"... in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other

would amount to arbitrary assumption for information which by hypothesis we do not have".

Devido a dificuldades de ordem prática e, até, de natureza essencial, para fazer o estudo do comportamento dum sistema, tem que recorrer-se a métodos de inferência estatística.

Por vezes, possui-se uma informação acerca de um sistema, que permite atribuir uma certa probabilidade a cada um dos vários estados que lhe são acessíveis. No entanto, nem sempre se verificam estas condições. Toda a informação que se tem acerca dum sistema é dada sob a forma de valores médios de propriedades observáveis, que o caracterizam macroscopicamente e do ponto de vista fenomenológico.

Então, designemos por p_i , a probabilidade dum microestado i e seja R uma propriedade que serve para identificar o estado do sistema que toma o valor R_i no estado i . Suponhamos que se conhecem os valores expectáveis de R_i , que representamos por $\langle R_i \rangle = r_i$, que se supõe reproduzirem a "verdade" obtida pela experiência, através dos valores observáveis. Consideremos agora uma propriedade genérica do sistema, $g_r(R)$, de que só é conhecido o valor médio dessa propriedade, que passamos a representar sistematicamente, por $\langle g_r \rangle$.

Tem-se então designando por p_i a probabilidade de estado i :

$$\sum p_i = 1$$

$$\sum p_i g_r(R_i) = \langle g_r \rangle \quad \text{em que } r = 1, 2 \dots l$$

Designando por l o número de graus de liberdade do sistema (variância do sistema), vê-se que se dispõe de $(l+1)$ equações, cujo número é, em geral, bastante inferior ao número de incógnitas p_i . Logo, o sistema anterior não permite determinar os p_i . Do ponto de vista determinístico o problema não tem solução e, pelo contrário, é indeterminado.

Temos, portanto, que atribuir à distribuição da probabilidade p_i estimativas plausíveis, verosímeis, coerentes e o "melhor" possível. Para isso, tem que utilizar-se um critério de escolha das infinitas hipóteses possíveis.

No estudo estatístico dum sistema deve usar-se somente a informação que **de facto** se possui e não se deve tirar partido de informação que se pensa, ou que se julga ter e que, não é dada!

Parece, portanto, razoável, escolher para valores das probabilidades p_i , aqueles que maximizem a incerteza (entropia) e que sejam, ao mesmo tempo, consistentes com o conhecimento parcelar que se tem acerca do sistema (propriedades médias $\langle g_r \rangle$).

De todos os valores p_i , que possam "dizer toda a verdade", aquele que implica a informação mínima, e que, portanto, corresponde ao que dá o máximo de falta de informação, é o que se afigura menos viciado e o mais próximo de conter só "a verdade". Por isso "atribuímos ao sistema o valor de p que reproduz toda a informação disponível e que, para sermos isentos, corresponde ao máximo de falta de informação". (Jaynes, 1958).

Procedendo assim, fica-se com a garantia de que, se for recebida uma mensagem referente ao estado do sistema, o ganho de informação (decréscimo de incerteza) contida nessa mensagem é máximo. Parte-se da posição mais desfavorável e menos "optimista" e de "maior humildade", o que elimina o uso de informação viciada. A lei de distribuição de p_i satisfaz a certas condições, mas não se sabe bem qual é. E pretende escolher-se a mais isenta de entre todas as distribuições possíveis. A distribuição mais razoável no estado de ignorância em que nos encontramos será aquela que corresponde ao máximo da nossa incerteza, sem quaisquer ideias preconcebidas.

A solução do problema, segundo Jaynes pode pôr-se assim:

A distribuição de probabilidades p_i , mais isenta e menos viciada, é a que maximiza a incerteza H (entropia S), e que é consistente com a informação certa disponível sobre o sistema.

Há portanto que maximizar a função

$$S = H = -k \sum p_i \log p_i$$

sujeita aos constrangimentos:

$$a) \sum p_i = 1 \quad \text{ou} \quad \sum dp_i = 0 \quad [- (\Omega + 1)]$$

$$b) \sum p_i g_r(R_i) = \langle g_r \rangle \quad \text{ou} \quad \sum g_r(R_i) dp_i = 0 \quad [(\lambda_r)]$$

Trata-se, portanto, de resolver um problema de máximos condicionados, cuja solução se pode obter pelo método dos multiplicadores de *Lagrange* aplicados às condições a) e b), que representaremos, respectivamente, por $-(\Omega + 1)$ e por λ_r . Toma-se $-(\Omega + 1)$ e não Ω , por comodidade de cálculo.

Tem-se então, seguindo a metodologia habitual:

$$\sum [\log p_i + 1 - \Omega - 1 + \sum \lambda_r g_r(R_i)] dp_i = 0$$

donde se conclui que:

$$\log p_i = \Omega - \sum \lambda_r g_r(R_i)$$

visto que dp_i é diferente de zero.

Logo, a lei de distribuição de probabilidade no formalismo de Jaynes (MEM), é dada por:

$$p_i = \exp (\Omega - \sum \lambda_r g_r (R_i))$$

Note-se que este tipo de solução (função exponencial) satisfaz automaticamente à condição de ser: $p_i \geq 0$

A função de probabilidade vem assim expressa em termos dos parâmetros indeterminados de Lagrange (Ω e λ_r) que, por sua vez, satisfazem aos constrangimentos. Logo, por ser:

$$\sum p_i = 1$$

vem:

$$\Omega = - \text{Log } \sum \exp [- \sum \lambda_r g_r (R_i)]$$

expressão que mostra que Ω e λ_r não são independentes.

Podíamos aqui introduzir uma função de partição, Z , constituída pelo argumento do logaritmo, como se fez anteriormente, mas preferimos continuar a escrever as expressões de forma explícita. De facto, é mais fácil, desta forma, pôr em destaque as relações entre os parâmetros de Lagrange e, por exemplo, mostrar que o parâmetro indeterminado Ω é uma função dos outros factores indeterminados, isto é, que:

$$\Omega = \Omega (\lambda_r)$$

As outras condições impõem que

$$\begin{aligned} \langle g_r \rangle &= \sum p_i g_r (R_i) / \sum p_i = \\ &= \sum \exp [\Omega - \sum \lambda_r g_r (R_i)] \cdot g_r (R_i) / \sum \exp \lambda_r g_r (R_i) = \\ &= e^\Omega / e^\Omega \cdot \sum [\exp (- \sum \lambda_r g_r (R_i)) g_r (R_i) / \sum \exp [- \sum \lambda_r g_r (R_i)] = \\ &= \partial \Omega / \partial \lambda_r \end{aligned}$$

isto é:

$$\langle g_r \rangle = \partial \Omega / \partial \lambda_r$$

o que mostra que: o valor expectável da grandeza g_r é a derivada de Ω em ordem a λ_r ; ou, ainda que, "o observável $\langle g_r \rangle$ é conjugado de λ_r em Ω ". (*)

Vejamos, agora, a variância da grandeza g_r (R_r), que constitui uma medida da flutuação desta grandeza. Como se sabe, a variância é dada por:

$$\sigma^2(g_r) = \langle g_r^2 \rangle - \langle g_r \rangle^2$$

Tem-se então:

$$\begin{aligned} \sigma^2(g_r) &= \sum p_i g_r^2 - \langle g_r \rangle^2 = \\ &= \sum g_r^2 \cdot \exp(\Omega - \sum \lambda_r g_r) / \sum \exp(\Omega - \sum \lambda_r g_r) - \\ &- [\sum g_r \exp(\Omega - \sum \lambda_r g_r) / \sum \exp(\Omega - \sum \lambda_r g_r)]^2 = \\ &= - [\partial / \partial \lambda_r \cdot \sum g_r \exp(\Omega - \sum \lambda_r g_r) / \sum \exp(\Omega - \sum \lambda_r g_r)] = \\ &= \partial^2 \Omega / \partial \lambda_r^2 \end{aligned}$$

o que mostra que a variância de g_r é dada pela segunda derivada de Ω em ordem a λ_r .

Suponhamos que se quer calcular a componente média do gradiente de uma grandeza genérica, que sendo função de R_r , se supõe ainda ser função duma variável qualquer Y . Tem-se, então:

$$\partial \Omega / \partial Y = 1 / \sum \exp \dots \times \sum \exp \dots \times (\lambda_r \cdot \partial g_r / \partial Y) = \lambda_r \langle \partial g_r / \partial Y \rangle$$

ou seja:

$$\langle \partial g_r / \partial Y \rangle = 1 / \lambda_r \cdot \partial \Omega / \partial Y$$

Esta expressão mostra que o gradiente da propriedade genérica, a menos do inverso do seu parâmetro indeterminado associado, se reduz sempre ao gradiente de uma única grandeza, que é o parâmetro Ω .

(*) Diz-se que a variável x é conjugada de y em u quando $x = \frac{\partial u}{\partial y}$

5.2. RELAÇÃO ENTRE GRANDEZAS $\Omega (\lambda_1, \lambda_2, \dots, \lambda_n)$ E $S (r_1, r_2, \dots, r_n)$

Dualidade formal

Vamos mostrar que as grandezas Ω e S (ou H) são transformadas de Legendre. Em vez da grandeza genérica, $g_r (R_i)$, passaremos a utilizar a própria grandeza R_i que tínhamos escolhido para caracterizar o estado do sistema.

Vimos que:

$$a) p = \exp (\Omega - \sum \lambda_i R_i) \quad (1)$$

$$b) \Omega (\lambda_1, \lambda_2, \dots, \lambda_n) = - \text{Log} \sum \exp [- \sum \lambda_i R_i] \quad (2)$$

$$c) \partial \Omega / \partial \lambda_i = r_i \quad (i = 1 \dots n) \quad (3)$$

onde r_i é o valor expectável, observável de R_i .

Daqui podemos tirar o valor de λ_i , que substituído em (2), conduz a um valor numérico de Ω e que, portanto, permite determinar p .

d) Com o valor de p pode calcular-se, a incerteza ou a entropia S (ou H):

$$\begin{aligned} H \equiv S &= -k \langle \text{Log } p \rangle = -k \langle \Omega - \sum \lambda_i R_i \rangle = \\ &= -k \Omega + \sum k \lambda_i r_i \end{aligned} \quad (4)$$

visto que

$$\langle \Omega \rangle = \Omega \quad e \quad \langle R_i \rangle = r_i$$

e) Por outro lado,

$$dH \equiv dS = -k \sum (\partial \Omega / \partial \lambda_i - r_i) d\lambda_i + k \sum \lambda_i dr_i$$

ou seja

$$dH = k \sum \lambda_i dr_i \quad (5)$$

e portanto:

$$\partial H / \partial r_i \equiv \partial S / \partial r_i = k \lambda_i \quad (6)$$

o que mostra que as variáveis λ_i são conjugadas das grandezas r_i em H (ou em S), a menos duma constante k.

f) A comparação de (3) com (6) mostra imediatamente a existência duma dualidade entre grandezas Ω e λ_i , por um lado, e entre a grandeza H (ou S) e r_i por outro.

g) A expressão (4) pode escrever-se:

$$H = -k \Omega + k \sum \lambda_i \cdot \partial \Omega / \partial \lambda_i \quad (7)$$

ou

$$k \Omega = -H + \sum r_i \cdot \partial H / \partial r_i \quad (8)$$

o que mostra que H (ou S) e Ω são transformadas de Legendre, a menos de uma constante, isto é:

$$\Omega = \mathcal{L} S \quad (9)$$

e

$$H \equiv S = \mathcal{L} \Omega \quad (10)$$

A grandeza Ω é, portanto, uma *função de Massieu*, isto é, uma transformada de Legendre no esquema de entropia.

h) Os restantes parâmetros indeterminados, $\lambda_1, \dots, \lambda_n$ são, por sua vez, grandezas intensivas e coincidem com as variáveis intensivas no esquema S da

entropia, $\Pi_i = \frac{\partial S}{\partial X_i}$. Convém relembrar que as variáveis intensivas no esquema da entropia Π_i estão relacionadas com as variáveis intensivas no esquema da energia $P_i = \frac{\partial U}{\partial X_i}$ pelas relações

$$\Pi_i \equiv \left(\frac{\partial S}{\partial X_i} \right)_U = \left(\frac{\partial U}{\partial X_i} \right)_S / \left(\frac{\partial U}{\partial S} \right)_{X_i} \equiv -P_i / T_i \quad (11)$$

em que T designa a temperatura absoluta, X_i uma propriedade extensiva (aditiva) do sistema e P_i os parâmetros intensivos no esquema da energia ($P_i \equiv \partial U / \partial X_i$).

Consideremos a título de exemplo um sistema termodinâmico aberto. Os " r_i " que conhecemos são a energia interna U , e o número de moles das várias espécies químicas que constituem o sistema e que designaremos por n_a, n_b, n_c . Então, tem-se para a expressão que dá a distribuição das probabilidades:

$$p_i = \exp (- \Omega - \beta U - \sum \alpha_j n_j) \quad (12)$$

Seguindo uma via análoga à anterior conclui-se, portanto, que:

$$H \equiv S = S (U, n_a, n_b, n_c) \quad (13)$$

Ora, esta expressão constitui a relação fundamental do sistema no *esquema da entropia*. Na forma explícita, como se depreende da expressão (4), pode escrever-se:

$$H \equiv S = k \Omega + k \beta \langle U \rangle + k \alpha_a \langle n_a \rangle + k \alpha_b \langle n_b \rangle + k \alpha_c \langle n_c \rangle + \dots \quad (14)$$

que é, afinal, um caso particular da expressão (4).

A análise das expressões (3) e (6) mostra imediatamente que a *grandeza* r_i é conjugada de λ_i em Ω ; e que λ_i é conjugada de r_i em H .

As expressões (3) e (6) são duais, visto que

$$H = \mathcal{L}(\Omega) \\ r_1 r_2$$

e

$$\Omega = \mathcal{L}(H) \\ \lambda_1 \lambda_2$$

Estamos, portanto, em face duma dualidade formal, que complementa e consubstancia a dualidade qualitativa referida no parágrafo anterior.

Por ser:

$$S \equiv H = -k \Omega + \sum k \lambda_i r_i$$

tem-se, para a parcela $k \lambda_a r_a$ seguinte expressão:

$$k \lambda_a r_a = H + k \Omega - \sum k \lambda_i r_i = S - k (-\Omega + \sum \lambda_i r_i)$$

o que demonstra que é também uma função de Massieu.

Como se sabe da Termodinâmica Clássica, estamos em face duma *dualidade homônima maximizante* (princípio de máxima da entropia $dS \leq 0$).

5.3. SIGNIFICADO FÍSICO DOS PARÂMETROS DE LAGRANGE, Ω E λ_r

Consideremos um sistema termodinâmico aberto.

Então Ω será uma função dos outros multiplicadores indeterminados $\lambda_1, \lambda_2, \dots$, que agora designaremos respectivamente por $\beta, \alpha_a, \alpha_b, \alpha_c \dots$ e doutras quaisquer possíveis variáveis extensivas, passivas:

$$\Omega = \Omega (\beta, \alpha_a, \alpha_b, \alpha_c \dots, x_1, \dots x_k) \quad (1)$$

Por outro lado, viu-se que se tem sucessivamente:

$$\langle U \rangle = \partial \Omega / \partial \beta \quad (2)$$

$$\langle n_\alpha \rangle = \partial \Omega / \partial \alpha_\alpha \quad (3)$$

$$\sigma^2(U) = - \partial^2 \Omega / \partial \beta^2 \quad (4)$$

$$\sigma^2(n_\alpha) = \partial^2 \Omega / \partial \alpha_\alpha^2 \quad (5)$$

Da expressão (1) vem por diferenciação:

$$d\Omega = \partial \Omega / \partial \beta \cdot d\beta + \partial \Omega / \partial \alpha_a \cdot d\alpha_a + \partial \Omega / \partial \alpha_b \cdot d\alpha_b + \dots + \\ + \sum \partial \Omega / \partial x_j \cdot dx_j$$

ou, atendendo aos significados das derivadas:

$$d\Omega = - \langle U \rangle d\beta - \langle n_a \rangle d\alpha_a - \langle n_b \rangle d\alpha_b + \dots + \\ + \beta \sum \langle P_j \rangle dx_j \quad (6)$$

Se substituirmos esta expressão na forma diferencial da entropia vem:

$$dS = k \beta < dU > + k \alpha_a d < n_a > + \dots + k \beta \Sigma < P_j > dX_j \quad (7)$$

Esta seria a expressão que dá a variação da entropia do sistema aberto, resultante de variações da energia, da massa, e de outros parâmetros exteriores.

A integração deste pfaffiano (7) conduzirá então à expressão da entropia para o sistema termodinâmico aberto, isto é:

$$S = k \beta < U > + k \alpha_a < n_a > + k \alpha_b < n_b > + k \alpha_c < n_c > + \dots + k \beta \Sigma < P_j > X_j \quad (8)$$

que é a conhecida fórmula de *Euler* e que, afinal, constitui a *forma explícita da relação fundamental* (13) do parágrafo anterior.

Da inspecção das expressões (1) a (8) conclui-se imediatamente, tendo em vista a relação (11) do parágrafo anterior, que:

$$\beta = 1/kT \quad (9)$$

é que

$$\alpha_c = - \mu_c / kT \quad (10)$$

em que

$$\mu_c = \partial < U > / \partial n_c \quad (11)$$

é o potencial químico do constituinte *c* do sistema. Finalmente, o parâmetro indeterminado de Lagrange Ω é, afinal, o grande potencial canónico, e que se reduz a:

$$\Omega = \beta \Sigma < P_j > X_j \quad (12)$$

Para qualquer outra propriedade extensiva do sistema termodinâmico, tinha-se ainda, como se viu anteriormente, ao tratarmos do caso geral:

$$< P_k > = 1/\beta \cdot \partial \Omega / \partial X_k \quad (13)$$

Para um mole de um gás ideal, com $\langle N_A \rangle$ moléculas, à temperatura T , à pressão P e ocupando o volume $\langle V \rangle$, é:

$$\beta = 1/kT = \langle N_A \rangle / P \langle V \rangle = \langle N_A \rangle / RT \quad (14)$$

O que confirma que a constante de *Boltzmann*, k , é dada por

$$k = R/N_A = 1,37 \times 10^{-16} \text{ erg/k} = 1,37 \times 10^{23} \text{ J/k} \quad (15)$$

em que R é a constante dos gases ideais e N_A o número de Avogadro.

A expressão (12) reduz-se neste caso a

$$\Omega = \beta P V \quad (16)$$

admitindo que as únicas forças externas que actuam são as que resultam da pressão P .

Já se tinha mostrado que o parâmetro indeterminado Ω é sempre um *potencial termodinâmico* no esquema da entropia, que se designa genericamente por função de *Massieu*. No caso anterior Ω seria o grande potencial canónico.

Quanto aos parâmetros λ_r , viu-se que são parâmetros intensivos conjugados de $\langle g_r \rangle$ em Ω . Coincidem, portanto, com os parâmetros intensivos Π_i do esquema da entropia S , relacionados com os parâmetros intensivos P_i do esquema U , pelas relações $\Pi_i = P_i/T$ em que T designa a temperatura.

No caso do exemplo presente o parâmetro λ associado à grandeza $\langle U \rangle$ é, afinal, o célebre parâmetro $\beta = 1/kT$ que figura na lei da *distribuição de probabilidade de Boltzmann*.

Os outros parâmetros λ_2, \dots estão afinal relacionados com os potenciais químicos μ_c . No caso dum sistema com um componente apenas (substância pura) seria:

$$\mu_c = - G/T$$

em que G é o potencial de *Gibbs*, dado por:

$$G = U + PV - TS = \mathcal{L}_{VS}(U)$$

Como se sabe na "difusão" do calor este transita de regiões em que β é menor para regiões em que β é maior.

Se falarmos da difusão da matéria veríamos que esta se daria das regiões em que α é menor para regiões em que α é maior. Analogamente ao que sucede com $-\beta$ que se pode considerar como a medida termodinâmica para a difusão da energia, assim também $-\alpha_c$ pode ser tomado como uma medida da tendência das partículas dum determinado tipo c se escaparem (difundirem) da região em que se encontram para outra com valor mais elevado de α .

6. O DEMÓNIO DE MAXWELL E A TEORIA DA INFORMAÇÃO

6.1. O DEMÓNIO DE MAXWELL E O SEGUNDO PRINCÍPIO

O Segundo Princípio Fundamental da Termodinâmica é muito mais subtil do que o Primeiro. Logo que aquele foi enunciado, apareceram várias versões, mas pior ainda, foram surgindo reparos à sua essência e invocadas várias limitações (aparentes). De entre todas estas destaca-se, pela sua persistência e pela sua agudeza, a questão levantada por *Maxwell* (1871).

Com efeito, *Maxwell* concebeu um mecanismo, segundo o qual um sistema isolado, a uma dada temperatura, poderia evoluir e conduzir a uma diferenciação de temperatura, o que constituía uma violação flagrante do Segundo Princípio. Deste modo se constituiria assim uma máquina perpétua de segunda espécie, permitindo obter trabalho de um único reservatório, que, inicialmente, estava a uma temperatura uniforme.

Imaginemos, como sugeriu *Maxwell*, um recinto isolado separado em dois compartimentos, A e B, por meio duma partição, e que estavam cheios de um gás, a igual temperatura. Ora, a teoria cinética indicava que algumas das moléculas se movem mais rapidamente que outras, mas a energia cinética média por molécula é a mesma nos dois compartimentos, porque estão à mesma temperatura (Fig. 7).

Maxwell discutia o que poderia acontecer se houvesse uma portinhola na partição, comandada por um pequeno ser inteligente, sagaz, suficientemente ágil para abrir e fechar a porta, sem dispêndio de energia e com extrema rapidez, capaz de seguir cada uma das moléculas no seu movimento e, ainda, de as identificar.

Este ser era o célebre "*demónio de Maxwell*". O *demónio* permitia a passagem das moléculas mais rápidas do compartimento A, que se aproximassem da porta,

para o compartimento B, enquanto que as moléculas mais lentas, deste compartimento, eram autorizadas a passar para o compartimento A. O *demônio* era um "ser prodigioso e notabilíssimo", ao ser capaz de fazer esta selecção em dois grupos de moléculas, uma das mais rápidas dum lado da partição e outro, o das lentas, do outro lado, o que, tendo em vista o significado, à luz da Teoria Cinética da temperatura, equivale a estabelecer uma diferença de temperatura entre eles.

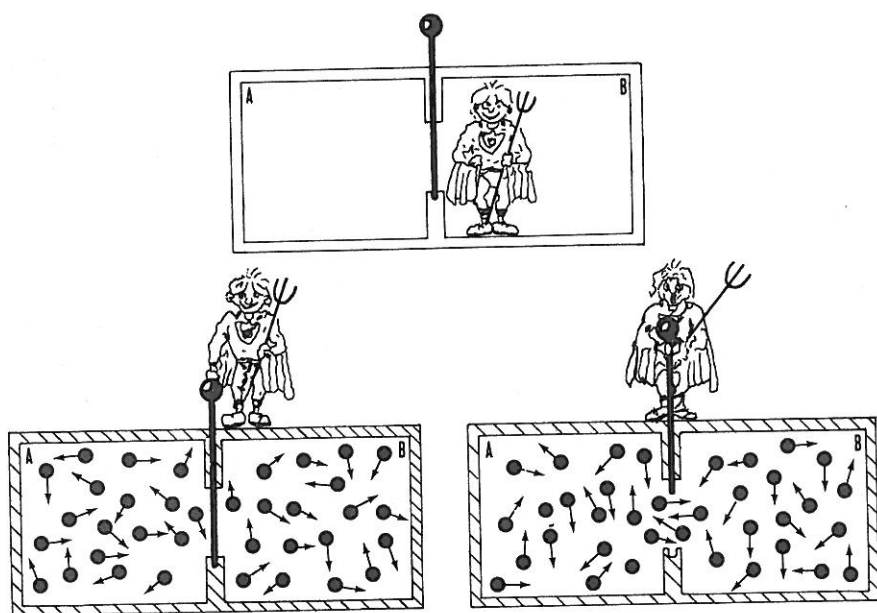


Fig. 7 - O Demônio de Maxwell.

Para proceder à selecção, o *demônio* só tinha que deixar correr o tempo ... e aguardar!

Ora, esta situação final constituiria uma violação do Segundo Princípio Fundamental da Termodinâmica, *porque implicava uma transferência de calor para a região de temperatura mais elevada.*

Além disso, o *demônio*, ao fazer a selecção, teria posto "ordem" na desordem inicial o que levaria à diminuição da entropia no recinto global. Não podia haver violação mais flagrante do Segundo Princípio, que no seu enunciado comum dizia: *"num sistema isolado a entropia só pode aumentar, atingindo o valor máximo no equilíbrio".*

O problema do *demônio* de *Maxwell* foi debatido durante quase um século e houve várias tentativas de "exorcização", as quais foram em vão, até surgir a Teoria da Informação.

Notaremos que, com o seu *demônio*, *Maxwell* pretendia acentuar que a transferência de energia dum corpo mais frio para um corpo mais quente pode dar-se, constantemente, mas em escala muito limitada. O Segundo Princípio tem um valor estatístico e pode não ser válido à escala microscópica e com pequenas populações. (Ver exemplos da I Parte)

A hipótese de *Maxwell* relativa às capacidades do demônio (o mais favoráveis possível) não é assim tão pacífica, ao admitir um mecanismo que fizesse actuar a portinhola, a bel prazer, sem qualquer dispêndio de energia.

6.2. A EXORCIZAÇÃO DO DEMÓNIO DE MAXWELL

Só, mais recentemente, *Léon Brillouin* (1956) deu a resposta fundamental ao paradoxo de *Maxwell*, ao pôr a seguinte questão fundamental: *Será possível "ver" cada uma das moléculas, individualmente?*

Quando o *demónio* olha para qualquer lado do sistema isolado, a temperatura constante, a uniformidade da radiação térmica predominante, não lhe deixa "ver" nada. Quanto muito, a uniformidade no interior do gás poderia permitir ao *demónio* aperceber-se da existência da radiação térmica dum corpo negro, em equilíbrio, e das suas flutuações, mas nunca ver e distinguir as moléculas, individualmente.

Somos levados a concluir que o *demónio* necessita de uma fonte própria de energia (uma "candeia") para romper o equilíbrio (uniformidade) da radiação no interior do sistema, que o vai "iluminar" e permitir ver e distinguir, assim as moléculas. Esta fonte de radiação não pode ficar em equilíbrio com o sistema. É a energia, que a "candeia" do *demónio* lança no interior do sistema, que lhe fornece a informação necessária para ele poder actuar a válvula, separando as moléculas de maiores velocidades, das de velocidades menores, como vamos mostrar.

Como o funcionamento desta fonte pressupõe consumo de energia, a actividade do *demónio* não é gratuita, ao contrário do que admitia *Maxwell*. Se dividirmos a energia radiante emitida, pela temperatura da "tocha" do *demónio*, determinamos uma grandeza, que é a entropia lançada no sistema, e que designamos por *neguentropia*. É esta *neguentropia* que poderia permitir ao *demónio* obter informação sobre a velocidade e a posição das moléculas.

Segundo a Teoria da Informação, o papel do *demónio* consiste em transformar a informação em "entropia negativa". Pode, no entanto, mostrar-se que o

aumento global de entropia é superior ao decréscimo que o demónio pode originar com a sua actuação. Afinal, para o sistema global: gás - demónio - "candeia", verificar-se-á aumento líquido de entropia, tal como exige o Segundo Princípio Fundamental da Termodinâmica.

A neguentropia mede a qualidade da energia. Um sistema contém neguentropia quando tem a possibilidade de realizar trabalho. Sistemas que tenham uma distribuição não uniforme de pressão, ou de temperatura ou de concentração, ou de potencial eléctrico, podem realizar trabalho e, portanto, contêm uma certa quantidade de *energia disponível*, que nesta concepção é proporcional à neguentropia.

Devemos ainda acentuar que a Teoria da Informação impõe limites nas observações e nas medições, descontando mesmo aqueles, que são inevitáveis, pelo princípio de incerteza de *Heisenberg* (a posição e a velocidade duma molécula não podem ser, simultaneamente, determinadas com precisão ilimitada)*. Em média, o aumento de entropia é superior à informação obtida.

Suponhamos que o demónio, mesmo sem ver as moléculas à distância, as pode detectar por intermédio de um dispositivo sensível aos campos de força moleculares. Como sabemos, estes campos variam na razão inversa do quadrado da distância. Ora, para que a abertura da válvula se possa fazer, sem trabalho, é necessário que sobre ela não actuem forças, o que se torna manifestamente impossível neste processo, porque logo que o "demónio" selecciona uma molécula, ele e a válvula ficam também sob a acção do campo de forças moleculares.

Voltemos à neguentropia

Segundo *Brillouin*, para que o raparo proposto por *Maxwell* tenha sentido, é necessário admitir que o "demónio" se informe da posição e da velocidade das moléculas, quando estas se encontram suficientemente distantes da válvula, visto que, só nestas condições, se pode esperar que a acção dos campos de força seja desprezável. Ora, tal operação requer, como vimos, o emprego de fonte de radiação auxiliar, consideravelmente diferente da radiação de fundo. Por isso, o "demónio" tem que munir-se de um dispositivo qualquer: bateria lâmpada eléctrica, "candeia", etc., que forneça a radiação indispensável para "iluminar" as moléculas. Todavia, quaisquer destes dispositivos passam a fazer parte integrante do sistema isolado, cuja entropia o "demónio" pretendia diminuir.

* Por exemplo, para determinar a posição duma molécula num gás, com grande precisão, seria necessário um feixe de radiação de alta energia. A interacção desta luz com a molécula alteraria a sua quantidade de movimento, de forma significativa. Logo, a posição e a quantidade de movimento não se podem determinar simultaneamente com precisão ilimitada.

