Universidade de Aveiro Departamento de Matemática
2003

Ana Helena Marques de Pinho Tavares

Aspectos Matemáticos da Entropia

Universidade de Aveiro Departamento de Matemática
2003

Ana Helena Marques de Pinho Tavares

Aspectos Matemáticos da Entropia

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática, realizada sob a orientação científica do Prof. Doutor António José Batel Anjo, Professor Auxiliar do Departamento de Matemática da Universidade de Aveiro.

o júri

presidente	Doutora Maria Rosália Dinis Rodrigues, Professora Associada da Universidade de Aveiro
vogais	Doutor Francisco José Lage Campelo Calheiros, Professor Associado da Faculdade de Engenharia da Universidade do Porto
	Doutor António José Batel Anjo, Professor Auxiliar da Universidade de Aveiro (Orientador)

agradecimentos

Ao meu orientador, Professor Doutor Batel Anjo, pela orientação científica, apoio e disponibilidade, assim como, pelas sugestões e ideias que lançou.

À Professora Doutora Paula Oliveira, pelo interesse que manifestou pelo meu trabalho, pela forma como se prontificou a discutir alguns assuntos e, acima de tudo, pela sua amizade.

Aos professores do Departamento de Matemática da Universidade de Aveiro e aos amigos que de alguma forma me apoiaram na parte escolar e na realização deste trabalho.

Aos meus pais, por tudo.

Ao meu maior amigo, João André, pelo apoio e compreensão demonstrados ininterruptamente. Pela alegria com que me fez acreditar neste trabalho.

A todos, obrigado pela confiança depositada em mim.

resumo

Com o presente trabalho pretende-se fazer uma abordagem ao desenvolvimento do conceito Entropia.

Inicia-se esta dissertação com a apresentação de uma perspectiva histórica da evolução do conceito.

Enunciam-se as propriedades mais importantes das entropias de Shannon, Rényi e de Tsallis, assim como do ganho de informação de Kullback-Leibler. Apresentam-se as demonstrações de algumas destas propriedades.

São expostas e analisadas as definiçõpes axiomáticas da entropia de Shannon e de Tsallis. Menciona-se ainda a definição axiomática do ganho de informação de Rényi, que conduz à definição da entropia de Rényi.

abstract

This work intends to be an approach of the development of the concept of Entropy.

We start with an historical perspective of the evolution of this concept.

Several properties of the Shannon, Rényi and Tsallis entropies are presented, likewise Kullback-Leibler's gain of information. Some of this properties are followed by it's proof.

We present and discuss the axiomatic definition of Shannon and Tsallis entropies. We also refer axiomatic definition of Rényi's gain of information, which defines Rényi's entropy.

Entropy

(...)
Without external force
All things must pass,
Order into disorder,
Heat into cold,
Motion into stillness.

But hold on!
This is only true for closed systems.
Surely a poem cou ld'nt
Beclassed as a cl
Osed system, ccou
Ld it? Sure ly
Apoeeeem hass ome exxterrr
Nalfor ce? Youdddon't th ink
Thiss poemmight bbbbeeeeesssubject
Toeeeenttrroppy, doyoui? Yyttimd
,asjin8ekp[?kigie5???
jh9y7364gni]sf' ? k-ewk?
J9j? ...wir399jee8ih?on9...

Ian McDonough

Índice Geral

Introdução

O conceito de entropia surge pela primeira vez no contexto da Termodinâmica, no século XIX, sendo atribuído ao alemão Rudolph Clausius. Desde então, o destrinçar do seu significado tem obcecado Físicos, Químicos, Matemáticos, Engenheiros, Biólogos e muitos outros, que o procuram aplicar nos diferentes contextos.

Um dos objectivos da presente dissertação é o de reunir num único documento uma abordagem a diversas definições de entropia, permitindo uma visão global da evolução do conceito e criando alguns dos alicerces necessários a uma análise crítica e fundamentada de diversas aplicações da entropia.

O carácter multidisciplinar do conceito de entropia dificultou a escolha do rumo a seguir. A frequência de um mini-curso sobre Mecânica Estatística Não Extensiva, que ocorreu no Departamento de Matemática desta Universidade, veio alargar o já vasto campo de escolha. Porém, em vez de aumentar a dificuldade de decisão, tornou inequívoco que o enfoque principal deste trabalho seria o estudo da entropia de Tsallis.

Foi através das palavras do próprio Constantino Tsallis que tomei conhecimento da entropia de Tsallis e consequente Mecânica Estatística Não Extensiva. A eloquência com que apresentou a sua teoria, assim como desenvolvimentos e aplicações que muitos autores têm elaborado, despertou a curiosidade e o interesse sobre o tema.

O conceito de entropia permitiu a estruturação do corpo teórico da Termodinâmica de equilíbrio e de processos irreversíveis, constituindo o conceito fundamental da Mecânica Estatística. Mais tarde exerceu um papel central na Teoria da Informação, criada por Shannon, em 1948. Para além disso encontrou terreno fértil em áreas científicas tão diversas como a Estatística, as Ciências Computacionais, a Psicologia e a Linguística. Dentro da Teoria da Informação foram formuladas algumas propostas de generalização da entropia, uma das mais conhecidas é a de Rényi, em 1961.

A formulação da Mecânica Estatística foi iniciada por Boltzmann, em meados de 1870, ao associar uma variável macroscópica (entropia) a conceitos microscópicos. Posteriormente Gibbs fez contribuições fundamentais a esta teoria, pelo que esta é conhecida como Mecânica Estatística de Boltzmann–Gibbs.

A Mecânica Clássica formulada por Newton foi, no século XIX, considerada por praticamente todos como universal. No entanto, quando as massas envolvidas são muito pequenas, como a de um electrão, sabe–se que a Mecânica Newtoniana tem de ser substituída pela sua forma quântica. Se as massas forem muito grandes, como a do Sol ou de uma galáxia, tem de se utilizar a forma relativística. É possível que o mesmo aconteça com a Mecânica Estatística de Boltzmann–Gibbs. Hoje sabe–se que esta Mecânica Estatística não é universal, não é válida para todos os fenómenos, nem ubíqua, não é válida em toda a parte.

Em 1988, Constantino Tsallis [21] propôs uma fórmula de entropia, cuja grande diferença com a fórmula usual é a não aditividade. A sua expressão fundamenta uma generalização da Mecânica Estatística usual – a Mecânica Estatística Não Extensiva.

Nos últimos anos muitos trabalhos têm mostrado a aplicabilidade do formalismo de Tsallis nos chamados "sistemas complexos", como por exemplo a turbulência e as colisões muito energéticas de partículas elementares.

A resenha histórica que acabamos de apresentar encontra–se tratada de forma mais pormenorizada no primeiro capítulo, cujo objectivo é o de proporcionar uma visão global e sintetizada da evolução do conceito de entropia. Por restrições de tempo e de espaço, contemplamos apenas as áreas onde o conceito teve mais notoriedade, como é o caso da Termodinâmica, Mecânica Estatística e Teoria da Informação. Este capítulo foca e contextualiza cronologicamente cada um dos temas dos capítulos seguintes.

Com o título do primeiro capítulo pretendemos suscitar a dúvida sobre a unicidade, ou não, das diversas entropias que têm surgido ao longo dos tempos. Apesar das fórmulas serem diferentes, algumas delas estão claramente relacionadas, chegando mesmo a reduzirem–se à mesma expressão nalguns casos particulares. Recorrendo à ideia relativista da realidade, será que se trata sempre de um mesmo conceito, olhado segundo diferentes "referenciais"? Em relação às entropias que são estudadas neste trabalho, nomeadamente as de Shannon, de Rényi e de Tsallis, ficámos convictos que todas traduzem um mesmo conceito. Contudo, a questão fica em aberto em relação a uma extensa lista de entropia(s) que apresentamos no final do referido capítulo.

No segundo capítulo é apresentada a entropia de Shannon assim como algumas das suas principais características. Shannon definiu axiomaticamente as propriedades que uma medida de informação deve obedecer. A axiomática culmina com o Teorema da Unicidade de Shannon e a medida por este definida, a entropia. Seguindo a abordagem efectuada por Shannon em [20], apresentamos a demonstração deste teorema, tecendo algumas considerações finais sobre a mesma.

Ainda no contexto da Teoria de Informação, o terceiro capítulo é dedicado a outras medidas de informação, com especial enfoque na medida de ganho de informação de Kullback–Leibler, ou entropia de Kullback–Leibler. Esta quantidade, apesar de não ser simétrica nem respeitar a desigualdade triângular, é muitas vezes interpretada como uma distância entre duas distribuições. São enunciadas e demonstradas algumas das principais propriedades desta medida.

No quarto capítulo apresentamos a entropia de Rényi, uma generalização da entropia de Shannon. A sua fórmula funcional para além de depender de uma distribuição, caracteriza–se por um parâmetro escalar, sendo apontada como a primeira fórmula paramétrica de entropia. Ao contrário da entropia de Shannon, que é definida de raíz por uma axiomática, a entropia de Rényi é formulada a partir de uma medida do ganho de informação, esta sim, definida axiomaticamente. Neste capítulo apresentamos as definições do ganho de informação e da entropia de Rényi, enunciamos e demonstramos algumas das propriedades que esta última verifica.

Por fim, no quinto capítulo, é apresentado o tema de excelência desta dissertação – a entropia de Tsallis. Esta entropia, tal como a entropia de Rényi, caracteriza–se pela existência de um parâmetro escalar, um subíndice q, daí que também seja designada por q–entropia. Após a definição da entropia de Tsallis enunciamos e demonstramos algumas das suas propriedades fundamentais, evidenciando semelhanças e diferenças entre esta e a entropia de Boltzmann–Gibbs.

A entropia de Tsallis foi axiomatizada por Roberto Santos [19] em 1997, que enuncia e demonstra um Teorema de Unicidade (da entropia de Tsallis), com base na definição de quatro propriedades que esta medida deve satisfazer. Na apresentação dessa axiomática, considerámos conveniente alterar (fortalecer) uma das propriedades. Esta decisão encontra–se devidamente justificada após a demonstração da unicidade da fórmula de Tsallis.

O formalismo de Tsallis é uma proposta relativamente recente. Nas pesquisas efectuadas, apenas se encontrou um livro relacionado com este tema [23]. Grande parte do trabalho elaborado baseou–se em artigos relativamente recentes, publicados em diversas revistas internacionais.

As referências descritas em nota de rodapé referem–se a documentos não consultados por inacessibilidade. Pelo facto não constarem da bibliogarfia final, é sempre indicado o documento de onde foi retirada essa referência.

No decorrer do trabalho optou–se por analisar com um certo rigor e pormenor as propriedades das entropias de Shannon, Rényi e Tsallis, para melhor apreender as alterações introduzidas por esta última.

Temos consciência que uma compreensão mais estruturada das potencialidades de aplicação da entropia de Tsallis, passa por uma análise e estudo da própria Mecânica Estatística Não Extensiva, pelo que o trabalho desenvolvido ao longo deste ano, e aqui apresentado, não é mais do que uma base de conhecimentos que servirá de suporte para algo mais elaborado.

Capítulo 1

Entropia(S)

> "Existem conceitos que resistem à passagem do tempo, uns pela sua importância, outros pela sua capacidade de se adaptar a novos domínios da ciência. A entropia, mais do que qualquer outro conceito, mantém bem viva a sua importância, a sua utilidade e, porque não, o seu mistério."
>
> Batel Anjo, [4]

A Entropia é uma ideia que nasceu no seio da Termodinâmica Clássica, não como algo fundamentalmente intuitivo, mas sim fundamentalmente quantitativo definido por uma equação. Para Clausius e seus contemporâneos a entropia era uma propriedade muito importante, mas do domínio estrito da Termodinâmica, fora deste contexto simplesmente não tinha significado. No entanto, os fundamentos matemáticos da Mecânica Estatística são aplicáveis a qualquer sistema estatístico, o que permitiu a aplicação desta quantidade noutros contextos.

A maior incursão da Entropia em novos domínios talvez tenha sido a impulsionada por Shannon no campo da Teoria da Comunicação, nomeadamente na Teoria da Informação.

Uma nova forma de entropia, a entropia de Tsallis, tem sido descrita por muitos como um dos maiores avanços da Física. Esta entropia generaliza a entropia clássica de Boltzmann e muitos trabalhos recentes têm mostrado a sua eficiência. Desde então a nova fórmula tem sido aplicada numa grande variedade de problemas e em diversas áreas, como por exemplo os problemas da turbulência.

Neste capítulo expomos de forma sucinta os aspectos, a nosso ver, mais importantes da evolução do conceito de entropia. Seccionando essa evolução pelas principais áreas envolvidas e dispondo-as por ordem cronológica, apresentamos, na secção 1, a interpretação Termodinâmica. Esta utiliza a entropia como uma medida da irreversibilidade dos processos, ou da capacidade de uma máquina realizar trabalho. Na secção 2, mostramos como é que a

Mecânica Estatística associa a entropia a uma quantidade de desordem. Na secção 3, apresentamos a fórmula de entropia adoptada na Mecânica Quântica, baseados no artigo [28] de Flamm. De seguida, na secção 4, focamos a entropia no contexto da Teoria da Informação, onde é relacionada com uma medida da quantidade de informação ou de incerteza. Ao longo dos anos, foram criadas muitas generalizações do conceito de entropia, algumas das quais são mencionadas na secção 5.

1.1 Termodinâmica

A afirmação de que a energia não pode ser criada nem destruída tornou-se conhecida como a Primeira Lei da Termodinâmica ou Lei da Conservação da Energia. A descoberta deste princípio, nos inícios do século XIX, pôs uma limitação importante aos processos que se podem produzir na natureza, sendo apenas possíveis aqueles em que a energia permanece inalterável. É através da Segunda Lei da Termodinâmica que surge o termo entropia [37].

A noção de entropia foi definida pela primeira vez por Rudolph Clausius em 1865, baseando-se no trabalho de Sadi Carnot e Lord Kelvin, no contexto da Termodinâmica, nomeadamente no estudo do engenho a vapor [17].

O engenho a vapor é uma construção teórica desenvolvida por Carnot ao estudar a termodinâmica de uma máquina a vapor. O rendimento de uma máquina a vapor é uma função da variação de temperatura, segundo a qual o engenho opera. Em 1824, Carnot descobriu que quando se fornece calor a uma máquina a vapor, apenas parte desta energia se transforma no movimento de rotação do veio do motor, o resto desperdiça-se inevitavelmente em forma de calor residual. Isto significa que apenas uma porção da energia calorífica da máquina pode ser convertida em trabalho (mecânico). Concluiu ainda que o calor poderia ser convertido em trabalho mecânico apenas se existir uma diferença de temperatura. Um engenho a vapor a uma temperatura constante não poderia converter nenhuma da sua energia calorífica em movimento. Este resultado indicia a existência de outra lei limitadora, a Segunda Lei da Termodinâmica. [17]

Clausius deu o passo seguinte na compreensão da transformação do calor noutras formas de energia. Este dirigiu a sua atenção para a distinção entre a fracção de calor que se pode transformar em energia mecânica e a que se deve desprezar como calor perdido. Descreveu a primeira por "energia livre" e aquela perda de energia (não pode ser transformada em trabalho útil) por meio de um novo termo "entropia", que tem origem numa palavra grega *entrepien* que significa degradação, e definiu-a como o quociente entre a quantidade de calor transferido e a temperatura absoluta a que o processo é realizado.

Esta entropia é o que permite enunciar o segundo princípio da Termodinâmica. Clausius sentiu que tanto a descoberta de Carnot como a Primeira Lei da Termodinâmica eram fundamentais e baseou uma lei baseando–se nesses resultados. Essa viria a ser conhecida como Segunda Lei da Termodinâmica e afirma que, na natureza, só podem ocorrer aqueles processos em que a entropia aumenta ou, quando muito, permanece constante.

A entropia não era medida em valor absoluto. Apenas se media a variação de entropia de um sistema quando este passa de um estado para outro. Considerando uma transformação reversível, a variação de entropia ΔS verifica a relação seguinte,

$$\Delta S = \frac{\Delta Q}{T}$$

onde ΔQ representa a variação de calor de um sistema a temperatura T constante.

As primeira e segunda Leis da Termodinâmica podem ser encerradas na famosa afirmação de Clausius,

> *"Die Energie der Welt ist Konstant. Die Entropie der Welt strebt einem Maximum zu."*[1]

A entropia surge como algo cuja variação se pode medir, mas da qual não se tem um conceito concreto. Uma das mais frutíferas interpretações iniciais de entropia, talvez seja a visão desta como medida da capacidade de realização de trabalho.

1.2 Mecânica Estatística

No século XIX prevalecia a ideia de que o calor era uma substância que fluía de um objecto para outro, o "calórico". Uma minoria de físicos sustentava a visão (agora universalmente aceite) de que o calor era a energia resultante do movimento aleatório dos átomos. Quanto mais elevada a temperatura de uma matéria, mais vigoroso é o movimento das moléculas que a formam. [33]

James Maxwell foi o primeiro físico a aceitar a existência dos átomos, tratando os gases como uma colecção de pequenas partículas movimentando–se a diferentes velocidades, colidindo umas com as outras. Maxwell determinava as propriedades de um gás, assumindo que este é composto por um grande número de átomos (ou moléculas) em movimento aleatório. Apesar de não ter sido o autor da Teoria Cinética dos Gases, foi o primeiro a aplicar métodos estatísticos e probabilísticos[2] para descrever as propriedades das moléculas dos gases, pelo que é considerado um dos fundadores da Teoria Cinética dos Gases.

[1] Traduzindo do alemão, "A Energia do Universo é constante. A Entropia do Universo converge para o seu máximo." [8]

[2] Criou a ideia de "função de distribuição" que governa as velocidades das moléculas do gás e analisou previsões concretas que podiam ser obtidas a partir da Teoria Cinética dos Gases.

Em 1859, Maxwell apresenta o seu primeiro trabalho em Teoria Cinética, que viria a ser publicado um ano mais tarde com o título "Ilustrations of Dynamical Theory of Gases" na *Philosofical Magazine*. Neste, propôs um tratamento estatístico dos gases, dando os primeiros passos em direcção à interpretação estatística da Termodinâmica.

Anos mais tarde, em 1872, Ludwing Boltzmann utilizou os métodos estatísticos de Maxwell para determinar um novo e mais geral tratamento da entropia. Utilizou as probabilidades para descrever as propriedades da matéria a partir das propriedades dos átomos.

Ao reduzir o calor a um movimento aleatório colectivo de muitos milhões de átomos, Boltzmann transpôs a entropia para o mundo reversível da Física Newtoniana. De acordo com a Mecânica Newtoniana, cada uma das pequenas partículas que formam a matéria pode retroceder no seu percurso e voltar ao lugar onde começou, revertendo exactamente o seu movimento. Assim, a totalidade do sistema também deve poder voltar ao seu estado inicial, o que contradiz o facto de um sistema termodinâmico envolver irreversibilidade no sentido do estado de máxima entropia e permanecer nesse estado. Boltzmann resolve esse dilema interpretando a entropia como uma propriedade estatística. Em princípio não há nada que impeça todas as partículas de reverter os seus percursos e irem espontaneamente de um estado de desordem para um estado de ordem, porém, o grande número de partículas envolvidas, tornaria o evento muito raro. Segundo Boltzmann, a desordem irreversível seria apenas provável, não certa. Ao introduzir a Teoria das Probabilidades na Segunda Lei da Termodinâmica, Boltzmann apresenta pela primeira vez uma lei da natureza diferente de todas as outras. Em vez de determinista, esta lei é probabilística.

Em 1877, Boltzmann formula uma expressão de entropia no artigo "On the relation betwenn the second law of mechanical theory of heat and probability calculus with respect to the theorems on thermal equilibrium"[3]. Num sistema fechado a entropia S do sistema seria proporcional ao volume do espaço de fase Ω ocupado pelo macroestado do sistema, ou seja,

$$S \propto \log \Omega$$

actualmente é usual escrever na notação

$$S = k \ln \omega \tag{1.1}$$

onde k é a chamada constante de Boltzmann e ω é o número de microestados do sistema compatíveis com o macroestado observado[4]. Apesar da fórmula ser atribuída a Boltzmann, esta nunca foi escrita por ele. De facto, a fórmula aparece pela primeira vez no livro "Theory of heat radiation"[5] de Max Planck em 1906, ano da morte de Boltzmann [34].

---

[3] Título traduzido do alemão. A referência completa pode ser encontrada no artigo [28].

[4] Em sistemas atómicos, um macroestado é um estado do sistema observável a grande escala. Um microestado é um arranjo preciso de átomos num sistema. Em geral, ao mesmo macroestado corresponde um número astronómico de microestados. [33]

[5] Traduzido do alemão "Theorie der Wärmestrahlung", [38].

Túmulo de Boltzmann em Viena[6]

A interpretação estatística veio revelar a natureza da entropia como um crescimento da desordem entre as partículas. De facto, o número de microestados pode ser considerado como uma medida de desordem do sistema. Na equação de Boltzmann (1.1) quanto maior for o número de microestados possíveis, maior é a desordem e maior é a entropia.

Implícito na versão de Boltzmann de entropia está a noção de que existem limites para o que um observador pode saber sobre um sistema onde a energia está em transformação. Pode–se saber mais sobre um sistema ordenado, tem entropia baixa, do que sobre um sistema desordenado, tem entropia elevada. A relação entre aumento de entropia e diminuição de conhecimento não se tornou explícita durante muitos anos. Boltzmann estava ciente desta ligação, mas apenas a referiu casualmente. Em 1884, afirmou que "a entropia está relacionada com a diminuição de conhecimento", [33]. Com esta frase, cruzou a ponte entre informação e entropia, relação tão harmoniosa ao pensamento do final do século XX.

Willard Gibbs percebeu que os trabalhos de Maxwell e de Boltzmann iniciavam uma nova disciplina cujo objectivo era o estudo de corpos de complexidade arbitrária que se movem segundo as leis da mecânica.

Gibbs introduziu uma expressão mais abrangente para descrever a entropia,

$$S = -k \sum_{i=1}^{\omega} p_i \ln p_i \tag{1.2}$$

onde p_i é a probabilidade do sistema estar no microestado i e ω o número total de microestados. A máxima entropia corresponde ao estado de equilíbrio, aquele em que existe a máxima desordem (microscópica) possível. Quando todos os microestados do sistema são equiprováveis, esta equação reduz–se à equação (1.1).

O desenvolvimento da Mecânica Estatística é actualmente associado a Gibbs devido aos

[6]Fotografado por Constantino Tsallis.

inúmeros trabalhos que publicou. Desde 1884 estudou a formulação de uma estrutura geral para a Mecânica Estatística publicando, em 1902, o livro "Elementary Principles in Statistical Mechanics", [28]. Boltzmann e Gibbs estabeleceram a ligação entre a Termodinânica e Mecânica por meio da Mecânica Estatística, pelo que a equação (1.1) ficou conhecida como fórmula de entropia de Boltzmann–Gibbs.

1.3 Mecânica Quântica

Tanto a Mecânica Estatística como a Mecânica Quântica lidam intimamente com a estatística e as probabilidades. A fórmula de entropia utilizada no contexto da Mecânica Quântica, que se deve a John von Neumann, é idêntica à utilizada na Mecânica Estatística (1.2), a única diferença está na forma como as probabilidades são calculadas.

Em 1897, Boltzmann e Planck terão tido uma disputa sobre a irreversibilidade do fenómeno de radiação em corpos negros, que terá estimulado Planck na sua descoberta da Teoria Quântica, em 1900. Esta viria a permitir o desenvolvimento de uma teoria sobre a interacção entre a matéria e a radiação, conhecida como Mecânica Quântica, que generaliza a Mecânica Clássica.

Planck utilizou os métodos estatísticos de Boltzmann para explicar as suas leis de radiação de corpos negros. Para além disso, assumiu que a absorção de energia não podia ser num processo contínuo. A energia seria absorvida em quantidades discretas, os *quantum* de energia. No trabalho de Boltzmann já era comum introduzir níveis discretos de energia para obter um conjunto numerável de estados. Planck tornou–se num dos proponentes mais importantes das ideias de Boltzmann. Como já foi referenciado atrás, a chamada fórmula de Boltzmann foi pela primeira vez escrita num livro de Max Planck em 1906.

A Mecânica Quântica generaliza a estrutura da Mecânica Estatística. Em 1932 foi apresentado pela primeira vez um formalismo matemático para a Mecânica Quântica, no livro "Mathematical Foundations of Quantum Mechanics" de John von Neumann. Este associou a quantidade de entropia a um operador estatístico, em 1927. A entropia de von Neumann é definida por

$$S(\rho) = -k\,\mathrm{Tr}(\rho \ln \rho) \tag{1.3}$$

onde ρ é um operador de densidade e Tr denota o traço da matriz. Se, ao determinar a entropia, forem usados os valores próprios λ_i da matriz densidade, a equação (1.3) toma a forma

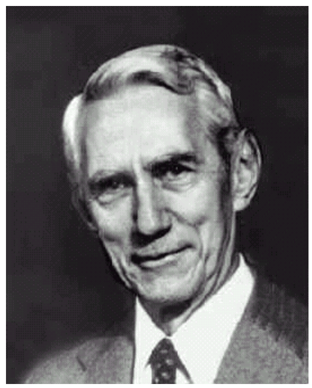
$$S(\rho) = -k \sum_i \lambda_i \ln \lambda_i$$

ou seja, reduz–se à forma de Gibbs (1.2).

1.4 Teoria da Informação

A Teoria da Informação surge na década de 40, com origem na telegrafia e telefonia. Posteriormente foi utilizada pela Cibernética, no estudo da troca de informação num organismo vivo ou mecânico.

Nos finais de 1940, Claude Shannon associou a entropia a uma quantidade de informação. Ao procurar um sistema telefónico que transferisse o máximo de quantidade de informação e uma forma de rectificar distorções nas mensagens, criou uma teoria matemática para a comunicação.



Claude Shannon, pai da Teoria da Informação[7]

O seu trabalho "A Mathematical Theory of Communication" foi publicado pela primeira vez em 1948 na revista *Bell System Technical Journal*[8] e nele é apresentada uma expressão matemática para a quantidade de informação transmitida numa mensagem. Estalebeceu então um esquema de transmissão de informação. Uma mensagem que parte de uma fonte, é codificada e emitida por um transmissor, passa por um canal de comunicação, sofre perturbações, designadas por ruídos, e chega ao receptor passando por um sistema de descodificação. Ao falar de uma "mensagem seleccionada", Shannon refere–se a uma sequência informativa que pode ser escolhida entre muitas outras e define a quantidade de informação com base na incerteza ou dificuldade de previsão dessa mensagem.

Independentemente da mensagem ou do seu impacto sobre o receptor, a quantidade de informação está relacionada com a possibilidade dessa mensagem ocorrer. Se a possibilidade de uma mensagem ocorrer for pequena então ela contém muita informação. Se, pelo contrário, a mensagem for predizível então conterá pouca informação. Para medir a quantidade de informação, Shannon criou o conceito de entropia, algo diferente do conceito homónimo em

[7] Imagem retirada do endereço http://www.bell-labs.com/news/2001/february/26/1.html, em Março de 2003.

[8] Em 1949, o mesmo trabalho viria a seu publicado, conjuntamente com um trabalho de Weaver, no livro "The Mathematical Theory of Communication", University Illinois Press.

Termodinâmica. Ao que parece, esta denominação foi sugerida a Shannon por Jonh von Neumann, argumentando que a mesma função matemática já era utilizada em Termodinâmica com esse nome. Teria dito, ironicamente, que

> *"No one really understands entropy (in the sense of Physics). Therefore, if you know what you mean by it and you use it when you are in an argument, you will win every time."*[9]

Shannon apresentou a quantidade de informação de uma mensagem como uma medida estatística que obedece a um conjunto de axiomas e chamou-lhe entropia. No seu trabalho define três axiomas que essa medida deve satisfazer e deduz uma expressão para a entropia, que prova ser única, nomeadamente

$$H = -k \sum_{i=1}^{n} p_i \ln p_i$$

sendo $p_1, p_2, ..., p_n$ as probabilidades associadas aos resultados possíveis de um acontecimento e k uma constante positiva. A escolha da base do logaritmo corresponde à escolha da unidade para medir a informação, em particular, a unidade resultante da utilização da base 2 é o chamado dígito binário ou simplesmente "bit".

Na sua obra [6], Shannon desenvolve também um método a partir do qual é possível codificar a mensagem por forma a que maximize a capacidade do canal e, simultaneamente, o grau de distorção da mensagem se aproxime de zero. Este método envolve o conceito de máxima entropia.

Naquela época, alguns matemáticos viam a Teoria da Informação simplesmente como um ramo da Matemática. Consideravam que o mais importante seria tornar a teoria de Shannon matematicamente correcta, através de critérios de rigor matemático. Um exemplo típico desta formalização é o livro de Khinchin (1957), "Mathematical foundations of information theory" [15].

A Teoria da Informação surgiu essencialmente completa, com as definições e conceitos básicos formulados, com os teoremas e resultados fundamentais já estabelecidos. Contudo, os primeiros estudos no sentido de encontrar uma medida de informação tinham já surgido na década de 20, pela mão de Harry Nyquist. Nos trabalhos "Certain Factors Affecting Telegraph Speed"(1924) e "Certain Topics in Telegraph Transmission Theory"(1928), analisa a relação entre a velocidade de um sistema telegráfico e o número de sinais usados pelo sistema. O artigo "Transmission of Information" de Robert Hartley, publicado em 1928, foi possivelmente o único precedente do trabalho de Shannon no campo da Teoria da Informação.

[9] Citado por E.T. Jaynes, [30].

Hartley reconheceu claramente que a recepção de um símbolo possui informação apenas se há outras possibilidades para o seu valor e propôs uma fórmula para quantificar a informação,

$$I(X) = \log \omega$$

onde X é uma variável discreta e ω é o número de possíveis valores que X pode assumir[10]. Quanto à base do logaritmo, nas próprias palavras de Hartley, a sua escolha determina a unidade de medida da informação. Estes três trabalhos são citados no primeiro parágrafo do clássico trabalho de Claude Shannon "A Mathematical Theory of Communication" (1948), onde são claramente reconhecidos como prolíficos no desenvolvimento da Teoria da Informação.

No mesmo ano em que é publicado "A Mathematical Theory of Communication", Norbert Wiener publica, independentemente, o trabalho intitulado "Cybernetics" [25], onde apresenta resultados semelhantes aos de Shannon. Tanto Shannon como Wiener consideram a comunicação como algo em que um um sinal é transmitido por um canal. Porém, no modelo de Shannon as mensagens são codificadas antes de serem transmitidas, enquanto que no modelo de Wiener o sinal é comunicado directamente através do canal, sem antes ser codificado.

Em 1951, Kullback e Leibler estudaram uma medida de informação envolvendo duas distribuições de probabilidade, que tem recebido diversos nomes como entropia cruzada, entropia relativa e divergência de Kullback–Leibler, entre outros[11]. Simultaneamente, estudaram uma medida de divergência, chamada J–divergência[12], já estudada por Jeffreys em 1946.

A entropia de Shannon e a informação relativa de Kullback–Leibler são medidas de informação clássicas. Nos últimos anos tem aparecido na literatura várias medidas que generalizam as anteriores, como por exemplo a medida de divergência de Taneja (1995) e suas generalizações. A caracterização e o estudo de generalizações de medidas de informação e de divergência podem ser encontrados em [31].