Tudo se passa como se a radiação da "tocha" fornecesse ao sistema entropia negativa, ou neguentropia. Logo, quando o demónio "ilumina" o sistema inunda-o com neguentropia e é desta neguentropia que o demónio obtém informação. Com esta informação opera a válvula, origina diferenças de temperatura, com a separação das moléculas, e cria, portanto, à sua conta, neguentropia (ordem), completando um ciclo. Este ciclo de operações comporta sucessivamente um fornecimento de neguentropia (pela radiação estranha) uma transformação desta em informação e finalmente a transformação inversa (com a separação). (Fig. 8)

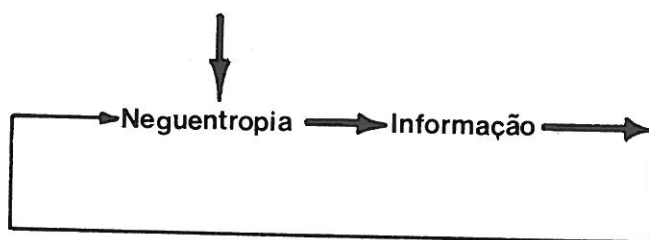


Fig. 8 - Exorcização do Demónio de Maxwell. Não se pode obter coisa alguma a partir do nada, nem mesmo uma observação!

A neguentropia assim produzida, é inferior à que foi fornecida ao sistema, afinal de acordo com o Segundo Princípio.

Neguentropia é, como se vê, um conceito muito global que, fundamentalmente, se identifica com ganho de informação, ordenamento, etc. Devemos ainda insistir no facto da emissão, da transmissão e de recepção exigirem, sempre, dispêndio de energia.

Informação, tal como entropia, está associada à energia.

A entropia surge, umas vezes, como um conceito subjectivo, constituindo uma medida da nossa ignorância, outras vezes sob a forma objectiva, como acontece na concepção da Termodinâmica Clássica, à *Clausius*.

Brillouin mostrou que qualquer informação, resultante duma observação de natureza física, por exemplo, se paga com um aumento da entropia do laboratório. Este aumento de entropia, em média, é superior à informação obtida, quando medidas no mesmo sistema de unidades. Esta condição impõe, portanto, uma nova limitação às possibilidades da observação. Um observador qualquer (físico, poeta, ecólogo) requer fontes de neguentropia, porque cada observação é feita à custa de neguentropia do seu universo complementar.

7. **BALANÇO ENERGÉTICO DA COMPUTAÇÃO. LIMITE COMPUTACIONAL**

Os computadores são máquinas que, mediante a aceitação de linguagem, geram informação, para o que têm que consumir energia.

A necessidade em energia para se obter aquela informação poderá indagar-se escrutinando o processo mais simples para realizar cálculo. Em 1982, Toffoli concebeu um computador em que, em vez de utilizar os estados quânticos de electrões, o cálculo era realizado com bolas de bilhar. Por bolas de bilhar quer significar-se bolas completamente rígidas. A presença num determinado local de uma bola de bilhar corresponde no alfabeto binário à, digamos, letra 1, a sua ausência corresponde à letra 0.

Nestas condições, colisões entre bolas rígidas sobre um plano servem para proceder a cálculo. Os momentos \vec{p} e as posições \vec{q} são sempre obtidos com erros $|\Delta\vec{p}| = \Delta p_1 = \Delta p_2$ e $|\Delta\vec{q}| = \Delta q_1 = \Delta q_2$ que consideramos iguais para simplificação do manuseamento do cálculo.

Se no computador de Toffoli houver N bolas o erro, V_o , no estabelecimento da posição do computador no espaço das fases é dado por:

$$V_o = \prod_{j=1}^N |\Delta p_j|^2 |\Delta q_j|^2 = (\Delta p \ \Delta q)^{2N} \quad (1)$$

De acordo com o princípio de incerteza de Heisenberg, no mínimo, para cada bola a imprecisão no posicionamento, no espaço das fases, será:

$$(\Delta p \ \Delta q) = h \quad (2)$$

em que h é a constante de Planck. Então, a entropia termodinâmica, S , criada por acumulação de todas as N bolas no computador é dada por:

$$\begin{aligned} S &= k \log (V_o / h^{2N}) \\ &= 2N k \log (\Delta p \Delta q / h) \end{aligned} \quad (3)$$

em que k é a constante de Boltzmann. Ou seja, à medida que o tempo passa, as colisões encarregam-se de aumentar o erro quer dos momentos quer das posições e, a quantidade de informação, a desordem aumentará inevitavelmente.

De facto, por cada colisão, o acréscimo do erro no ângulo e saída $\Delta\phi'$ é, em primeira aproximação, dado por:

$$\Delta\phi' = \Delta\phi (1 + l / r) \quad (4)$$

em que l é o livre percurso médio e r o raio de cada bola rígida. O erro acumulado na posição do sistema no espaço das fases após n colisões é de:

$$V_n = V_o \prod_{i=1}^n (1 + l_i / r) \quad (5)$$

Na hipótese mais favorável $l_i = r$ o que significa que nestas condições, no cálculo, por cada passo executado no computador, por cada bit de informação obtido, a entropia termodinâmica nunca será inferior a:

$$S_{\min} = k \log 2 \quad (6)$$

O que implica um balanço entre informação e energia (não tendo em conta a agitação térmica que à temperatura T introduziria um factor de kT na última equação) traduzido na troca de 1 bit de energia por 1 bit de informação.

∴

Esta equivalência impõe um limite para a capacidade de cálculo de um computador.

Consideremos um computador de massa, m . O máximo de energia possível, E , esse computador é dado pela relação de Einstein.

$$E = m c^2$$

em que c é a velocidade da luz no vácuo.

Por outro lado, o mínimo de energia discernível entre estados deste computador, ΔE , é definido pelo princípio de incerteza de Heisenberg:

$$\Delta E \cdot \Delta t = h$$

em que Δt é um intervalo de tempo e h a constante de Planck.

Então o número máximo de bits, N , calculados no computador de massa m no intervalo de tempo Δt é dado por:

$$N = \frac{E}{\Delta E} = m \frac{c^2}{h} \Delta t$$

que constitui o limite da computação.

Se m for a massa do planeta Terra $m \cong 10^{24} \text{Kg}$ e Δt for o tempo de existência da Terra, $\Delta t \cong 10^{17} \text{s}$, então:

$$N \cong 10^{93} \text{ bit}$$

constitui o limite de capacidade de computação do Planeta.

**A LARGUEZA DO CONCEITO DE
ENTROPIA**

III

1. UMA VISÃO NOVA DA ENTROPIA

A experiência com as limitações da conversão de calor em trabalho levou a vários enunciados do Segundo Princípio Fundamental da Termodinâmica, todos equivalentes (enunciados de Carnot, de Kelvin, de Clausius, de Planck, etc.). A formulação deste Princípio, em termos de linguagem matemática, conduziu Clausius (1865) à versão equivalente do Princípio da Entropia. O conceito de entropia surgiu, assim, como uma necessidade da matemática, mas logo se verificou tratar-se dum conceito muito mais profundo (Boltzmann, Maxwell, Gibbs, etc.). O conceito revelou-se de grande utilidade imediata em vários ramos da Física, na Química - Física, em Meteorologia, em Geofísica, em Engenharia, em Astronomia, em Medicina, etc.

O Princípio da Entropia passou a ser considerado como uma das leis Fundamentais da Natureza, como se depreende da citação de Eddington no início destas lições. A sua aplicação em vários domínios da Ciência e da Tecnologia e a sua extensão a tantos domínios do conhecimento, veio demonstrar a sua universalidade e, ultimamente, a sua aceitação rápida nas Humanidades e na Ciência.

Em 1948, E.T. Shannon utilizou o conceito de entropia num contexto completamente novo. E, ao surgir na Teoria da Comunicação uma expressão matemática que já existia em Termodinâmica, apercebeu-se logo da importância da sua descoberta, como tivemos ocasião de referir e como se depreende das várias citações que fomos fazendo no decurso deste trabalho (II Parte). O conceito de entropia podia, afinal, estender-se a qualquer processo aleatório. E foi dentro desta concepção, que surgiu a Teoria da Informação e que se abriu um mundo novo de aplicações noutros domínios do saber.

Em Matemática o conceito de entropia foi introduzido na Análise, na Teoria dos Números e na Teoria das Probabilidades. Na Lógica foi introduzido na Teoria

de Inferência. Na Linguística está a ser utilizado na análise da Linguagem. Em Economia, há muito que se liga a inflação à entropia e o Segundo Princípio da Entropia está intimamente relacionado com o processo económico.

Fundamentalmente, o processo económico consiste na transformação de produtos de baixa entropia em produtos de entropia mais elevada. Os economistas sabem agora que o desperdício (em matéria e energia) é um resultado inexorável do Segundo Princípio, a que o processo económico se não pode furtar. Produzir mais máquinas e mais automóveis; explorar mais minas; fazer mais vias de comunicação, tudo conduz a uma contaminação do nosso universo complementar o que, em termos de linguagem comum, constitui a poluição do ambiente. A poluição do ambiente é, apenas, uma confirmação da universalidade do Segundo Princípio Fundamental da Termodinâmica: a Lei da Entropia. A gestão do ambiente impõe-nos a todos, deveres e responsabilidades para que o não tornemos inviável para as gerações vindouras. E esses deveres, em princípio, são pautados pelo modo como saberemos perscrutar e administrar a Lei da Entropia.

Do ponto de vista epistemológico a Lei da Entropia introduziu uma das maiores modificações na Física. É o reconhecimento da existência duma mudança qualitativa no Universo.

A fim de consubstanciar as considerações que fizemos, vamos apresentar algumas aplicações fora dos domínios da Física e da Meteorologia (onde "entropia" se lançou, imediatamente com a introdução da temperatura potencial com Sir Napier Shaw, Helmholtz e outros). Para isso, vamos apresentar algumas aplicações à Hidrologia, à Ecologia, à Detecção Remota, à Análise Espectral, à Linguística, à Análise de Imagens e a três grandes sectores industriais.

2. ENTROPIA DUMA REDE HIDROLÓGICA

Como sabemos, a quantidade de precipitação que cai apresenta uma grande variabilidade espacial e temporal. A variabilidade depende do intervalo de tempo que se considera. Assim, por exemplo, a variabilidade das quantidades anuais da precipitação é menor do que a das quantidades mensais ou diárias.

Como se sabe, a quantidade de precipitação avalia-se em litros por metro quadrado ou em mm. A precipitação mede-se com udómetros, que devidamente localizados e distribuídos, constituem uma rede udométrica. Muitas vezes, a rede é excessivamente densa, outras é esparsa de mais. Tanto um caso, como o outro, são de evitar. No primeiro caso a qualidade da informação não aumenta e no segundo caso a informação que se obtém é incompleta e não satisfaz.

O que se pretende é uma rede regular, convenientemente espaçada, e que com o número mínimo de postos de observação permita obter o máximo da informação sobre a quantidade de precipitação caída e a sua distribuição espacial.

Note-se que os udómetros são independentes uns dos outros, e que se excluem mutuamente, no sentido de que a água precipitada e recolhida por um deles não pode ser recolhida por outro qualquer.

Vamos mostrar como se pode utilizar o conceito de entropia no dimensionamento e na densidade de uma rede hidrológica (ou meteorológica).

Seja $f(r)$ o valor da quantidade de precipitação caída no ponto r , e $f(r + l)$ o valor da precipitação caída à distância l de r , no mesmo intervalo de tempo. Convém definir um "conjunto de acontecimentos" $\{f\}$, tal que

$$f^l = f(r) - f(r + l)$$

isto é, cada elemento de $\{f\}$ é constituído pela diferença simultânea de valores da quantidade de precipitação referida ao mesmo período, designadamente de totais diários, mensais, ou outros.

A variável f tem uma distribuição de probabilidade que depende da distância l . Como é que se calcula a "distribuição de probabilidade"? Esta pode fazer-se a partir do histograma da variável aleatória f , que permite obter "frequências relativas", que se tomam por "probabilidades empíricas". Para construir o histograma de f começa-se por fixar os "intervalos de classe", que dependem da natureza da grandeza a estudar. Por exemplo, no caso de valores da quantidade de precipitação mensal, podem tomar-se intervalos de 0 a 150 mm. Para cada classe conta-se o número de casos apropriados e o resultado divide-se pelo número total de ocorrências possíveis. As classes do histograma devem ser as mesmas para todas as distâncias consideradas.

Uma vez determinadas as probabilidades, calcula-se a entropia, ou o grau de incerteza, através da fórmula de Shannon:

$$H(f) = \sum p_i(f) \log p_i(f)$$

em função das distâncias l .

É evidente que para $l = 0$ a função $H(f)$ é nula e à medida que l aumenta a entropia aumenta. De facto, verifica-se que há um aumento de entropia rápido nos primeiros quilómetros, para depois passar a aumentar mais lentamente, com um aumento quase linear.

A entropia varia com a estação do ano que se considera, isto é, com a ordenação dos meses, por exemplo.

Os valores absolutos da entropia H assim obtidos são de pouca utilidade para a determinação da "densidade racional" da rede, porque pode variar abruptamente. Por isso, deve-se utilizar o conceito de "entropia relativa" ou "percentagem de entropia".

Esta pode calcular-se do seguinte modo: determina-se a distância máxima possível, l_{\max} , na região a estudar e, para cada caso, determinam-se os valores de $H(l_{\max})$ que são considerados os valores máximos empíricos dos campos respectivos. Depois define-se uma relação entropia - distância:

$$H(\beta)_1 = 100 \times H(l) / H(l_{\max}), \%$$

Para um dado valor de l esta expressão dá informação sobre a "percentagem de incerteza", da distância, l , em relação à distância máxima, l_{\max} . Este cálculo deve fazer-se para vários valores atribuídos à distância l .

Os resultados assim obtidos, permitem-nos calcular a distância óptima, l_c da rede em função de redução da entropia máxima $H(l_{\max})$ considerada. Em geral, fixa-se essa redução em 50% o que permite obter um valor razoável para l_c e assim

estabelecer uma relação entre o número de estações regularmente dispostas e a redução da entropia.

A introdução do conceito de entropia estatística permite determinar a densidade racional de uma rede meteorológica, mesmo quando o campo meteorológico não é homogêneo, nem isotrópico.

Os valores que se obtêm, com uma rede entropicamente planeada, conduzem a estimativas médias espaciais das quantidades de precipitação muito aceitáveis em estudos hidrológicos, fundamentais para a avaliação dos recursos hídricos.

Num estudo sobre a rede udométrica de S. Miguel (Açores) Rocha Faria, utilizando os dados mensais da precipitação do decénio de 1960- 1970 de 16 estações actualmente existentes, calculou as entropias de pares de estações de precipitação para as várias estações do ano. Verificou que a entropia aumenta rapidamente com a distância nos primeiros quilómetros, para depois aumentar linearmente com uma taxa muito lenta e que a entropia é maior na Primavera e menor no Outono.

Para valores de l_{\max} , dadas as dimensões de S. Miguel arbitrou-se 80 Km, o que permitiu calcular os valores da "percentagem de entropia". É claro que para $H(80 \text{ Km})$ a percentagem é 100%.

Os resultados obtidos mostram, por exemplo, para que haja uma redução de 50% do valor máximo da entropia é necessário que as estações fiquem distanciadas de 6 Km, aproximadamente.

Admitindo que se pretende cobrir S. Miguel com uma rede de udómetros dispostos segundo uma malha de triângulos equiláteros (que é, em geral, das formas mais convenientes para a "cobertura regular" de uma região) de lado l , o número necessário de udómetros variará, para diferentes valores de l . Assim, para $l = 2,5 \text{ Km}$ seriam necessários 147 udómetros, para uma redução de entropia de 80%, enquanto para $l = 5 \text{ Km}$ seriam necessários 42 udómetros, se a redução de entropia fosse 51%. Logo, o aumento do número de estações de 42 para 147 (aumento de 350%) provoca apenas uma redução de 29% na entropia.

Conclui-se, assim, que o aumento indiscriminado do número de postos udométricos não aumenta, substancialmente, a quantidade de informação da rede.

3. A ENTROPIA E A ECOLOGIA: UM EXEMPLO

Vejamos agora um exemplo de aplicação do conceito da entropia à ecologia.

Como se sabe, os efluentes provenientes da actividade agrícola, de esgotos domésticos e industriais, que desaguam em muitos lagos e estuários, são ricos em produtos fosfatados, azotados e noutras matérias orgânicas. Esta abundância em produtos ricos em fósforo e em azoto, nos lagos, conduz à sua *eutroficação* (que em grego quer dizer "bem nutrido"). Na acepção actual, eutroficação significa a deterioração das condições e das propriedades dos lagos e dos estuários (que permitem e suportam a vida natural) causada por uma fertilização excessiva provocada pelos efluentes.

A eutroficação é devida, portanto, a um afluxo abundante de nutrientes de que resulta uma taxa elevada de produção de matéria orgânica vegetal.

As algas e as plantas aquáticas desenvolvem-se rapidamente, alastrando-se a regiões não afectadas, inicialmente. Esta proliferação conduz a uma decomposição da matéria orgânica, com o aparecimento de produtos pouco desejáveis.

A eutroficação reflete-se na composição das espécies, no número de géneros existentes, nas dimensões das populações e na produtividade de vários grupos de organismos do ecossistema aquático.

Se o aumento da produção de matéria orgânica (algas, plantas, etc), no sistema aquático, for excessivo, verifica-se um grande consumo do oxigénio dissolvido cujo teor pode baixar rapidamente, e, por vezes, provocar o seu desaparecimento completo, principalmente nas águas de maior profundidade, que assim, se transformam em "águas mortas". O desaparecimento do oxigénio dissolvido, ou mesmo a sua diminuição, impede a vida animal e conduz ao desaparecimento de muitas espécies piscícolas e outras. De facto, as águas de escoamento, passando através de regiões tratadas com vários fertilizantes,

constituem mananciais de nutrientes, mais ou menos ricos, produzindo floras específicas muito diferentes nas águas dos lagos, com uma grande variedade de algas, de microorganismos e de plantas aquáticas.

Existe, pois, uma correlação entre a natureza dos fertilizantes utilizados e os tipos de flora predominante. Pode, assim, organizar-se um processo indirecto de estudar as consequências da poluição dos vários tipos de fertilizantes. Por isso, têm surgido vários estudos em que se comparam os microorganismos dos lagos eutrofizados, com a natureza dos produtos arrastados. E esses estudos são baseados na fórmula de Shannon.

Seja n_i população da espécie i e N a população total da comunidade. Pode assim definir-se uma probabilidade empírica $p_i = n_i/N$ de ocorrência da espécie i .

O significado ecológico das águas pode avaliar-se em termos da diversidade das espécies de algas e de microorganismos existentes, através da fórmula de Shannon:

$$H = - \sum p_i \log p_i$$

a que os biólogos chamam, por vezes, "índice de diversidade".

O valor de H varia entre zero, para populações unialgais, e um valor próximo da unidade, para comunidades muito diversificadas, visto que H é uma função crescente do número das espécies dominantes. Um "índice de diversidade" de valor um, requer a presença de muitos géneros de algas (superior a 10), correspondendo a uma grande "riqueza" e a uma "distribuição equalitária" das espécies em presença.

A diversidade da comunidade algal diminui com a eutroficação. Os lagos oligotrópicos têm valores da entropia entre 0,7 e 1,0, mas nos lagos eutrofizados os valores de H descem a 0,3. Os "índices de diversidade" variam com a concentração e com a natureza dos nutrientes. Os números de géneros e as suas populações individuais são características dos tipos e teores dos poluentes (fertilizantes, detergentes, pesticidas, etc.) das águas.

O "índice de diversidade", ou a entropia de Shannon, constitui uma forma expedita e adequada para descrever as comunidades algais e é um instrumento muito útil para comparar estes sistemas entre si e, portanto, estimar os efeitos das variáveis ensaiadas (neste caso tipos e teores de fertilizantes ou de detergentes). É interessante notar que a entropia permite avaliar de forma quantitativa a qualidade das águas através de parâmetros biológicos.

Vê-se, pois, que, a entropia de Shannon constitui uma ferramenta preciosa na investigação do ambiente, de que o caso tratado constitui, apenas, um exemplo.

4. ANÁLISE E ENTROPIA DUMA IMAGEM

4.1. ESTATÍSTICAS DUMA IMAGEM

Consideremos a figura 9 duma representação bidimensional. Seja p_{ij} a probabilidade de observar uma certa grandeza na célula (i, j) da matriz $P \times Q$. Notaremos que $\sum \sum p_{ij} = 1$.

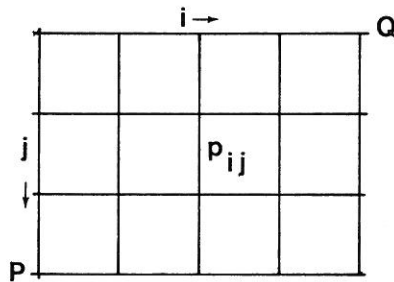


Fig. 9 - Representação bidimensional duma imagem.

Esta "quadrícula" poderia representar uma imagem bidimensional digitalizada. Suponhamos que os dados representam grandezas arbitrárias $U(u)$ e $V(v)$ dos valores fronteira $u_i = \sum p_{ij}$ e de $v_j = \sum p_{ij}$ que correspondem a percursos nas direcções "norte-sul" e "leste-oeste" respectivamente. Além disso, as grandezas u_i e v_j consideram-se independentes, isto é, as estruturas nas duas direcções ignoram-se, e os dados "norte-sul" não interferem com os dados "leste-oeste". Notaremos que, neste caso:

$$p_{ij} = u_i v_j$$

Podemos assim descrever as propriedades estatísticas de primeira ordem do sinal, em termos da função de distribuição de níveis de cinzento. A cada pixel* correspondem 8 bits de informação e, portanto, a digitalização conduz a um máximo de $2^8 = 256$ níveis de cinzento. Com a informação sobre a percentagem de pixels, que têm um determinado nível de cinzento ao longo duma direcção i , ou j , podemos construir histogramas como se faz em estatística elementar. Mais, em vez de considerar os histogramas das duas orientações normais, por classes, os níveis de cinzento, em função das "frequências".

Se for N_i o número de pixels com o nível de cinzento i , "a fracção" p_i de pixels da imagem, cujo nível de cinzento é i , é dada por:

$$p_i = N_i / \sum N_i$$

A partir das características e da forma envolvente do histograma pode construir-se um conjunto de parâmetros estatísticos, que servem para caracterizar a imagem. Assim, introduziremos os seguintes parâmetros:

I) *média*, definida por:

$$\langle i \rangle \equiv \sum p_i i$$

e que constitui um parâmetro global de localização;

II) *desvio médio quadrático*, dado por:

$$\sigma = [\sum p_i (i - \langle i \rangle)^2]^{1/2}$$

e que constitui um indicador de dispersão em relação à média;

III) *assimetria* (skewness) definida por:

$$K_3 = 1/\sigma [\sum p_i (i - \langle i \rangle)^3]^{1/3}$$

e mede assimetria da distribuição em torno da média;

IV) *curtose* determinada a partir da expressão:

$$K_4 = 1/\sigma [\sum p_i (i - \langle i \rangle)^4]^{1/4}$$

e representa o "achatamento", em relação a uma distribuição normal (gaussiana) conceptual, com a mesma média e o mesmo desvio médio quadrático.

(*) pixel - picture element

V) *energia* dada por:

$$E \equiv \sum p_i^2$$

Como se sabe, a soma dos quadrados duma grandeza é igual à soma dos quadrados da amplitude das suas componentes espectrais (Teorema de Parseval). Mas, a última grandeza é interpretada como uma medida da energia, o que justifica o significado da grandeza E anterior.

VI) *entropia* dada pela fórmula de Shannon:

$$H = - \sum p_i \text{Log } p_i$$

A entropia é máxima para distribuições desordenadas ou caóticas, com ausência completa de estrutura; é mínima para distribuições bem ordenadas, com localizações bem definidas e com uma estrutura regular.

A entropia pode ser utilizada como um discriminante na interpretação de mapas temáticos. Assim, podem fixar-se de entropia para obter informação duma imagem, por exemplo, sobre a natureza do ordenamento de culturas. Se tivermos uma região com olival, com soutos de castanheiros e com um carvalhal, não é difícil obter a sua identificação, através dos níveis de entropia que definem os seus limites. O olival, muito ordenado, tem uma entropia menor do que o soto que, em geral é menos ordenado, ao passo que o carvalhal, com uma distribuição quase caótica, tem uma entropia muito maior. O critério da separação depende da fixação dos limites da entropia, que definem a fronteira de cada um dos três tipos de florestação.

4.2. INDICADORES DE CONTINUIDADE E DE ESTRUTURA DUMA IMAGEM

Na análise duma imagem, além dos parâmetros anteriores, devemos considerar outros, que dão conta da uniformidade, ou da continuidade da imagem, como os operadores gradiente e laplaciano e ainda outros, que são indicadores da estrutura, como a transformada de Fourier, etc.

Gradiente, dá a variação espacial duma grandeza escalar que define um campo. Suponhamos que a grandeza escalar é o tom de cinzento f . Então, $\text{grad } f$ é um vector que só o incremento de f num dado ponto ao longo da direcção em que f experimenta a variação máxima e cuja intensidade é a taxa de variação de f ao longo dessa direcção. Se considerarmos a direcção genérica s , a componente do gradiente de f ao longo da direcção s , que se designa por derivada direccional, é dada por:

$$| \text{grad } f |_s \cong [f(s + \Delta s) - f(s)] / \Delta s$$

e mede taxa de variação do nível de cinzento ao longo da direcção s .

Laplaciano: o operador $\text{Lap } f$ é uma medida da diferença entre o valor dum escalar f num dado ponto e o valor médio do mesmo escalar f na vizinhança desse ponto, $\langle f \rangle$, isto é:

$$\text{Lap } f = C [\langle f \rangle - f]$$

em que C é uma constante que depende da geometria da rede do discreto. Ora, no caso bidimensional (x, y) , é:

$$\text{Lap } f \equiv \text{div grad } f \equiv \partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$$

e é fácil obter a expressão anterior, a partir dos desenvolvimentos de $f(x \pm h, y)$ e de $f(x, y \pm h)$.

Em geral, no caso duma imagem, considera-se a vizinhança do ponto central dum pixel, constituída pelos oito pixels vizinhos.

Podemos assim definir o valor médio que é dado por $1/8$ da soma dos f_i dos pixels vizinhos.

Note-se que, se for atribuído ao pixel central o valor médio faz-se a "suavização" da imagem e esbatem-se as heterogeneidades dum campo variável no espaço. O laplaciano permite discriminar regiões de natureza diferente. Em particular faz acentuar os contornos da imagem.

O laplaciano dum campo é uma medida da falta de "harmonia" do campo. Mas harmonia não quer dizer uniformidade. Uma distribuição é harmónica quando o seu laplaciano é nulo, isto é, quando o valor médio do campo duma propriedade nas vizinhanças de um ponto coincide com o valor pontual dessa propriedade.

O operador laplaciano permite realçar os contornos dos limites e fazer a discriminação entre as características temáticas de uma imagem.

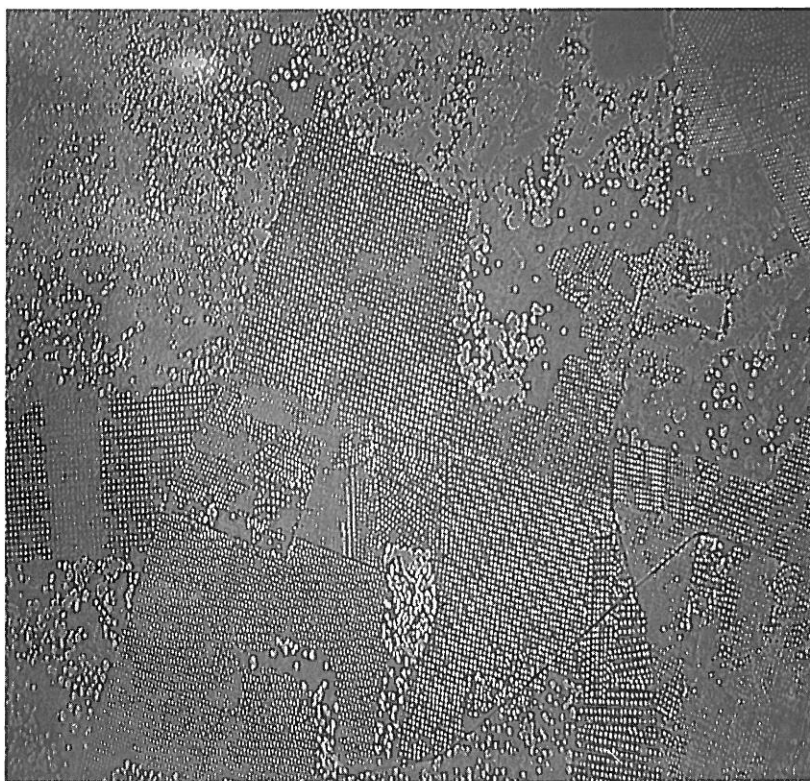


Fig. 10 - O operador laplaciano permite discriminar o olival relativamente ao sobreiral.

Transformada de Fourier permite avaliar a estrutura duma distribuição de uma imagem, no caso presente, bidimensional. A transformada de Fourier duma dada função no domínio espacial x , constitui uma representação no domínio das frequências K . A transformada de Fourier no domínio das frequências, contém exactamente a mesma informação, que a função original. As duas representações só diferem na maneira de apresentar a informação. De facto, a transformada de Fourier torna possível examinar a função de outro ponto de vista - o das frequências (número de ondas). É o mesmo dizer que uma sebe de 1,5 Km é formada por 4.500 arbustos, ou dizer que a densidade da sebe é de 3 arbustos por metro.

A transformada de Fourier de uma função $f(x)$ definida em (x) no domínio do número de ondas (frequência) k , é dada por:

$$F(k) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx \quad (1)$$

Em geral, a transformada de Fourier é uma grandeza complexa que se pode escrever sob as formas equivalentes:

$$F(k) = F_1(k) - i F_2(k) = |F(k)| e^{i\theta(k)}$$

em que $|F(k)|$ é a amplitude ou o espectro de Fourier de $f(x)$ e é dada por $F_1(k)^2 + F_2(k)^2$; e $\theta(k)$ é o ângulo de fase da transformada de Fourier e é dado por $\theta(k) = \arctan F_2(k)/F_1(k)$. É evidente que $i = \sqrt{-1}$.