Em 1957, Jaynes reexamina o princípio da máxima entropia de Shannon aplicando-o a problemas caracterizados por informação insuficiente. Através deste princípio é possível determinar a distribuição de um acontecimento, do qual apenas são conhecidas medidas macroscópicas, tais como a média e a variância. Este autor escreveu inúmeras páginas que estão disponíveis apenas em formato electrónico [30], nas quais aborda vários aspectos da Teoria das Probabilidades e da Teoria da Comunicação.

[10]Sejam A e B duas experiências aleatórias independentes e seja C a experiência que consiste na realização de A e B. Cada resultado de C não é mais do que um resultado de A e um resultado de B. É natural impôr que a informação contida em C seja a soma das informações contidas em A e em B. A única função real de variável real, não identicamente nula, que verifica essa propriedade é a função logaritmo.

[11]Traduzido do inglês "crossed entropy", "relative entropy" e "Kullback–Leibler divergence".

[12]Traduzido do inglês "J–divergence".

Nos últimos sessenta anos, a Teoria da Informação tem encontrado aplicações em diferentes domínios, como por exemplo, Economia, Estatística, Linguística, Psicologia, Ecologia, reconhecimento de padrões, Ciências Computacionais, etc.

1.5 Formas Generalizadas

Vários autores, e inicialmente Shannon, mostraram que a medida de quantidade de informação é unicamente determinada por alguns postulados naturais. A evidência de que a medida de informação de Shannon é a única possível permanece válida em relação aos problemas de codificação considerados por Shannon, mas não necessariamente noutro tipo de problemas. O matemático Alfréd Rényi evidenciou que outras quantidades poderiam ajustar–se igualmente, ou ainda melhor, a uma medida de informação. Tal seria sustentado pelo seu significado [de medida de informação] operacional ou por um conjunto de postulados que a caracterizasse, surgindo então na literatura a ideia de entropia generalizada.

Em 1961, Rényi propôs uma entropia paramétrica (parâmetro escalar) que contém a entropia de Shannon como caso limite,

$$S_\alpha = \frac{\ln\left(\frac{\sum_{i=1}^{n}(p_i)^\alpha}{\sum_{i=1}^{n} p_i}\right)}{1-\alpha}$$

sendo $0 < \sum_{i=1}^{n} p_i \leq 1$ e $\alpha \in \mathbb{R}$. Independentemente do valor deste parâmetro, a entropia é sempre aditiva, no entanto, nem sempre é côncava (ou convexa) em relação à distribuição de probabilidade.

Alfred Rényi[13]

A entropia de Rényi é uma forma funcional mais geral que a de Shannon, porém esta medida não tem encontrado tantas aplicações como a de Shannon. Na Física, estudos recentes

[13] Imagem retirada do endereço:
http://www1.cs.columbia.edu/∼ sanders/graphtheory/people/Renyi.A.html, em Março de 2003.

têm revelado que a entropia de Rényi é apropriada na análise de sistemas com estruturas multifractais. [27]

Fora do contexto da Termodinânica e da Teoria da Informação, a fórmula de Boltzmann–Gibbs mostrou ter limitações. Ela falha, por exemplo, ao tentar explicar a complexidade de fenómenos como um ciclone, ou a geometria fractal das moléculas de ADN (ácido desoxirribonucleico) e de outras macromoléculas. Basicamente a fórmula clássica mostrou–se inadequada quando a quantidade de entropia do sistema é não aditiva [36].

Em meados da década de oitenta, Constantino Tsallis definiu uma forma generalizada de entropia. A sua ideia deu origem ao artigo "Possible Generalization of Boltzmann–Gibbs Statistics", publicado em 1988 no *Journal of Statistical Physics* [21].

A entropia de Tsallis tem como caso particular a de Boltzmann–Gibbs. Estudos recentes mostram que esta é adequada a muitos casos onde a entropia clássica parece não funcionar, como no estudo de raios cósmicos, movimento de Galáxias, impulsos cerebrais e cardíacos, análise da sequência de genoma no ADN e até em ciências humanas como a linguística. Segundo Tsallis, os trabalhos mais promissores são os relacionados com a turbulência, um dos problemas mais complicados da Física [35]. Esta entropia já inspirou centenas de trabalhos em todo o mundo[14].

Constantino Tsallis[15]

A expressão proposta para generalizar a Mecânica Estatística de Boltzmann–Gibbs é caracterizada por um parâmetro "q". No caso limite em que $q = 1$ a fórmula reproduz a entropia tradicional. A grande diferença com a expressão de Boltzmann–Gibbs é a sua não aditividade.

[14] No endereço http://tsallis.cat.cbpf.br/biblio.htm encontra–se uma lista com cerca de um milhar de trabalhos. A lista é regularmente actualizada por Tsallis.

[15] Imagem retirada do endereço http://www.abc.org.br/org/aca.asp?codigo=tsallis da Academia Brasileira de Ciências, em Março de 2003.

A expressão introduzida por Tsallis para a entropia generalizada é

$$S_q \equiv k\,\frac{1-\sum_{i=1}^{\omega} p_i^q}{q-1} \tag{1.4}$$

onde k é uma constante positiva cujo valor depende da unidade a utizar, $q \in \mathbb{R}$ caracteriza a estatística particular e ω é o número total de configurações microscópicas cujas probabilidades são $\{p_i\}$.

A expressão (1.4) já tinha sido anteriormente introduzida noutros contextos tais como a Cibernética e a Teoria da Informação (utilizando um factor diferente) por Harvda e Charvat[16] em 1967. Em 1970, Daroczy[17] redescobre esta fórmula. Em 1975, Sharma e Mittal[18] definem uma forma biparamétrica que tem como caso particular a entropia de Tsallis [23]. Contudo, enquanto que estas tentativas foram elaboradas sem nenhum objectivo concreto para a Física, Tsallis teve sempre como objectivo a generalização da Mecânica Estatística e da Termodinâmica usuais.

Em 1997, é publicado um artigo da autoria de Roberto Santos [19], onde é apresentada uma interessante discussão sobre a unicidade da equação da entropia de Tsallis. Santos estabelece um conjunto de quatro axiomas e, seguindo as linhas da demonstração apresentada por Shannon [20], demonstra a unicidade da fórmula de Tsallis. Na axiomática de Santos procede–se a uma generalização natural dos axiomas de Shannon para sistemas não extensivos, estabelecendo de alguma forma um paralelismo com a axiomática de Shannon.

Na mesma linha de pensamento, Sumiyoshi Abe publica, em 2000, o artigo "Axioms and uniqueness theorem for Tsallis entropy" [1], no qual generaliza os axiomas de Khinchin baseando–se no conceito de entropia condicional não extensiva. Seguindo essencialmente o raciocínio apresentado em [15], expõe uma prova completa da unicidade do teorema de Tsallis, completando assim a discussão de Santos.

Existem outras formas de generalizadas de entropia para além destas. Contudo, a entropia de Tsallis e a de Rényi parecem ser as que, actualmente, despertam maior interesse na comunidade científica. No trabalho de Taneja [31] encontra–se uma longa lista com cerca de trinta fórmulas de entropias generalizadas, que vão desde as paramétricas, trigonométricas até às ponderadas[19].

[16] Harvda e Charvat (1967), "Quantification Method of Classification Processes: Concept of Structural a–Entropy", Kybernetica, vol.3, pp.30–35. [23]

[17] Daroczy, "Generalized Information Measures", Information and Control, vol.16, pp.36–51. [23]

[18] Sharma e Mittal (1975), "New Nonadditive Measures of Inaccuracy", J. Math. Sci., vol. 10, pp.122–133. [23]

[19] A ideia de uma entropia paramétrica surgiu pela primeira vez através de Rényi em 1961. Dois anos mais tarde, surgem as entropias trigonométricas por Aczél e Daróczy. Em 1968 as entropias ponderadas são iniciadas por Belis e Guaisu.[31]

1.5.1 Lista de Fórmulas de Entropia

Em [31], Taneja organiza uma lista com diversas fórmulas de Entropia, H_i. Com a devida autorização do autor transcreve–se aqui parte dessa lista. As fórmulas são diferenciadas por um índice, que não possui qualquer significado matemático.

No que se segue considere–se que os logaritmos são de base 2 e que $P = (p_1, \ldots, p_n)$ é uma distribuição de probabilidades no sentido usual, isto é,

$$0 \leq p_i \leq 1 \quad \text{e} \quad \sum_{i=1}^{n} p_i = 1 \ .$$

Contudo, em algumas das fórmulas apresentadas é necessário assumir que $p_i > 0$. Para simplificar a notação, denota–se $\sum_{i=1}^{n}$ apenas por $\sum_i$.

Shannon (1948):

$$H_1(P) = -\textstyle\sum_i p_i \log p_i$$

Rényi (1961):

$$H_2(P) = \frac{1}{1-r} \log \left(\textstyle\sum_i {p_i}^r\right) \qquad r > 0;\ r \neq 1$$

Aczél e Daróczy (1963):

$$H_3(P) = \frac{-\sum_i {p_i}^r \log p_i}{\sum_i {p_i}^r} \qquad r > 0$$

$$H_4(P) = \frac{1}{s-r} \log \left(\frac{\sum_i {p_i}^r}{\sum_i {p_i}^s}\right) \qquad r \neq s;\ r, s > 0$$

$$H_5(P) = \frac{1}{s} \arctan \left(\frac{\sum_i {p_i}^r \sin(s \log p_i)}{\sum_i {p_i}^r \cos(s \log p_i)}\right) \qquad s \neq 1;\ r, s > 0$$

Varma (1966):

$$H_6(P) = \frac{1}{m-r} \log \left(\textstyle\sum_i {p_i}^{r-m+1}\right) \qquad m \geq 1;\ m-1 < r < m$$

$$H_7(P) = \frac{1}{m(m-r)} \log \left(\textstyle\sum_i {p_i}^{r/m}\right) \qquad m \geq 1;\ 0 < r < m$$

Kapur (1967):

$$H_8(P) = \frac{1}{1-t}\log\left(\frac{\sum_i {p_i}^{t+s-1}}{\sum_i {p_i}^{s}}\right) \qquad t \neq 1;\ t > 0;\ s \geq 1$$

Harvda e Charvat[20] (1967):

$$H_9(P) = \frac{1}{2^{1-s}-1}\left(\sum_i {p_i}^{s} - 1\right) \qquad s \neq 1;\ s > 0$$

Belis e Guiasu (1968):

$$H_{10}(P) = -\sum_i p_i\, w_i \log p_i \qquad w_i > 0$$

Rathie (1970):

$$H_{11}(P) = \frac{1}{1-r}\log\left(\frac{\sum_i {p_i}^{r+s_i-1}}{\sum_i {p_i}^{s_i}}\right) \qquad r \neq 1;\ r > 0;\ s_i \geq 0$$

Arimoto (1971):

$$H_{12}(P) = \frac{1}{2^{t-1}-1}\left[\left(\sum_i p_i^{1/t}\right)^t - 1\right] \qquad t \neq 1;\ t > 0$$

Sharma e Mittal[21] (1975):

$$H_{13}(P) = \frac{1}{2^{1-s}-1}\left(2^{\{(s-1)\sum_i p_i \log p_i\}} - 1\right) \qquad s \neq 1;\ s > 0$$

$$H_{14}(P) = \frac{1}{2^{1-s}-1}\left[\left(\sum_i {p_i}^{r}\right)^{\frac{s-1}{r-1}} - 1\right] \qquad r,s \neq 1;\ r,s > 0$$

Sharma e Taneja (1975, 1977):

$$H_{15}(P) = -2^{r-1}\sum_i {p_i}^{r}\log p_i \qquad r > 0$$

$$H_{16}(P) = \frac{1}{2^{1-r}-2^{1-s}}\ \sum_i\left({p_i}^{r} - {p_i}^{s}\right) \qquad r \neq s;\ r,s > 0$$

$$H_{17}(P) = -\frac{2^{1-r}}{\sin s}\ \sum_i {p_i}^{r}\ \sin(s\log p_i) \qquad s \neq k\pi;\ k \in \mathbb{N};\ r > 0$$

[20] Repare–se que a expressão é igual à da entropia de Tsallis, a menos de uma constante.

[21] Se em H_{14} se tomar $s = r = 2$, obtém–se $H_{14}(P) = -2\left(\sum_i p_i^2 - 1\right)$, que é a entropia de Tsallis para o

Picard (1979):

$$H_{18}(P) = \frac{-\sum_i v_i \log p_i}{\sum_i v_i}$$

$$H_{19}(P) = \frac{1}{1-r} \log \left(\frac{\sum_i {p_i}^{r-1} v_i}{\sum_i v_i} \right) \quad r \neq 0;\, r > 0$$

Sant'anna e Taneja (1985):

$$H_{20}(P) = \sum_i p_i \log \left(\frac{\sin(s\, p_i)}{2 \sin(s/2)} \right) \qquad 0 < s < \pi$$

Kapur (1988):

$$H_{21}(P) = -\sum_i \log \Gamma(1 + p_i) \qquad \text{, onde } \Gamma \text{ é a função gama}$$

A primeira das entropias aqui listada, a entropia de Shannon, foi definida axiomaticamente. O capítulo seguinte é dedicado ao estudo dessa definição e das suas propriedades.

---

caso de $q = 2$, a menos de uma constante.

Capítulo 2

Entropia de Shannon

> "Suppose we have a set of possible events whose probabilities of occurrence (...) are known, but that is all we know concerning which event will occur. Can we find a measure of how much choice is involved in the selection of the event or of how uncertain we are of the outcome?"
>
> Claude Shannon, [20]

No capítulo anterior vimos que o conceito de Entropia surgiu pela primeira vez no contexto da Termodinâmica em 1865 e, quase um século depois, é associado por Shannon a uma quantidade de informação.

A definição axiomática de entropia elaborada por Shannon, foi publicada pela primeira vez em 1948. A expressão que traduz a entropia é definida a partir de três axiomas, dando origem ao conhecido *Teorema de Unicidade* de Shannon. Na literatura encontram–se diferentes formulações desta axiomática, por exemplo, as de Khinchin [15] e de Ash [3]. Este último assume quatro condições na sua axiomática, uma delas, conhecida por "axioma de grupo", pode ser considerada como uma generalização do axioma III de Shannon.

Na secção 1, apresentamos a axiomática de Shannon, o consequente Teorema de Unicidade e sua demonstração (seguindo essencialmente os passos descritos por Shannon em [20]). A demonstração deste teorema recorre a uma forma mais geral do axioma III de Shannon. Por esse motivo, propomos um raciocínio que leva à generalização do dito axioma, com base em argumentos intuitivos. Na secção 2, enumeram–se algumas das propriedades da entropia. Na secção seguinte definem–se entropias multivariadas, nomeadamente a entropia conjunta e condicionada e enumeram–se algumas das suas propriedades. No contexto da Teoria da Informação, Shannon aborda também o caso das variáveis contínuas definindo a entropia diferencial. Na secção 5, apresentamos a definição desta entropia, assim como as das entropias condicionada e conjunta (para variáveis contínuas). Na secção seguinte faz–se uma breve referência ao Princípio da Máxima Entropia.

2.1 Definição Axiomática de Shannon

Um conceito fundamental da Teoria da Informação é o de entropia que, muitas vezes, pode ter um significado matemático como quantidade numérica mensurável, com base num modelo probabilístico. Esta interpretação matemática permite que a solução de muitos problemas importantes de armazenamento e transmissão de informação, possam ser formulados em termos desta medida de quantidade de informação. A entropia de uma variável aleatória é definida em função da sua distribuição de probabilidade e revela–se como uma medida de incerteza ou de aleatoriedade.

2.1.1 Teorema da Unicidade

Considere–se uma experiência aleatória com n resultados possíveis cujas probabilidades de ocorrência são $p_1, p_2, \ldots, p_n$. Segundo Shannon, se existir uma medida $H(p_1, \ldots, p_n)$ que indique a incerteza sobre o resultado da experiência, é razoável impor-lhe as seguintes propriedades:

Propriedade I.
A função H deve ser contínua nos p_i.

Propriedade II.
Se $p_1 = \ldots = p_n = \frac{1}{n}$ então $H(n)$ deve ser uma função monótona crescente. Para acontecimentos equiprováveis há uma maior incerteza sobre o resultado, quando o número de resultados possíveis aumenta.

Propriedade III.
Se uma escolha for subdividida em duas escolhas sucessivas então a entropia original deve ser a soma ponderada das duas entropias individuais.[1]

As propriedades descritas permitem enunciar o Teorema de Shannon:

Teorema: *A única função H que satisfaz as propriedades I–III é da forma*

$$H(p_1, \ldots, p_n) = -k \sum_{i=1}^{n} p_i \ln p_i$$

onde k é uma constante positiva.

[1] No original, "If a choice be broken down into two succcessive choices, the original H should be weighted of the individual values of H", [20].

Esta quantidade H é designada por entropia do conjunto de probabilidades $(p_1, \ldots, p_n)$. Apesar de não estar enunciado no original, convenciona–se que $x \ln x$ é igual a zero quando $x = 0$. Esta convenção é justificada pelo facto de $\lim_{x \to 0} x \ln x = 0$, pelo que se preserva a continuidade de H. Por abuso de linguagem escreve–se $0 \ln 0 = 0$.

A cada experiência aleatória com espaço de resultados discreto e finito pode ser associada uma variável aleatória X que traduz os n resultados possíveis dessa experiência $(x_1, \ldots, x_n)$ e, consequentemente, a uma função de probabilidade $(p_1, \ldots, p_n)$, isto é,

$$p_i = P(X = x_i) \quad , \quad 0 \leq p_i \leq 1 \quad \text{e} \quad \sum_{i=1}^{n} p_i = 1\,.$$

Ao longo deste texto usam–se as notações $H(X)$, $H(P)$ ou $H(p_1, \ldots, p_n)$, para denotar a entropia,

$$H(X) = H(P) = H(p_1, \ldots, p_n) = -k \sum_{i=1}^{n} p_i \ln p_i\,. \tag{2.1}$$

A entropia pode ser interpretada como a quantidade de informação que se obtém ao realizar a experiência, ou equivalentemente, como a quantidade de incerteza que se tem sobre o resultado da experiência. Esta incerteza não depende dos valores que a variável aleatória X assume, nem de qualquer outra característica a não ser das probabilidades associadas a esses valores.

A função entropia de Shannon é facilmente representada graficamente, no caso de se terem dois resultados possíveis. As probabilidade p_1 e p_2, verificam a relação $p_2 = 1 - p_1$, pelo que H depende apenas de uma das probabilidades. O gráfico da função $H(p_1, p_2)$, ou seja, $H(p_1, 1 - p_1)$, é o seguinte:

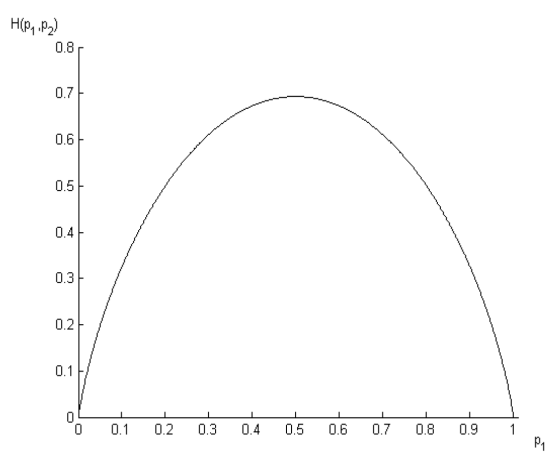


Gráfico da função entropia de Shannon $H(p_1, p_2)$.

A escolha da base do logaritmo corresponde meramente à definição de uma unidade para a medida de informação. Ao utilizar a base 2 a unidade resultante é o dígito binários ou abreviadamente *bit*, o que significa que a unidade de medida é a quantidade de incerteza associada a uma experiência com dois resultados possíveis. Na base 10 obtém–se o dígito decimal e na base e o dígito natural. A mudança da base a para a base b requer apenas a multiplicação por $\log_b a$, pelo que a existência da constante positiva k permite a passagem para qualquer base logarítmica (superior a 1). No livro de Shannon [20] é utilizada a notação log na definição da entropia, contudo, da leitura do livro, subentende–se que este símbolo denota o logaritmo neperiano[2]. Por esse motivo adopta–se, nesta dissertação, a notação neperiana na definição da entropia de Shannon.

2.1.2 Generalização do Axioma III

Antes de proceder à demonstração do teorema, é conveniente tornar claro o sentido da propriedade III. Em [20], Shannon apresenta uma situação particular por forma a exemplificar o seu significado. A expressão que resulta desse exemplo é a seguinte:

$$H(p_1, p_2, p_3) = H(p_1 + p_2, p_3) + (p_1 + p_2)H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) \tag{2.2}$$

A propriedade III é enunciada para o caso de se efectuar uma sequência de duas escolhas, conquanto é generalizável ao caso de se tomar mais do que duas sucessões de escolhas. A generalização encontra–se formulada, por exemplo, por Jaynes em [30].

Optou–se por descrever um processo que leva à generalização de (2.2), com base em raciocínios intuitivos (mas não imediatos) apresentados de forma construtiva, como se passa a descrever.

Considere–se um conjunto W formado por n elementos com probabilidades de ocorrência $p_1, \ldots, p_n$. A selecção de um elemento desse conjunto pode ser efectuada por uma sequência de duas escolhas, da seguinte forma:

1. distribuem–se os elementos de W por k conjuntos $C_1, \ldots, C_k$, sendo $q_1, \ldots, q_k$ as probabilidades associadas à ocorrência destes conjuntos;

2. cada conjunto C_i é formado por n_i elementos, sendo $a_{i,1}, \ldots, a_{i,n_i}$ as probabilidades associadas a cada um quando ocorre C_i (ou seja, as probabilidades condicionadas por C_i)

Para seleccionar um elemento escolhe–se em primeiro lugar um conjunto e desse conjunto escolhe–se um elemento. Esquematicamente, tem–se

[2] Em [20, pp. 89] encontra–se a passagem de $\frac{1}{2}$ para $\log \sqrt{e}$, pelo que se trata do logaritmo neperiano.

1^a escolha (conjunto i)	P(C_i)	2^a escolha (elemento j de C_i)	P(elemento j$\|$ C_i)	Prob. ocorrer elemento j de C_i
C_1	q_1	1	$a_{1,1}$	p_1
		2	$a_{2,1}$	p_2
		$\ldots$	$\ldots$	$\ldots$
		n_1	$a_{n_1,1}$	p_{n_1}
C_2	q_2	1	$a_{2,1}$	p_{n_1+1}
		2	$a_{2,2}$	p_{n_1+2}
		$\ldots$	$\ldots$	$\ldots$
		n_2	a_{2,n_2}	$p_{n_1+n_2}$
$\vdots$		$\vdots$	$\vdots$	$\vdots$
C_k	q_k	1	$a_{k,1}$	$p_{n_1+\ldots+n_{k-1}+1}$
		2	$a_{k,2}$	$p_{n_1+\ldots+n_{k-1}+2}$
		$\ldots$	$\ldots$	$\ldots$
		n_k	a_{k,n_k}	$p_{n_1+\ldots+n_k}$
k conjuntos		$\sum_{i=1}^{k} n_i = n$ elementos		1

sendo,

$$q_i = P(C = C_i) \qquad \text{e} \qquad a_{i,j} = P(X = x_j | C = C_i)$$

sendo C e X as variáveis aleatórias que indicam, respectivamente, o conjunto e o elemento seleccionados.

Repare–se que os elementos a ser seleccionados (pertencentes ao conjunto W) têm probabilidades de ocorrência $p_1, \ldots, p_n$, pelo que a probabilidade de seleccionar o elemento j do conjunto C_i, que é igual a $q_i \cdot a_{i,j}$, tem de verificar

$$q_i \cdot a_{i,j} = p_{n_1+\ldots+n_{i-1}+j} \tag{2.3}$$

ou seja,

$$\begin{cases} p_m = q_i \cdot a_{i,j} \\ m \;=\; \displaystyle\sum_{k=0}^{i-1} n_k + j \\ n_0 = 0 \end{cases} .$$

A condição III afirma que a entropia associada à escolha do elemento num só passo, isto é,

$$H(p_1 \ldots, p_n)$$

deve coincidir com a soma ponderada das entropias individuais (da sequência de escolhas).

Na primeira escolha tem de se optar por um de entre k possíveis conjuntos com probabilidades $q_1, \ldots, q_k$, pelo que a entropia associada é $H(q_1, \ldots, q_k)$. A entropia associada à escolha de um dos elementos do conjunto C_i, que ocorre com probabilidade q_i, é dada por $H(a_{i,1}, \ldots, a_{i,n_i})$. Dado que antes da segunda escolha pode ter ocorrido qualquer um dos conjuntos $C_1, C_2, \ldots, C_k$, então a entropia associada à segunda escolha é

$$q_1 \, H(a_{1,1}, \ldots, a_{1,n_1}) + \ldots + q_k \, H(a_{k,1}, \ldots, a_{k,n_k}) \, .$$

E da condição III resulta que

$$H(p_1, \ldots, p_n) = H(q_1, \ldots, q_k) + q_1 \, H(a_{1,1}, \ldots, a_{1,n_1}) + \ldots + q_k \, H(a_{k,1}, \ldots, a_{k,n_k}) \, .$$

Se algum dos q_i for nulo, a parcela $q_i \, H(a_{i,1}, \ldots, a_{i,n_i})$ é também nula, pelo que pode ser eliminada da expressão. Pode–se então supor que todos os q_i são não nulos e aplicar a relação (2.3) na igualdade anterior, obtendo–se

$$\begin{aligned} H(p_1, \ldots, p_n) &= H(q_1, \ldots, q_k) + q_1 \, H\left(\frac{p_1}{q_1}, \frac{p_2}{q_1}, \ldots, \frac{p_{n_1}}{q_1}\right) \\ &\quad + q_2 \, H\left(\frac{p_{n_1+1}}{q_2}, \frac{p_{n_1+2}}{q_2}, \ldots, \frac{p_{n_1+n_2}}{q_2}\right) + \ldots \\ &\quad + q_k \, H\left(\frac{p_{n_1+\ldots+n_{(k-1)}+1}}{q_k}, \frac{p_{n_1+\ldots+n_{(k-1)}+2}}{q_k}, \ldots, \frac{p_{n_1+\ldots+n_k}}{q_1}\right) \\ &= H(q_1, \ldots, q_k) + \sum_{j=1}^{k} q_j \, H\left(\frac{p_{\sum_{k=0}^{i-1} n_k+1}}{q_i}, \frac{p_{\sum_{k=0}^{i-1} n_k+2}}{q_i}, \ldots, \frac{p_{\sum_{k=0}^{i} n_k}}{q_i}\right) \, . \end{aligned}$$

Pelo que se pode generalizar a condição III da seguinte forma:

> Considerando um conjunto W com n resultados possíveis, sendo $p_1, \ldots, p_n$ a distribuição de probabilidade associada, e particionando esse conjunto nos subconjuntos $W_1, W_2, \ldots, W_k$ de modo a que o conjunto W_i tenha n_i resultados possíveis ($\sum_{i=1}^{k} n_i = n$) com probabilidades
>
> $$p_{b_i+1}, \, p_{b_i+2}, \, \ldots, \, p_{b_i+n_i} \qquad \text{, sendo } b_i = \sum_{j=0}^{i-1} n_j \text{ e } n_0 = 0$$
>
> e consequentemente a probabilidade de ocorrência do conjunto W_i é
>
> $$q_i = \sum_{m=b_i+1}^{b_i+n_i} p_m$$
>
> então, exige–se que
>
> $$H(p_1, \ldots, p_n) = H(q_1, \ldots, q_k) + \sum_{\substack{i=1 \\ q_i \neq 0}}^{k} q_i \, H\left(\frac{p_{b_i+1}}{q_i}, \frac{p_{b_i+2}}{q_i}, \ldots, \frac{p_{b_i+n_i}}{q_i}\right) \, . \qquad (2.4)$$

Este resultado é conhecido por axioma de grupo e encontra–se formalizado, por exemplo, por Robert Ash em [3].

2.1.3 Demonstração do Teorema da Unicidade

Atendendo à interpretação da entropia como uma quantidade da incerteza inerente ao resultado de uma experiência, é óbvio que, quando existe um único resultado, a entropia seja nula. Da expressão (2.2) facilmente se conclui que $H(1) = 0$. De facto, ao considerar $n = 2$, $p_1 = 1$ e $p_2 = 0$, resulta que

$$H(1, 0) = H(1) + 1 \cdot H(1, 0)$$

e portanto,

$$H(1) = 0 \; . \tag{2.5}$$

Consequentemente, pode–se provar que

$$H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n) \; . \tag{2.6}$$

De facto, basta observar que

$$H(p_1, \ldots, p_n, 0) = H(\; (p_1 + \ldots + p_n) + 0) \; + \; (p_1 + \ldots + p_n)\, H(p_1, \ldots, p_n)$$

$$H(p_1, \ldots, p_n, 0) = H(1) + H(p_1, \ldots, p_n)$$

pelo que,

$$H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n) \; .$$

Na demonstração do teorema de Shannon utilizam–se dois resultados, que podem ser deduzidos a partir das propriedades I–III, nomeadamente,

(i) A entropia associada à selecção de um elemento, de entre um conjunto com s^m elementos equiprováveis ($s, m \in \mathbb{N}$), numa única escolha é $H(\frac{1}{s^m}, \ldots, \frac{1}{s^m})$. Ao particionar essa escolha em m escolhas cada uma com s resultados equiprováveis verifica–se que

$$H\left(\frac{1}{s^m}, \ldots, \frac{1}{s^m}\right) = m\, H\left(\frac{1}{s}, \ldots, \frac{1}{s}\right) \; . \tag{2.7}$$

A dedução deste resultado encontra–se em [2].

(ii) O segundo resultado a utilizar é um caso particular da equação (2.4), generalização da propriedade III. Considere–se um conjunto com n resultados equiprováveis, isto é, com função de probabilidade $(p_1, \ldots, p_n) = (\frac{1}{n}, \ldots, \frac{1}{n})$. Se este conjunto for dividido em k

conjuntos $W_1, \ldots, W_k$, cada um com n_i resultados equiprováveis ($\sum_{i=1}^{k} n_i = n$), então a probabilidade de ocorrência de cada conjunto W_i é

$$q_i = \frac{n_i}{n} \, . \tag{2.8}$$

Da equação (2.4) resulta

$$H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = H(q_1, \ldots, q_k) + \sum_{i=1}^{k} q_i H\left(\frac{1/n}{q_i}, \ldots, \frac{1/n}{q_i}\right)$$

e como $n = \sum_{i=1}^{k} n_i$, então

$$H\left(\frac{1}{\sum_i n_i}, \ldots, \frac{1}{\sum_i n_i}\right) = H(q_1, \ldots, q_k) + \sum_{i=1}^{k} q_i \, H\left(\frac{1}{n_i}, \ldots, \frac{1}{n_i}\right) \tag{2.9}$$

sendo os q_i definidos por (2.8). Repare–se que $\frac{1}{n_i}$ é exactamente a probabilidade de selecção de um determinado elemento do conjunto W_i, condicionada pela ocorrência de W_i.

Tendo enunciado estas propriedades, procede–se agora à dedução da fórmula de entropia, seguindo as linhas apresentadas por Shannon em [20].