A transformada inversa é definida por:

$$f(x) = \int_{-\infty}^{+\infty} F(k) e^{ikx} dk \quad (2)$$

e permite reconstituir a função $f(x)$ a partir das componentes espectrais $F(k)$. As equações (1) e (2) constituem um par de Fourier e, formalmente, só diferem no sinal da exponencial. As funções exponenciais de Fourier são fechadas para multiplicação, isto é, da multiplicação de duas funções de base resulta outra função de base, que formam um sistema ortonormado, completo. Esta propriedade simplifica muito a manipulação matemática destas funções.

Suponhamos agora uma função $f(x, y)$ definida em (x, y) . A sua transformada de Fourier no espaço dos kk será:

$$F(k_x, k_y) = \iint f(x, y) e^{-i(k_x x + k_y y)} dx dy$$

em que k_x e k_y são os números de onda (frequência), segundo os eixos dos xx e dos yy .

Sabe-se que uma lente delgada, convergente, tem uma propriedade singular, que é a de gerar transformadas de Fourier entre a distribuição da amplitude da luz paraxial coerente nos planos focais objecto e imagem. Assim, quando um objecto bidimensional está colocado no plano focal objecto e é iluminado por luz coerente paraxial forma-se no plano focal imagem a transformada de Fourier.

Esta propriedade é, por isso, muito útil na análise de distribuições bidimensionais e, em particular, na análise de imagens no domínio das frequências e, portanto, na avaliação da "arquitectura" da imagem em (x, y) .

A representação da transformada de Fourier no plano focal aparece representada por um par de pontos simétricos em relação à origem. O vector vem representado segundo $\pm k_x$ e $\pm k_y$. O seu módulo constitui uma medida da

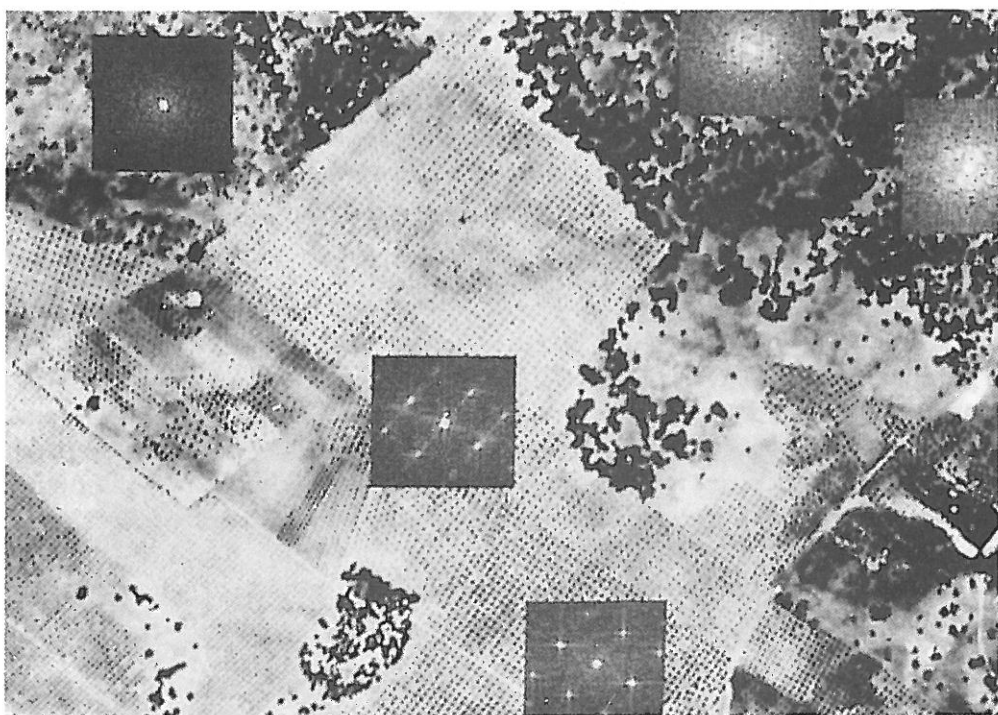


Fig. 11 - A transformada de Fourier permite avaliar o grau de ordenamento dum olival e de um sobreiral.

probabilidade duma distribuição com as frequências $|k_x|$ e $|k_y|$ ao longo da direcção correspondente.

A uma distribuição ordenada em (x, y) , isto é, de baixa entropia, corresponde um número limitado e, em geral, reduzido de componentes no domínio transformado (k_x, k_y) e com orientações bem definidas.

Uma distribuição desordenada, isto é, de entropia elevada, mais ou menos caótica, em (x, y) , representa várias transformadas de Fourier constituídas por um feixe de vectores com uma distribuição circular, com intensidades que enchem toda a vizinhança da origem no plano (k_x, k_y) formando como que uma rosácea (ver figura 11).

Se, por outro lado, a forma for uma elipse, então há orientações privilegiadas e a distribuição em (x, y) apresenta algum ordenamento ao longo dos eixos da elipse. No caso limite do eixo menor tender para zero, cai-se no caso anteriormente descrito.

5. O MÉTODO DA ENTROPIA MÁXIMA (MEM)

5.1. FORMALISMO DA ENTROPIA MÁXIMA

O método da entropia máxima (MEM) constitui uma técnica poderosa para o processamento e a análise de dados. As aplicações vão desde a recuperação de imagens à teoria da inferência estatística; desde a análise espectral de séries temporais de observações, ao seu emprego em medicina, como na tomografia axial computurizada.

Neste parágrafo apresentam-se, apenas, alguns aspectos da utilização do método e dão-se alguns exemplos de aplicação.

A generalidade do método assenta na possibilidade de obter uma estimativa da distribuição da probabilidade a partir de dados bastante incompletos, ou afectados de erros. O método da entropia máxima escolhe sempre o resultado mais simples, partindo do mínimo da estrutura possível com a única condição de satisfazer os dados observados e desprezando pormenores espúrios. Assenta, apenas, na verdade que se conhece: os dados das observações, o conhecimento físico contido no enunciado do problema e as leis das probabilidades.

A situação é semelhante à de um doente que vai ao médico e lhe transmite os sintomas que sente. O médico não começa a pensar em sintomas que o doente, por ventura, poderia ter, mas que não tem.

O que ele faz é analisar os sintomas que existem, de facto, e inferir as causas que os provocam. E, assim, baseado em factos e na "sua ciência", vai procurar identificar as várias perturbações possíveis que podem causar a doença, revelada pelos sintomas, começando por aquele que é o mais provável de acordo com a história médica do doente.

O princípio é sempre o mesmo: maximiza-se a entropia $H = - \sum p_i \log p_i$, sujeita apenas às restrições impostas pelos dados e às condições ditas pelas

leis da Teoria das Probabilidades. Para a utilização do método, as probabilidades devem ser substituídas pelas "probabilidades físicas", avaliadas pelas frequências.

No § 5 da II Parte, apresentámos o fundamento do Método da Entropia Máxima (Jaynes, 1978) e analisámos, com pormenor, os seus aspectos matemáticos. Por isso, agora, faremos apenas uma síntese.

Suponhamos uma grandeza x , que pode assumir os valores $\{x_1 \dots x_n\}$ a que corresponde a distribuição de probabilidade $\{p_1 \dots p_n\}$, de forma a representar informação parcial que dispomos, acerca de x .

A fórmula de Shannon

$$H = - \sum p_i \log p_i \quad (5.1)$$

dá uma medida da incerteza (ou entropia) contida nos acontecimentos, cuja distribuição de probabilidade deveria ser $\{p_1 \dots p_n\}$. A distribuição que descreve de forma "mais isenta" o que sabemos, sem introduzir quaisquer hipóteses, ou elementos estranhos, é aquela que maximiza H , sujeita aos constrangimentos impostos pelos dados das observações.

Consideremos então m funções $\{A_1(x) \dots A_m(x)\}$ de que, segundo "o enunciado do problema", se conhecem valores, representados por um conjunto de números $\{A'_1 \dots A'_m\}$. A partir destes valores, pretendemos determinar, ou prever, a distribuição geral da probabilidade $\{p_1 \dots p_n\}$ a que satisfazem as funções $\{A_1(x) \dots A_m(x)\}$, sem introduzir quaisquer hipóteses suplementares. Para ajustar a distribuição aos "dados" de que se dispõe, tem que se verificar os m constrangimentos simultâneos,

$$\sum p_i A_k(x_i) = A'_k \quad (k = 1 \dots m) \quad (5.2)$$

na maximização da entropia $H = - \sum p_i \log p_i$

A solução, que vamos apresentar, é uma duplicação da análise feita no § 5, II Parte. Começamos por considerar a função de partição, $Z(\lambda)$:

$$Z(\lambda_1 \dots \lambda_m) = \sum \exp [- \lambda_1 A_1(x_i) - \dots - \lambda_m A_m(x_i)] \quad (5.3)$$

A distribuição que corresponde ao máximo da entropia é, como já sabemos, dada por:

$$p_i = 1/Z(\lambda_1 \dots \lambda_m) \cdot \exp [- \lambda_1 A_1(x_i) - \dots - \lambda_m A_m(x_i)] \quad (5.4)$$

em que os factores indeterminados de Lagrange, λ_k , se podem obter a partir das expressões:

$$A'_k = - \partial \text{Log} Z / \partial A_k \quad (k = 1 \dots m) \quad (5.5)$$

Estas expressões conduzem a um sistema de m equações, às incógnitas $\{\lambda_1 \dots \lambda_m\}$, cujos valores, assim, se podem determinar.

Podemos então, calcular $Z(\lambda)$, a partir de (5.3) e a distribuição de p_i , recorrendo à expressão (5.4).

Algumas vezes, o método da entropia máxima (MEM) é apresentado de forma alternativa, recorrendo ao formalismo da análise combinatorial (§ 3.3, I Parte).

Então, a partir dos acontecimentos em que a variável se observa N_i vezes, x_i , pode definir-se a frequência $f_i = N_i/N$. Uma dada distribuição de frequências $\{f_1 \dots f_n\}$ tem uma multiplicidade $W(f)$ dada por:

$$W(f) = N! / (Nf_1)! \dots (Nf_n)! \quad (5.6)$$

que corresponderia ao número de modos que se observariam em condições normais. Para N suficientemente grande, pode recorrer-se à fórmula de Stirling, e vem (§ 3.3, I Parte):

$$1/N \cdot \text{Log } W(f) = - \sum f_i \text{Log } f_i = H(f) \quad (5.7)$$

ou seja:

$$W(f) = \exp NH(f) \quad (5.7a)$$

A igualdade (5.7) é a fórmula de Shannon, mas agora dependendo das "probabilidades físicas" ou objectivas, f_i , (frequências), em vez das probabilidades subjectivas p_i .

Não se conhece a sequência específica $\{x\}_i$ que foi gerada. O que conhecemos são os valores médios de m grandezas $\{A_1(x) \dots A_m(x)\}$:

$$\sum f_i A_k(x_i) = A'_k \quad (k = 1 \dots m) \quad (5.8)$$

O formalismo da análise combinatorial, conduz a uma lei exponencial $W = \exp(NH)$ da entropia, quando N for suficientemente grande (5.7 a).

Logo, o máximo da entropia corresponde ao máximo da multiplicidade que, por seu turno, se pode identificar com a probabilidade máxima.

5.2. A ENTROPIA DUMA DISTRIBUIÇÃO NORMAL DE PROBABILIDADE

Vamos, a título de exercício, calcular a entropia duma distribuição normal de probabilidade (distribuição gaussiana). Nestas circunstâncias,

$$p(x) = 1 / \sqrt{2 \pi \sigma} \cdot e^{-x^2/2 \sigma^2}$$

Logo, $H(x)$ será dada por:

$$H(x) = - \int p(x) \log p(x) dx$$

Ora:

$$- \log p(x) = \log \sqrt{2 \pi \sigma} + x^2 / 2 \sigma^2$$

e, portanto:

$$H(x) = \int p(x) \log \sqrt{2 \pi \sigma} dx + \int p(x) x^2 / 2 \sigma^2 dx$$

$$= \log \sqrt{2 \pi \sigma} + \sigma^2 / 2 \sigma^2 = \log \sqrt{2 \pi \sigma} + \log \sqrt{e}$$

$$= \log \sqrt{2 \pi e \sigma} = \log \sigma + \text{const.}$$

Esta expressão mostra que a entropia duma distribuição normal é, a menos duma constante, dada pelo logaritmo do desvio quadrático. Este resultado confirma a ideia, várias vezes expressa, de que a entropia é uma medida da dispersão.

Na dedução admitiu-se que se trata duma distribuição estacionária, em que, a variável flutua em torno da média, com uma variância constante.

Este resultdo é um caso particular duma relação mais geral, que dá uma relação entre a entropia H e a densidade espectral $P(f)$ num processo gaussiano estacionário:

$$H = 1/4 f_N \int \text{Log } P(f) df$$

em que

$$f_N = 1/2 \Delta$$

é a frequência de Nyquist, que é metade da frequência de discretização, Δ ; em radianos será $f_N = 2 \pi / 2 \Delta t = \pi / \Delta t$. A frequência de Nyquist é a frequência mais elevada com significado físico, que é resolúvel pela série $\{x_t\}$.

∴

Vamos, agora, demonstrar que a *distribuição normal de probabilidade corresponde ao valor máximo da entropia duma variável aleatória $\{x\}$, cujo valor médio é nulo e cuja variância é constante*.

Então vamos determinar a distribuição de probabilidade, $p(x)$, que corresponde ao máximo de $H(x)$:

$$H(x) = - \int p(x) \text{Log } p(x) dx$$

sujeita às condições

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

e

$$\int p(x) x^2 dx \equiv \sigma^2 = \text{const.}$$

Utilizando o método dos multiplicadores de Lagrange, tem que se maximizar a expressão:

$$\int [- p(x) \text{Log } p(x) + \lambda_1 p(x) x^2 + \lambda_2 p(x)] dx \quad (1)$$

Vamos recorrer ao Cálculo das Variações. Para que a variação da expressão anterior seja nula, tem que se verificar a equação de Euler-Lagrange^(*)

No nosso caso a função F da equação Euler é a função integrante de (1) e $y=p(x)$.

Somos assim conduzidos à equação:

$$-1 - p(x) + \lambda_1 x^2 + \lambda_2 = 0$$

Depois de alguns ajustamentos obtém-se para a probabilidade $p(x)$ uma expressão da forma:

$$p(x) = 1/\sqrt{2\pi}\sigma \cdot e^{-x^2/2\sigma^2}$$

Fica assim demonstrado que a distribuição normal é a distribuição de probabilidade que corresponde ao valor máximo da entropia numa série aleatória, cujo valor médio é nulo e a variância é constante.

^(*) Para uma função $F(x,y)$ assume a forma:

$$\partial F/\partial y - d/dx (\partial F/\partial y') = 0$$

em que $y' = \frac{dy}{dx}$

5.3. RECONSTITUIÇÃO DE IMAGENS

Suponhamos que se pretende reconstituir uma fotografia de que se tem apenas um exemplar deteriorado, ou um exemplar desfocado.

Suponhamos ainda que a fotografia está digitalizada. Sendo assim, vamos supor que:

f_i representa a intensidade da imagem na célula que queremos determinar;

D_i a intensidade dos dados observacionais na célula i ;

σ o desvio médio quadrático dos dados, devido ao ruído.

A imagem é constituída por N células e i é um índice bidimensional que cobre as N células.

Seja b_j a função que caracteriza a deterioração da câmara ou da instrumentação, que dá a intensidade na célula j . Então os valores expectáveis que deveriam ter sido produzidos pela imagem f são:

$$F_i = \sum b_j f_{i:j} = (f * b)_i$$

em que a asterisco (*) representa a operação de *convolução*.

Para avaliar a diferença entre os dados de que dispomos, D , e os que deveríamos ter se não houvesse ruído usa-se como estatística de prova o teste de qui quadrado (χ^2), definido por:

$$\chi^2 = \sum (F_i - D_i)^2 / \sigma_i^2$$

cuja lei de distribuição se conhece em função do número de observações independentes e do nível de confiança desejado.

Depois maximiza-se a entropia

$$H(x) = - \int p(x) \log p(x) dx$$

ajustando-a à invariância de χ^2 .

A lei de distribuição de probabilidade será:

$$p = \exp N [H(F) - WQ(F)]$$

em que $W = 1/N \sigma^2$

$$e \quad Q(F) = 1/2 \sum [D_k - \sum A_{ki} f_i]^2$$

O factor $\exp(NH)$ representa a informação anterior, acerca das multiplicidades das diferentes "cenas"; o factor $\exp(-N W Q)$, na sua dependência em $\{f_1 \dots f_n\}$ é a "função de verosimilhança" que nos diz o que aprendemos dos dados disponíveis.

A cena mais provável $\{f_1, \dots, f_n\}$ corresponde ao máximo de $H - W Q$ sujeito à condição $\sum f_i = 1$. E portanto,

$$f_i \cong \exp \left[- W \frac{\partial Q}{\partial f_i} \right]$$

A cena mais provável é encontrada através da maximização da entropia configuracional sem qualquer cometimento a dados espúreos que resultem de conjecturas por mais plausíveis que possam parecer.

5.4. ANÁLISE ESPECTRAL DE SÉRIES TEMPORAIS

5.4.1. O ESPECTRO DE DENSIDADE ESPECTRAL E O MÉTODO DA ENTROPIA MÁXIMA

Em muitas aplicações é indispensável obter estimativas do espectro de potência duma série temporal discreta de que se conhece apenas um troço limitado. É um caso muito frequente em Climatologia, em Geofísica, em Oceanografia, em Hidrologia, em Econometria, etc. As séries de que se dispõe são de duração curta e no mundo real não existem séries infinitamente longas (M. Almeida, 1979).

A análise espectral das séries tem sido baseada na transformada de Fourier. Mas, a aplicação simples da análise de Fourier exige o conhecimento das funções em todo o seu domínio, condição que as séries finitas de que se dispõe não têm. Por isso, há necessidade de introduzir hipóteses que correspondem a fornecer informação complementar, que os dados das séries não contêm. As estimativas espectrais vêm, portanto, envolvidas numa certa incerteza. Basta notar que as determinações espectrais variam com a dimensão da amostra.

Há vários métodos que procuram ultrapassar estas limitações. O cálculo do espectro de densidade de potência pode fazer-se a partir dos valores discretizados da função de autocorrelação, tendo em vista o *teorema de Wiener* que diz que a *função de autocorrelação e o espectro formam um par de Fourier*.

Mas, a análise de Fourier está sujeita a determinadas limitações, sobretudo, quando aplicada a séries temporais. As séries infinitas ou com continuidade cíclica, a que se aplica a análise de Fourier têm que ser adaptadas às amostras disponíveis, evidentemente finitas. É, por isso, que os procedimentos clássicos, tradicionais, assentam em hipóteses sobre o comportamento dos dados fora da amostra. Num caso admite-se, que os dados fora da amostra, se repetem

periódica e indefinidamente nos dois sentidos. Noutro caso admite-se, que fora do troço da série, os valores são identicamente nulos. A série disponível seria portanto uma série truncada que iguala a série original nos pontos onde esta é conhecida e é igual a zero fora deles. Mas a truncatura origina um fenómeno característico das séries de Fourier, conhecido por fenómeno de Gibbs, que se traduz por uma oscilação e uma sobrevalorização das intensidades nas bandas das frequências correspondentes aos pontos de truncatura.

Outra restrição das séries temporais é inerente ao processo de discretização, o que limita as frequências com significado físico. A frequência permissível, mais elevada, a frequência de Nyquist, é dada, como já se referiu, pelo inverso do dobro do intervalo de discretização: $f_N = 1/2 \Delta t$ em s^{-1} ou $f_N = \pi/\Delta t$ em radianos/s. Esta é a frequência mais elevada que é resolúvel pela série temporal $\{x_i(t)\}$. Para frequências superiores a f_N verifica-se uma autorepetição espúria, sem significado físico.

Para obviar as limitações expostas e minimizar os efeitos da truncatura na análise espectral, Blackman e Tuckey (1957) usaram uma técnica que consistia em afectar os dados por "funções - peso" que no domínio das frequências são transformadas em "janelas espectrais". Torna-se assim possível obter estimativas espectrais mais aproximadas e com maior precisão.

Vejamos agora como é que o método da entropia máxima permite resolver algumas das limitações impostas pelos métodos convencionais na determinação da densidade espectral de potência.

Todos os métodos utilizados na análise espectral de dados experimentais, incluindo o da entropia máxima, estendem a todo o domínio, duma forma ou doutra, a informação obtida numa amostra finita. Ora, a superioridade do método da entropia máxima, reside no facto da extrapolação ser o mais imparcial possível.

De facto, o que se pretende é um método que seja, em primeiro lugar, consistente com a informação prévia disponível; e em segundo lugar, que permita obter uma lei de distribuição de probabilidade que descreva a informação prévia existente, sem fazer quaisquer suposições e que seja o mais isenta possível.

Essa distribuição é a que corresponde à entropia máxima permitida pela informação objectiva disponível. Logo, admitimos que fora da série finita esta é prolongada artificialmente por um processo aleatório estacionário em que a média e a variância são constantes e iguais aos valores correspondentes do troço conhecido. Doutra maneira, a extrapolação é feita por um sinal de "ruído branco". Esta extrapolação é muito mais razoável do que a que assentava na repetição periódica dos dados, ou a da truncatura nos extremos da série (completada com "zeros" a montante e a jusante).

5.4.2. FUNDAMENTO E APLICAÇÃO DO MÉTODO

Consideremos uma série temporal $\{x_i(t)\} \equiv \{x_0, x_1, \dots, x_N\}$, que para facilitar se supõe ser de média nula e de variância constante. A série $\{x_i(t)\}$, com um intervalo de discretização Δt , resulta dum processo aleatório, a que está associada uma dada distribuição de probabilidade $p(x)$.

A hipótese fundamental da análise espectral, com base no princípio da entropia máxima, é a de aceitar que o processo que gera a série temporal é o mais aleatório e o menos previsível possível, mas que é consistente com constrangimentos conhecidos. Neste caso, os constrangimentos são os valores da função de autocorrelação R_k definidos por:

$$R_k = \sum x_j x_{k+j} \quad (j = 0, \pm 1, \dots, \pm N) \quad (1)$$

que se podem calcular a partir dos dados de uma amostra da série para vários desfasamentos.

O princípio do método é o de aceitar que o espectro de potência deve ser o que tem a entropia máxima de entre todos os espectros possíveis que satisfazem aos mesmos constrangimentos. Neste sentido, este método não implica quaisquer hipóteses, nem impõe quaisquer restrições à extensão e à extrapolação dos dados da função da autocorrelação fora do intervalo da amostra.

A função de autocorrelação, R_n e o espectro de potência $P(f)$ constituem, como dissemos, uma par de Fourier, isto é:

$$R(n) = \int_{-f_N}^{+f_N} P(f) \exp 2 \pi i f n \Delta t \cdot dt \quad (2a)$$

$$P(f) = \frac{1}{2f_N} \sum_{-\infty}^{+\infty} R(n) \exp - 2 \pi i f n \Delta t \quad (2b)$$

em que Δt é o intervalo de discretização, $f_N = 1/2 \Delta t$ é a frequência de Nyquist e a sequência $\{ R(n) \}$ é constituída pelas funções de autocorrelação para vários desfasamentos $n \Delta t$.

Por outro lado, sabemos que a entropia pode ser expressa em termos do espectro de potência pela fórmula:

$$S \equiv \int_{-f_N}^{+f_N} \text{Log } P(f) \cdot df \quad (3)$$

Vemos assim que $P(f)$ surge como um elemento de ligação entre a entropia S e a função de autocorrelação $R(n)$. O problema que nos propusemos resolver reduz-se, portanto, a determinar o espectro de potência $P(f)$ que maximiza a entropia S dada pela expressão (3), sujeita aos constrangimentos impostos pelos valores de $R(n)$ expressa por (2a) seguindo o tratamento geral. Burg (1967) desenvolveu um algoritmo que evita o cálculo directo dos R_k para obter $P(f)$.

∴

As equações anteriores são exactas para N suficientemente grande e sem "efeitos nos limites"; se necessário, pode admitir-se que $N \rightarrow \infty$.

A dificuldade reside em que a série é limitada e nem todos os R_k são conhecidos. Os dados reais de que dispomos só nos permitem obter um subconjunto $R'_{-m} \dots R'_m$ em que $m < N$.

O nosso problema consiste portanto em estimar $P(f)$ desta informação incompleta, que representaremos por $P(f)$.

Consideremos, então, o problema do ponto de vista da entropia máxima. Os dados de que dispomos são:

$$\{ R'_{-m} \dots R'_0 \dots R'_m \}$$

Representam em notação real $m+1$ constrangimentos, visto que $R_{-k} = R_k$.

A distribuição, que corresponde à entropia máxima e está sujeita aos constrangimentos R_k é, como se viu, dada por:

$$p(x_0 \dots x_N) = \exp [-\sum \lambda_k R_k] \quad (4)$$

Escrevendo por extenso o expoente, temos:

$$\sum \lambda_k R_k = \lambda_0 (x_0^2 + x_1^2 + \dots + x_N^2) + \lambda_1 (x_0 x_1 + x_1 x_2 + \dots + x_{N-1} x_N) +$$

$$\begin{aligned}
& + \lambda_2 (x_0 x_2 + x_1 x_3 + \dots + x_{N-2} x_N) + \dots + \\
& + \lambda_m (x_0 x_m + \dots + x_{N-m} x_N) \\
& + \lambda_{-1} (x_1 x_0 + x_2 x_1 + \dots + x_N x_{N-1}) + \dots + \\
& + \lambda_{-m} (x_m x_0 + \dots + x_N x_{N-m})
\end{aligned} \quad (5)$$

Ora, estes produtos e estas somas podem ser organizados e escritos sob a forma dum produto de matrizes:

$$\sum \lambda_k R_k = \sum \Lambda_{ij} x_i x_j$$

em que Λ_{ij} é uma matriz de Toeplitz dos λ_k :

$$\Lambda_{ij} = \begin{cases} \lambda_{j-i} & \text{se } |j-i| \leq m \\ 0 & \text{se } |j-i| > m \end{cases}$$

em que os multiplicadores de Lagrange se repetem em bandas paralelas à diagonal principal. Portanto:

$$p(x_0, \dots, x_N) \propto \exp [-\sum \Lambda_{ij} x_i x_j]$$

A lei de distribuição de probabilidade, resultante da aplicação da entropia máxima é, afinal, uma lei gaussiana, que, assim, surge como a lei de distribuição que representa de forma mais isenta os dados disponíveis.

A função de partição Z é proporcional a $[\det \Lambda]^{1/2}$, por isso;

$$\text{Log } Z = -1/2 \sum \text{Log } e_j + \text{const.}$$

em que os $\{e_j\}$ são os valores próprios de $[\Lambda]$.

Os valores de λ_k podem obter-se a partir da solução do sistema de equações (5). No limite de $N \rightarrow \infty$ os valores próprios podem obter-se a partir da teoria das matrizes de Toeplitz.

O algoritmo de *Burg* procede através duma factorização de *Wiener - Hopf* do polinómio

$$\sum \lambda_k Z_k = \left| \sum a_k 2^k \right|^2$$

e o cálculo dos coeficientes $\{a_k\}$ é feito de forma recursiva. Estes coeficientes podem interpretar-se como um filtro de predição de *Wiener*. É evidente que, do ponto de vista do método de entropia máxima, não é indispensável que se calculem os $\{a_k\}$; o que é indispensável é o cálculo dos λ_k .

A partir da expressão de $\hat{P}(f)$ pode usar-se a extrapolação óptima $\hat{R}_{m+1} \dots \hat{R}_N$ da covariância, para além dos limites dos nossos dados, recorrendo à lei da distribuição de probabilidade, estabelecida anteriormente.

Depois de algumas manipulações matemáticas obtém-se o resultado, hoje clássico de *Burg* (1967). Para o valor estimado do espectro de potência:

$$\hat{P}(f) = 1/2 f_N \cdot 1 / \sum \lambda_k e^{2 \pi i f k \Delta t}$$

ou

$$\hat{P}(f) = 1/2 f_N \cdot 1 / \left| \sum a_k e^{2 \pi i f k \Delta t} \right|^2$$

Pode assim obter-se uma estimativa do espectro de potência $P(f)$ utilizando apenas a informação contida nos dados de que dispomos e que corresponde à estimativa óptima.

6. A ENTROPIA NA LINGUAGEM

6.1. CARACTERIZAÇÃO ESTATÍSTICA DA LÍNGUA

Tem havido sempre uma preocupação, mais ou menos acentuada, pela caracterização estatística das línguas. Basta atentar nas notícias da diferenciação entre as linguagens dos diferentes povos, de que se fazem eco os primeiros documentos de linguagem escrita e que se assinalam nos grandes livros de todas as religiões e, em especial, na *Bíblia*.

Estes estudos tomaram, recentemente, um grande incremento, devido ao desenvolvimento da Teoria da Informação. Desde a descoberta do telégrafo que se têm verificado progressos no tratamento matemático das comunicações e que culminaram, como dissemos, com a formulação da Teoria de Comunicação e, depois, com a Teoria da Informação. O "alfabeto" de Morse não foi escolhido ao acaso, mas depois de uma profunda reflexão sobre a natureza da Língua Inglesa. Não é por acidente que o E, a letra mais frequente na Língua Inglesa, é representado apenas por um ponto (.). Mas, muitas das propriedades estatísticas da língua foram aflorando com a experiência. Todos nos lembramos de que as divisórias das caixas "do tipo" dos tipógrafos tinham dimensões diferentes, que variavam na razão inversa da frequência de utilização dos vários tipos na composição. No ensino infantil são muito usados os "jogos com letras" em que as crianças vão aprendendo a fixá-las e que se baseiam nas frequências com que aparecem.

Ficou célebre a série de conferências de divulgação que Sir Arthur Eddington fez na Universidade de Edimburgo, em 1927, sobre a Mecânica Estatística. Foram as célebres "Gifford Lectures".