Seja A a função que define a entropia no caso da equiprobabilidade, isto é,

$$A(n) \equiv H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \, . \tag{2.10}$$

Seja $s \in \mathbb{N}\backslash\{1\}$ e $k \in \mathbb{N}$. O resultado (i) garante que existe uma relação entre a quantidade de entropia envolvida numa escolha de s^k resultados equiprováveis e a envolvida numa série de k escolhas todas equiprováveis, nomeadamente

$$A(s^k) = k \, A(s) \;\; , \;\; \forall \, k \in \mathbb{N} \, . \tag{2.11}$$

Seja $t \in \mathbb{N}\backslash\{1\}$. Para um inteiro n suficientemente elevado é possível[3] tomar um número inteiro m que satisfaça as desigualdades

$$s^m \leq t^n \leq s^{m+1} \, . \tag{2.12}$$

Dado que $s > 1$, pode–se escrever a seguinte sequência de desigualdades,

$$m \ln s \leq n \ln t \leq (m+1) \ln s$$

[3] De facto, a reunião dos intervalos da forma $[s^i, \, s^{i+1}]$ contém todos os números reais superiores a s, isto é, $\bigcup_{i \in \mathbb{N}} [s^i, \, s^{i+1}] = [s, +\infty[$. Para n suficientemente elevado verifica–se que $t^n \geq s$, pelo que a quantidade t^n terá obrigatoriamente de pertencer a um intervalo da forma $[s^i, s^{i+1}]$, para algum $i \in \mathbb{N}$.

$$\begin{aligned}\frac{m\ln s}{n\ln s} &\le \frac{n\ln t}{n\ln s} \le \frac{(m+1)\ln s}{n\ln s}\\ \frac{m}{n} &\le \frac{\ln t}{\ln s} \le \frac{m}{n}+\frac{1}{n}\,. \end{aligned} \tag{2.13}$$

Por outro lado, atendendo à propriedade II e à desigualdade (2.12) tem–se que

$$A(s^m) \le A(t^n) \le A(s^{m+1})$$

e pela equação (2.11),

$$m\cdot A(s) \le n.A(t) \le (m+1)A(s)\,. \tag{2.14}$$

Sendo A uma função estritamente positiva[4] pode–se concluir que

$$\frac{m}{n} \le \frac{A(t)}{A(s)} \le \frac{m}{n}+\frac{1}{n}\,. \tag{2.15}$$

Das desigualdades (2.13) e (2.15) pode–se afirmar que

$$\left|\frac{A(t)}{A(s)}-\frac{\ln t}{\ln s}\right| \le \frac{1}{n}\,.$$

Dado que a desigualdade acima se verifica para um inteiro n arbitrariamente elevado, é possível tornar o valor de $1/n$ arbitrariamente pequeno. No limite com n obtém–se

$$\frac{A(t)}{\ln t}=\frac{A(s)}{\ln s}\,.$$

Da arbitrariedade de t resulta que o quociente $A(t)/\ln t$ é uma constante, pelo que a função A é do tipo

$$A(t)=k\ln t \quad ,\ \forall\, t\in\mathbb{N}\,.$$

Desta igualdade e da definição de A em (2.10), resulta que

$$H\left(\frac{1}{t},\ldots,\frac{1}{t}\right)=k\ln t \quad ,\ \forall\, t\in\mathbb{N} \tag{2.16}$$

Para que a propriedade II se verifique, isto é, para que a função H seja monótona crescente no caso de equiprobabilidade, a constante k tem de ser positiva.

Considerando agora o resultado (2.9), isto é,

$$H\left(\frac{1}{\sum_{j=1}^{n} n_j},\ldots,\frac{1}{\sum_{j=1}^{n} n_j}\right)=H(p_1,\ldots,p_n)+\sum_{i=1}^{n} p_i\cdot H\left(\frac{1}{n_i},\ldots,\frac{1}{n_i}\right)$$

[4] Na realidade a equação (2.14) não permite garantir que $A(s)>0$, pois

$$m\,A(s)\le(m+1)\,A(s)\,,\ \forall m\in\mathbb{N} \Rightarrow A(s)\ge 0$$

Não foi considerado o caso $A(s)=0$, pelo motivo que se explica na subsecção (2.1.4).

onde os p_i são da forma $p_i = n_i/\sum_{j=1}^{n} n_j$ e aplicando a igualdade (2.16) em ambos os membros, obtém-se

$$k\,\ln\left(\sum_{j=1}^{n} n_j\right) = H(p_1,\ldots,p_n) + \sum_{i=1}^{n} p_i\, k\, \ln n_i \tag{2.17}$$

ou seja,

$$H(p_1,\ldots,p_n) = k\ln\left(\sum_{j=1}^{n} n_j\right) - k\,\sum_{i=1}^{n} p_i\,\ln n_i\;.$$

Dado que $\displaystyle\sum_{i=1}^{n} p_i = 1$, então

$$\begin{aligned}
H(p_1,\ldots,p_n) &= k\left(\sum_{i=1}^{n} p_i\right)\,\ln\left(\sum_{j=1}^{n} n_j\right) - k\sum_{i=1}^{n} p_i \ln n_i \\
&= k\sum_{i=1}^{n} p_i\,\ln\left(\sum_{j=1}^{n} n_j\right) - k\sum_{i=1}^{n} p_i \ln n_i \\
&= -k\sum_{i=1}^{n}\left(-p_i\ln\sum_{j=1}^{n} n_j + p_i\ln n_i\right) \\
&= -k\sum_{i=1}^{n}\left(p_i\left[-\ln\sum_{j=1}^{n} n_j + \ln n_i\right]\right) \\
&= -k\sum_{i=1}^{n} p_i\ln\left(\frac{n_i}{\sum_{j=1}^{n} n_j}\right)\,.
\end{aligned}$$

Por fim, aplicando a relação (2.8) cujos q_i são agora p_i obtém–se

$$H(p_1,\ldots,p_n) = -k\sum_{i=1}^{n} p_i \ln p_i\,.$$

onde os p_i são números fraccionários. Devido ao axioma da continuidade, a expressão permanece válida para o caso geral.

Fica assim demonstrado que as propriedades I–III conduzem a uma única função.

2.1.4 Comentário à demonstração

Na passagem da equação (2.14) para a equação (2.15) surge o problema de se ter, eventualmente, um denominador nulo, razão pela qual considerámos que $A(s) \neq 0$. No original [20], não se encontra nenhuma observação sobre este facto.

Tendo em consideração a interpretação da entropia como quantidade de incerteza associada ao resultado da experiência, torna–se óbvio que $A(s) = 0$ apenas quando $s = 1$ (o que não é o caso). Com efeito parece–nos que os axiomas não são suficientes para concluir que $A(s) \neq 0$ quando $s > 1$, pelo que não é legítimo adoptar este facto sem o ter definido previamente.

Para contornar este aspecto, propomos, por exemplo:

- que se considere a monotonia da função A no sentido estrito; ou

- que a unidade de medida de entropia seja definida por $H(\frac{1}{2}, \frac{1}{2})$.

No primeiro caso, devido ao facto que $H(1) = 0$, resulta que

$$A(n) = H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) > 0 \quad \forall n \in \mathbb{N} \backslash \{1\} .$$

No segundo caso, devido a $H(\frac{1}{2}, \frac{1}{2}) = 1$ (unidade) e à monotonia, resulta que

$$A(n) = H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \geq 1 \quad \forall n \in \mathbb{N} \backslash \{1\} .$$

Eliminando por completo a possibilidade de se ocorrer $A(s) = 0$ quando $s > 1$.

Evidenciamos que não é só na demonstração de Shannon [20] que se encontrou a "possibilidade" de ocorrer um denominador nulo. No trabalho de Khinchin [15], também não se vislumbra qualquer justificação para o facto de ser viável a divisão por $A(s)$.

2.2 Propriedades da Entropia de Shannon

A função H, definida por (2.1), é uma função real de várias variáveis reais, definida no conjunto

$$\Delta_n = \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n : \sum_{i=1}^{n} p_i = 1 \ , \ 0 \leq p_i \leq 1 \text{ para } i = 1, \ldots, n \right\} .$$

Esta medida satisfaz muitas propriedades interessantes. Apresentam–se de seguida algumas dessas propriedades. Por simplicidade de notação, nos enunciados que se seguem, considera–se que $k = 1$ e por vezes denota–se $H(p_1, \ldots, p_n)$ apenas por $H(P)$.

S1. $H(p_1, \ldots, p_n) \geq 0$ e a igualdade verifica–se se e só se $\exists^1 \ i \in \{1, \ldots, n\}$: $p_i = 1$ (pelo que $p_j = 0 \ , \ \forall j \neq i$).

S2. $H(p_1, \ldots, p_n)$ é contínua em Δ_n.

S3. $H(P)$ é uma função simétrica nos seus argumentos, ou seja,

$$H(p_1,\ldots,p_n) = H(p_{\delta(1)},\ldots,p_{\delta(n)})$$

onde δ denota uma permutação de 1 até n.

S4. Tem–se $H(p_1,\ldots,p_n,0) = H(p_1,\ldots,p_n)$.

S5. $H(p_1,\ldots,p_n)$ é uma função côncava em Δ_n.

S6. A função tem um máximo no caso de equiprobabilidade, isto é,

$$H(p_1,\ldots,p_n) \leq H\left(\frac{1}{n},\ldots,\frac{1}{n}\right) \quad , \ \forall\,(p_1,\ldots,p_n) \in \Delta_n \,.$$

S7. Tem–se

$$H(p_1,\ldots,p_n) = H(p_1+p_2,p_3,\ldots,p_n) + (p_1+p_2)H\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2}\right) .$$

S8. Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_m)$ duas distribuições e $P * Q$ a distribuição produto directo[5]. Então,

$$H(P * Q) = H(P) + H(Q)\,.$$

É de notar que algumas das propriedades enunciadas acima resultam directamente dos axiomas, como é o caso de (S2.) e (S7.). A demonstração da concavidade (S5.) e da existência de máximo (S6.) não foi encontrada na literatura consultada, pelo que a demonstração destas propriedades resultam unicamente do trabalho desenvolvido pelos autores deste texto.

Procede–se à demonstração de algumas das propriedades enunciadas.

S1. Verifica–se que $H(P) \geq 0$ devido ao facto de $-p_i \ln p_i > 0$, $\forall i$ e verifica–se a igualdade quando a função probabilidade é igual a $P = (1,\ldots,0)$, ou a uma sua permutação.
Por outro lado, se $H(P) = 0$, então $p_1 \ln p_1 + \ldots + p_n \ln p_n = 0$. Como as parcelas têm todas o mesmo sinal, a soma só é nula quando estas forem todas iguais a zero. Assim,

$$p_i \ln p_i = 0 \ \ , \ \forall i \in \{1,\ldots,n\}$$

o que implica $p_i = 0 \quad \vee \quad p_i = 1$ para cada $i \in \{1,\ldots,n\}$. Devido à condição $\sum_{i=1}^{n} p_i = 1$, resulta que

$$\exists^1 \ i \in \{1,\ldots,n\} \ : \ p_i = 1 \,.$$

[5] Define–se distribuição produto directo de $P = (p_1,\ldots,p_n)$ e $Q = (q_i,\ldots,q_m)$ por $P * Q = \{p_i q_j\}_{i,j}$, com $1 \leq i \leq n$ e $1 \leq j \leq m$.

S5. Para provar que a função H é côncava[6], considerem–se duas funções probabilidade $(p_1, \ldots, p_n)$ e $(q_1, \ldots, q_n)$ e $\lambda \in [0,1]$. Então,

$$H(\{\lambda p_i + (1-\lambda) q_i\}_i) - [\, \lambda\, H(\{p_i\}_i) + (1-\lambda)\, H(\{q_i\}_i)\,] =$$

$$= \; -\sum_{i=1}^{n} (\, \lambda p_i + (1-\lambda)\, q_i\,) \ln(\, \lambda p_i + (1-\lambda)\, q_i\,) + \lambda \sum_{i=1}^{n} p_i \ln p_i + (1-\lambda) \sum_{i=1}^{n} q_i \ln q_i$$

$$= \; \sum_{i=1}^{n} -(\lambda p_i + (1-\lambda)\, q_i\,) \ln(\, \lambda p_i + (1-\lambda)\, q_i\,) + \lambda p_i \ln p_i + (1-\lambda)\, q_i \ln q_i\, .$$

Considere–se $\varphi(x) = x \ln x$, definida em $\mathbb{R}^+$. Como $\varphi''(x) > 0$ em $\mathbb{R}^+$, então φ é uma função convexa. Portanto, para todos os $a, b \in \mathbb{R}^+$ e $0 \leq \lambda \leq 1$ tem–se,

$$\varphi(\, \lambda\, a + (1-\lambda)\, b\,) \leq \lambda\, \varphi(a) + (1-\lambda)\, \varphi(b)\, .$$

Substituindo φ pela sua expressão e fazendo $a = p_i$ e $b = q_i$, obtém–se

$$(\, \lambda p_i + (1-\lambda)\, q_i\,) \ln(\, \lambda p_i + (1-\lambda)\, q_i\,) \; \leq \; \lambda p_i \ln p_i + (1-\lambda)\, q_i \ln q_i \quad (i = 1, \ldots, n)$$

ou seja,

$$-(\lambda p_i + (1-\lambda)\, q_i\,) \ln(\lambda p_i + (1-\lambda)\, q_i\,) + \lambda p_i \ln p_i + (1-\lambda) q_i \ln q_i \; \geq \; 0 \quad (i = 1, \ldots, n)\, .$$

Conclui–se que o somatório anterior é não negativo e, portanto, H é uma função côncava.

S6. Pelo facto de Δ_n $(n \in \mathbb{N})$ ser um subconjunto fechado e limitado de $\mathbb{R}^n$ e de H ser uma função contínua, é garantido que H atinge um valor máximo e um valor mínimo em Δ_n.

Considerem–se duas funções probabilidade quaisquer, $\{p_i\}_i$ e $\{q_i\}_i$. Através do Teorema de Lagrange (ou Teorema do Valor Médio) é possível demonstrar que $\ln x \leq x - 1$ em $\mathbb{R}^+$. Logo,

$$\ln \frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1 \qquad , \; i = 1, \ldots, n$$

$$p_i \ln \frac{q_i}{p_i} \leq q_i - p_i \qquad , \; i = 1, \ldots, n$$

$$\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^{n} (q_i - p_i) = 1 - 1 = 0$$

Portanto,

$$\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leq 0$$

[6] Seja f uma função definida num subconjunto convexo B de um espaço vectorial. Diz–se que f é uma função côncava em B se, e só se, para todo o $x, y \in B$, se verificar

$$f(\, \lambda\, x + (1-\lambda)\, y\,) \geq \lambda\, f(x) + (1-\lambda)\, f(y) \quad , \; 0 \leq \lambda \leq 1\, .$$

No caso de se verificar a desigualdade contrária, diz–se que f é uma função convexa em B.

ou seja,

$$-\sum_{i=1}^{n} p_i \ln p_i \leq -\sum_{i=1}^{n} p_i \ln q_i \,.$$

Em particular, tomando $q_i = \frac{1}{n}$, resulta que

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln p_i \leq -\sum_{i=1}^{n} p_i \ln \frac{1}{n} = \ln n \sum_{i=1}^{n} p_i = \ln n \,.$$

Dado que

$$\begin{cases} H(p_1, \ldots, p_n) \leq \ln n \\ H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \ln n \end{cases}$$

conclui–se que o máximo de H é atingido no caso de equiprobabilidade.

S8. Considerem–se duas distribuições $P = (p_1, \ldots, p_n)$ e $Q = (q_1, \ldots, q_m)$. A distribuição produto directo é $P * Q = (\{p_i q_j\})$ com $1 \leq i \leq n$ e $1 \leq j \leq m$.

$$\begin{aligned} H(P*Q) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \ln(p_i q_j) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j (\ln p_i + \ln q_j) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \ln p_i + p_i q_j \ln q_j = -\sum_{i=1}^{n} p_i \ln p_i \sum_{j=1}^{m} q_j - \sum_{i=1}^{n} p_i \sum_{j=1}^{m} q_j \ln q_j \\ &= -\sum_{i=1}^{n} p_i \ln p_i - \sum_{j=1}^{m} q_j \ln q_j = H(P) + H(Q) \end{aligned}$$

Esta propriedade é conhecida por “aditividade”.