Eddington estava a analisar problemas ligados à termodinâmica, a moléculas, e probabilidades, quando trouxe, a título de ilustração, o problema dos macacos dactilógrafos, que hoje constitui o célebre "problema de Eddington":

"if an army of monkeys were strumming on typewriters they might write all the books in the British Museum".

("Se um exército de macacos batesse os teclados de máquinas de escrever aqueles poderiam dactilografar todos os livros da Biblioteca do Museu Britânico"):

Sem querer, Eddington estava a abrir caminho à análise entrópica e quantitativa do problema da linguagem. Numa simulação matemática, os macacos e as máquinas de escrever podem ser substituídos (e com vantagem!) por uma fonte geradora de números aleatórios, com uma correspondência bijectiva entre os inteiros e o teclado das máquinas (letras, símbolos, pontuação, etc.).

As técnicas estatísticas para a determinação da entropia estão, presentemente a ser aplicadas à linguagem escrita e com grande proveito. A entropia, neste contexto, é o parâmetro estatístico que mede o quanto de inesperado contém uma fonte de mensagens, qualquer que seja a sua natureza. Uma página impressa pode ser analisada no que se refere à ordem ou à desordem do texto. Ora, uma língua tem uma estrutura bem definida. As palavras são constituídas por um certo ordenamento dos símbolos, para formarem as sílabas, os ditongos, etc ... Aquelas são ainda separadas umas das outras por espaços e organizadas, segundo normas ("gramática"), para constituírem as frases.

Surge, assim, na caracterização estatística de uma língua, a necessidade de calcular a probabilidade de ocorrência de várias ordens de sequências coerentes e lógicas de letras, ou de símbolos, em relação ao número total de arranjos aleatórios, da mesma ordem, que se poderiam formar com a mesma colecção de letras e de símbolos. Este cálculo pode ser expresso em termos de negantropia, ou de redução de entropia, que resulta duma selecção inteligente da colecção aleatória das letras de que se dispõe. As entropias que se usam nestes cálculos, ao contrário da entropia termodinâmica, são definidas como medidas matemáticas, sem dimensões, dessas probabilidades.

Antes de entrar no cálculo destas determinações vamos considerar as várias ordens de estatísticas da linguagem, neste caso, escrita.

6.2. A APROXIMAÇÃO DE PRIMEIRA ORDEM

Começemos por fixar o número de caracteres necessários para a linguagem escrita. Deveremos contar com o alfabeto (23 letras) mais o k, o y e o w, que também contam na linguagem científica, o que dá ao todo 26 caracteres. Mas já se pensou na dificuldade que seria, se as palavras não fossem separadas, isto é, se não houvesse tipos para os espaços entre elas? Vejamos como seria:

"é só experimentar para ver como seria difícil compreender o que a quise escrever".

À semelhança do que sucede com a língua inglesa, em que estes estudos são muito abundantes, vamos considerar um "alfabeto alargado" com 28 símbolos, incluindo o espaço e o apóstrofo ('). Em rigor, deveríamos considerar além deste "alfabeto", os números dígitos, as letras de caixa alta, o tipo itálico, os sinais de pontuação, o travessão, os acentos, etc.. Seriam, ao todo, mais de 100 caracteres.

Suponhamos, no entanto, que nos detemos no nosso alfabeto alargado de 28 caracteres. Com as 28 possibilidades, se todas fossem igualmente prováveis e independentes, a probabilidade de cada símbolo seria $1/28$, a que corresponderia uma informação por símbolo de:

$$h_0 = \text{Log}_2 28$$

Esta seria a aproximação de ordem zero, que se obtém atribuindo a cada símbolo a mesma probabilidade, e a mais desfavorável possível.

Voltaremos a este ponto mais tarde.

Ora, nós sabemos, que numa língua há letras mais frequentes que outras e que, devido á estrutura da língua, as letras não se agrupam indistintamente. Por isso, devemos considerar momentos estatísticos de ordem mais elevada.

No Quadro I, apresentam-se os valores das frequências reais das línguas inglesa e portuguesa, baseadas no alfabeto alargado A, B, C ... X, Y, Z a que correspondem, respectivamente, os números 1, 2, 3 ... 26, 27, 28. As frequências da língua inglesa são baseadas em Hamlet e as da língua portuguesa são baseadas na análise da obra Literatura no Brasil de Afrânio Coutinho, 1969.

QUADRO I

Frequências referidas a 1000 caracteres

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Inglês	58	12	17	31	93	18	14	50	49	1	7	35	25	49
Português	106	3	41	46	102	8	10	4	66	1	0	24	37	50

	O	P	Q	R	S	T	U	V	W	X	Y	Z	""	,
Inglês	73	12	1	45	53	73	29	2	20	1	22	0	197	6
Português	99	23	5	54	62	44	28	12	0	1	0	3	163	0

Deste quadro conclui-se que as frequências do inglês e do português são, substancialmente, diferentes. Há, no entanto, um resultado semelhante: o espaço é o tipo mais frequente em ambas as línguas (197 e 163 por 1000, respectivamente). As palavras inglesas são mais curtas do que as portuguesas.

Mas comentemos o que se passa com o Português.

Como se vê, as probabilidades dos diferentes símbolos não são iguais. Por isso, a aproximação de ordem zero, que atribui a todos os símbolos a mesma probabilidade, é a mais desfavorável possível, ainda que seja a mais isenta de pressupostos. É a aproximação de entropia máxima.

A probabilidade $P(I)$ do símbolo, I , é dada, no caso geral, por:

$$P(I) = M(I) / \sum M(I)$$

em que $M(I)$ designa a ocorrência do símbolo I e $\sum M(I)$ o número total de possibilidades. No nosso caso $\sum M(I) = 1000$ caracteres.

O espaço é, de longe, o carácter de maior probabilidade (0,163); logo a seguir vem o A (0,106) e depois, por ordem decrescente, o E, O, I, S, R, N, D, T, C, M, U, L, P, V, G, F, H, B, X, Z.

Se dividirmos o número total de caracteres (1000) pelo número de espaços (número de palavras), obtemos o número médio de $6,02 - 1 = 5,02$ letras por

palavra (subtraímos 1, porque incluímos o símbolo "espaço" no alfabeto alargado). Notaremos que em Inglês as palavras são mais curtas, porque, em média, tem 4 letras por palavra.

A aproximação de primeira ordem na determinação da entropia, por carácter, h_1 , assenta nas frequências reais dos caracteres individuais, que figuram no QUADRO I.

6.3. MATRIZES DE CORRELAÇÃO

A aproximação de segunda ordem começa a introduzir a estrutura da língua. Depois de escolher uma letra, a outra é seleccionada de acordo com as frequências com que as letras se seguem na linguagem escrita real. Ora, em termos de linguagem matemática, passamos, de facto, a lidar com matrizes de covariância ou de correlação, como se vai mostrar.

Vamos, então estabelecer as matrizes de correlação entre os vários símbolos, que contêm as propriedades estatísticas da linguagem. O problema pode pôr-se assim. Vamos analisar o número de vezes em que o carácter I é seguido pelo carácter J, depois pelo símbolo K, etc.. Esta análise pode fazer-se a partir da consideração da matriz multidimensional, cujo elemento genérico é

$$M(I, J, K, L, \dots) \quad (**)$$

Claro que o problema, posto com esta generalidade, levar-nos-ia a matrizes de ordem cada vez mais elevada, de que se perderia o significado. Estas matrizes contêm informação estatística de ordem diferente e estão relacionadas entre si. Assim,

$$M = \sum_I M(I) = \sum_{I, J} M(I, J) = \sum_{I, J, K} M(I, J, K \dots)$$

A matriz de ordem zero é a matriz com o número total de "acontecimentos".

$$M = \sum_I M(I)$$

(**)Fundamentalmente trata-se de um problema de probabilidades condicionadas (Ver Apêndice II)

A matriz de primeira ordem é a matriz coluna, que contém as frequências de ocorrência:

$$M(I) = \sum_J M(I, J)$$

A matriz de segunda ordem, que dá as correlações entre pares de caracteres, pode determinar-se a partir da matriz de terceira ordem:

$$M(I, J) = \sum_K M(I, J, K)$$

A matriz de segunda ordem é a mais usual e muitas vezes conduz, na representação gráfica, ao diagrama de Shannon, ou a um diagrama de dispersão (estatística). Analogamente, podemos definir as matrizes de probabilidades. Assim:

$$P(I) = M(I) / \sum M(I) = M(I) / M$$

representa a probabilidade de ocorrência do símbolo genérico, I;

$$P(I, J) = M(I, J) / M(I)$$

dá o elemento genérico da matriz de probabilidade, que, como se vê, se pode obter da matriz de correlação. Representa a probabilidade de que o carácter J se siga ao carácter I. Por outro lado,

$$P(I, J, K) = M(I, J, K) / M(I, J)$$

representa a probabilidade de que o carácter K se siga à sequência (I, J), etc..

Em cada língua as probabilidades de diferentes ordens têm valores bem definidos.

Já fizemos algumas considerações sobre a aproximação de primeira ordem (frequências reais, ou probabilidades de símbolos isolados) e sobre a aproximação de ordem zero (igual probabilidade).

Vejamos, agora, alguns aspectos importantes sobre a matriz de correlação $P(I, J)$ que é do tipo 28 x 28. Assim, $P(27, J)$, é o termo da coluna 27 e da linha J da matriz e representa a probabilidade das palavras que começam por J, em que J designa uma letra qualquer.

Por outro lado:

$$\sum_J M(27, J) = M(27) = 1000 \text{ caracteres}$$

O elemento $M(1, 27)$, ou $P(1, 27)$, representa a probabilidade de o A ser seguido por um "espaço", isto é, a probabilidade das palavras que terminam por A; o elemento $P(27, 1)$, representa a probabilidade de um "espaço" ser seguido por um A, isto é, a probabilidade das palavras que começam por A. Os elementos $P(17, 21)$ e $P(21, 17)$ são muito diferentes: enquanto o primeiro representa a probabilidade de que ao Q se siga o U (o que é muito comum), o outro representa a correlação de que ao U se siga um Q, o que não é nada comum. As matrizes $M(I, J)$ e $P(I, J)$ não são em geral simétricas: por exemplo, o ditongo AI é diferente de IA tempo do verbo ir. Logo, $P(1, 10) \neq P(10, 1)$. Há na língua portuguesa alguns casos em que os elementos das matrizes podem ser iguais. É o caso de $M(18, 18)$ e de $M(19, 19)$, que correspondem à repetição do R, e do S, respectivamente. Em geral, os elementos $P(J, J)$ são muito pequenos, ou nulos, a não ser nos dois casos anteriores. Claro que a probabilidade de ocorrência de ditongos é relativamente, elevada na língua portuguesa e os elementos correspondentes na matriz $M(I, J)$ em que I, J formam um ditongo, assumem valores também elevados. Por exemplo, $M(5, 9)$ tem valores elevados, porque o ditongo EI é muito frequente na língua portuguesa. No caso de se considerar o elemento $M(27, 27) = 0$, significa que se evitam sequências longas de espaços, etc..

Podíamos prosseguir estas considerações, recorrendo a matrizes de terceira ordem, que contêm muitos mais elementos da estrutura da língua. Mas, as matrizes de terceira ordem podem obter-se a partir de matrizes de segunda ordem, como vamos mostrar. Consideremos a sequência ordenada de três caracteres I, J, K em que ambos I e K são previamente especificados. A probabilidade de obter o J entre o I e o K é dada por:

$$P(I, J, K) = M(I, J, K) / \sum M(I, J, K) = M(I, J, K) / N(I, K),$$

em que $N(I, K)$ designa a matriz de correlação entre caracteres alternados. Por outro lado, a probabilidade de obter o símbolo J a seguir ao I é:

$$P(I, J) = M(I, J) / \sum M(I, J) = M(I, J) / M(I)$$

e a probabilidade de obter um J antes de um K é:

$$P(J, K) = M(J, K) / \sum M(J, K) = M(J, K) / M(K)$$

Pelo teorema das probabilidades compostas, sabemos que:

$$P(I, J, K) = P(I, J) P(J, K)$$

e, portanto:

$$P(I, J, K) = N(I, K) [M(I, J) / M(I)] \cdot [M(J, K) / M(K)]$$

Logo, a matriz de terceira ordem $P(I, J, K)$ pode obter-se a partir de matrizes de segunda ordem, como pretendíamos mostrar.

6.4. A ENTROPIA DE VÁRIAS ORDENS NA LINGUAGEM

Consideremos, mais uma vez, a fórmula de Shannon:

$$H = -\sum P(I) \log_2 P(I)$$

em que, por utilizar a base 2, H vem expressa em *bits* (ver Apêndice II).

A grandeza H corresponde, como se sabe, à medida estatística do número de bits por "tipo", necessários para transmitir mensagens em código. Notaremos que:

$$-\log_2 P(I) = \log_2 [1/P(I)]$$

e representa o número de bits necessários para especificar uma lista de $1/P(I)$ caracteres. Em especial quando há N símbolos na lista:

$$P(I) = \text{const.} = 1/N$$

a equação de Shannon representa o número de bits necessários para especificar N "acontecimentos" igualmente prováveis.

Mas, em geral, os $P(I)$, isto é, as probabilidades de ocorrência de um dado tipo são diferentes. Tem que se conhecer as várias probabilidades reais de ocorrência $P(1), P(2) \dots P(N)$ para obter informação, mais apurada, porque, a fórmula de *Shannon* dá-nos o *número médio mínimo de bits necessários para especificar o número de escolhas possíveis, que podem ocorrer de entre N possibilidades diferentes, com as probabilidades $P(1), P(2), P(3) \dots$*

Aplicando estas considerações a um texto de linguagem escrita, aquela expressão dá-nos o valor médio da entropia por carácter e este parâmetro

constitui uma das características da língua. Os valores obtidos são independentes da ordem pela qual o texto é lido, ou está escrito. Vamos, então, determinar o número mínimo médio de bits, por carácter, para transmitir uma mensagem numa dada língua. Consideremos o Inglês e o Português, por exemplo.

Os resultados obtidos dependerão, como é natural, das propriedades estatísticas da língua. Além disso, devido às variações de estilo dos diferentes autores, não se pode esperar uma resposta única e precisa. Vemos que há razões para admitir que as línguas reais obedecem ao Segundo Princípio Fundamental da Termodinâmica. As respostas dependerão, sempre, do texto particular que se analisa, mas as diferenças são relativamente pequenas. A estrutura da língua, é, de longe, o factor predominante e as pequenas variações constituem até um índice para a identificação de um autor.

Vamos rever como se calcula a entropia por carácter, numa aproximação de "ordem zero". Suponhamos que dispomos dum alfabeto de 28 caracteres (23 letras + 3 (K, W e Y) + 1 espaço entre palavras + 1 apóstrofo). Neste caso $N=28$. Supondo que todos são equiparáveis, a probabilidade $P(I)$ de cada carácter é $1/28$.

Logo, a equação anterior dá:

$$h_0 = \text{Log}_2 28 = 4,80735 \quad \text{bits/carácter}$$

Claro que, neste caso, o número mínimo de caracteres necessários é: $C_0 = 2^{h_0} = 28$. Com este alfabeto, e nesta aproximação, todas as línguas seriam equivalentes: não haveria diferença entre o Português e o Inglês.

Numa aproximação de primeira ordem, a entropia por símbolo, que representamos por h , requer o conhecimento das probabilidades reais de ocorrência dos elementos individuais do "alfabeto".

Aceitamos que os símbolos são independentes entre si e que não há influência de letras adjacentes, nem preferências, por enquanto, para se formarem pares, tão comuns na língua portuguesa tais como QU; CH; NH; AI; AU. As probabilidades reais podem determinar-se da observação da língua real. No caso de 28 caracteres a probabilidade de ocorrência do símbolo I , isto é, $P(I)$ pode obter-se somando as linhas da matriz de correlação de pares $M(I, J)$:

$$P(I) = \sum M(I, J) / \sum \sum M(I, J)$$

Os valores das probabilidades da ocorrência dos vários símbolos em Inglês e em Português são muito diferentes, como se depreende da análise do QUADRO I. Aplicando a equação de Shannon, obtém-se para a língua

portuguesa um valor de $h_1 = 3,91$ bits/caracter, a que corresponde um número mínimo de caracteres dado por $2^{3,91} = 15,03$. Para a língua inglesa teríamos um valor de $h_1 = 4,106$ bits/caracter a que corresponde um número mínimo de caracteres dado por $2^{4,106} = 17,21$. Em qualquer dos casos o "alfabeto" reduzido necessário é inferior ao alfabeto disponível de 28 símbolos. Como se vê, ao nível de uma aproximação de primeira ordem, há uma redundância visto que há um "excedente" de 11 caracteres. Em termos de entropia, a redundância é (Apêndice III):

$$R = \frac{4,80 - 3,91}{4,80} = 1 - \frac{3,91}{4,80} = 0,185$$

Numa aproximação de segunda ordem começa a introduzir-se a estrutura da língua. Depois de escolhida uma letra (ou um caracter), a outra é seleccionada de acordo com as frequências com que as letras se seguem. Assim, na aproximação de segunda ordem, a determinação, h_2 , da entropia por caracter, tem que tomar em consideração a probabilidade $P(I,J)$ do caracter J vir a seguir ao caracter I.

Suponhamos que se fixa o sinal I. A probabilidade de que o tipo que se lhe segue seja J pode obter-se a partir de correlação, $M(I,J)$, notando que:

$$P(I,J) = M(I,J) / \sum M(I,J), \text{ tal que } \sum P(I,J) = 1$$

A entropia, em bits, necessária para especificar um número de escolha é uma função de I, isto é, depende da história passada, ou seja, do caracter inicial que se fixou:

$$H(I) = - \sum P(I,J) \log_2 P(I,J)$$

Finalmente, pretende-se encontrar o valor médio de $H(I)$ para todos os caracteres iniciais $I = 1$ a 28. Este valor médio será mínimo, porque $H(I)$ já é mínimo. Logo, a aproximação de segunda ordem para o cálculo da entropia por caracter, h_2 , é:

$$h_2 = - \sum P_i \sum P(I,J) \log_2 P(I,J)$$

em que as probabilidades se podem obter a partir da matriz de "pares de correlação" $M(I,J)$, como vimos.

Usando uma matriz de 28 x 28, obtém-se para as línguas portuguesa e inglesa, os valores respectivamente, de:

$$h_2 = 3,11 \text{ bits/caracter}$$

e

$$h_2 = 3,308$$

a que corresponde para a língua inglesa um número mínimo de caracteres:

$$2^{3,308} = 9,905$$

inferior aos dois casos anteriores ("alfabetos" de 17, de 21 e de 28 caracteres).

A determinação da entropia pode, em princípio, estender-se a ordens cada vez mais elevadas de correlação. Por exemplo, na aproximação de terceira ordem, cada novo símbolo é escolhido tomando em consideração a probabilidade com que ocorrem os dois anteriores. Nestas condições, tem-se para a entropia:

$$h_3 = -\sum P(I) \sum P(I,J) \sum P(I,J,K) \log_2 P(I,J,K)$$

Quanto mais elevada for a ordem de aproximação, menor é a entropia (expressa em bits) e menor é a lista de símbolos ("alfabeto") necessários para transmitir uma mensagem.

6.5. ALGUNS COMENTÁRIOS

A análise estatística da Língua Inglesa tem vindo a ser feita por vários autores desde os trabalhos iniciais de Brillouin e de Shannon. Ao contrário, poucos estudos semelhantes se têm realizado para a Língua Portuguesa. As referências que utilizamos baseiam-se, como já referimos, na análise estatística da obra de Afrânio Coutinho, *A Literatura no Brasil* (1969). Na cadeira de Termodinâmica Generalizada temos sugerido este tema entre outros, aos Srs. Alunos para o Trabalho Final Complementar da Cadeira. E os resultados são muito interessantes. Damos em Apêndice, a título de exemplo, um resumo de um desses trabalhos.

Uma análise estatística, mais refinada, permite avaliar a probabilidade de ocorrência de sílabas e avaliar o número médio de sílabas por palavra e, depois, relacioná-lo com a entropia.

Uma das interpretações mais simples que se pode apresentar é a de que, quanto menor for o valor da entropia, tanto maior é a informação por sílaba. É interessante assinalar, que todos nós temos ideia de que as palavras inglesas são mais "curtas" do que as portuguesas, ou do que as das línguas novilatinas. Com efeito, a análise da Epístola de Salustio apresenta um número médio de sílabas por palavra de 2,48 e uma entropia de 0,64 bits, em contraste com os valores de 1,29 e de 0,29, respectivamente, resultantes da análise de Otelio de Shakespeare.

Mas há problemas a outro nível e que serão referidos oportunamente. As letras formam sílabas e estas palavras, que estão organizadas em frases, devendo satisfazer às normass da gramática. Impõem-se assim, constrangimentos à formação das palavras e à sua disposição no discurso.

A análise estatística duma língua é um problema eminentemente pluridisciplinar, que envolve muitos domínios do saber.

Deve acentuar-se que os resultados obtidos assentam, exclusivamente nas propriedades estatísticas das duas línguas.

Ora, por causa do estilo dos vários autores, não se podem esperar respostas absolutamente precisas. Há, portanto, razão para aceitar que a linguagem real é uma "vítima" do Segundo Princípio da Termodinâmica. É certo que a estrutura de uma língua é o factor predominante e as pequenas variações constituem muitas vezes um indício para a identificação de um autor. (Ver Apêndice IV)

Antes de encerrar esta pequena incursão pelos campos da linguística matemática devemos acentuar que o conceito quantitativo de informação não contém juízos de valor sobre a utilizabilidade, ou sobre a relevância do que está escrito. O mesmo conteúdo de informação pode dizer coisas muito diferentes conforme a inteligência, a sensibilidade e a cultura do escritor que a produz e a do destinatário que a recebe.

7. ENTROPIA COMO DISCRIMINANTE

Entropia da Informação passou desde a sua criação por Shannon, com a cumplicidade de Von Neumann, a ter um conjunto de significados que tornaram o conceito fugidio. Este facto, tem como consequência um alheamento continuado em relação à possibilidade de medir e portanto utilizar um conceito basilar na teoria da informação.

No entanto, temos a possibilidade da medição expedita, eficaz, precisa e rigorosa da entropia de um sistema quando a informação que emite é suportada pelo espectro electromagnético, designadamente, pela luz.

Vale a pena então rever brevemente o conceito da medida da entropia, utilizando sistemas de visão convencionais e ilustrar a capacidade do método de medição pela aplicação em três grandes sectores industriais e um das indústrias dos serviços: têxtil, curtumes, papel e em cartografia.

7.1. MEDIDA DA ENTROPIA DA INFORMAÇÃO NO CONCEITO DE SHANNON

A questão, fica enquadrada pela definição de Shannon ou se se quizer pela utilização que Shannon faz para conseguir uma quantidade de informação calculada a partir de um conjunto de mensagens. Para Shannon a quantidade de informação não fica agarrada às mensagens e a todas as ligações entre receptor e emissor no plano da semântica e do relativismo, antes fica definida pelas probabilidades que essas mensagens têm ou não de ser emitidas. Se se quizer trata-se de informação despojada de conhecimento.

É neste âmbito que a medida da quantidade de informação de Shannon traduz uma propriedade intrínseca dos sistemas que se traduz, como repetidamente vimos na célebre equação, da entropia da informação, H :

$$H = - \sum_{i=1}^N p_i \log p_i \quad (1)$$

em que p_i é a probabilidade da mensagem i . O suporte da mensagem é irrelevante para a definição.

O valor de H é característico do sistema num determinado estado macroscópico. A alteração quer da forma (na variante da informação), quer do estado, produz modificações na medida que porventura se faça de H . Neste sentido a entropia da informação, como caracterizadora de um sistema, pode ser utilizada como discriminante. Aliás, do mesmo modo que no domínio espaço - tempo, o conceito de acontecimento, num determinado instante, divide o passado e o futuro, e na variável espaço podemos aferir a variação nas dimensões do sistema e na energia as alterações do seu comportamento, também a informação pode ser empregue para aquilatar a variabilidade de qualquer sistema. Para o fazer

basta que, tal como acontece para as medidas do tempo, espaço e energia, exista um método que leve à medida da quantidade de informação, sinónimo de entropia de informação definida pela expressão anterior.

7.2. MEDIDA DA ENTROPIA DA INFORMAÇÃO

Para medir a entropia da informação de um sistema, de acordo com aquela expressão, basta medir a probabilidade de cada uma das mensagens que ele emite. Então para a medição, a natureza do suporte das mensagens é de grande importância, uma vez que a detecção da mensagem e a medida da sua probabilidade são os factores essenciais.

Nestas circunstâncias a medida da entropia da informação, seja de que sistema fôr, será sempre dentro de uma gama de mensagens da mesma natureza, uma vez que não haverá sensores que permitam reconhecer instantaneamente as mensagens suportadas por todo o tipo de emissão corpuscular ou radiactiva e avaliar a probabilidade de ocorrência de cada uma delas. Assim sendo, a medida da entropia da informação, só poderá ser utilizada para efectuar comparações quando diz respeito ao mesmo elenco de mensagens.

Uma vez assegurada a uniformidade referida, a medida da entropia da informação fica muito facilitada. Basta que exista um sensor que atribua uma probabilidade para cada mensagem.

7.3. MEDIDA DA ENTROPIA DA INFORMAÇÃO NO ESPECTRO VISÍVEL

Quando as mensagens, que são escrutinadas, têm por veículo a luz, a tarefa da medida da quantidade de informação fica bastante simplificada.

Uma câmara CCD (charge coupled device), ligada a um sistema de digitalização de imagem, figura 12, permite medir a probabilidade da quantidade de luz emitida num espectro largo por uma área definida pela ampliação do sistema

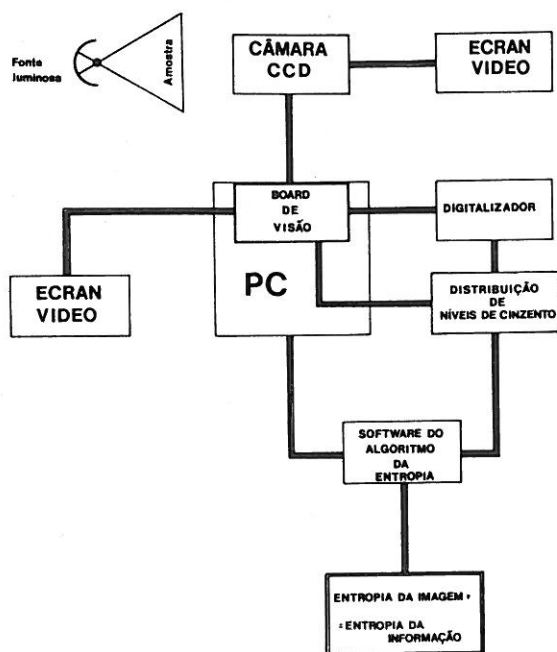


Fig. 12 - Diagrama de blocos para a medida da entropia de informação.

óptico e pelo pixel definido pela resolução do CCD. A quantidade luz na imagem, ou seja, o nível de cinzento de cada pixel é registado e obtem-se de imediato uma função de distribuição de níveis de cinzento, figura 13.

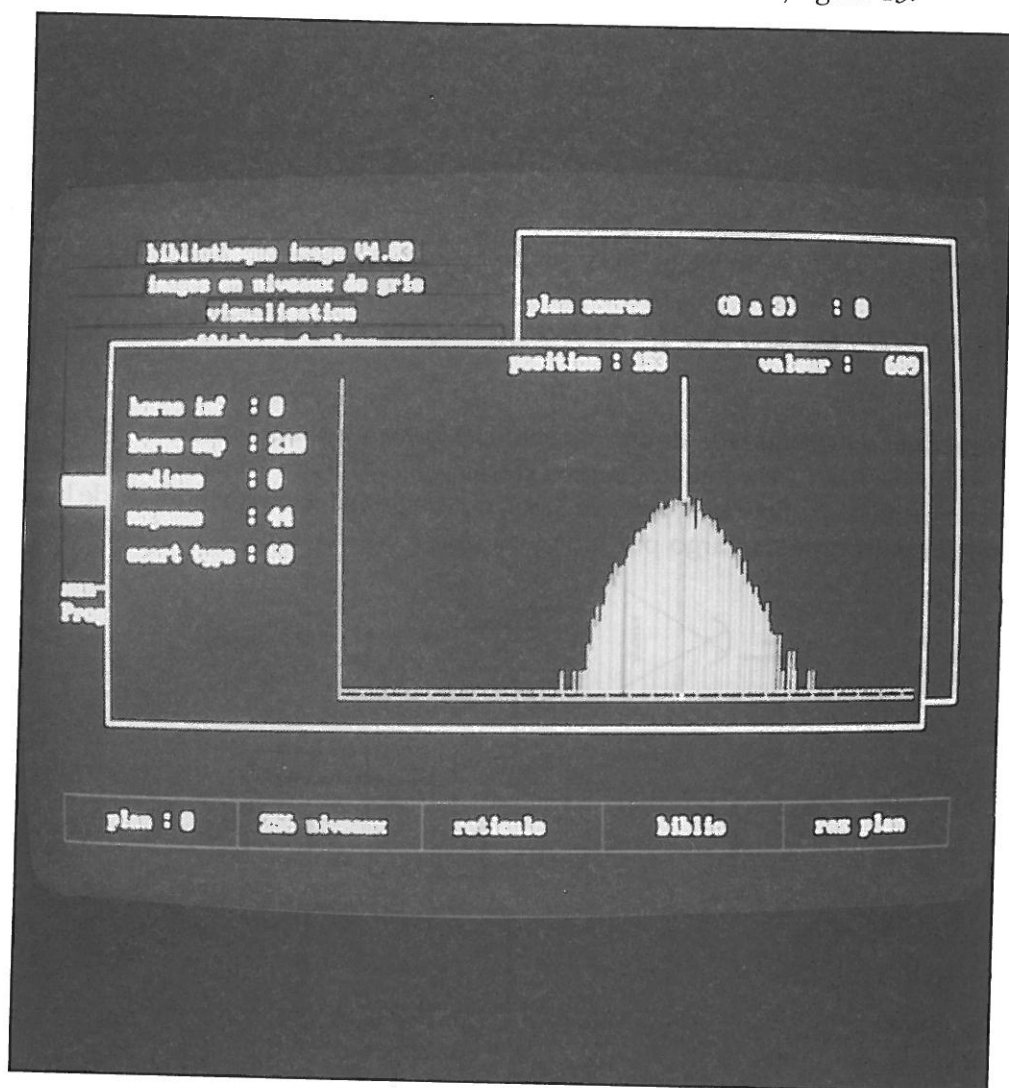


Fig. 13 - Distribuição típica de níveis de cinzento de uma imagem. Entropia: 2,243

A partir da função de distribuição das mensagens, constituídas neste método pela iluminação da imagem do sistema, o emprego da expressão de definição de entropia da informação permite, de imediato, o seu cálculo.

7.4. APLICAÇÕES DA ENTROPIA DA INFORMAÇÃO COMO DISCRIMINANTE

A medida da entropia da informação de um sistema, a partir da medida da mesma quantidade referida a uma imagem, permite seguir aquele parâmetro, em tempo real, mesmo para sistemas com estruturas altamente complexas. Quer pela complexidade, quer pela importância económica dos sectores industriais, os processos têxtil, do papel e dos curtumes são exemplos da aplicação ilustrativos da aplicabilidade da entropia de informação como discriminante na detecção de anomalias nos processos de fabrico e de defeitos no produto final.

7.4.1. SECTOR TÊXTIL

7.4.1.1. PENTEÇÃO

Processos de fabrico, como é o caso da penteação de lã, passaram a poder ser completamente aferidos. Neste caso a interpretação, sempre ligada à entropia da informação, da ordem e da desordem que atribui à desordem um valor mais elevado da medida entrópica tem uma realidade palpável. Com a energia gasta na penteação almeja-se tornar um conjunto de fibras orientadas caoticamente num conjunto de fibras paralelizadas. Com o processo de penteação pretende-se diminuir a entropia do sistema.

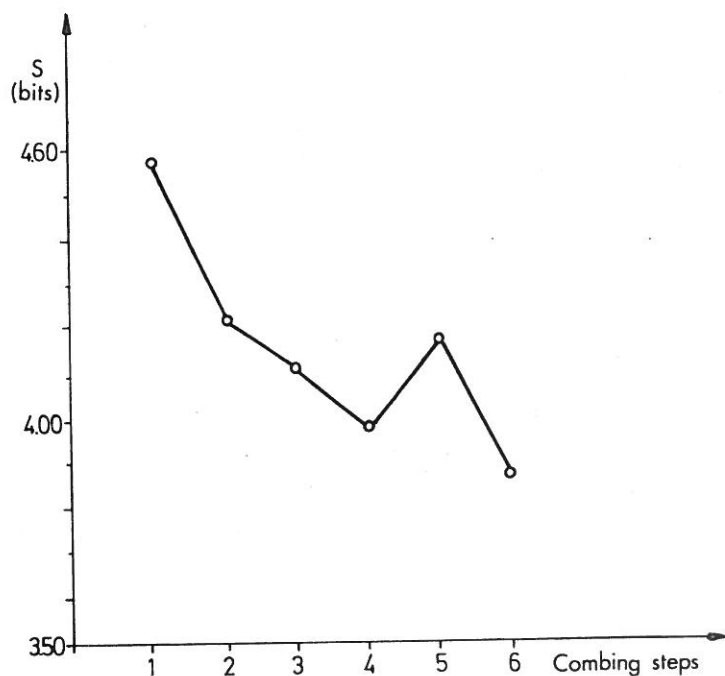


Fig. 14 - Evolução da entropia de informação ao longo da penteação de lã

O resultado do seguimento do valor entrópico da lã penteada, utilizando o método descrito, ao longo das etapas do processo mostra claramente uma evolução no sentido de uma maior ordem no sistema com a consequente diminuição da entropia de informação.

7.4.1.2. MEDIDAS DE PILOSIDADE DE FIOS TÊXTEIS

A caracterização do grau de pilosidade de fios é um dos exemplos mais paradigmáticos da capacidade da utilização do método de medida de entropia aqui descrito.

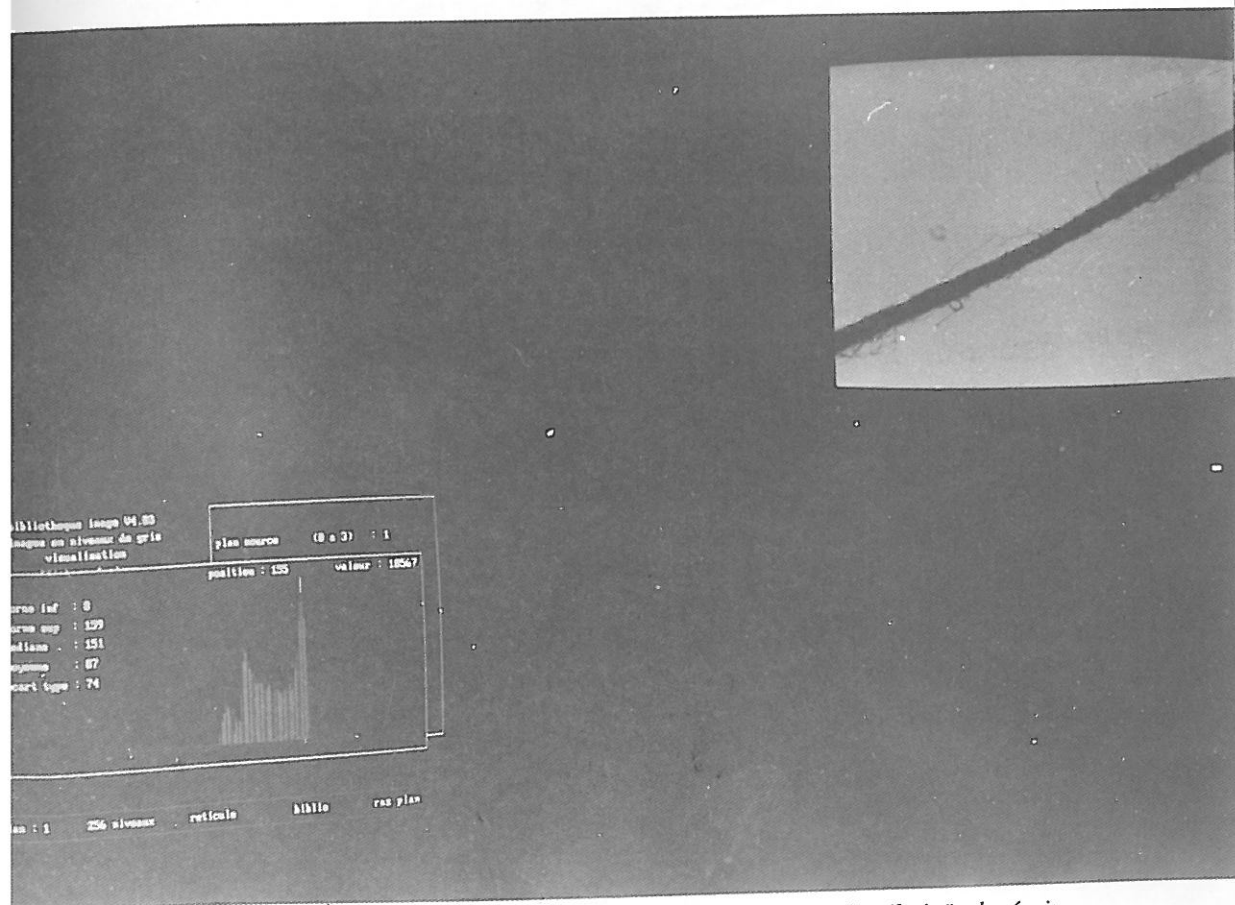


Fig. 15 - (a) Imagem de um fio têxtil sem pilosidade e (b) a correspondente distribuição de níveis de cinzento. Entropia: 0,678

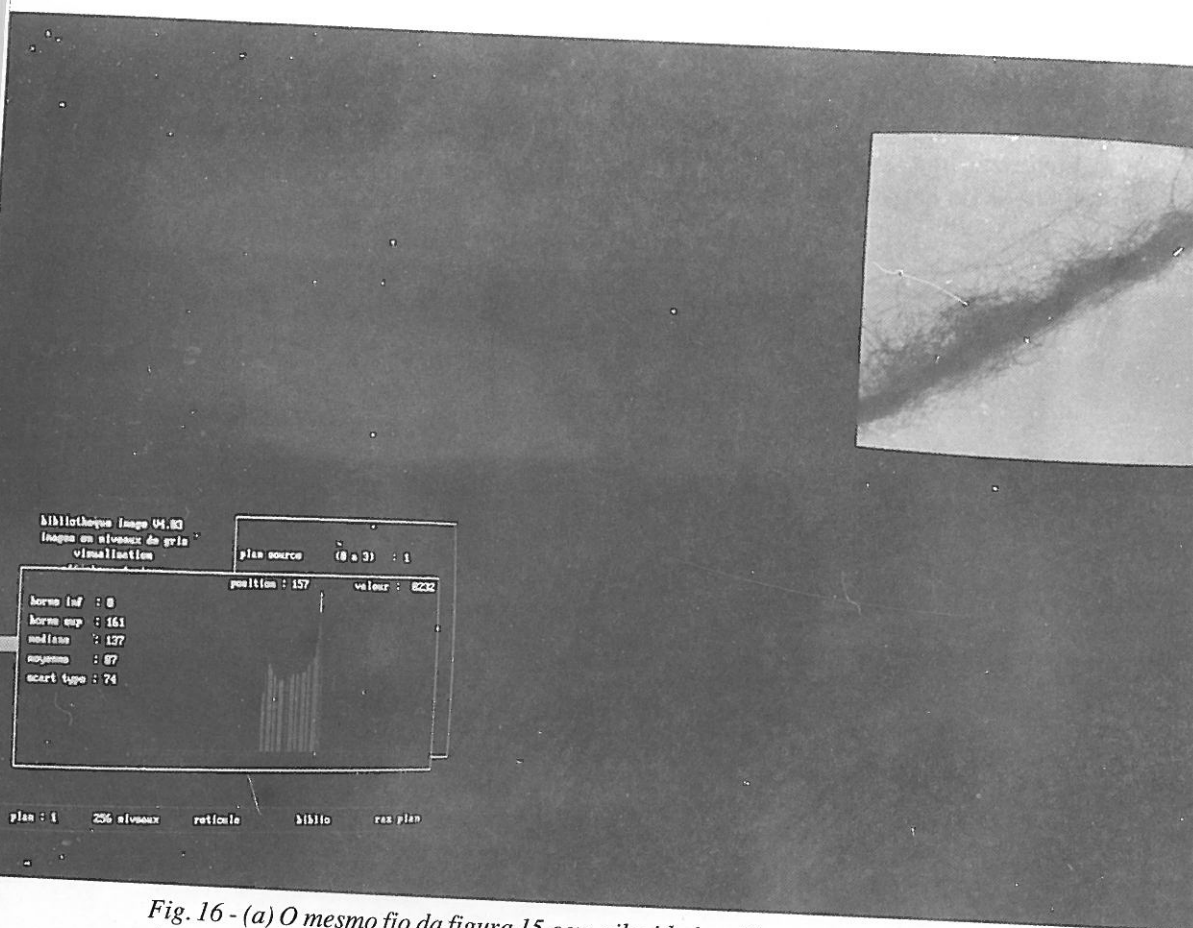


Fig. 16 - (a) O mesmo fio da figura 15 com pilosidade e (b) a respectiva distribuição de níveis de cinzento. Entropia: 0,7069

Para os dois casos expressos nas figuras 15 e figura 16 a variação de entropia de informação é no sentido que a teoria prevê. A uma maior pilosidade associa-se uma maior desordem a que corresponde uma maior entropia.

7.4.1.3. CARACTERIZAÇÃO DE DEFEITOS EM TECIDOS

A determinação de forma automática de defeitos em tecidos constitui uma tarefa de substituição de uma actividade humana que melhor se pode descrever como sabedoria. Recorrer aos sistemas computacionais para resolver tais casos traduz-se, quase sempre, por uma grande limitação de resultados. A detecção de defeitos em tecidos é um destes casos.

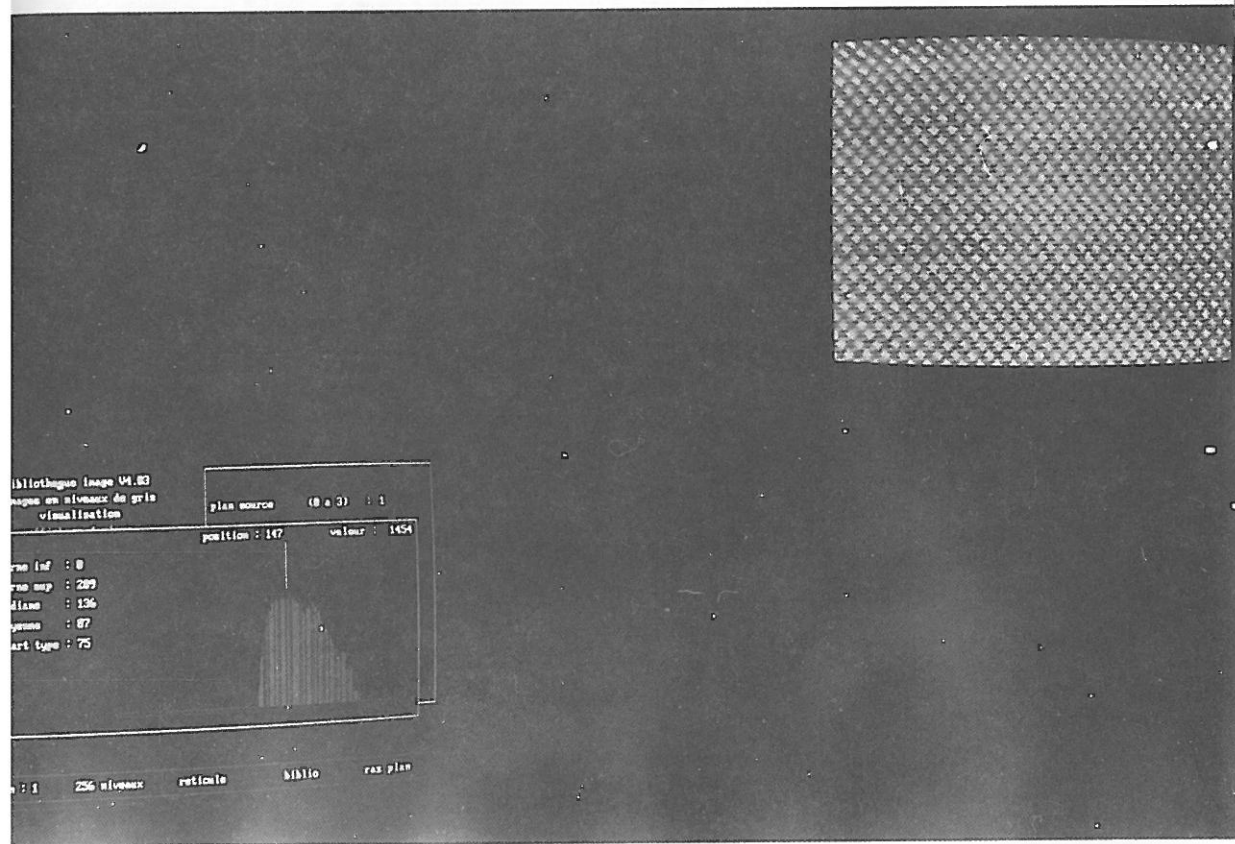


Fig. 17 - (a) Imagem de uma malha sem defeitos e (b) a respectiva distribuição de níveis de cinzento. Entropia: 3,1673

Parece, no entanto, que a caracterização de tecidos por um parâmetro de medida da informação constituirá um avanço no sentido da tomada de decisão baseada em apenas um parâmetro, a entropia da informação.

Assim, no domínio da informação um defeito é qualquer variação da entropia da informação do tecido relativamente ao valor que é medido para o tecido sem defeitos.

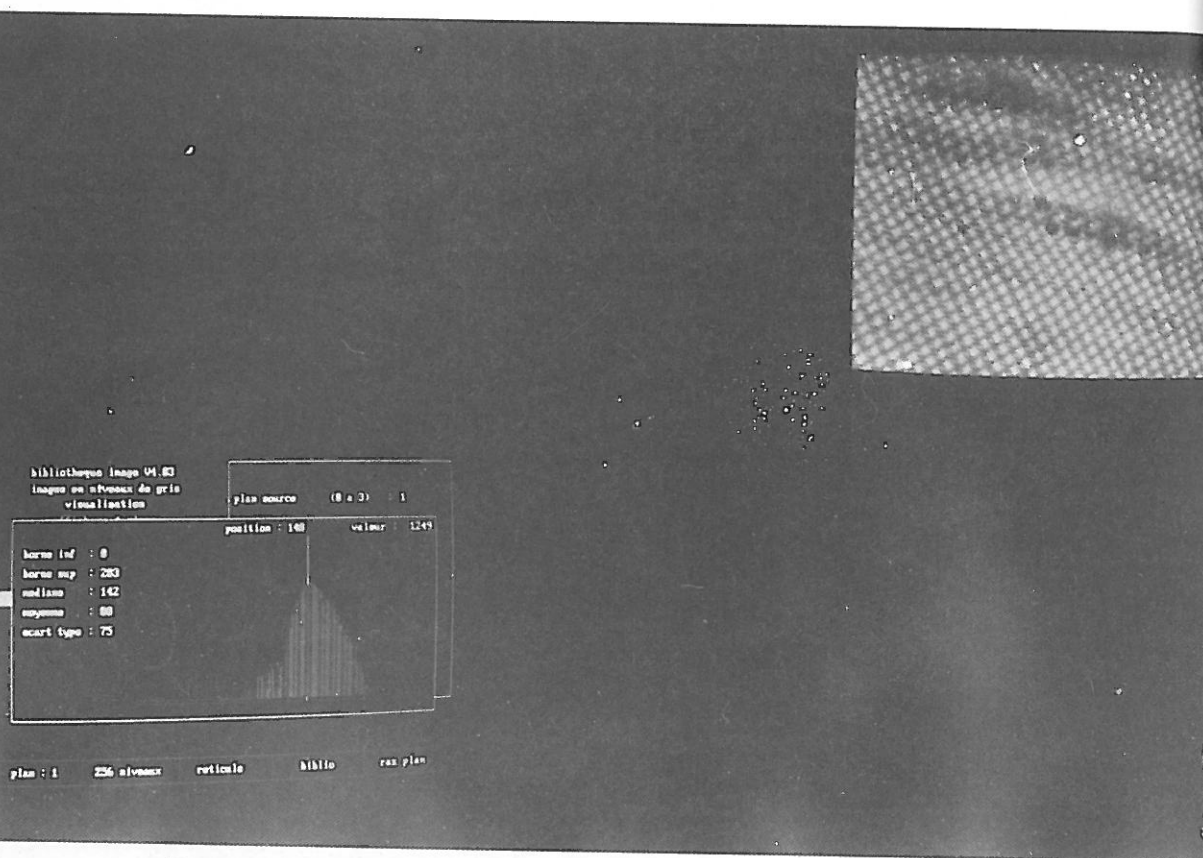


Fig. 18 - (a) A mesma malha que na figura 17 mas com defeitos e (b) a respectiva distribuição de níveis de cinzento. Entropia: 3,2806

Os exemplos ilustrados nas figuras 17 e 18 representam casos típicos e concludentes.

7.4.2. SECTOR DO PAPEL

A distribuição espacial das fibras no papel dita um conjunto de propriedades físicas e químicas determinantes do aproveitamento que dele se faz.

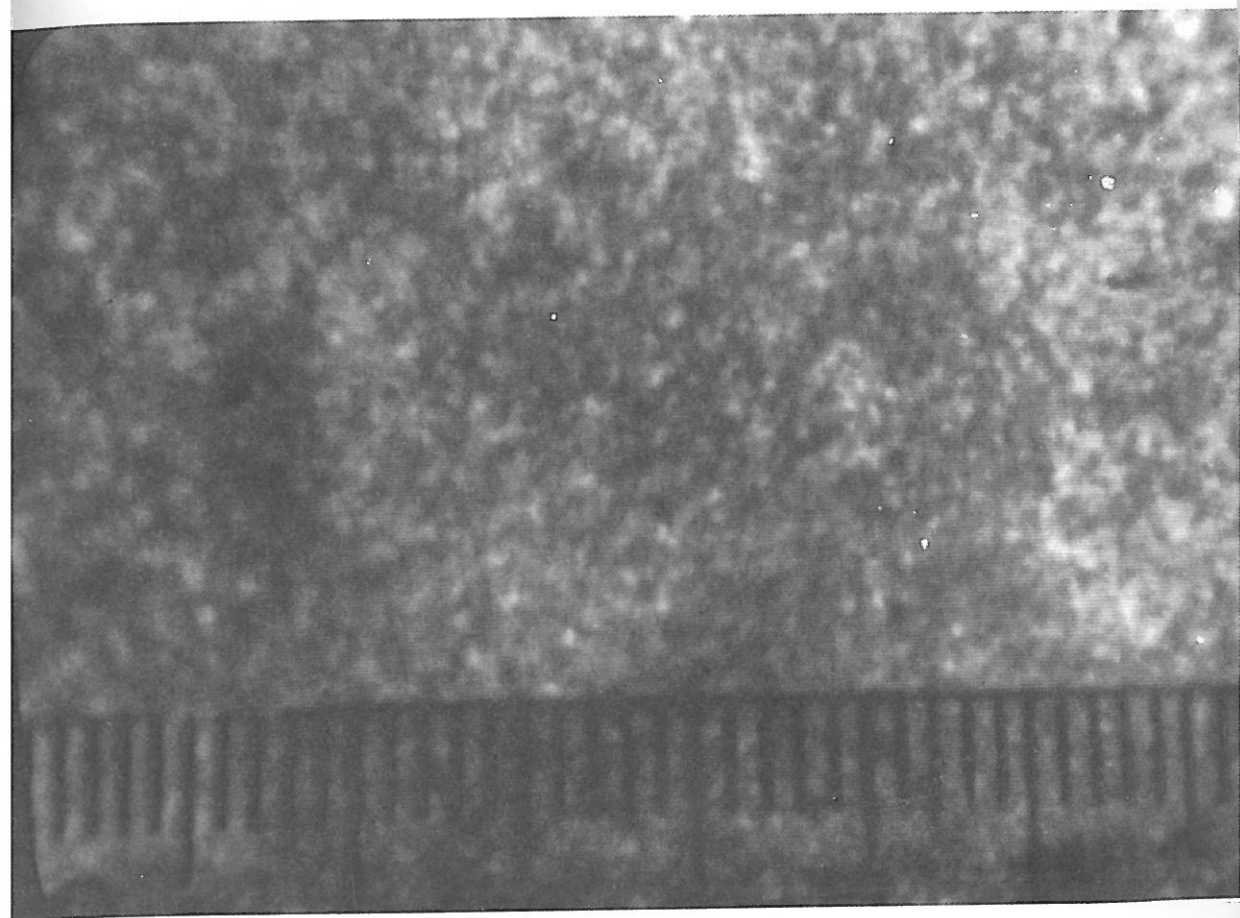


Fig. 19 - Imagem de papel Kraft. Entropia: 2,243

Antes da possibilidade de determinação do grau de desordem na distribuição espacial das fibras que a medida da entropia da informação trouxe, a medida da heterogeneidade na distribuição requeria métodos baseados na difracção óptica. Estes implicam a utilização de lasers, que têm sempre uma vida limitada, e o tratamento em paralelo de grande número de dados uma vez que a medida tem que ser feitas numa área suficientemente extensa.

A diferença em entropia da informação mostra, para o exemplo da figura 19 claramente uma grande diferença caso se trate de um papel Kraft ou papel normal (Fig. 20).

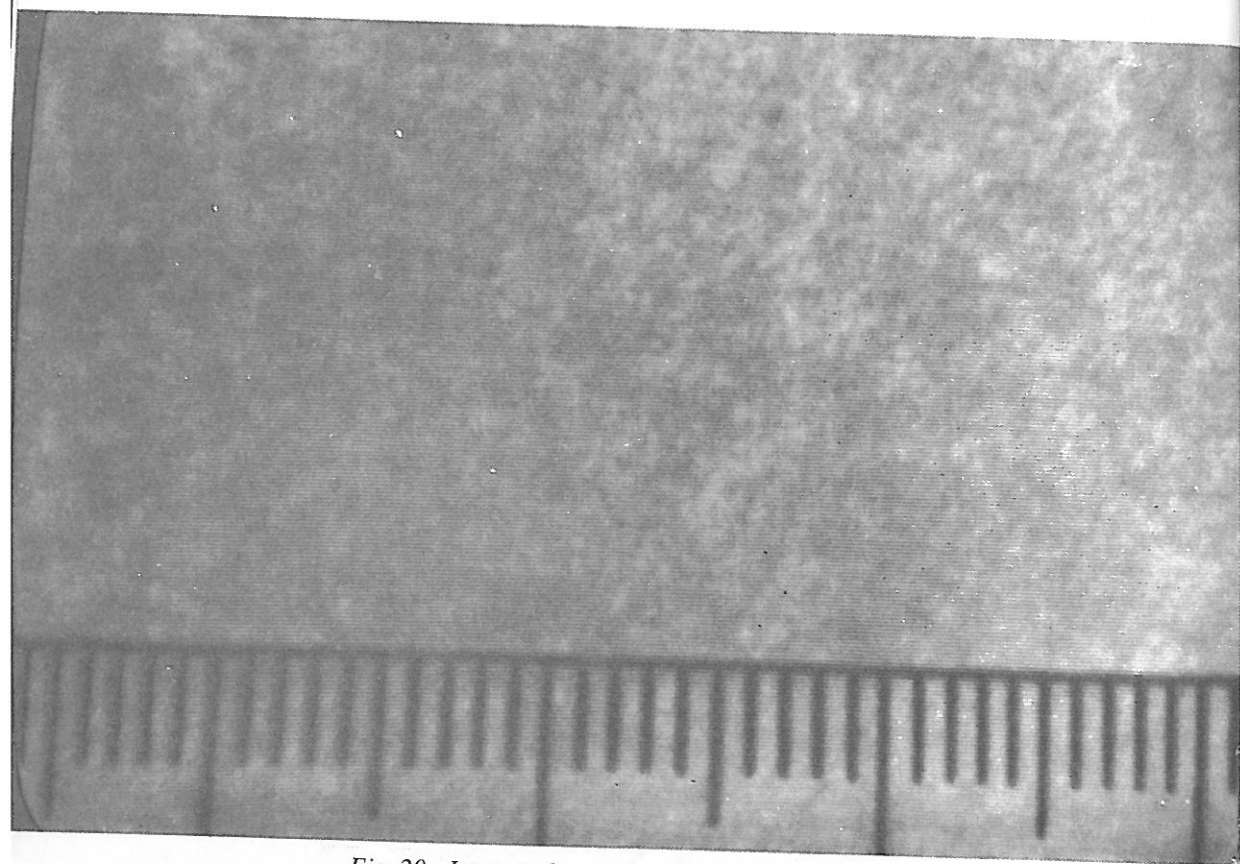


Fig. 20 - Imagem de papel para impressão. Entropia: 2,649

O primeiro corresponde a uma distribuição espacial mais caótica enquanto que o segundo apresenta uma distribuição espacial bastante mais ordenada uma vez que neste tipo de papel mercê do processo de fabrico, as fibras tendem a organizar-se segundo uma direcção preferencial.

7.4.3. *SECTOR DOS CURTUMES*

Mais uma vez a caracterização de uma pele e a consequente monitoração de desvios ou defeitos não é um problema a que a ciência de computação tenha dado resposta cabal.

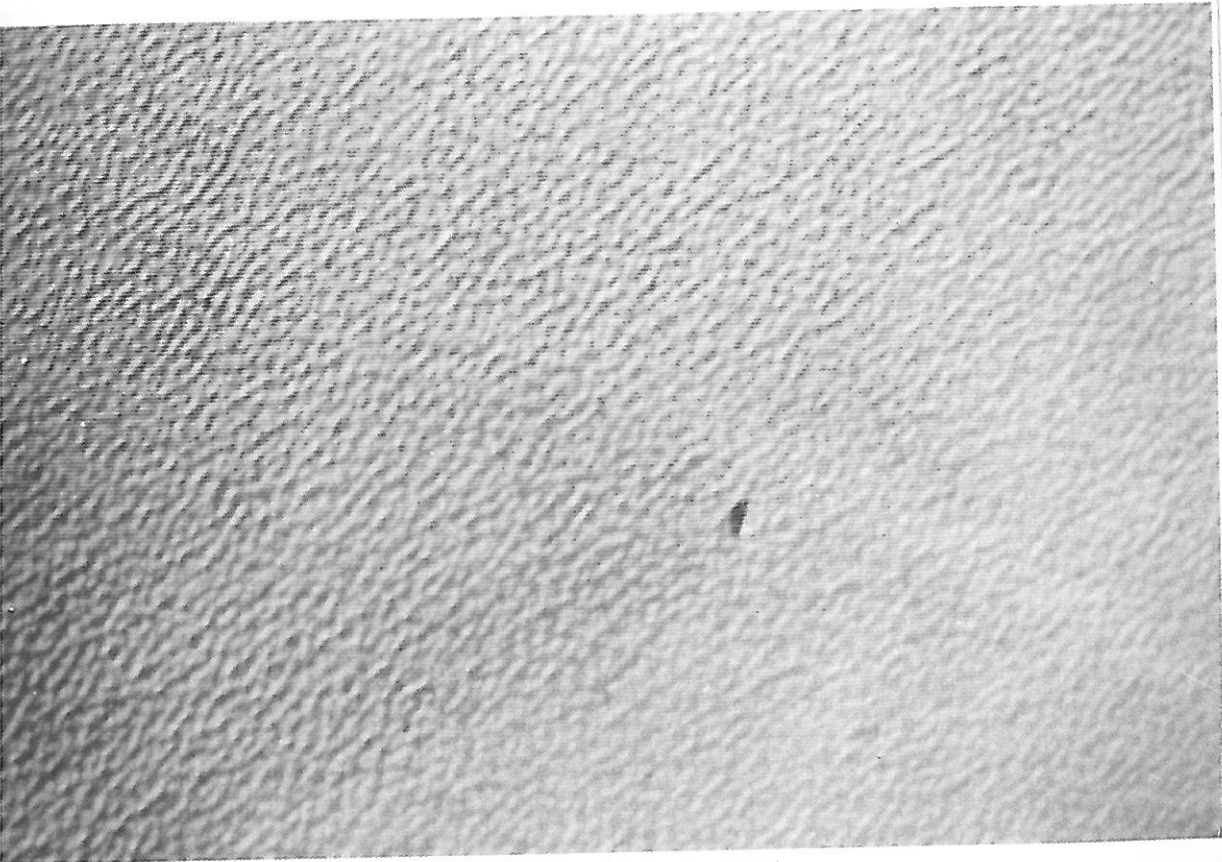


Fig. 21 - Cabedal sem defeitos. Entropia: 6,284

A medida da entropia da informação permite, contudo, ter um método para a avaliação de desvios da pele relativamente ao padrão que fôr definido, considerado, como o mais apropriado. Qualquer variação, em relação ao valor da entropia da informação da pele sem defeitos, para um determinado fim, é indicadora da presença de defeitos em relação à pele considerada.