Da definição de esperança matemática pode–se ainda observar que a entropia de Shannon não é mais do que a esperança matemática[7] dos $\{-k \ln p_i\}_i$. De facto,

$$\begin{aligned} H(p_1, \ldots, p_n) &= \sum_{i=1}^{n} p_i \left(-k \ln p_i\right) \\ &= \langle -k \ln p_i \rangle \,. \end{aligned}$$

A entropia é uma medida da quantidade de informação que se obtém ao realizar uma experiência, dependendo de todos os resultados possíveis da experiência e não apenas daquele que efectivamente se concretiza. Tenha–se então em mente que esta medida traduz uma quantidade média da informação associada a cada um dos resultados possíveis e não a quantidade de informação do resultado particular que se concretiza, aquando da realização da experiência.

[7] A esperança matemática de uma variávela aleatória discreta $X = (x_1, \ldots, x_n)$ com função probabilidade f é definida por $\langle X \rangle = \sum_i^n x_i f(x_i)$. O valor esperado de uma função $\phi(X)$ é definido por $\langle \phi(X) \rangle = \sum_i^n \phi(x_i)\, f(x_i)$

2.3 Entropias Multivariadas

Na obra "A Mathematical Theory of Communication", Shannon define os conceitos de entropia conjunta e entropia condicionada, que envolvem duas variáveis aleatórias. Na Teoria da Informação estas quantidades multivariadas são utilizadas na análise do ruído e da capacidade do canal. Efectivamente, numa mensagem, cada símbolo emitido é condicionado pelos símbolos precedentes, ou seja, as sequências de símbolos são produzidas de acordo com certas probabilidades que dependem dos símbolos anteriores.

Sejam X e Y duas variáveis aleatórias discretas e finitas que assumem os valores $x_1, \ldots, x_n$ e $y_1, \ldots, y_m$, respectivamente, com funções probabilidade individuais e função probabilidade conjunta dadas por

$$p(x_i) = P(X = x_i) \qquad \text{com} \quad \sum_{i=1}^{n} p(x_i) = 1$$

$$p(y_j) = P(Y = y_j) \qquad \text{com} \quad \sum_{j=1}^{m} p(y_j) = 1$$

$$p(x_i, y_j) = P(X = x_i, Y = y_j) \quad \text{com} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1\ .$$

A probabilidade de $X = x_i$ condicionada pela ocorrência de $Y = y_j$ é denotada por

$$p(x_i|y_j) = P(X = x_i|Y = y_j) \quad \text{com} \quad \sum_{i=1}^{m} p(x_i|y_j) = 1\ .$$

Intuitivamente, não faz sentido definir a probabilidade de X tomar um valor x_i condicionado pela ocorrência de $Y = y_j$, ou seja, a probabilidade $p(x_i|y_j)$, quando não há a possibilidade de ocorrer $Y = y_j$, isto é, $p(y_j) = 0$. De facto, a probabilidade condicionada não está definida quando o acontecimento condicionante tem probabilididade nula.

Para cada $i = 1, \ldots, n$ e $j = 1, \ldots, m$, verificam–se as seguintes relações

$$\begin{aligned} p(x_i) &= \sum_{j=1}^{m} p(x_i, y_j) \\ p(y_j) &= \sum_{i=1}^{n} p(x_i, y_j) \end{aligned}$$

$$p(x_i|y_j) = \frac{p(x_i, y_j)}{p(y_j)} \qquad \text{com } p(y_j) \neq 0$$

$$p(y_j|x_i) = \frac{p(x_i, y_j)}{p(x_i)} \qquad \text{com } p(x_i) \neq 0$$

e, em particular, se X e Y forem variáveis aleatórias independentes tem–se

$$\begin{aligned} p(x_i|y_j) &= p(x_i) \\ p(y_j|x_i) &= p(y_j)\,. \\ p(x_i,y_j) &= p(x_i)\,p(y_j) \end{aligned}$$

A entropia conjunta de (X,Y) é definida por

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\ln p(x_i,y_j) \tag{2.18}$$

e as entropias individuais de X e Y são, respectivamente,

$$\begin{aligned} H(X) &= -\sum_{i=1}^{n} p(x_i)\ln p(x_i) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\ln p(x_i) \end{aligned}$$

e

$$\begin{aligned} H(Y) &= -\sum_{j=1}^{m} p(y_j)\ln p(y_j) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\ln p(y_j) \end{aligned}$$

convencionando que "$0\ln 0 = 0$" (abuso de linguagem já referido e justificado na página 19).

A entropia de X condicionada por um valor particular de Y, por exemplo $Y = y_j$, é definida por,

$$H(X|Y=y_j) = -\sum_{i=1}^{n} p(x_i|y_j)\ \ln p(x_i|y_j)$$

desde que a probabilidade condicionada esteja definida, isto é, $p(y_j) \neq 0$. A entropia conjunta $H(X,Y)$ pode ser interpretada como uma quantidade de incerteza sobre o resultado da experiência conjunta XY ou, equivalentemente, como a quantidade de informação que se obtém após a realização desta experiência.

A entropia de X condicionada por Y é, por definição, a média dos $H(X|Y=y_j)$ ponderada pelos $p(y_j)$, ou seja

$$H(X|Y) = \sum_{j=1}^{m} p(y_j)\,H(X|Y=y_j)$$

ou equivalentemente,

$$\begin{aligned} H(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(y_j)\,p(x_i|y_j)\ \ln p(x_i|y_j) && (2.19) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\ \ln p(x_i|y_j)\,. && (2.20) \end{aligned}$$

Esta traduz a quantidade de incerteza sobre o resultado de X (ou, equivalentemente, a quantidade de informação que se obtém ao realizar X), quando é conhecido o resultado de Y. Se algum dos $p(x_i|y_j)$ for nulo, a respectiva parcela em (2.20) toma o valor zero devido à convenção "$0 \ln 0 = 0$". Assim, desde que a probabilidade condicionada esteja definida, a equação (2.20) tem significado.

As medidas de entropia conjunta e entropia condicionada apresentadas para duas variáveis aleatórias podem ser generalizadas a três ou mais variáveis. As definições encontram–se, por exemplo, no livro de Ash [3].

Algumas das propriedades da entropia condicionada e da entropia conjunta são as seguintes:

S9. $H(X,Y) \geq 0$ e $H(X|Y) \geq 0$

S10. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

S11. $H(X,Y) \leq H(X) + H(Y)$

S12. $H(X,Y) = H(X) + H(Y)$ se, e só se, X e Y são variáveis independentes

S13. $H(X|Y) \leq H(X)$

S14. $H(X|Y) = H(X)$ se, e só se, X e Y são variáveis independentes

S15. $H(X,Y) \geq \max\{H(X), H(Y)\}$

As demonstrações resultam quase de imediato das definições anteriores. Repare–se que a aditividade referida anteriormente (S8.), coincide exactamente com a igualdade da propriedade (S12.).

2.4 Entropia Diferencial

A entropia de Shannon, pensada para variáveis aleatórias discretas, é também extendida a variáveis contínuas.

Para Shannon, a entropia de uma transmissão que envolve um conjunto contínuo de símbolos pode ser obtida por um processo limite do caso discreto, dividindo a mensagem contínua por um grande número (finito) de pequenas regiões e determinando os vários parâmetros numa base discreta. À medida que o tamanho dessas regiões diminui, esses parâmetros aproximam-se, em geral, dos valores limites do caso contínuo. [20]

Seja X uma variável aleatória contínua com função densidade $p(x)$. Shannon define a entropia da distribuição contínua por

$$H(X) = -\int_{-\infty}^{\infty} p(x) \, \ln p(x) \, dx \tag{2.21}$$

desde que o integral exista. Esta medida é conhecida por entropia diferencial.

De forma análoga ao caso discreto, também se definem entropias diferenciais multivariadas. Se X e Y forem duas variáveis aleatórias com função densidade conjunta $p(x,y)$, então a entropia conjunta é definida por

$$H(X,Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\ln p(x,y)\,dx\,dy \tag{2.22}$$

desde que o integral exista, e a entropia de X sendo conhecido Y, isto é, a entropia condicionada, é definida por

$$H(X|Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\ln p(x|y)\,dx\,dy\ , \tag{2.23}$$

desde que o integral exista, sendo

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x,y)}{\int_{-\infty}^{+\infty} p(x,y)dx}$$

a função densidade de X condicionada por Y.

A entropia diferencial verifica muitas das propriedades da entropia discreta, como por exemplo,

S'10. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

S'11. $H(X,Y) \leq H(X) + H(Y)$

S'12. $H(X,Y) = H(X) + H(Y)$ se, e só se, X e Y são variáveis independentes

S'13. $H(X|Y) \leq H(X)$

S'14. $H(X|Y) = H(X)$ se, e só se, X e Y são variáveis independentes

Uma das maiores diferenças entre a entropia de uma variável discreta e a entropia diferencial está no facto desta última poder ser infinitamente grande, positiva ou negativa.

Apresenta–se de seguida um exemplo de uma variável aleatória contínua com entropia negativa:

> Seja X uma variável aleatória com distribuição gausseana de média nula e variância igual a σ^2, então a sua função densidade é
>
> $$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}$$
>
> e facilmente se conclui que
>
> $$-\ln p(x) = \ln(\sigma\sqrt{2\pi}) + \frac{x^2}{2\sigma^2}\,.$$
>
> Assim,
>
> $$\begin{aligned} H(X) &= -\int_{-\infty}^{+\infty} p(x)\ln p(x)\,dx \\ &= \int_{-\infty}^{+\infty} p(x)\,\ln(\sigma\sqrt{2\pi})\,dx + \int_{-\infty}^{+\infty} p(x)\,\frac{x^2}{2\sigma^2}\,dx \\ &= \ln(\sigma\sqrt{2\pi})\int_{-\infty}^{+\infty} p(x)\,dx + \frac{1}{2\sigma^2}\int_{-\infty}^{+\infty} p(x)\,x^2\,dx\,. \end{aligned}$$

Como $p(x)$ é uma função densidade então $\int_{-\infty}^{+\infty} p(x)dx = 1$ e por definição de variância tem–se $\int_{-\infty}^{+\infty} p(x)x^2dx = \sigma^2$(pois X tem média nula). Logo,

$$H(X) = \ln(\sigma\sqrt{2\pi}) + \frac{1}{2} = \ln(\sigma\sqrt{2\pi}) + \ln(\sqrt{e}) = \ln(\sigma\sqrt{2\pi e})\,.$$

Portanto, basta que $\sigma < 1/\sqrt{2\pi e}$ para que a entropia de X seja negativa. Por exemplo, para $\sigma = 0,01$ tem–se $H(X) \approx -3,7$.

2.5 Princípio de Máxima Entropia

Algumas distribuições de probabilidade podem ser caracterizadas através do princípio de máxima entropia, desde que se conheça uma sua característica como a média ou a variância.

Seja X uma variável aleatória contínua com função densidade $p(x)$. Supondo que se conhece a média μ de X, isto é

$$\mu = \int_{-\infty}^{+\infty} x\,p(x)dx\,,$$

então existe um grande número de funções que satisfazem essa igualdade. Segundo Jaynes [30], se a função densidade $p(x)$ é desconhecida, então a incerteza sobre os valores de X é máxima, pelo que $p(x)$ deve ser uma função que maximiza a entropia

$$H(X) = -\int_{-\infty}^{+\infty} p(x)\,\ln p(x)\,dx\,.$$

É este princípio que permite restringir o conjunto inicial de funções a uma única função densidade.

O princípio de máxima entropia já tinha sido abordado por Shannon. No capítulo *Continuous Information* de [20], é enunciada a seguinte propriedade da entropia diferencial:

> A distribuição de probabilidade $p(x)$ que maximiza a entropia, sujeita à condição de que o desvio padrão da variável X tome um determinado valor σ, é a distribuição gausseana.

O autor explica que, para maximizar a entropia $H(X)$, com as restrições

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2\,p(x)dx \quad \text{e} \quad 1 = \int p(x)dx$$

é necessário resolver um problema de maximização que recorre ao cálculo de variações. Uma demonstração completa desta propriedade encontra–se em [4].

Shannon foi o primeiro a estabeler uma relação entre informação e entropia, podendo esta última ser interpretada como uma medida da primeira. Contudo, existem outras medidas de informação igualmente importantes na Teoria da Informação. No próximo capítulo são estudadas duas dessas medidas.

Capítulo 3

Informação

"Information must not be confused with meaning."

Warren Weaver, [20]

A informação foi introduzida na estatística por Fisher em 1925, num trabalho sobre teoria de estimação[1]. Hartley, em 1928 definiu uma medida de informação, no contexto da comunicação, para aplicar na engenharia. Em 1948, Shannon e Wiener publicam, independentemente, trabalhos onde descrevem medidas logarítmicas de informação, para utilizar na Teoria da Comunicação.

Para além da entropia, Shannon definiu outra medida de informação que quantifica a informação que um processo contém sobre outro, a qual chamou de razão de transmissão[2]. Nos trabalhos de Ash [3] e Yaglom [26] esta quantidade de informação é interpretada como uma redução na incerteza, quantificando a informação contida numa variável aleatória sobre outra. Esta é conhecida como informação mútua (ou simplesmente informação).

Em 1951, Kullback e Leibler[3] definiram também uma medida de informação associada a duas distribuições de probabilidade.

Neste capítulo são utilizadas as definições de entropia de Shannon, e respectivas entropias condicionada e conjunta, definidas no capítulo anterior.

Na secção 1, apresentamos a medida de informação mútua, introduzindo–a tal como Ash em [3]. Na secção 2, definimos o ganho de informação de Kullback–Leibler (também conhecida por entropia relativa), assim como algumas das suas propriedades. Dado que a função do

[1]Fisher (1925), "Theory of Statistical Estimation", Proc. Camb. Phil. Soc., vol.22, pp.700–725. [16]

[2]Tradução de *rate of transmition*, [20, pp.67–70].

[3]Kullback e Leibler (1951), "On information and Sufficiency", The annals of mathematical statistics, vol.22, 79–86. [31]

ganho de informação não é simétrica, apresenta–se uma sua versão simétrica que também foi estudada por Kullback e Leibler, a J–divergência.

3.1 Informação Mútua

Segundo Ash [3], a medida de informação deve ser definida como uma redução na incerteza, sendo esta última quantificada pela entropia.

Considerem–se duas experiências aleatórias α e β, cujos espaço de estão associados às variáveis aleatórias discretas X e Y, respectivamente. A incerteza sobre o resultado da experiência α pode, eventualmente, ser reduzida pela realização da experiência β. Este facto, encontra–se reflectido na propriedade

$$H(X|Y) \leq H(X)$$

já referida no capítulo anterior. A incerteza inicial sobre o resultado de α é $H(X)$ e após a realização de β é $H(X|Y)$. Ocorre portanto, um decréscimo de incerteza igual a $H(X) - H(X|Y)$. Define–se a quantidade de informação que Y contém sobre X por

$$I(X|Y) = H(X) - H(X|Y) \tag{3.1}$$

traduzindo–se assim a informação numa quantidade numérica.

Sejam X e Y duas variáveis aleatórias que tomam os valores $(x_1, \ldots, x_n)$ e $(y_1, \ldots, y_m)$, respectivamente. Sejam

$$\{\, p(x_i, y_j) \,\}_{i,j} \quad , \quad \{\, p(x_i) \,\}_i \quad \text{e} \quad \{\, p(x_i|y_j) \,\}_{i,j}$$

as funções probabilidade conjunta, probabilidade marginal de X e probabilidade condicionada de X por Y, respectivamente. Atendendo às equações (2.18) e (2.20), a informação mútua $I(X|Y)$ pode ser reescrita por

$$I(X|Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \, \ln \frac{p(x_i)}{p(x_i|y_j)} \; . \tag{3.2}$$

A informação mútua (ou simplesmente informação) é uma medida da informação partilhada entre duas variáveis, caracterizando a dependência entre duas variáveis. Se o valor for nulo as variáveis são independentes.