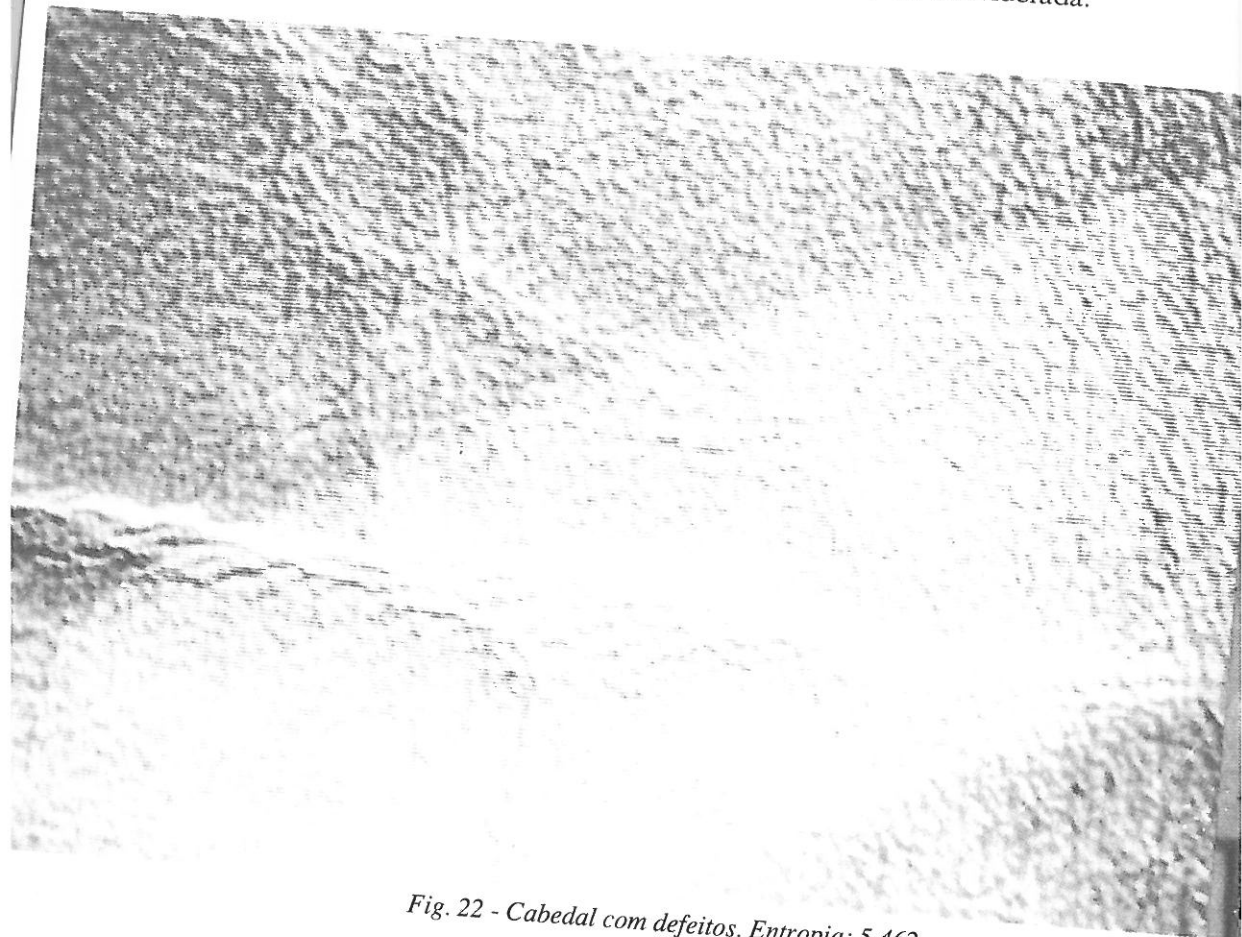


Fig. 22 - Cabedal com defeitos. Entropia: 5,462

A pele que se considera sem defeitos apresenta uma entropia superior à da pele a utilizar. É de salientar que neste caso a pele a utilizar tem um grande grau de ordenamento que não é o que mais convém para o fim em vista.

Este exemplo mostra que a utilização da entropia da informação como discriminante não implica que a utilidade esteja sempre associada a um decréscimo da entropia da informação.

Tal como "chegar a horas", não é chegar antes nem depois, um defeito é qualquer variação, para mais ou para menos, do valor estipulado para a entropia da informação.

7.4.4. CARTOGRAFIA

As espécies agrícolas quando localizadas numa forma densa no terreno apresentam características de distribuição espacial que lhe são peculiares. É o caso de um olival em que oliveiras que ocorrem, muitas vezes, em espaços onde há outras espécies agrícolas nomeadamente sobreiros.

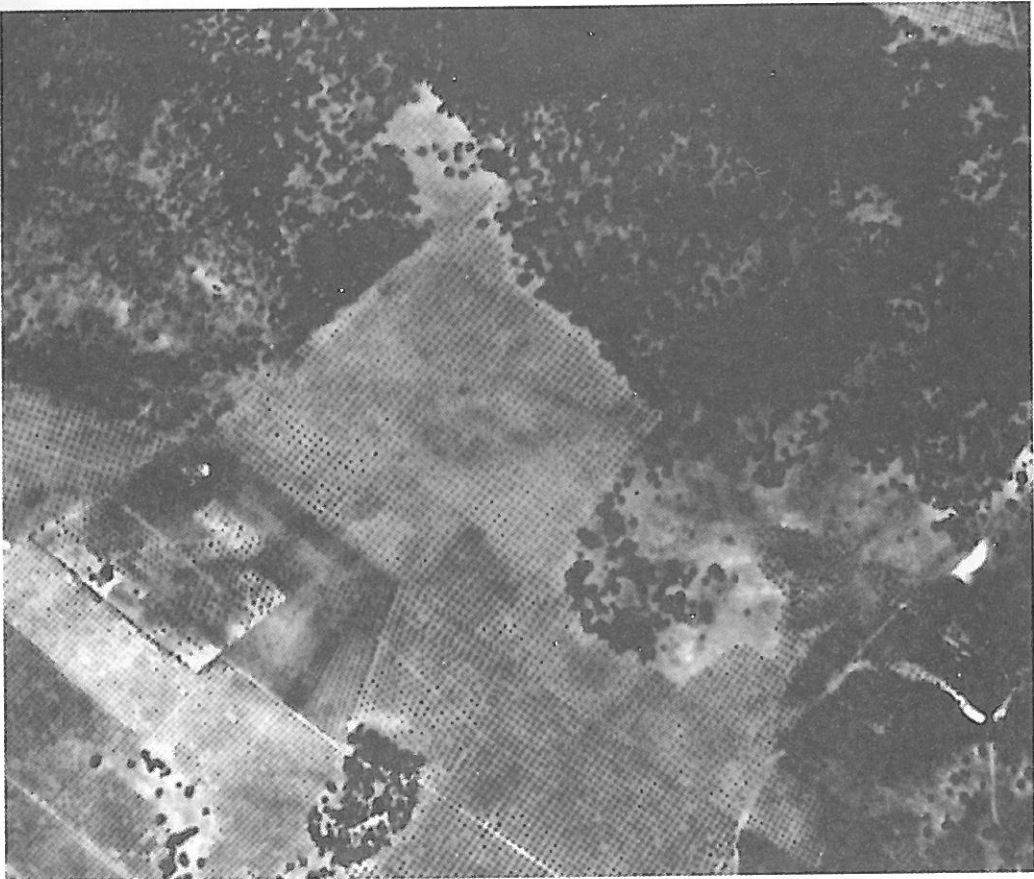


Fig. 23 - Fotografia aérea da região agrícola com sobreiros e oliveiras.

Como se esperava a figura 23 mostra que há mais ordem posta na distribuição espacial das oliveiras do que na disseminação das restantes plantações. Por divisão em janelas sobre a imagem (figura 23) pode constituir-se um mapa de medidas de entropia da informação daquela região agrícola (figura 23).

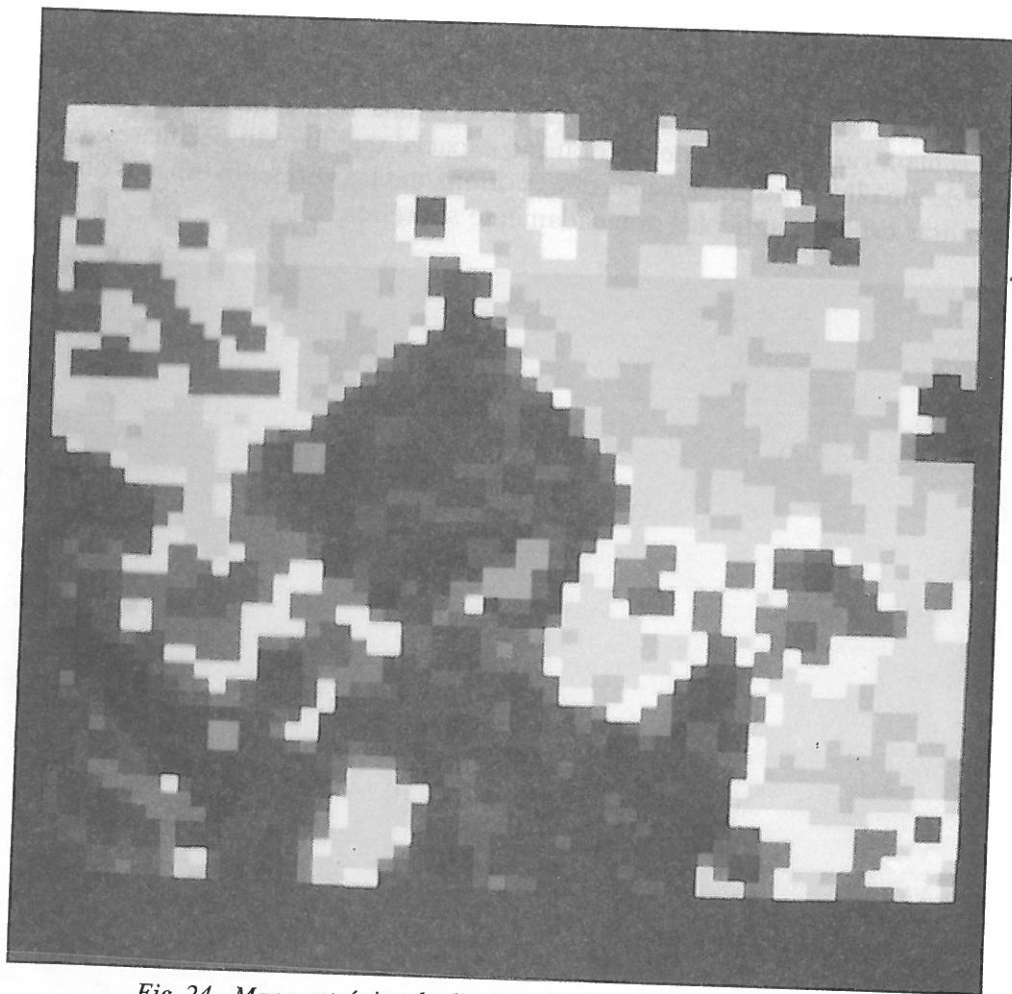


Fig. 24 - Mapa entrópico da distribuição das oliveiras e sobreiros.

Para este caso, tomando como valor discriminante da entropia de informação 4,5 bits constrói-se para a zona não ocupada por oliveiras, a partir da figura 24, uma máscara que deixa ficar na imagem apenas o olival.

A contagem das oliveiras e a sua cartografia a partir desta fase são imediatas, como mostra a figura 25. A contagem das oliveiras (1761) verificou-se ter apenas um desvio de cinco por cento em relação ao seu número efectivo.

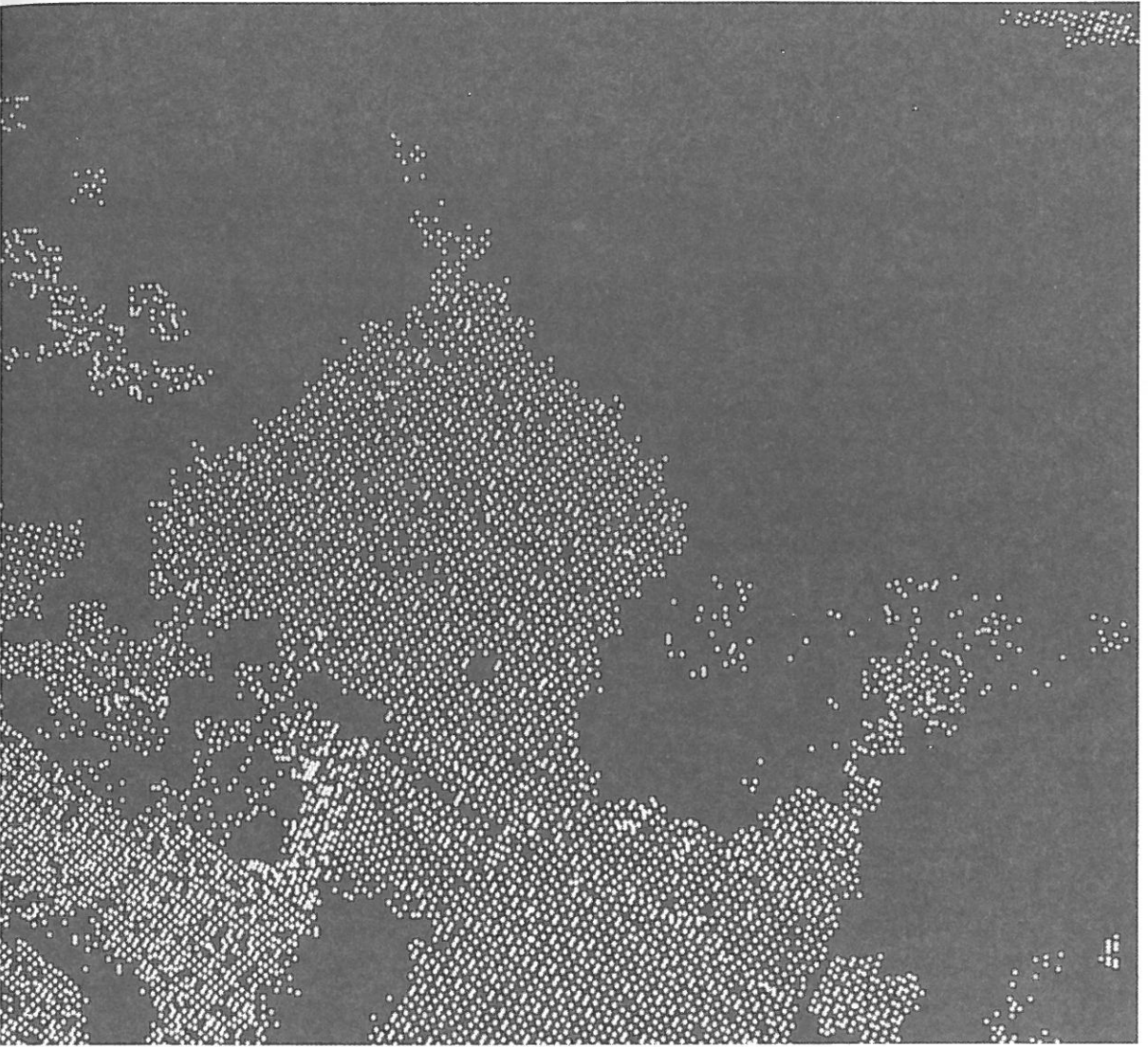


Fig. 25 - Imagem obtida da figura 23 depois de filtrada pela máscara entrópica da figura 24.

7.5. CONCLUSÕES

A entropia de informação é uma grandeza de fácil medida na zona do espectro electromagnético quando é possível organizar a informação numa imagem.

A utilização desta grandeza como discriminante revelou-se eficaz na resolução de um conjunto de problemas, sem solução anterior, em domínios tão complexos e díspares como são o têxtil, o papel, os curtumes e a cartografia temática de espécies agrícolas.

A entropia da informação sendo um parametro global indicador de textura, arquitectura e ordenamento (estrutura) facilita a tomada de decisões.

A entropia de informação fica, para a informação, em pé de igualdade com parametros globais inferidos do espaço, do tempo e da energia.

COMPLEXIDADE.INFORMAÇÃO.COESÃO

IV

1. COMPLEXIDADE E COMPLEXIDADE DE COMPORTAMENTO

Complexidade é quase sempre confundida, ou simplesmente tida como complexidade de comportamento.

Existe, contudo, uma grande diferença. Há sistemas complexos, isto é, sistemas que encerram ou abrangem muitos elementos, ou partes, com comportamentos bem previsíveis, diríamos mesmo simples, e existem sistemas muito simples com poucas partes ou elementos com enorme complexidade de comportamento.

A complexidade de comportamento de um sistema não tem que ver com a sua complexidade intrínseca. A complexidade de um sistema é independente do respectivo comportamento entendido como a sua conduta em termos da predictabilidade da sua evolução.

A complexidade provém primordialmente, do número de componentes enquanto que a complexidade de comportamento se deve ao tipo e ao grau de ligações. Assim sendo, a complexidade é função do número de constituintes, considerados elementares, do sistema.

A complexidade do sistema U , $C(U)$, é tal C que se:

$$U = U_1 \cup U_2$$

então:

$$C(U) = C(U_1) + C(U_2)$$

Sem dúvida que há um grande número de medidas do sistema capazes de satisfazer às condições anteriores. Mas, certamente, a mais conveniente é o próprio número de constituintes do sistema, N .

Assim, a medida mais eficaz da complexidade de um sistema em que os subsistemas U_1 e U_2 têm respectivamente N_1 e N_2 constituintes é dada por

$$C(U) = N_1 + N_2 = N$$

Esta medida de complexidade não está, de maneira alguma, directamente relacionada com a complexidade de comportamento. Basta notar que a complexidade é uma propriedade extensiva e linear; um comportamento altamente complexo é não linear.

No entanto, tem que existir entre eles alguma relação. Essa relação é feita pela quantidade de informação do sistema, que não depende do número de constituintes, mas do tipo de ligações e da sua diversidade. Assim, um sistema complexo em que existe simplicidade de informação não pode ter comportamento complicado. Se o mesmo sistema tiver uma grande quantidade de informação torna-se num sistema de comportamento complexo.

A complexidade de um supermercado arrumado e a complexidade de um supermercado desorganizado é a mesma. No entanto, a variação da complexidade de comportamento do primeiro, relativamente ao segundo é muito grande. A diferença reside na quantidade de informação do supermercado desorganizado ser maior do que a do supermercado em que os produtos estão no seu lugar.

Também um cristal que funde tem a mesma complexidade que o líquido em que se transformou. Mas, o líquido tem uma complexidade de comportamento muito maior quando comparado com a do cristal. A variação da quantidade de informação é responsável pela alteração da complexidade do comportamento.

Por outro lado, é possível incrementar a complexidade do sistema mantendo uma baixa complexidade de comportamento. Basta para tal que a quantidade de informação seja mantida a níveis baixos.

A evolução biológica é o exemplo mais acabado deste facto. A evolução tem sido sempre no sentido de sistemas mais complexos, associada a uma cada vez maior simplicidade de informação.

Tudo se passa como se se pudessem construir diagramas de fase de comportamento. Em vez de ser, por exemplo, pressão, P , e temperatura, T , nos eixos passaríamos a ter complexibilidade e quantidade de informação.

Como ilustração consideremos o caso da água nas suas três fases. Se na figura 26 onde está pressão P , e temperatura T , passarem a estar complexidade e a quantidade de informação, o incremento da complexidade de comportamento coincide com as diferentes fases.

A complexidade de comportamento para o uso de moléculas da água na fase vapor é maior do que na água líquida e a desta é, por exemplo, inferior à do gelo.

Não quer isto dizer que, internamente, a cada um destes estados não correspondam, em sentido inverso, ligações cada vez mais intrincadas. De facto, as ligações que mantêm o gelo coeso são mais difíceis de descrever do que as que se estabelecem no vapor de água. Isto mostra a relatividade da noção de complexidade de comportamento.

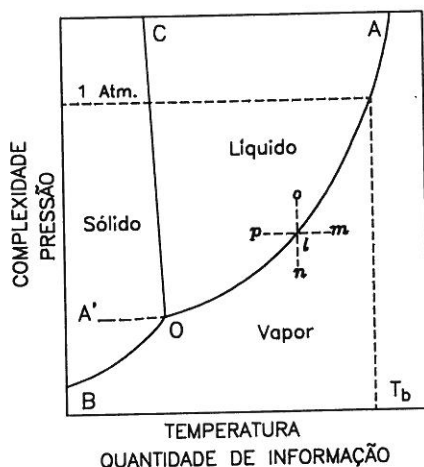


Fig. 26 - Diagrama das fases do vapor de água.

Para o observador externo, o vapor de água tem um comportamento mais imprevisível do que o gelo. No entanto, para a molécula de OH_2 é mais difícil manter as ligações no cristal de gelo, com permuta de electrões, vibrações de rede cristalina, acerto de ângulos entre moléculas do que à mesma molécula vagar ao sabor de correntes de convecção e agitação térmica no vapor de água.

Neste sentido, a complexidade definida é em termos absolutos enquanto que a complexidade de comportamento aparece como um fenómeno altamente relativista.

A relação que possa medir, para um observador externo ao sistema, a sua complexidade de comportamento e a sua manutenção tem que ver com a quantidade de informação medida pelo observador.

Sendo a estrutura de um sistema a forma como as diferentes partes que o compõem estão dispostas umas em relação às outras, podemos definir propriedades dessa estrutura.

Se chamarmos coesão à capacidade de uma estrutura de um sistema se manter com um determinado comportamento, o estabelecimento de uma função de coesão para uma estrutura de um sistema corresponde ao estabelecimento dos parâmetros que a mantêm e a destroem.

2. FUNÇÃO DE COESÃO DE ESTRUTURAS LIGADAS POR INFORMAÇÃO

2.1. DA INFORMAÇÃO À COESÃO

Consideremos a possibilidade de conhecer o grau de coesão de estruturas mantidas por informação. É o caso de todas as estruturas de sistemas, em particular as geradas pelos seres vivos designadamente as estruturas sociais. A sua manutenção depende de um fluxo apropriado e ajustado de informação. O seu progresso requer um acréscimo de possibilidades do número de fontes de informação. A fragilidade de tais estruturas humanas é, historicamente, bastante evidente. Os sociólogos falam, entre outros conceitos, em compreensão, espírito do corpo, solidariedade.

A quantificação do grau de coesão de uma estrutura, que não é mantida pelas forças normalmente consideradas quando se trata de obtenção das leis físicas de sistemas cujos constituintes são ligados pelo espaço, tempo e energia pode ser explorada a partir da quantidade de informação.

O problema consiste no cálculo da coesão de uma estrutura, que existe num sistema, numa situação de alternativa a outra também possível para o mesmo sistema. O problema aqui tratado diz respeito às situações em que há duas estruturas alternativas presentes. Para as estruturas sociais implementam-se, conforme o sistema em causa, conceitos para significar as duas alternativas sempre presentes num sistema.

Assim, na sociedade humana existem a cidade e a anarquia, o exército e o bando de homens, a evolução no sentido do bem estar e a degradação.

Sabemos que existe uma variedade apreciável de definições do que é informação. Há, mesmo, um conjunto numeroso de medidas de quantidade de

informação. Sendo a informação uma variável tão fundamental como o espaço, o tempo e a energia, existe uma grande dificuldade na sua definição tal como acontece quando se quer dar uma definição do que é o espaço e o tempo. A informação, está como estas últimas variáveis, imersa na realidade.

A informação é medida pela quantidade de informação tal como foi definida por Shannon. A função de coesão obtém-se por comparação entre a variação da quantidade de informação a que o sistema se sujeita quando se dá uma transição do estado dos seus elementos constituintes e a quantidade de informação que um observador exterior obtém quando tal acontece.

2.2. DEFINIÇÃO DA FUNÇÃO DE COESÃO

Um sistema construído por N elementos todos discerníveis, e independentes, num determinado estado, tem potencialmente em si uma entropia máxima, $H_{\max}(N)$ dada por:

$$H_{\max}(N) = \log N \quad (1)$$

Se σ desses elementos sofrerem uma transição de estado, que pode ser a sua retirada do sistema, o novo máximo de entropia para o sistema, $H_{\max}(N, \sigma)$, é dado por:

$$H_{\max}(N, \sigma) = \log(N - \sigma) \quad (2)$$

A diferença $H_{\max}(N) - H_{\max}(N, \sigma)$ representa a perda de quantidade de informação quando o sistema passa de N para $N - \sigma$ constituintes.

$$\Delta H_{\max}(N, \sigma) = -\log\left(1 - \frac{\sigma}{N}\right) \quad (3)$$

$\Delta H_{\max}(N, \sigma)$ é uma grandeza própria do sistema.

Quanto maior for $\Delta H_{\max}(N, \sigma)$ maior é a dificuldade do sistema para actuar como fonte de informação (maior dificuldade em se fazer notar).

Mas quando num sistema existe uma estrutura, a quantidade de informação observada $H_{\text{obs}}(N, \sigma)$, quando σ constituintes mudam de estado é forçosamente superior a $\Delta H_{\max}(N, \sigma)$. Essa diferença é a marca da existência da estrutura.

Então, quando se verifica a condição:

$$H_{\text{obs}}(N, \sigma) > \Delta H_{\max}(N, \sigma) \quad (4)$$

existe no sistema uma estrutura coesa. Nestas condições a quantidade de informação que a transição de estado gera quando há estrutura é superior à que seria de admitir se todos os constituintes não estivessem ligados numa estrutura com coesão. Assim sendo, a função,

$$\chi(N, \sigma) = \frac{H_{\text{obs}}(N, \sigma)}{\Delta H_{\text{max}}(N, \sigma)} - 1 \quad (5)$$

descreve a evolução da coesão de uma estrutura mantida por informação com σ . Quando:

$$\chi(N, \sigma) > 0 \quad (6)$$

a estrutura mantida por informação está coesa. Quanto maior for $\chi(N, \sigma)$, maior é o seu grau de coesão.

Pelo contrário, sempre que,

$$\chi(N, \sigma) < 0 \quad (7)$$

a estrutura coesa é substituída por uma estrutura sem coesão.

2.3. QUANTIDADE DE INFORMAÇÃO OBSERVADA EM DUAS ESTRUTURAS ALTERNATIVAS DE UM SISTEMA

As estruturas alternativas num sistema aparecem pela complementariedade dos estados dos elementos constituintes que mutuamente se excluem. Sempre que haja dois estados possíveis, estado 1 e estado 2, com probabilidades p_1 e p_2 de existirem, se forem alternativos,

$$p_1 + p_2 = 1 \quad (8)$$

Geram-se, assim, duas estruturas cujas quantidades de informação observadas têm o comportamento que mostra a figura 27 (a) e a figura 27 (b)

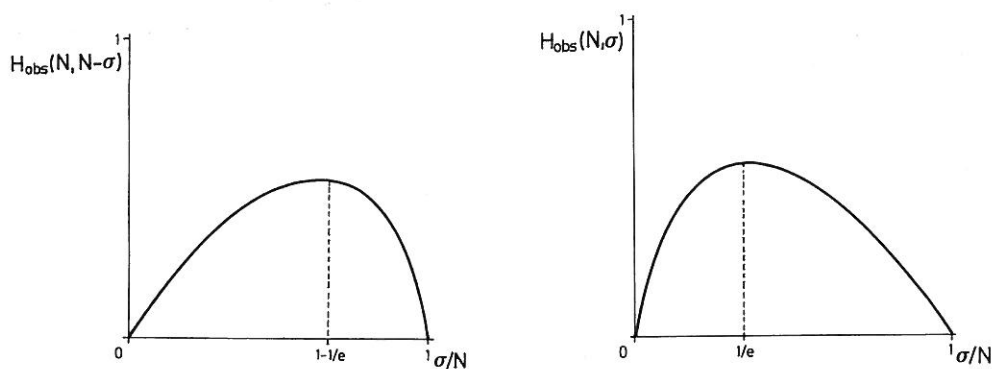


Fig. 27 - Quantidade de informação da estrutura com probabilidade de elementos constituintes p_1 (figura 27 (a)) e p_2 (figura 27 (b)).

Da soma das curvas da figura 27 resulta a bem conhecida curva de Shannon (Fig. 28);

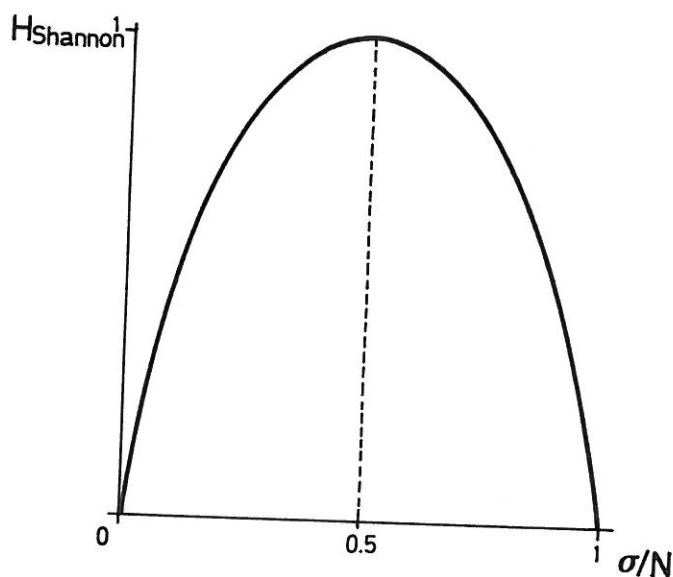


Fig. 28 - A soma das quantidades da função resultante das duas estruturas alternativas dá a curva de Shannon.

No caso de haver σ elementos removidos do conjunto N de elementos no estado 1, a quantidade de informação observada da estrutura de que faziam parte é dada por:

$$H_{\text{obs}}(N, \sigma) = - \frac{(\sigma)}{N} \log \frac{(\sigma)}{N} \quad (9)$$

A quantidade de informação observada da estrutura alternativa para os mesmos elementos é dada por:

$$H_{\text{obs}}(N, N - \sigma) = - \frac{N - \sigma}{N} \log \frac{N - \sigma}{N} \quad (10)$$

3. COMPORTAMENTO DA FUNÇÃO DE COESÃO PARA O CASO DE DUAS ESTRUTURAS ALTERNATIVAS

Substituindo na função de coesão, definida na equação (5), a função H_{obs} pela quantidade de informação definida pelas equações (9) e (10) resultam as seguintes funções de coesão para as duas estruturas alternativas:

$$\chi(N, \sigma) = \frac{\frac{\sigma}{N} \log \frac{\sigma}{N}}{\log \left(1 - \frac{\sigma}{N}\right)} - 1 \quad (11)$$

e

$$\chi(N, N - \sigma) = \frac{\frac{N - \sigma}{N} \log \left(1 - \frac{\sigma}{N}\right)}{\log \left(1 - \frac{\sigma}{N}\right)} - 1 = - \frac{\sigma}{N} \quad (12)$$

A equação 12 dá para a função de coesão valores positivos para probabilidades de perda dos elementos constituintes do sistema inferiores a 0,3. Para taxas de perdas superiores a este valor a função de coesão toma valores negativos (Figura 29).

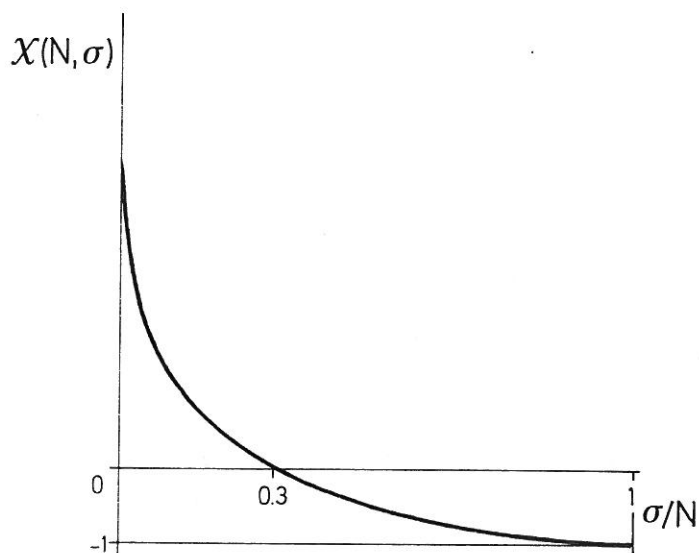


Fig. 29 - A função de coesão fica negativa para valores de σ/N superiores a 30%.

Da equação 13 resulta que $\chi(N, N - \sigma)$ tem sempre valores negativos seja qual fôr o valor de σ e de N , desde que $\sigma \neq N$ (Figura 30).

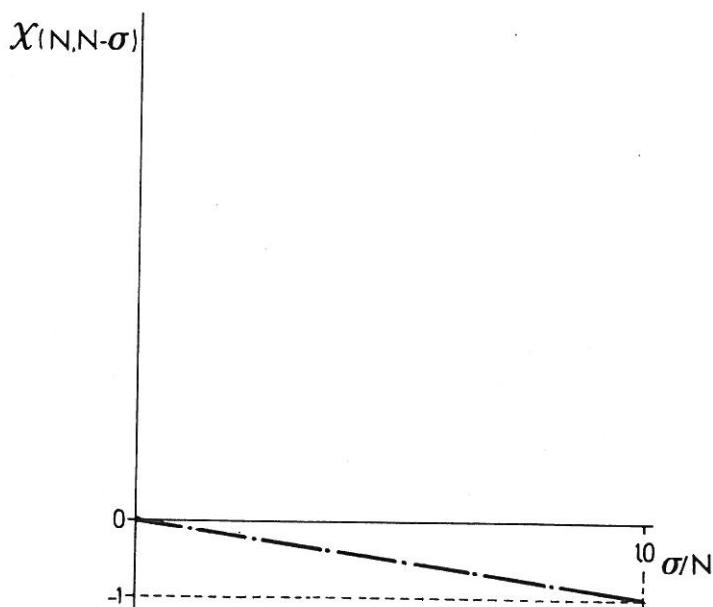


Fig. 30 - A função de coesão da estrutura alternativa resultante da perda de elementos do sistema é sempre negativa.

A comparação da coesão $\chi(N, \sigma)$ e de $\chi(N, N - \sigma)$ (figura 31) mostra que a coesão da estrutura ligada por informação a partir de uma perda de 50% dos elementos fica irreversivelmente sujeita à degradação.

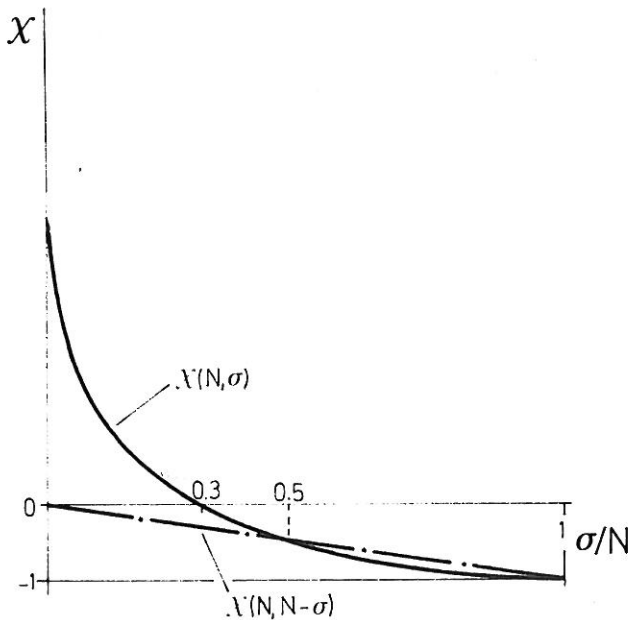


Fig. 31 - Comparação entre $\chi(N, \sigma)$ e $\chi(N, N - \sigma)$

A função de coesão proposta mostra que se um sistema fôr constituído por elementos que podem ocupar dois estados que se auto-excluem, no sentido de que os elementos constituintes podem ser retirados definitivamente da estrutura, a informação gera uma estrutura com coesão. Essa estrutura existe simultaneamente, com outra, sem coesão. Esta é alternativa da estrutura com coesão e substituí-la-à logo que a percentagem de perdas de elementos constituintes seja superior a trinta por cento.

A função de coesão mostra, também, o carácter irreversível desta transição num sistema de uma estrutura coesa para a estrutura sem coesão. De facto, logo que a função se torna negativa é impossível recuperar a estrutura com coesão, a menos que novos elementos sejam trazidos para o sistema do seu Universo complementar.

O essencial, para manter coesa uma estrutura ligada por informação, é que o número total de elementos constituintes seja mantido pelo menos inalterável. Para aumentar o grau de coesão aquele número deve crescer.

Portanto, as equações da entropia referidas nas equações 9 e 10 podem ser entendidas como medidas da entropia da coesão de estruturas.

APÊNDICE - I

NÚMEROS EXTREMAMENTE GRANDES E EXTREMAMENTE PEQUENOS E A SUA REPRESENTAÇÃO EM POTÊNCIAS DE 10

Temos vindo a usar, nesta exposição, números extremamente grandes e números extremamente pequenos. Parece que será conveniente dedicar alguns momentos de reflexão à sua análise, para obter uma noção, mais clara, das ordens de grandeza que eles representam.

Aqueles números são, geralmente, expressos em potências de 10.

Na análise recorre-se a exemplos da vida comum, para melhor elucidação.

Suponhamos quatro bolas iguais, animadas de movimento e distribuídas uniformemente numa caixa. Poderão as bolas espontaneamente, concentrar-se todas na mesma metade? Qual é a probabilidade de que tal aconteça?

Ora, a probabilidade de que uma bola esteja numa metade da caixa é de $1/2$; logo, a probabilidade de que as quatro estejam na mesma metade é de $1/2 \times 1/2 \times 1/2 \times 1/2 = 1/16 = (1/2)^4$. Suponhamos que em vez de bolas se tem moléculas. A probabilidade de as N moléculas ficarem na mesma metade da caixa seria $1/2 \times 1/2 \times 1/2 \times \dots = (1/2)^N$ isto é, um só caso provável em 2^N casos possíveis. Suponhamos que $N = 10^{20}$ (moléculas contidas num volume aproximado de 4 cm^3). Então a probabilidade neste caso seria 1 contra;

$$2^N = 2^{10^{20}} = 10^{3 \times 10^{19}}$$

Este número é tão grande, que desafia a imaginação humana. Este problema foi posto por Eddington (1927).

Por isso, dentre $2^4 = 16$ fotografias, tiradas com intervalos de tempos iguais (filmagens com 16 imagens por segundo, por exemplo), em média, só há uma em que as quatro bolas estariam na mesma metade da caixa.

E se fossem 40 bolas? Agora o caso já é diferente e é muito raro. Em média, só uma dentre $2^{40} = 10^{12}$ fotografias é que mostrariam todas as bolas na mesma metade da caixa.

E se se tratar de um gás ordinário? Neste caso, em média, só uma "fotografia" dentre as $2^{10^{23}}$ é que mostraria todas as moléculas na mesma metade da caixa. De facto, o número total de "bolas" seria, não de $4 \times 10^1 (= 40)$, mas da ordem de $6,0 \times 10^{23}$ (número de Avogadro). Para "apanhar" a imagem em que as moléculas estivessem todas na mesma metade, teríamos que correr um filme que, feito nas condições actuais de filmagem, daria uma sessão extremamente longa; duraria várias vezes a idade da Terra.

∴

Sabendo que o número de Avogadro é de $6,02 \times 10^{23}$ moléculas por mole; calculemos o número de moléculas de azoto contidas num litro de ar (1,15 g). Este é dado por:

$$6,02 \times 10^{23} \times 1,15/28 = 2,47 \times 10^{22} \text{ moléculas por litro.}$$

Quando se tratou do número de microestados acessíveis, os números eram ainda muitíssimo elevados.

∴

Mas que ideia fazemos nós da grandeza de tais números? As potências de 10, tão usuais e tão úteis na Física, podem ser muito enganadoras, se não forem meditadas.

Vamos ver. Suponhamos que se tem 25 000 contos em moedas de 25\$00. Qual será a altura da pilha de todas as moedas? Temos portanto 10^6 moedas; se a espessura de cada uma fôr de 2 mm a altura da pilha é de 2 000 m. E se em vez de 10^6 moedas tivéssemos 10^{12} moedas? A pilha teria uma espessura de 2 000 000 Km, isto é, cinquenta vezes o perímetro do Equador...

Vejamus outro exemplo, elucidativo, sobre a "traição" das potências de 10. Consideremos uma caixa cúbica de 2,5 cm x 2,5 cm x 2,5 cm. Pois bem, suponhamos que, cheia de ervilhas, leva 100 ervilhas. Logo, 10^6 ervilhas encheriam uma arca razoável; 10^9 ervilhas, uma casa; 10^{12} uma cidade de 1 000 casas; 10^{15} uma cidade com 1 000 000 de casas, com dimensões superiores à de uma grande Lisboa. Um aguaceiro de 10^{18} grânulos de granizo com as dimensões de grãos de ervilhas, que caíssem sobre a Península Ibérica e se distribuisse

uniformemente, cobri-la-ia, toda, com a altura média de 4 m. E se fosse um aguaceiro de 10^{21} grânulos? Cobriria todos os Continentes da Terra, com uma espessura da ordem de 3 m. E com um aguaceiro de 10^{24} ? Cobriam-se 250 planetas como a Terra. Com um aguaceiro de 10^{27} cobrir-se-iam 250 000 planetas como a Terra.

Agora pode fazer-se uma ideia do número de Avogadro. Equivale a um aguaceiro de ervilhas capaz de cobrir $6 \times 25 = 150$ planetas como a Terra. De forma análoga se podia concluir que o número de átomos do corpo humano ($\approx 10^{27}$ átomos) corresponderia a um aguaceiro capaz de cobrir 250 000 planetas.

Se as moléculas de ar contidas num copo vulgar ($0,25 \times 6 \times 10^{23} : 28,9 = 5 \times 10^{22}$) se transformassem em grãos de areia, seriam suficientes para cobrir a área dos Estados Unidos da América com uma espessura da ordem de 30 m.

Se para escrever as coordenadas duma molécula se gastasse 1 μ s, seriam necessários 6×10^{17} segundos ou $\approx 7 \times 10^{12}$ anos, para fazer o registo das moléculas que existem num mole da substância. Mas esta é a ordem de grandeza da idade do Universo!

∴

Devemos ter ainda uma palavra de cautela para números grandes que parecem muito semelhantes, como por exemplo, os números $10^{6,02 \times 10^{24}}$ e $10^{6,01 \times 10^{24}}$ mas que no fundo, são muito diferentes. Basta tentar escrever os expoentes, por extenso, e verificar que o primeiro número é 10^{1022} vezes maior do que o segundo!

Uma inalação humana é da ordem de 400 cm^3 e contém aproximadamente 10^{22} moléculas. A atmosfera tem cerca de 10^{44} moléculas. Portanto, uma molécula está para uma inspiração, como uma inspiração está para toda a atmosfera.

($10^{22} : 1 = 10^{44} : 10^{22}$). Ficou famosa a asserção que se divulgou e tanto se popularizou:

"Se o último suspiro de Aristóteles se difundiu por toda a atmosfera, há uma certa probabilidade de que inalemos uma molécula de Aristóteles cada vez que fazemos uma inspiração..."

∴

Vejamos, por fim, alguns exemplos de números pequenos.

O "diâmetro efectivo" dum electrão livre é da ordem de 10^{-11} a 10^{-12} cm. Com uma amplificação de 10^{12} ficaria, portanto, com as dimensões de uma bola de golf.

O espectro das partículas, que constituem o aerosol da atmosfera, distribui-se assim, se exprimirmos as suas dimensões em termos dum raio médio:

moléculas - 10^{-8} cm; iões médios - 10^{-7} cm; livre percurso médio das moléculas - 10^{-4} cm; gotas das nuvens 10^{-3} a 10^{-2} cm.

Notaremos que o limite da visão é expresso por uma resolução da ordem de 10^{-3} cm que define o pixel humano.

APÊNDICE - II

MEDIDA DA INFORMAÇÃO

1 - INFORMAÇÃO CONDICIONADA. ALGUNS EXEMPLOS

Quando lançamos ao ar uma moeda sabemos que pode sair "cara ou cunho". Estamos em presença de um problema que comporta duas possibilidades, em princípio igualmente prováveis e mutuamente exclusivos. Se, em vez duma moeda, considerarmos um dado, o número de respostas possíveis é agora 6, sempre igualmente prováveis. Se se tratar dum baralho de cartas, quando se tira uma certa carta "à sorte", há 52 possibilidades.

Em qualquer dos casos, o exame do resultado do lançamento da moeda, ou do dado, ou da tiragem de uma carta, fornece uma certa informação, porque se adquire um certo conhecimento.

Vejamos alguns exemplos, muito simples, recorrendo ao jogo de dados.

Quando se lança um dado a incerteza *a priori* desde que não se possua qualquer outra informação, é dada por:

$$H_1 = -k \log 1/6 = k \log 6$$

Suponhamos agora que depois do lançamento do dado se recebia a mensagem A de que o número que saiu "não é a sena". Logo, deduz-se que só podem ter saído as faces 1, 2, 3, 4 e 5 (todas com igual probabilidade). A incerteza com que se fica é agora:

$$H_2 = -k \log 1/5 = k \log 5$$

que é menor do que a incerteza do estado inicial. A quantidade de informação recebida com a mensagem A é então:

$$I(A) = H_1 - H_2 = k \log 6/5$$

Suponhamos, agora, que recebemos mais outra mensagem B que diz que a face é par. Então, como já se sabia que não tinha saído a sena, só há duas possibilidades, que são as faces 2 e 4, às quais corresponde uma probabilidade de $1/2$. A incerteza passa a ser:

$$H_3 = k \log 2$$

em vez do valor anterior:

$$H_2 = k \log 5$$

A informação da mensagem B, conhecida A, e que representaremos por $I(B/A)$, é, então, representada por:

$$I(B/A) = H_2 - H_3 = k \log 5/2$$

A informação, fornecida pela sucessão das mensagens A e B, isto é, $I(AB)$ é diferente de $I(B/A)$, e é dada por:

$$I(AB) = H_1 - H_3 = k \log 3$$

É fácil verificar que:

$$I(AB) = I(A) + I(B/A)$$

Suponhamos agora que recebemos a mensagem B antes da mensagem A. A incerteza seria, neste caso:

$$H'_2 = -k \log 1/3 = k \log 3$$

visto que as faces que poderiam sair (faces pares) seriam 2, 4 e 6. A informação associada à recepção da mensagem B, é agora:

$$I'_B = H_1 - H'_2 = k \log 2$$

que é diferente de $I(B/A)$, se a mensagem A já fosse conhecida. Das faces 2, 4 e 6, deve eliminar-se agora a face 6, devido à recepção da mensagem A. Depois desta mensagem, a incerteza será:

$$H'_3 = -k \log 1/2 = k \log 2$$

E verificamos, como dissemos, que $H'_3 = H_3$

A informação contida na mensagem A, quando vem depois de B, isto é, $I(A/B)$, é tal que:

$$I(A/B) = H'_2 - H'_3 = k \log 3/2$$

que não é a mesma que $I(B/A)$.

A informação contida na sucessão das mensagens B e, depois A, é

$$I(BA) = H_1 - H'_3 = k \log 3$$

Por isso, verificamos que a expressão

$$I(AB) = I(A) + I(B/A)$$

se há-de interpretar assim:

$I(AB)$ é a informação que se obtém quando se recebem as mensagens A e B e $I(B/A)$ é a informação condicional, que se obtém com a mensagem B, se se tiver observado A (é a técnica da dedução e da indução usada pelos detectives ... *"se isto se deu, então ... aquilo ..."*). Por exemplo $I(B/A) = 0$ significa que B é impossível se se tiver verificado A; se $I(B/A) = 1$ significa que há a certeza que B se realizará se se verifica A. É fácil ver:

$$I(BA) = I(AB) = I(B) + I(A/B)$$

o que nos permite concluir que a quantidade de informação contida com a recepção das mensagens A e B é independente da ordem pela qual as mensagens são recebidas.

2 - UNIDADES DE INFORMAÇÃO

Em regra, não se especificou o valor da constante arbitrária k que figura na fórmula que dá a incerteza. Essa escolha pode ser guiada por condições de comodidade e convém que seja uma grandeza sem dimensões.

Vamos tomar para unidade de incerteza, u_n , a incerteza associada à realização de um acontecimento de entre n acontecimentos possíveis e equiprováveis, isto é, em que $p = 1/n$. Então, por definição:

$$k \log_e n = 1$$

o que corresponde a fazer:

$$k = 1/\log_e n$$

Por isso, a informação $I(x)$ é dada, em unidades $[u_n]$ por:

$$I(x) = 1/\log_e n \cdot \log_e x = \log_n x [u_n]$$

visto que, como se sabe da mudança de base de logaritmos:

$$\log_e x = \log_e n \cdot \log_n x$$

Habitualmente, toma-se para base n a base 2. Neste caso as unidades $[u_2]$ designam-se por bits (unidades logarítmicas binárias), designação introduzida por Tuckey.

A introdução do sistema binário foi ditada, principalmente, pelo problema de transmissão de informações, que é realizada recorrendo à combinação de dois

sinais elementares, em que as regras de combinação são definidas por um código. Os dois sinais estavam associados a dois estados possíveis dum circuito eléctrico: passagem ou ausência de corrente, isto é, ligar e desligar um circuito; passagem duma intensidade a outra; inversão da tensão da corrente; passagem duma frequência a outra (em corrente alternada), etc. Por outro lado, a escolha da base 2 convém ainda porque, em muitas circunstâncias, temos que escolher entre a realização de dois acontecimentos alternativos de igual probabilidade (cara ou cruz; par ou ímpar; preto ou branco; sim ou não; + ou -, etc.). Daqui resulta que, se uma mensagem é composta por N símbolos, como cada um destes pode ter dois estados diferentes, que muitas vezes se designam por 0 e 1, o número total das possibilidades é:

$$P = 2^N$$

e, por consequência, a quantidade de informação contida na mensagem de N símbolos é:

$$I = k \log P = N k \log 2$$

Para simplificar, faz-se:

$$(k = \frac{1}{\log 2})$$

donde resulta:

$$I = N \log_2 P$$

Vemos, assim, que se introduz, duma forma natural, a numeração binária na medida da informação, que, portanto, passa a vir expressa em bits. Nos exemplos anteriores temos, no caso do lançamento ao ar duma moeda:

$$I = \log_2 2 = 1 \text{ bit};$$

no caso do lançamento do dado:

$$I = \log_2 6 = 2,6 \text{ bits};$$

no caso das cartas, quando retiramos uma carta dum baralho de 32 cartas, a quantidade de informação será:

$$I = \log_2 32 = 5 \text{ bits}$$

Como referimos, foram os problemas relativos à técnica de transmissões que levaram a adoptar, para a quantificação da informação, uma unidade baseada na numeração binária. Mas, é evidente, que se podem utilizar outras unidades $[u_n]$ com toda a legitimidade.

Assim, se mantivermos na expressão de I os logaritmos naturais e se fizermos $k = 1$, vê-se que:

$$I = \text{Log } P [u_e]$$

em que:

$$[u_e] = \text{Log}_2 e = 1,44 \text{ bits}$$

Se usássemos o sistema decimal seria

$$[u_{10}] = 3,32 \text{ bits, porque}$$

$$\text{Log}_2 P = \text{Log}_{10} P / \text{Log}_{10} 2 = 3,32 \text{ Log}_{10} P$$

Portanto, quando $k = 1$ as unidades $[u_n]$ são tais que para se exprimirem em bits basta multiplicar por $\text{Log}_2 n$.

Ora, como k pode ser arbitrariamente escolhido, podemos atribuir-lhe (e porque não?) o valor da constante k de Boltzmann. As grandezas S e H exprimem-se pela mesma fórmula e as variáveis P e W são análogas, visto que P mede o número de possibilidades ($P = 1/p$), correspondentes a um dado estado macroscópico. A unidade de informação vem, assim, idêntica à unidade da entropia e as grandezas S e H (ou I) exprimem-se por números iguais. A informação vem medida em unidades de entropia termodinâmica ("Boltzões") (energia/temperatura).

Se atendermos às expressões que dão a incerteza (entropia), expressas em Boltzões e em bits, verifica-se, atendendo a que $k = 1,38 \times 10^{-16} \text{ erg/k}$, que são dadas pelo mesmo número (puro), a menos dum factor 10^{-16} . De facto,

$$\begin{aligned} 1 \text{ boltz} &= k \text{ Log}_e 2 \text{ Log}_2 (2) = k \text{ Log}_e 2 \text{ bit} = 1,38 \times 0,6931 \times 10^{-16} \text{ bit} = \\ &= 9,565 \times 10^{-17} 10^{-16} \text{ bit} \end{aligned}$$

Logo:

$$1 \text{ bit} = 1,045 \times 10^{16} \text{ boltz}$$

APÊNDICE - III

REDUNDÂNCIA

Quando se está a receber uma mensagem, a probabilidade da parte que ainda não foi recebida é em geral dependente do que já se sabe da mensagem. Por exemplo, em português, sabe-se que a seguir à letra *q* vem sempre um *u* e, portanto, não há probabilidade de vir qualquer outra letra. Logo, a informação recebida com a recepção de um *u* a seguir ao *q* é, evidentemente, nula. Mas pode acontecer que haja casos menos extremos, ainda que se esteja a receber menos informação do que os que o precedem, se fossem independentes.

Suponhamos que os sinais ("notícias") emitidos por uma fonte de informação são independentes, e que se calculou a sua entropia (ou informação, ou liberdade de escolha). Podemos comparar este valor com a entropia máxima sujeita apenas à condição da fonte emitir os mesmos sinais. Se não forem independentes o valor médio informativo será S' , em geral, inferior a S . O valor da redundância é dado por:

$$R = (S - S') / S = 1 - S' / S$$

A redundância é, portanto, igual à unidade menos a entropia relativa, que é a razão entre a entropia actual da fonte e a entropia máxima. A redundância é a fracção da estrutura duma mensagem que não pode ser determinada pela livre escolha dos símbolos e que é condicionada pelas "regras" que regem o emprego dos símbolos em questão. Em geral, a fracção redundante duma mensagem é desnecessária do ponto de vista da teoria da comunicação. É o que se faz quando se redige um telegrama em que se procura evitar a redundância e usar o máximo de dados, pagando o preço mínimo possível.

APÊNDICE - IV

**A EVOLUÇÃO DA LÍNGUA PORTUGUESA
(TRABALHO PARA A CADEIRA DE TERMODINÂMICA
GENERALIZADA EM 1987 POR NUNO BARRADAS)**

PROPOSIÇÃO

" ... Dispus portanto de uma oportunidade de executar a tarefa monótona e longa que é escrever longos textos num computador, ao qual deixei a tarefa ainda mais monótona, mas mais rápida, que é de calcular as frequências das letras que aparecem nesses textos, de modo a calcular a sua entropia, numa primeira ordem de aproximação, através da conhecida fórmula $S = -\sum f_i \log f_i$ (que não irei demonstrar nem tão pouco explicar; é demasiado bem conhecida, e poupo assim uma outra tarefa aborrecida, a mim que me dispense de o fazer, e ao caro e presumível leitor que se dispensa de a rever). Não me preocupei em considerar as correlações existentes na língua portuguesa (a seguir a um q vem quase sempre um u, antes de um b ou p quase nunca vem um n, etc.).

Assim, medi a entropia de textos portugueses (ou quase), escritos por autores portugueses (ou quase), ao longo dos séculos, tracei uma curva entropia vs tempo desde o séc. IX até aos nossos dias, obtendo uma curva contínua (ou quase). Suponho que estes "(ou quase)" sejam tão intrigantes como a existência de textos portugueses do séc. IX, mas tudo será explicado a seu tempo.

NARRAÇÃO

Contar à mão as frequências das várias letras seria demasiado longo, aborrecido e fatigante. Portanto, usei um computador para o qual escrevi um pequeno programa, e no qual introduzi os textos.

Quanto às próprias letras, usei um alfabeto alargado; 27 minúsculas e 26 maiúsculas, diversos sinais e pontuação (: ; , ` ^ ~ ! ? " , etc.).

Como é óbvio, usei os textos originais. Os resultados estão dados em anexo. Naturalmente que a entropia de uma cantiga (de um amigo) é diferente conforme se usar português arcaico ou moderno; no entanto, não tive acesso aos textos respeitantes a um largo período de tempo (entre 1550 e 1900). A curva que traduz os resultados obtidos foi no entanto apenas ligeiramente afectada; o seu aspecto é razoavelmente claro. Por outro lado considere também textos em latim bárbaro, em uso antes de haver um Reino de Portugal propriamente dito, textos esses que ajudam um pouco à visualização do gráfico como um todo (e isto explica ou "ou quase").

Também é óbvio que numa mesma época nem todos os textos têm a mesma entropia; há variações devidas à individualidade de cada autor, tamanho dos textos, ou mesmo mero acaso; obtêm-se não uma linha, mas uma estrada, uma banda de entropia.

Para tamanho dos textos escolhi mil palavras por texto, onde existissem textos tão grandes, número que por experiência verifiquei ser suficiente para não aparecerem grandes variações nos valores da entropia (2a. decimal constante). Eis um exemplo típico:

José Cardoso Pires:

O hóspede de Job

Palavras	Letras	Entropia
83	502	2,891
198	1196	2,900
290	1811	2,945
450	2731	2,961
523	3215	2,985
616	3779	2,981
728	4436	2,981
844	5102	2,990
934	5573	2,993
1000	5951	2,992

No entanto, em muitos casos os textos disponíveis eram pequenos, e então usei os textos completos; por exemplo, na poesia trovadoresca, período Alfonsino (1245-1280), usei três textos:

Palavras	Letras	Entropia
145	713	2,824
95	500	2,828
220	1036	2,810

E, na representação gráfica apareceria um rectângulo com lados entre 1245 e 1280 e entre 2,81 e 2,83. Isto porque não teria significado referir autores individuais com base em tão poucos dados.

Os textos que não têm referido o autor são textos diversos, que se podem encontrar no livro de consulta 6 (cfr. apêndice das obras usadas), por exemplo, leis da época ou cantigas d'amigo de autores de uma mesma época. Note-se ainda que procurei sempre usar, na medida do possível, textos semelhantes (romances, poemas).

Por esta altura já os leitores terão imaginado o gráfico, e reparado que no período 1550-1850 a banda traçada está abaixo dos valores obtidos; isto é, devido a ser esse o período em que usei as versões modernas dos textos da época, e o português moderno obviamente modificou a sua entropia. Não o suficiente para não nos apercebermos da continuidade da curva que aparece se reduzirmos Camões aos valores da época, o que é duplamente justificado por

a elevada entropia do português de hoje levar a um aumento da entropia desses textos de antanho.