Na Teoria da Informação, a importância da medida de informação deve–se à sua aplicação em transmissões fidedignas de mensagens através de canais com ruído, [3].

Algumas propriedades da medida de informação mútua, que facilmente se provam a partir das propriedades da entropia, são as seguintes:

IM1. $I(X|Y) \geq 0$, verificando–se a igualdade se, e só se, X e Y são independentes.

IM2. $I(X|X) = H(X)$

IM3. $I(X|Y) = H(X) + H(Y) - H(X, Y)$

IM4. $I(X|Y) = I(Y|X)$

IM5. $I(X|Y) \leq min\{H(X), H(Y)\}$

A propriedade (IM2.) evidencia a estreita relação entre quantidade de informação e entropia. Sendo X a variável aleatória associada aos resultados de uma experiência aleatória α, a redução de incerteza que se obtém sobre α quando se conhece o seu resultado, "$I(X|X)$", é exactamente igual à quantidade de informação que se obtém ao realizar α, "$H(X)$". É então natural que a entropia seja também definida como a quantidade de incerteza associada ao resultado da experiência, [10].

A medida de informação é simétrica (IM4.). Intuitivamente, é surpreendente que a quantidade de informação que uma variável aleatória X contém sobre outra variável aleatória Y seja igual à quantidade que Y contém sobre X. Devido a esta simetria, a quantidade $I(X|Y)$ pode ser simplesmente chamada de informação mútua entre X e Y, sem se ter de identificar a ordem das variáveis.

3.2 Ganho de Informação de Kullback–Leibler

Em 1951, Kullback e Leibler definiram uma medida para o ganho de informação que quantifica a variação média da informação associada a duas distribuições de probabilidade. De seguida apresenta–se uma dedução da expressão de ganho de informação, com base no exposto no artigo [7], de Tsallis e outros.

Considere–se uma variável aleatória discreta X que toma os valores $x_1, \ldots, x_n$, com função probabilidade $P = (p_1, \ldots, p_n)$. A quantidade $\sigma_i = -\ln p_i$ pode ser interpretada como a informação contida no resultado i, assim a entropia de Shannon

$$S(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln p_i$$

pode ser interpretada como a esperança matemática dos σ_i.

Pretende–se determinar uma expressão que traduza a variação da quantidade de informação que se obtém quando se substitui a distribuição P por uma distribuição Q. Assumindo que se efectua um conjunto de medições que levam a um novo conjunto de valores para as probabilidades $Q = (q_1, \ldots, q_n)$, a variação de informação sobre o resultado i é então igual a

$$\Delta\sigma_i = -\ln p_i - (-\ln q_i)$$

e a variação média de informação é obtida calculando a esperança matemática dos $\Delta\sigma_i$ (sobre a nova distribuição Q), ou seja,

$$K(Q,P) = \sum_{i=1}^{n} q_i\,\Delta\sigma_i = \sum_{i=1}^{n} (\,-q_i\ \ln p_i + q_i \ln q_i\,) \tag{3.3}$$

ou ainda,

$$K(Q,P) = \sum_{i=1}^{n} q_i\ \ln \frac{q_i}{p_i}\,. \tag{3.4}$$

A quantidade definida por (3.4) traduz o ganho de informação resultante da substituição da distribuição (de referência ou *a priori*) P pela distribuição (*a posteriori*) Q. Na literatura encontram–se muitas designações e notações para o ganho de informação. A ideia foi estruturada por Kullback e Leibler, pelo que a quantidade é vulgarmente designada por ganho de informação de Kullback-Leibler[4] e denotada por $K(Q,P)$, notação esta que será a adoptada neste texto.

A dedução da expressão do ganho de informação de Kullback–Leibler que aqui se apresenta teve por base o artigo [7]. Contudo, nem neste artigo nem na restante literatura consultada, se encontrou qualquer referência sobre as restrições a que o argumento de K deve obedecer para que a quantidade esteja bem definida. Observe–se no entanto que, impondo que a distribuição P não tome qualquer valor nulo e adoptando a convenção[5] "$0\ln 0 = 0$", a expressão (3.4) fica bem definida.

Algumas das propriedades do ganho de informação de Kullback–Leibler são as seguintes:

K1. $K(Q,P) \geq 0$ e a igualdade verifica–se se e só se $Q = P$.

K2. $K(Q,P)$ é uma função simétrica nos seus argumentos Q e P, no seguinte sentido,

$$K(\,(q_1,\ldots,q_n)\,,\,(p_1,\ldots,p_n)\,) = K(\,(q_{\delta(1)},\ldots,q_{\delta(n)})\,,\,(p_{\delta(1)},\ldots,p_{\delta(n)})\,)$$

onde δ é uma permutação de 1 até n. No entanto, não é simétrica no par (Q,P), pois pode acontecer que

$$K(Q,P) \neq K(P,Q)$$

K3. Tem–se $K(\,(q_1,\ldots,q_n,0)\,,\,(p_1,\ldots,p_n,0)\,) = K(\,(q_1,\ldots,q_n)\,,\,(p_1,\ldots,p_n)\,)$.

[4]Também conhecida por entropia relativa, distância ou divergência de Kullback–Leibler.

[5]Convenção já referida e justificada na definição de entropia de Shannon (página 19).

K4. $K(Q,P)$ é uma função convexa do par (Q,P).

K5. Verifica–se que $K(Q_1 * Q_2, P_1 * P_2) = K(Q_1, P_1) + K(Q_2, P_2)$.

Na propriedade (K5.) o símbolo $*$ denota a operação produto directo entre duas distribuições, conforme definido na página 28 do capítulo anterior.

A maior parte das propriedades enunciadas tem demonstração quase imediata, razão pela qual apenas de apresenta a demonstração da não negatividade (K1.) e da convexidade (K4.). A demonstração desta última propriedade não foi encontrada em nenhuma da literatura consultada, pelo que esta é de exclusiva autoria dos autores deste trabalho.

Seguem–se as demonstrações:

K1. Através do Teorema de Lagrange é possível demonstrar que $\ln x \leq x - 1$ (verificando–se a igualdade apenas quando $x = 1$) em $\mathbb{R}^+$. Logo,

$$\ln \frac{p_i}{q_i} \leq \frac{p_i}{q_i} - 1 \qquad , \ i = 1, \ldots, n$$

$$q_i \ln \frac{p_i}{q_i} \leq p_i - q_i \qquad , \ i = 1, \ldots, n$$

$$\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} \leq \sum_{i=1}^{n} (p_i - q_i) = 1 - 1 = 0 \tag{3.5}$$

verificando–se a igualdade apenas quando $q_i/p_i = 1$ $(i = 1, \ldots, n)$. Por fim, dado que

$$K(Q,P) = \sum_i q_i \ln \frac{q_i}{p_i} = -\sum_i q_i \ln \frac{p_i}{q_i}$$

então $K(Q,P) \geq 0$, verificando–se a igualdade apenas quando $P = Q$.

K4. Considerem–se distribuições de igual dimensão $P = \{p_i\}_i$, $Q = \{q_i\}_i$, $\overline{P} = \{\overline{p}_i\}_i$ e $\overline{Q} = \{\overline{q}_i\}_i$, quaisquer. Para que a função K seja convexa, é necessário que para todo o $\lambda \in [0,1]$ se verifique o seguinte

$$K\left(\lambda\,(Q,P) + (1-\lambda)\,(\overline{Q},\overline{P})\right) \leq \lambda\, K(Q,P) \,+\, (1-\lambda)\, K(\overline{Q},\overline{P})$$

ou seja,

$$K(\,\{\lambda q_i + (1-\lambda)\overline{q}_i\}_i\,,\,\{\lambda p_i + (1-\lambda)\overline{p}_i\}_i\,) \leq \lambda\, K(\,\{q_i\}_i\,,\,\{p_i\}_i\,) + (1-\lambda)\, K(\,\{\overline{q}_i\}_i\,,\,\{\overline{p}_i\}_i\,)\,.$$

Substituindo K pela sua expressão, pretende–se provar que a desigualdade

$$\sum_{i=1}^{n} [\lambda q_i + (1-\lambda)\overline{q}_i] \ln \frac{\lambda q_i + (1-\lambda)\overline{q}_i}{\lambda p_i + (1-\lambda)\overline{p}_i} - \lambda \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i} - (1-\lambda) \sum_{i=1}^{n} \overline{q}_i \ln \frac{\overline{q}_i}{\overline{p}_i} \leq 0. \tag{3.6}$$

é válida para todo o $\lambda \in [0,1]$.

Da desigualdade (3.5) pode-se concluir que para duas funções probabilidade $\{w_i\}$ e $\{z_i\}$ quaisquer, se verifica

$$\sum_i w_i \ln z_i \leq \sum_i w_i \ln w_i \ .$$

Em particular, se para cada $i = 1, \ldots, n$ se fizer $w_i = \lambda\, q_i + (1-\lambda)\,\overline{q}_i$ e $z_i = \dfrac{w_i}{\lambda\, p_i + (1-\lambda)\,\overline{p_i}}$, pode–se concluir que

$$
\begin{aligned}
& \sum_{i=1}^{n} [\lambda\, q_i + (1-\lambda)\,\overline{q}_i] \ln \frac{\lambda\, q_i + (1-\lambda)\,\overline{q}_i}{\lambda\, p_i + (1-\lambda)\,\overline{p_i}} - \sum_{i=1}^{n} \lambda\, q_i \ln \frac{q_i}{p_i} - \sum_{i=1}^{n} (1-\lambda)\,\overline{q}_i \ln \frac{\overline{q}_i}{\overline{p_i}} \\
\leq \; & \sum_{i=1}^{n} [\lambda\, q_i + (1-\lambda)\,\overline{q}_i] \ln(\lambda\, q_i + (1-\lambda)\,\overline{q}_i) - \sum_{i=1}^{n} \lambda\, q_i \ln \frac{q_i}{p_i} - \sum_{i=1}^{n} (1-\lambda)\,\overline{q}_i \ln \frac{\overline{q}_i}{\overline{p_i}}
\end{aligned}
$$

$$
\begin{aligned}
= \; & \sum_{i=1}^{n} \{ \, [\lambda\, q_i + (1-\lambda)\,\overline{q}_i] \ln(\lambda\, q_i + (1-\lambda)\,\overline{q}_i) - \lambda\, q_i \ln q_i - (1-\lambda)\overline{q}_i \ln \overline{q}_i \, \} + \\
& + \sum_{i=1}^{n} \{ \, \lambda q_i \ln p_i + (1-\lambda)\overline{q}_i \ln \overline{p_i} \, \}
\end{aligned} \tag{3.7}
$$

Por um lado sabe–se que a função $\varphi(x) = x \ln x$ é convexa, pelo que o valor do primeiro somatório de (3.7) é não positivo. Por outro lado, devido ao facto de $\lambda, p_i, \overline{p}_i, q_i, \overline{q}_i \in [0,1]$, resulta que as parcelas do segundo somatório são todas não positivas. Assim, a desigualdade (3.6) verifica–se, pelo que K é uma função convexa.

O ganho de informação de Kullback–Leibler é uma quantidade que pode ser considerada como uma espécie de distância entre duas funções probabilidade, uma vez que é não negativa e é nula se, e apenas se, as duas funções forem iguais. Contudo, esta não é uma distância no sentido usual, uma vez que não é simétrica e não satisfaz a desigualdade triangular [13].

A versão simétrica, conhecida por J–divergência (estudada por Jeffreys[6] em 1946 e por Kullback e Leibler em 1951) é dada por

$$
\begin{aligned}
J(Q,P) &= K(Q,P) + K(P,Q) \\
&= \sum_k (q_k - p_k) \ln \frac{q_k}{p_k}\,.
\end{aligned}
$$

A J–divergência verifica todas as propriedades de uma distância, excepto a desigualdade triangular.

A informação mútua, definida por (3.2), pode ser considerada como um caso particular do ganho de informação de Kullback–Leibler. Efectivamente, no caso de se terem duas variáveis aleatórias discretas X e Y, sendo $Q = \{p(x_i, y_j)\}$ a função probabilidade conjunta e $P = \{p(x_i)p(y_j)\}$ o produto directo das funções probabilidade marginais, o ganho de informação de Kullback–Leibler reduz–se à informação mútua:

[6] Jeffreys (1946), "An invariant form for the prior probability in estimation problems", Proc. Roy. Soc (London), Série A, vol.186, pp.453–461. [16]

$$
\begin{aligned}
K(Q,P) &= \sum_i \sum_j p(x_i,y_j) \ln \frac{p(x_i,y_j)}{p(x_i)p(y_j)} \\
&= \sum_i \sum_j p(x_i,y_j) \ln \frac{p(x_i|y_j)}{p(x_i)} \\
&= -\sum_i \sum_j p(x_i,y_j) \ln \frac{p(x_i)}{p(x_i|y_j)} \\
&= I(X|Y) \,.
\end{aligned}
$$

Portanto, a informação mútua, pode ser interpretada como uma "distância" entre a distribuição conjunta e a distribuição produto, indicando se duas variáveis são, ou não, independentes.

O ganho de informação de Kullbach–Leibler também se encontra definido para o caso contínuo. Seja X uma variável aleatória contínua com função densidade de probabilidade p e seja q uma função densidade estimada, o ganho de informação de Kullback–Leibler é dado por,

$$
K(q,p) = \int_{-\infty}^{+\infty} q(x) \, \ln \frac{q(x)}{p(x)} \, dx
$$

quando o integral existe, sendo p designada por distribuição de referência. Esta definição é referida em diversa literatura, como é o caso de [14] e [22].

A título de curiosidade apresenta–se uma expressão, muito semelhante à da entropia de Shannon, mas que envolve duas distribuições de probabilidade. A medida de Kerridge[7] é definida da seguinte forma

$$
H(Q,P) = -\sum_i q_i \ln p_i \,.
$$

Esta medida verifica as propriedades enunciadas para o ganho de informação de Kullback–Leibler, excepto a de convexidade (é convexa apenas na segunda variável, P).

Neste capítulo foram apresentadas algumas medidas de informação cuja definição envolve o conceito de entropia de Shannon que, como se viu no segundo capítulo, foi definida axiomaticamente. No próximo capítulo, onde é apresentada a entropia de Rényi, é invertida a ordem de definição de entropia e de ganho de informação: é a partir da definição axiomática de ganho de informação que é deduzida uma expressão para a entropia de Rényi.

[7] A medida, também conhecida por *inaccuracy* (que se traduz por falta de exactidão), foi definida por Kerridge em 1961 no artigo "Inaccuracy and Inference" do Journal of the Royal Statistical Society (série B, vol.23, pp.184–194). [31]

Capítulo 4

Entropia de Rényi

> "(...) we give another characterization of the information; this approach will show what other quantities can be considered as measures of information besides that of Shannon's."
>
> Alfred Rényi, [18]

O matemático Rényi evidenciou que existem quantidades que podem ajustar–se a uma medida de informação tão bem como a entropia de Shannon, ou ainda melhor do que esta. Em 1961[1], apresentou uma fórmula para a entropia que, para além de depender da função probabilidade, está sujeita a um parâmetro real.

Como vimos no capítulo anterior, a partir da entropia de Shannon é possível definir uma quantidade para o ganho de informação, como é o caso da de Kullback–Leibler. Rényi propõe que se faça o raciocínio contrário. Como ponto de partida, define uma medida para o ganho de informação e, só depois, uma medida para a informação (entropia).