Vamos então ao exame, e um pouco, apenas um pouco, de interpretação dos resultados obtidos:

- 850 - 1100 Tive acesso a muito poucos textos de latim bárbaro da zona do futuro Portugal; não tendo razões em contrário, pareceu-me que um valor constante da banda de entropia (2,79/2,85) é compatível com os resultados obtidos, integrando-se bem como os pontos seguintes; corresponde a não haver praticamente alteração na língua escrita desta época.
- 1100 - 1375 A formação de Portugal e consequente desenvolvimento da língua deu origem a uma maior diversificação das formas escritas e consequente aumento da entropia, a uma taxa razoavelmente constante.
- 1375 - 1500 Atingido o máximo entre 1350 - 1400, a banda de entropia tem então um ligeiro decréscimo até 1500/25, que eu não sei explicar (talvez reflectindo o esgotamento das possibilidades da poesia e cultura trovadoresca).
- 1500 - 1800 Com o Renascimento e Cultura Clássica a língua portuguesa escrita diversifica-se (por exemplo, a pontuação desenvolve-se), e a banda de entropia sobe com uma taxa de crescimento um pouco superior à do período 1100 - 1375.
- 1800 - 1900 O crescimento continua mas com uma taxa mais elevada, reflectindo a literatura Romântica extremamente desordenada, cheia de "ais" e de "uis", pontos de exclamação e reticências. O máximo absoluto é atingido com "Os Maias" do Eça, pós-Romântico, mas extremamente vivo e imprevisível.
- 1900 - 1988 A banda de entropia desce tão rapidamente quão rapidamente subiu no período anterior; não me perguntem porquê: não tenho resposta.

Tenho ainda alguns reparos a fazer:

1. Gil Vicente apresenta um valor de entropia claramente superior aos valores correntes da época; eis duas explicações alternativas, à escolha:

a) Gil Vicente evidenciou-se entre os seus contemporâneos, tanto na escrita como na entropia, e, b) o teatro está mais próximo da língua falada que da língua escrita; ora ninguém escreve como fala, e vice-versa.

2. Fernando Pessoa apresenta valores de entropia para diversos textos que alcançam uma gama muito grande; isto dever-se-á à conhecida versatilidade do poeta.

Note-se que as duas excepções encontradas à banda de entropia são dois dos nossos maiores escritores. Se tem que haver excepções, é bom que assim seja. Seria interessante calcular as entropias de muitos outros autores, e verificar se obedecem a esta regra.

3. 1800 é um "caso bicudo"; a sua banda de entropia tem o dobro da amplitude usual; por outro lado, se reduzir a entropia de Camões aos valores da época é ponto pacífico, o mesmo não se passa com o facto de, em consequência disso, o declive entre 1500 e 1900 ser maior que o que obteríamos usando estritamente os pontos a negro; note-se que esse maior declive vem do meu desejo de tornar a curva contínua. Ponho à consideração dos meus ilustríssimos leitores a seguinte alternativa: existiria uma descontinuidade na banda de entropia por volta de 1800, decorrendo do corte que a literatura Romântica fez com a Clássica. Reduzindo Camões ao valor da época, mas mantendo o declive da banda de entropia até 1800, obtém-se a representação da figura 32.

A indefinição da transição é aumentada pelo simples facto de que houve um certo período razoavelmente longo em que coexistiram as duas correntes literárias; o pré-Romântico Bocage estava rodeado de Clássicos, e mesmo durante o ultra-Romantismo havia ainda quem se agarrasse à escola velha. Para esclarecer a situação seria necessário, em estudo posterior, calcular a entropia de muitos autores, desde 1700 até 1900, usando o português de então, marcando em separado, por um lado, os escritores clássicos e por outro os românticos.

4. No período moderno, é de notar que há valores no limite superior da banda de entropia, valores no limite inferior, e nenhum no inferior; faz pensar se não haverá duas correntes principais na literatura portuguesa, uma de alta entropia (proveniente dos expansivos Românticos?) e outra de baixa entropia (ligada aos mais sóbrios Clássicos?).

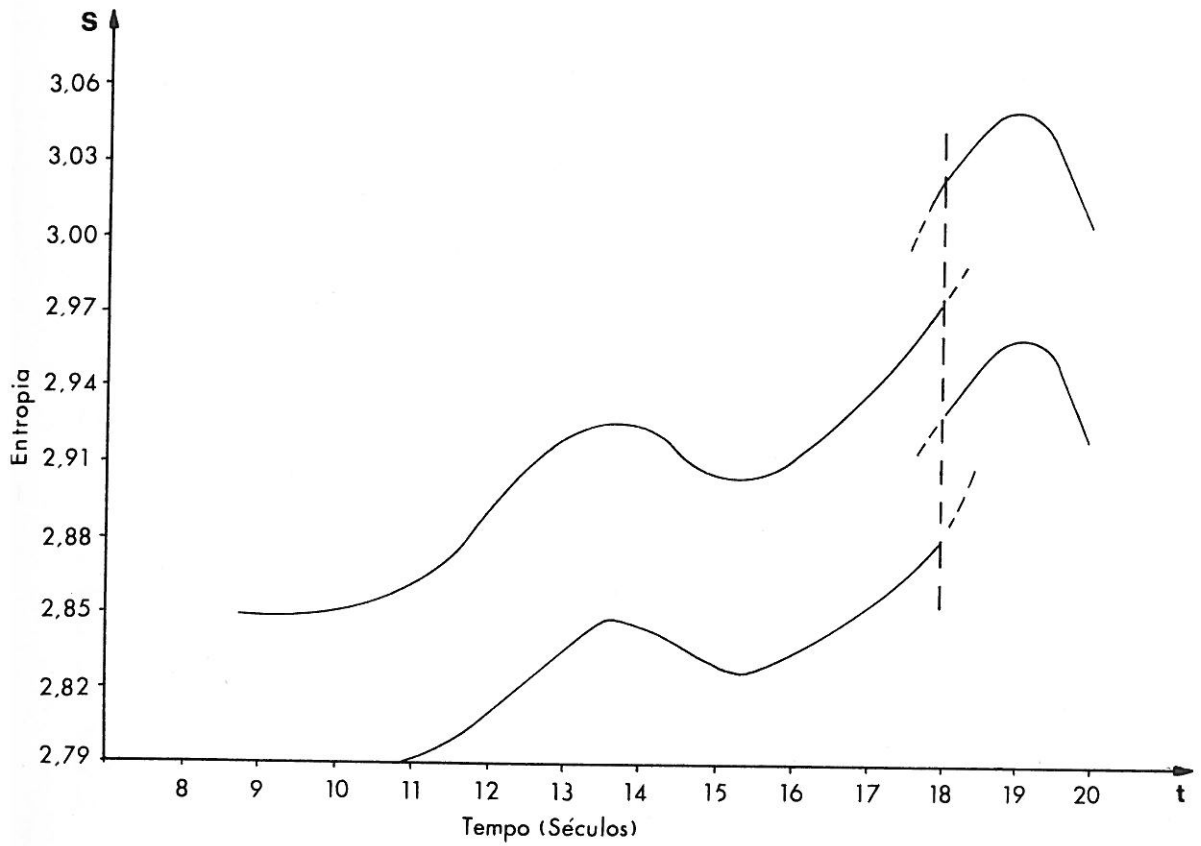


Fig. 32 - Mais uma vez apenas um estudo mais completo, calculando a entropia das obras de muitos autores deste século considerados representativos poderá eventualmente solucionar o problema.

CONCLUSÃO (OU A MORAL DA HISTÓRIA)

Cheguei ao fim [...]

O essencial deste trabalho é poder-se construir uma representação gráfica. O modo como é feita a sua interpretação, as correcções a introduzir são aspectos secundários. O simples facto de se poder conceber este gráfico já é importante, e seria interessante fazer o mesmo com outras línguas, sujeitas a uma evolução cultural semelhante, e estabelecer uma comparação.

Quer dizer, importante propriamente dito não é, mas no mínimo é agradável ao espírito pensar que, até mesmo na evolução de uma língua a Termodinâmica tem uma palavra a dizer (ainda que seja uma palavra curta); é algo que está nos limites do seu domínio de aplicação.

Finalmente, fiquei com a curiosidade (parcialmente) satisfeita e as férias arruinadas.

RESULTADOS

TEXTO	Palavras	Letras	Entropia
870	412	2757	2,846
959	109	775	2,816
1026	211	1264	2,790
1090	82	1529	2,849
1192 1º Texto	197	1133	2,886
Período pré-Alfonsino (1190-1245)	117	600	2,812
	68	329	2,843
	105	554	2,855
1200 Notícia de torto	497	2632	2,890
1211 Lei	118	713	2,822
Período Alfonsino (1245-1280)	145	713	2,824
	95	500	2,828
	220	1036	2,810
Período dionisíaco (1280-1300)			
(D. Dinis)	224	1014	2,882
	149	695	2,898
	134	668	2,819
Período pós-dionisíaco (1300-1360)	194	944	2,866
	91	424	2,860
	191	956	2,900
1360-1390 (período galego-castelhano)	152	836	2,899
1365-1400 " " "	136	724	2,934
1375 " " "	155	858	2,890

TEXTO	Palavras	Letras	Entropia
1309	135	735	2,914
1357	243	1310	2,852
1407	505	2788	2,834
Fernão Lopes - Mestre d'Aviz (1434-1454)	658	3768	2,915
Pero Vaz de Caminha (1500)	1018	5403	2,831
Francisco de Sousa (1500-1530)	67	400	2,879
Garcia de Resende (1502-1516)	357	2019	2,884
	574	3435	2,860
Gil Vicente - Auto da Fé (1520-1530)	641	3721	2,986
Camões - Sonetos (1540-1557)	84	507	2,903
	90	510	2,930
	100	509	2,886
	87	502	2,853
Damião de Góis - Crónica de D. Manuel (1560)	879	4944	2,913
Rodrigues Lobo - Pastor Peregrino (1610)	500	2686	2,859
- O Jumento com pele de Leão	95	519	2,886
Frei Luís de Sousa (1620)	426	2468	2,926
D. Francisco Manuel de Melo - Sonetos (1630-1660)	94	526	2,947
	87	515	2,965
	68	404	2,962
P. António Vieira - Sermões (1640-1680)	138	763	2,914
P. Manuel Bernardes - Nova Floresta (1680/71)	224	1301	2,970
P. António Garção - Cantata de Dido (1750/70)	294	1890	2,940

TEXTO	Palavras	Letras	Entropia
Tomás Gonzaga – Marília de Dirceu (1770)	296	1668	2,936
Nicolau Tolentino – Autobiografia (1770/1800)	224	1288	2,998
– Contra os toucados	97	533	3,001
Francisco Manuel do Nascimento – A Língua Portuguesa e a Latina (1770/1810)	394	2439	2,963
Bocage – Crinaura (1780/1799)	531	3273	3,024
– Sonetos (A Camões)	90	527	2,947
(A Adamastor)	81	506	2,822
(Ditado da Agonia)	90	513	2,990
Marquesa de Alorna – Sonetos (1780/1820)	82	513	2,904
	87	522	2,930
Soares de Passos – Firmamento (1856)	779	4315	2,979
Camilo Castelo Branco – A Queda de um Anjo	1000	5928	3,021
Antero de Quental – Carta ao Duque d'Ávila (1871)	1002	6108	3,013
Eça de Queiroz – Os Maias (1880-1888)	1000	6109	3,048
Mário de Sá Carneiro – Últimos Textos (1913/16)	347	1941	3,034
Ricardo Reis (1916)	566	3107	2,892
Fernando Pessoa (1931)	195	1025	2,891
	109	545	2,854
	79	446	2,949
Manuel da Fonseca – Aldeia Nova (1942)	1000	5463	2,949
Miguel Torga – Novos Contos da Montanha (1945)	1000	6068	3,016

TEXTO	Palavras	Letras	Entropia
Armando Rodrigues – Odes ao Tejo (1951) 1ª	415	2275	2,930
10ª	201	1466	2,941
José Cardoso Pires – O hóspede de Job (1963)	1000	5951	2,993
– Alexandra Alpha (1987)	1000	4934	3,009
José Saramago – O ano da morte de Ricardo Reis (1984)	1000	5922	3,004
Mário de Carvalho – A Paixão do Conde de Fróis (1987)	1000	6316	2,923

OBRAS DE CONSULTA

1. Miguel Torga - *Novos contos da montanha* - 4ª. ed. 1959 Coimbra
2. Mário de Carvalho - *A paixão do Conde de Fróis* - 1987, Ed. Rolim
3. Pêro Vaz de Caminha - *Carta a El-Rei D. Manuel* - 1974 INCM
4. Manuel da Fonseca - *Aldeia Nova* - 6ª. ed. 1978 Forja
5. Fernão Lopes - *Quadros da crónica de D. João I* - 6ª. ed. 1945
6. C. de Oliveira e S. Machado - *Textos portugueses medievais* - 2ª. ed. 1957 Atlântica
7. Gil Vicente - *Obras Completas Vol. I* - 1958 Livraria Sá da Costa
8. Camilo Castelo Branco - *A queda dum Anjo* - 1980 Europa América
9. José Cardoso Pires - *Alexandra Alpha* - 1987 Dom Quixote
10. Fernando Namora - *Retalhos da Vida de um Médico* - 8ª. ed. - 1960
11. Bocage - *Poesias selectas* - 1944 Domingos Barreiro
12. Ricardo Reis - *Odes* - 1959 Arcádia
13. Fernando Pessoa - *Poesias* - 1961 Arcádia
14. Anthero de Quental - *Carta ao Exmo. Sr. António José d'Avila*
15. Mário de Sá Carneiro - *Poesias* - Atica
16. Armindo Rodrigues - *10 Odes ao Tejo* - 1951 Cancioneiro Geral
17. Rodrigues Lapa - *Lições de Literatura Portuguesa* - 9ª. ed. - 1977
18. D. Quintela, M. Silva, M. Barroso - *Temas de Língua e Cultura Portuguesa* - 1983 Editorial Presença

BIBLIOGRAFIA

- Almeida, Mário T.R., *Análise de Séries Cronológicas. Método de Entropia Máxima*, Rev. Inst. Nac. de Met. e Geof., Vol. 2, 1979.
- Anderson P.W., *Is Complexity Physics? Is it Science? What is it?*, *Physics Today*, 9-11, Julho 1991.
- Bell G.I., *Computers and DNA*, Santa Fé Institute, Vol. VII, Addison Wesley, Santa Fé, E.U.A., 1990.
- Bekenstein J.D., *Energy cost of information transfer*, *Phys. rev. Lett.*, 46, 10, 623-626, 1981.
- Bennet C.H., *Logical Reversibility of computation*, *IBM J. Res. Dev.*, 525-532, Novembro de 1973.
- Bennet C.H., *The Thermodynamics of computation - a Review*, *Int. J. of Theor Phys.* 21, 12, 905-940, 1982.
- Bennet C.H., *Notes on the history of reversible computation*, *IBM J. Res. Develop.*, 32, 1, 16-22, 1988.
- Benioff P.A., *Quantum Mechanical Hamiltonian models of discrete processes that erase own histories: Applications to Turing Machines*, *Int. J. Theor Phys.*, 21, 3/4, 177-201, 1982.
- Benioff P., *Quantum Mechanical models of Turing Machines that Dissipate no Energy*, *Phys. Rev. Lett.*, 48, 23, 1581-1585, 1982.
- Bookstein F.L., *Comment on a "Nonequilibrium" approach to evolution*, *Syst. Zool.*, 32 (3), 291-300 (1983).
- Brillouin, L., *Science and Information Theory*, 1956.
- Brooks, D.R., Wiley E.O., *Nonequilibrium thermodynamic and evolution: a response to Loutrup*, *Syst. Zool.* 32, (2), 209-219 (1983).
- Brooks D.R., Wiley E.O., *Nonequilibrium thermodynamic and evolution: responses to Bookstein and to Wicken*, *Syst. Zool.*, 34 (1), 89-97, (1985).
- Brooks D.R., Collien J., Wiley E.O., *Definition of Terms and the essence of theory: a reply to J.S. Wicken*, *Syst. Zool.*, 35 (4), 640-647 (1986).
- Brooks D.R., Wiley E.O., *Evolution as entropy*, Univ. Chicago Press, Chicago, 1986.
- Burg J.P., *Maximum entropy spectral analysis, in Modern Spectrum Analysis*, ed. por D.G. Wilders, Wiley, N.Y. 1967.
- Burg, J.P., *Ph. D. Thesis*, Stanford Univ., 1975.
- Chaitin, G.J., *Algorithmic Information Theory*, *IBM J. Res. Develop.* 350-358, Julho 1977.
- Chaitin, G.J., *Algorithmic Information Theory*, Cambridge University Press, Cambridge, 1987.
- Carvalho Rodrigues F., *A proposed Entropy Measure for Assessing Combat Degradation*, *J. Opl. Res. Soc. (UK)*, 40 (8): 789-93, 1989.
- Carvalho Rodrigues F., Pinto Peixoto J., Fernando D. Carvalho, Santos Silva M., *The hidden parameter in textile processes: Informational Entropy*, *J. Text Inst.*, 80, 4, 595-7, 1989.
- Carvalho Rodrigues F. *A entropia como discriminante*, *Memórias da Academia das Ciências de Lisboa*, 1 de Fevereiro de 1990.
- Carvalho Rodrigues F., *"Uma medida entrópica para as mudanças económicas"*, *Memórias da Academia das Ciências de Lisboa*, 15 de Março de 1990.
- Carvalho Rodrigues F., Dockery J.T., Woodcock A.E.R., *Casualty based entropy calculations as combat predictor*, *Int. CIS journal*, Dec. 1990.
- Carvalho Rodrigues F., Fernando D. Carvalho, Santos Silva M., Pinto Peixoto J., *Measurement of randomness in paper using computer vision*, *TAPPI J.*, 73, 12, 209-212 1990.
- Carvalho Rodrigues F., Fernando D. Carvalho, Pinto Peixoto J., Santos Silva M., *Application of information entropy to defect characterization in leather*, *J. Soc. Leather Techn. and Chemists*, 75, 10-14, 1990.
- Carvalho Rodrigues F., Rodrigues T., *"Uma medida entrópica para a perda da coesão de uma estrutura social devido à peste"*, *Memórias da Academia das Ciências de Lisboa*, 5 de Abril de 1990.

- Carvalho Rodrigues F., *As novas tecnologias, o futuro dos impérios e os quatro cavaleiros do Apocalipse*, Ed. Discórdia, Lisboa, 1991.
- Carvalho Rodrigues, F., *Complexidade, quantidade de informação e coesão*, Memórias da Academia das Ciências de Lisboa, 16 de Maio de 1991.
- Carvalho Rodrigues F., Dockery John T., Woodcock Alexander E.R., *Entropy in Combat Data*, Proceedings of the Basic Research Groups of the Joint Directors of Laboratories Symposium at the National Defense University, Washington, DC, 24-25 de Junho de 1991.
- Carvalho Rodrigues F., Dockery J.T., Rodrigues T. *Entropy of plagues*, a ser publicado no European Journal of the Operational Research (1991).
- Casti J., *Alternate Realities*, John Wiley & Sons New York, 1989.
- Cooper D., Emily M. A., Lynch M. F., Veates A. R., *Compression of continuous prose texts using variety generation*, J. of Amer. Soc. of Inf. Sci, Maio 1980.
- De Luca A., Termini S., *A definition of a non-probabilistic entropy in setting of fuzzy sets theory*, Laboratório di Cibernetica del CNR, Napoles, Itália, 1970.
- Denbigh K.G., Denbigh J.S., *Entropy in relation to incomplete Knowledge*, Cambridge University Press, Cambridge, 1985.
- Dilão R., *Aplicações dum intervalo: dinâmica simbólica entropia topológica e comportamento periódico*, Tese de doutoramento, I.S.T., 1985.
- Dockery, J., *The Design of a Fuzzy Military Information Processing System*, International J. of Man-Machine Studies, 16: 1-38, 1982.
- Eddington E.S., *The Nature of the Physical Reality*, 1927.
- Eigen M., Winkler R., *O jogo; as leis naturais que regulam o acaso*, Gradiva, Lisboa, 1989.
- Empotz H., *Nonprobabilistic entropy and indetermination measure in the setting of fuzzy sets theory*, Fuzzy Sets and Systems, 2, 307-317, 1981.
- Ferdinand A.E., *A theory of system complexity*, Intern. J. of Gen. Syst., 1, 19-33, 1974.
- Feshbach H., Matsui T., Niels Bohr: *Physics and World*, Harwood Acad., London, 1988.
- Fredkin E., Toffoli T., *Conservative Logic*, Int. J. theor. Phys., 21, 3/4, 219-252, 1982.
- Gallager R.G., *Information Theory and Reliable Communication*, N.Y., 1968.
- Georgescu - Roegen N., *The entropy law and the economic process*, Harvard Univ. Press, Cambridge, Mass., 1971.
- Gonzalez R. C., Wintz P., *Digital Image Processing*, Addison-Welsey, Reading, Mass, 1987.
- Hopfield J.J., and Tank D.W., *Computing with neural circuits: a model*, Science, 233, 601-692, 1986.
- Hopfield J.J., *Neural Networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci., USA, 79, 2554-2558, 1982.
- Hopfield J.J., *Neurons with graded response have collective computational properties like those of two - state neurons*, Proc. Nat. Acad. Sci., USA, 81, 3088-3092, 1984.
- Jaynes E.T., *Information Theory and Statistical Mechanics I, II*, Physical Review, 106 e 108, 1957.
- Jaynes, E.T., *On the rationale of maximum entropy methods*. Proc. IEEE, 70; 939-952, 1982.
- Jumarie G.M., *Subjectivity, Information, Systems; an introduction to a theory of relativistic cybernetics*, Gordon & Breach, N.Y., 1986.
- Kinchin A.U., *Mathematical Foundation of Information Theory*. Dover, New York, 1957.
- Klein, J. Martin, *Maxwell, His Demon, and the Second Law of Thermodynamics*, American Scientist, vol. 58, 1970.
- Klir G.J., Folger T.A., *Fuzzy sets, uncertainty and Information*, Prentice-Hall Int., Londres, 1988.
- Landauer R., *Irreversibility and heat generation in the computing process*, IBM Journal, 183-191, Julho de 1961.

- Landauer R., *Uncertainty Principle and Minimum Energy dissipation in the computer*, Int. J. of Theor. Phys., 21 3/4 283-297, 1982.
- Landauer R., *Information is physical*, Physics Today, 23 - 29, Maio 1991.
- Leff H. S., Rex A. F., *Maxwell's demon: entropy, information, computing*, Adam Hilger, Bristol, 1990
- Lynch M. F., Variety Generation - A reinterpretation of Shannon Mathematical Theory of Communication, and its implications for information science, J. of Amer. Soc. of Inf. Sci, Janeiro 1977.
- Nicolis G., Prigogine I., *Exploring Complexity*, W.H. Freeman and Company, N.Y., 1989.
- Palácios, Júlio, *Termodinâmica de Mecânica Estatística*, Espasa-Calpe, Madrid, 1949.
- Peixoto J. Pinto, *Introdução à Teoria de Informação*, Faculdade de Ciências de Lisboa, 1967.
- Peixoto, J. Pinto, *Curso de Termodinâmica: Algumas noções elementares sobre a teoria da informação*. Faculdade de Ciências de Lisboa, 1968.
- Peixoto, J. Pinto, *Entropia e Informação*. Seminário da Faculdade de Ciências da Universidade de Lisboa, 1972.
- Peixoto, J. Pinto, *Maxwell, a Teoria Cinética e a Termodinâmica*. Memórias da Academia de Ciências, Classe de ciências, Tomo XXIII, 1980.
- Peixoto, J. Pinto, *Entropia e Ainda Entropia*, Universidade do Algarve, Faro, 1984.
- Peixoto, J. Pinto, *Interpretação Mecanicista da Termodinâmica*, Universidade do Algarve, Faro, 1986.
- Prod W., Grondin R.D., Feny D.K., Porod G., *Dissipation in computation*, Phys. rev. Lett., 52, 3, 232-235, 1984.
- Prigogine I., Stengers F., *Order out of chaos*, Bantam Books, Londres, 1984.
- Prigogine I., in *O homem e a natureza, Balanço do Século*, INCM, Lisboa, 1990.
- Rocha Faria, J.M., *Utilização do conceito de entropia no planeamento de uma rede de udómetros*, Revista do Instituto Nacional de Meteorologia e Geofísica, vol. 2, nºs: 3 e 4, 1979.
- Serra R., Andretta M., Copiani M., Zanarini G., *Introduction to the Physics of Complex Systems, the mesoscopic approach to fluctuations, non linearity and self-organization*, Pergamon Press, 1986.
- Shannon E.C., Weaver W., *The mathematical theory of communications*, University of Illinois Press, Chicago, 1949.
- Shannon, C.E., *Prediction and Entropy of Printed English*, Bell System, 1951.
- Smith C. Ray, W.T., Grandy, *Maximum Entropy Bayesian Methods in Inverse Problems*. D. Reidel Comp., Boston, 1985.
- Stewart I., *The ultimate in undecidability*, Nature, 321, 115-116, 10 de Março de 1988.
- Stonier T., *Information and the internal structure of the Universe*, Springer-Verlag, Londres, 1990.
- Theil H., *Economics and Information Theory*. North-Holland, Amsterdam, 1967.
- Thom R., *Structural Stability and Morphogenesis*, Benjamin Cummings Publishing inc., Mass, 1975.
- Toffoli T., *Physics and Computation*, Plenum, N.Y., 165-175, 1981.
- Tribus M., *Thermostatistics and Thermodynamics*. Van Nostrand, Nova York, 1961.
- Von der Lubbe J.C., et al., *A Generalized Class of Certainty and information measures*, Information Science, 32, 187-215 (1984).
- Watanabe S., *Knowing and Guessing - A quantitative study of inference and information*, John Wiley & Sons, Inc., N.Y. 1968.
- Weber B.H., Depew D.J., Smith J.D., *Entropy, Information and Evolution*, Bradford Books, MIT Press, Cambridge Mass, 1988.
- Wicken J.S., *The generation of complexity in evolution: a thermodynamic and information theoretical discussion*, J. Theor. Biol., 77, 349-365, 1979.
- Wicken J.S., *A thermodynamics theory of evolution*, J. Theor. Bio., 87, 9-23, 1980.
- Wicken J.S., *Entropy Information and Nonequilibrium evolution*, Syst.Zool., 32 (4), 438-443 1983.

- Zurek W.H., *Reversibility and Stability of Information Processing Systems*, Phys. Rev. Letters, 53, 4, 391-394, 1984.
- Zurek W.H., *Thermodynamic cost of computation, algorithmic complexity and the information metric*, Nature, 341, 119-124, 14 de Setembro de 1989.
- Zurek W.H., *Complexity, Entropy and the Physics of information*, Santa Fé Institute, Addison - Wesley, Santa Fé, 1990.

AGRADECIMENTOS

Ao Sr. Eng^o Fernando Carvalho as discussões, a construção do equipamento optoelectrónico e a implementação tecnológica dos conceitos na aplicação prática aos diversos sectores industriais.

Ao Sr. Prof. Santos Silva todos os dados e os materiais relativos à aplicação dos conceitos e da tecnologia a problemas do sector têxtil, dos curtumes e do papel.

Ao Sr. Doutor José Rebordão os resultados obtidos no âmbito do projecto SATCART no domínio da cartografia temática a partir de fotografia aérea.

Ao Sr. Prof. Carlos Reis as sugestões dadas, os comentários apresentados e o cuidado que pôs na leitura do manuscrito.

EDIÇÃO APOIADA PELA SECRETARIA DE ESTADO DA CIÊNCIA E TECNOLOGIA

José Pinto Peixoto, Presidente da Academia das Ciências de Lisboa é Professor Catedrático da Universidade de Lisboa, colaborador, desde 1956, do Departamento de Meteorologia do Massachusetts Institute of Technology, Vogal do Comité Executivo da Fundação Europeia da Ciência, Presidente da Secção Portuguesa das Uniões Internacionais de Astronomia, Geodesia e Geofísica (SPUIAGG) e exerce as funções de Director do Instituto Geofísico do Infante D. Luís. Perito de renome internacional nos domínios da meteorologia, em particular no que diz respeito ao ciclo da água e à circulação geral da atmosfera, é autor de comunicações a diversos congressos e conferências internacionais sobre hidrologia e dinâmica da atmosfera.

Tem publicada a sua actividade científica em numerosas revistas de prestígio mundial, tendo colaborado em várias monografias da Organização Meteorológica Mundial, da UNESCO e da Secretaria de Estado do Ambiente. Pelo seu trabalho em Teoria da Informação recebeu em 1990 o Prémio da Boa Esperança.

Fernando Carvalho Rodrigues, investigador coordenador do LNETI, Professor Catedrático do Instituto Superior Técnico, sócio correspondente da Academia das Ciências de Lisboa, é Director do Instituto de Electromecânica e das Tecnologias de Informação do LNETI, Presidente do Capítulo Português da AFCEA - Armed Forces Communications and Electronics Association, vogal do Conselho de Administração da EID - Empresa de Investigação e Desenvolvimento de Electrónica, S. A. e Director do Consórcio SATCART para a área de detecção remota. É especialista em óptica e optoelectrónica com numerosos resultados publicados internacionalmente. Na área de processamento de imagem foi galardoado com o Prémio Pfizer em 1977. Em 1978 e 1982 recebeu o Prémio Gulbenkian de Ciência e Tecnologia por descobertas no domínio da óptica. Em 1989 a sociedade Histórica da Independência de Portugal conferiu-lhe o Prémio da Identidade Portuguesa pelo conjunto da sua obra científica. Pelo seu trabalho em Teoria de Informação recebeu em 1990 o Prémio da Boa Esperança.

DISCÓRDIA, imagem e exemplo do contraditório, assume o elogio do conflito e a recusa do consenso. Porque no mundo do complexo e do excesso de informação, a comunicação pode ser, em milhares de bits, uma só frase, uma só palavra, urge procurar em cada coisa o seu oposto, gerir a mudança, resolver contradições sem conciliar os contrários, viver os saberes que aspiram à sabedoria. Caminho para a Discórdia.