Neste capítulo propomo–nos a apresentar a entropia de Rényi, seguindo o raciocínio exposto por este em [18]. Na primeira secção, definimos alguns conceitos preliminares, nomeadamente, variável aleatória incompleta, distribuição incompleta e distribuição condicional completa. Na secção 2, descrevemos a proposta de Rényi para a definição da medida de ganho de informação. Seguidamente, na secção 3, é estabelecida a fórmula da entropia de Rényi, de acordo com o conceito de ganho de informação, para distribuições discretas e finitas. Na última secção apresentamos algumas propriedades da entropia de Rényi, apenas para o caso de distribuições completas, evidenciando as principais diferenças entre esta e a de Shannon.

No livro de Rényi [18] tanto a definição de ganho de informação como a de entropia envolvem o logaritmo de base 2. Contudo, vamos aqui adoptar a base neperiana, mantendo

[1] Rényi (1961), "On Measures of Entropy and Information", *Proc. 4th. Berk. Symp. Math. Statis. and Probl.*, University of California Press. [31]

assim a notação utilizada no capítulo "Entropia de Shannon". A base do logaritmo apenas tem influência sobre a unidade da medida de informação, pelo que se pode proceder a esta mudança de base.

4.1 Distribuições incompletas

Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade[2]. Rényi define variável aleatória incompleta como uma função $X = X(w)$ mensurável em relação à medida em $\mathcal{A}$ e definida num subconjunto Ω_1 de Ω, onde $\Omega_1 \in \mathcal{A}$ e $P(\Omega_1) > 0$. A única diferença entre uma variável aleatória usual e uma variável aleatória incompleta é que a última não está necessariamente definida em todo o espaço de resultados. Neste sentido, as variáveis aleatórias usuais podem ser consideradas como casos particulares de variáveis aleatórias incompletas.

A função probabilidade de uma variável aleatória incompleta X designa–se por função probabilidade incompleta (ou distribuição incompleta). Se a variável X assume os valores $(x_1, x_2, \ldots, x_n)$ com probabilidades $(p_1, p_2, \ldots, p_n)$ tem–se

$$0 < \sum_{i=1}^{n} p_i \leq 1$$

e não necessariamente $\sum_{i=1}^{n} p_i = 1$. Neste sentido, as distribuições usuais podem ser consideradas como casos particulares de distribuições incompletas. Repare–se que para este tipo de distribuições, a função distribuição de probabilidade, $F = P(X \leq x)$, verifica as seguintes propriedades

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{e} \quad \lim_{x \to +\infty} F(x) = s$$

sendo $\sum_{i=1}^{n} p_i = s \in [0, 1]$.

A cada função probabilidade incompleta $P = (p_1, \ldots, p_n)$ pode–se associar uma função probabilidade usual[3] $P' = (p'_1, \ldots, p'_n)$ fazendo

$$p'_i = \frac{p_i}{s} \quad (i = 1, \ldots, n)\,. \tag{4.1}$$

Assim, P' pode ser interpretada como uma função probabilidade condicional de X em relação à condição de que o resultado da experiência seja observável. Por isso se diz que P' é a função probabilidade condicionada completa da variável aleatória de X. A toda a função probabilidade incompleta P pode ser associada uma função probabilidade condicionada completa P' através das relações (4.1).

[2]Num espaço de probabilidades $(\Omega, \mathcal{A}, P)$, o conjunto Ω é um espaço de resultados, $\mathcal{A}$ é uma família de subconjuntos de Ω que constituem uma σ–álgebra de elementos de Ω e P é uma medida de probabilidade que a cada elemento de Ω faz corresponder um número real (que é a sua probabilidade).

[3]Isto é, tal que $\sum_i p_i = 1$

Define–se ainda o produto directo de duas funções probabilidade incompletas $P = (p_1, \dots, p_n)$ e $Q = (q_1, \dots, q_m)$, e denota–se por $P{*}Q$, a função que toma os valores $\{p_i\, q_j\}$ com $i = 1, \dots, n$ e $j = 1, \dots, m$.

4.2 Ganho de Informação de Ordem α

Rényi definiu axiomaticamente [18] uma quantidade para o ganho de informação que se obtém ao substituir uma distribuição por outra.

Seja X uma variável aleatória. Ao substituir a distribuição (*a priori*) incompleta P pela distribuição (*a posteriori*) incompleta Q, dá–se um "ganho" de informação que é denotado por $\mathrm{I}(Q||P)$. Rényi impõe que o ganho médio de informação inerente a esta substituição, satisfaça um conjunto de seis postulados, que não se transcreve aqui por estar fora do âmbito deste trabalho. Dos postulados resulta um teorema que define o ganho de informação.

Sejam $Q = (q_1, \dots, q_n)$ e $P = (p_1, \dots, p_n)$ distribuições incompletas com o mesmo número de termos, estando os seus índices relacionados por uma função biunívoca, sendo os p_i não nulos.

Teorema: Se $\mathrm{I}(Q||P)$ satisfaz os postulados I–VI[4], então ou existe um número real $\alpha \neq 1$ tal que

$$I_\alpha(Q||P) = \frac{1}{\alpha - 1} \ln \left(\frac{1}{\sum_{k=1}^{n} q_k} \sum_{k=1}^{n} \frac{(q_k)^\alpha}{(p_k)^{\alpha-1}} \right) \tag{4.2}$$

ou se tem $I(Q||P) = I_1(Q||P)$, com

$$I_1(Q||P) = \frac{1}{\sum_{k=1}^{n} q_k} \sum_{k=1}^{n} q_k \ln \left(\frac{q_k}{p_k} \right) \quad . \tag{4.3}$$

A prova deste teorema encontra–se em [18, pp. 574–578].

A quantidade $I_\alpha(Q||P)$ representa o ganho de informação de Rényi de ordem α e $I_1(Q||P)$ é o de ordem 1 ou, simplesmente, ganho de informação de Rényi. Facilmente se averigua que $I_1(Q||P)$ é um caso limite de $I_\alpha(Q||P)$, como se demonstra de seguida:

Por (4.2), tem–se

$$I_\alpha(Q||P) = \frac{1}{\alpha - 1} \ln \left(\frac{1}{\sum_k q_k} \sum_k \frac{q_k^\alpha}{p_k^{\alpha-1}} \right) = \frac{1}{\alpha - 1} \left[-\ln \left(\sum_k q_k \right) + \ln \left(\sum_k \frac{q_k^\alpha}{p_k^{\alpha-1}} \right) \right]$$

[4] Os postulados estão definidos em [18, pp. 570–573].

Defina–se

$$f(\alpha) = -\ln\left(\sum_k q_k\right) + \ln\left(\sum_k \frac{q_k^\alpha}{p_k^{\alpha-1}}\right)$$

Então, o limite procurado é dado por

$$\lim_{\alpha\to 1} I_\alpha(Q||P) = \lim_{\alpha\to 1} \frac{f(\alpha)}{\alpha - 1}$$

Aplicando a Regra de Cauchy, vem

$$= \lim_{\alpha\to 1} \frac{\dfrac{\partial}{\partial\alpha} f(\alpha)}{\dfrac{\partial}{\partial\alpha}(\alpha-1)} = \lim_{\alpha\to 1} \frac{\displaystyle\sum_k \frac{q_k^\alpha}{p_k^{\alpha-1}} \ln\left(\frac{q_k}{p_k}\right)}{\displaystyle\sum_k \frac{q_k^\alpha}{p_k^{\alpha-1}}} = \frac{1}{\displaystyle\sum_k q_k} \sum_k \left(q_k \ln\frac{q_k}{p_k}\right) = I_1(Q||P)\,.$$

Se P e Q forem distribuições completas, $I_1(Q||P)$ coincide com o ganho de informação de Kullback–Leibler, definido por (3.4).

Em [18] é exigido que os p_i sejam não nulos. Contudo, se $\alpha \leq 0$, o ganho de informação (4.2) apenas está definido quando os q_i são todos não nulos.

4.3 Entropia de Rényi

A partir da definição de ganho de informação – (4.2) e (4.3) – Rényi deduz uma fórmula para a medida de informação, isto é, para a entropia.

Seja $P = \mathcal{U}_n$ uma distribuição uniforme discreta de n termos[5], Q uma distribuição incompleta qualquer e $\alpha \neq 1$, então de (4.2) tem–se

$$I_\alpha(Q||\mathcal{U}_n) = \frac{1}{\alpha - 1} \ln\left(\frac{1}{\sum_{k=1}^n q_k} \sum_{k=1}^n \frac{(q_k)^\alpha}{\left(\frac{1}{n}\right)^{\alpha-1}}\right)$$

pelo que se verifica a seguinte relação

$$I_\alpha(Q||\mathcal{U}_n) = \ln n - \frac{1}{1-\alpha} \ln\left(\frac{\displaystyle\sum_{k=1}^n (q_k)^\alpha}{\displaystyle\sum_{k=1}^n q_k}\right)\,. \tag{4.4}$$

Para uma distribuição incompleta $P = (p_1, \ldots, p_n)$ qualquer[6], define–se

$$S_\alpha^R(P) = \frac{1}{1-\alpha} \ln\left(\frac{\displaystyle\sum_{k=1}^n (p_k)^\alpha}{\displaystyle\sum_{k=1}^n p_k}\right)$$

[5] Ou seja $\mathcal{U}_n = (\frac{1}{n}, \ldots, \frac{1}{n})$.

[6] Apesar de, em [18], não existirem nenhumas advertências neste sentido, considera–se aqui que, no caso de $\alpha \leq 0$, a distribuição terá de ser tal que $p_i \neq 0$, $\forall\, i = 1, \ldots, n$.

então, de (4.4) vem que

$$I_\alpha(Q||\mathcal{U}_n) = S^R_\alpha(\mathcal{U}_n) - S^R_\alpha(Q)\,. \tag{4.5}$$

Uma vez que $I(Q||\mathcal{U}_n)$ traduz o ganho de informação que se obtém ao substituir $\mathcal{U}_n$ por Q, e essa quantidade é igual ‘q diferença

$$S^R_\alpha(\mathcal{U}_n) - S^R_\alpha(Q)$$

então, é natural que se interprete S^R_α como uma medida da quantidade de informação associada à distribuição no seu argumnento. Define–se então $S^R_\alpha(P)$ como a entropia de Rényi de ordem α. No caso particular de P ser uma distribuição completa, a definição de $S^R_\alpha(P)$ reduz–se a

$$S^R_\alpha(P) = \frac{1}{1-\alpha}\ln\left(\sum_{k=1}^{n}(p_k)^\alpha\right)\,.$$

Define–se de seguida a entropia de Rényi para o caso limite de $\alpha = 1$. Da equação (4.3), facilmente se averigua que

$$I_1(Q||\mathcal{U}_n) = \ln n - \frac{\displaystyle\sum_{k=1}^{n} q_k\, \ln\left(\frac{1}{q_k}\right)}{\displaystyle\sum_{k=1}^{n} q_k}\,.$$

Por forma a estabelecer uma expressão similar a (4.5) para o caso $\alpha = 1$, define–se

$$S^R_1(P) = \frac{\displaystyle\sum_{k=1}^{n} p_k\, \ln\left(\frac{1}{p_k}\right)}{\displaystyle\sum_{k=1}^{n} p_k}$$

sendo $P = (p_1, \ldots, p_n)$ uma distribuição incompleta qualquer. Assim,

$$I_1(Q||\mathcal{U}_n) = S^R_1(\mathcal{U}_n) - S^R_1(Q)\,.$$

Fazendo uma interpretação análoga ao caso anterior, define–se S^R_1 como a entropia de Rényi de ordem 1. Se P for uma distribuição completa, $S^R_1(P)$ reduz–se à entropia de Shannon. De facto,

$$S^R_1(P) = -\sum_{k=1}^{n} p_k\, \ln p_k\,.$$

Repare–se ainda que a entropia de Rényi de ordem 1 traduz exactamente o caso limite de S^R_α para $\alpha = 1$. Para toda a função distribuição completa ou incompleta, verifica–se que

$$\lim_{\alpha\to 1} S^R_\alpha(P) = S^R_1(P)\,.$$

Em suma, a entropia de Rényi pode tomar as seguintes formas:

(i) Para uma distribuição incompleta $P = (p_1, \ldots, p_n)$ e $\alpha \neq 1$, a entropia de Rényi de ordem α é dada por

$$S_\alpha^R(P) = \frac{1}{1-\alpha} \ln \left(\frac{\sum_{k=1}^{n} (p_k)^\alpha}{\sum_{k=1}^{n} p_k} \right) . \tag{4.6}$$

No caso limite $\alpha = 1$, tem–se

$$S_1^R(P) = \frac{\sum_{k=1}^{n} p_k \ln \left(\frac{1}{p_k} \right)}{\sum_{k=1}^{n} p_k} . \tag{4.7}$$

(ii) Para uma distribuição completa $P = (p_1, \ldots, p_n)$, a entropia de Rényi de ordem 1 é dada por

$$S_\alpha^R(P) = \frac{1}{1-\alpha} \ln \left(\sum_{k=1}^{n} (p_k)^\alpha \right) . \tag{4.8}$$

No caso limite $\alpha = 1$, obtém–se a entropia de Shannon

$$S_1^R(P) = \sum_{k=1}^{n} p_k \ln \left(\frac{1}{p_k} \right) . \tag{4.9}$$

É evidente que em (4.6) e em (4.8), as probabilidades não podem ser nulas quando $\alpha \leq 0$, pois em tal situação as expressões não têm significado (não estão definidas). Apesar de tal não ser referido em [18] considera–se aqui que, quando $\alpha \leq 0$, a função é aplicável apenas a distribuições em que os p_i são todos positivos.

No caso de $\alpha > 0$, o domínio da função S_α^R é

$$\Delta_n = \left\{ (p_1, \ldots, p_n) : 0 \leq p_i \leq 1 \wedge 0 < \sum_{i=1}^{n} p_i \leq 1 \right\} .$$

Como se acabou de expor Rényi não define a sua entropia de forma axiomática[7] (ao contrário de Shannon). A entropia de Rényi resulta de algumas considerações naturais que são tomadas depois de ter definido axiomaticamente uma medida para o ganho de informação. Em 1964, esta entropia foi axiomatizada por Daróczy[8] e em 2002, Arimitsu e Jizba [27], propõem uma axiomática que caracteriza simultaneamente a entropia de Rényi, para distribuições completas, e a entropia de Shannon.

[7] Em [27] é afirmado que Rényi define a entropia axiomaticamente, facto com o qual não concordamos.

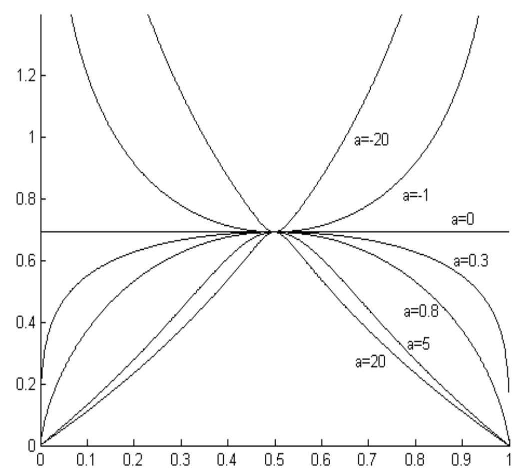
[8] Daróczy (1963), *Acta Mathematica Academiae Scientiarium Hungaricae*, vol.15, pp.203. [27]

Apesar da entropia de Rényi estar definida para distribuições incompletas, a maior parte da literatura consultada refere–se à entropia de Rényi apenas no caso de distribuições completas. Tendo em conta esse aspecto, daqui em diante trata–se apenas este caso particular, isto é, a entropia de Rényi na forma (4.8).

4.4 Propriedades da Entropia de Rényi

A entropia de Rényi, podendo ser interpretada como uma generalização da entropia de Shannon verifica muitas das suas propriedades, nomeadamente, é não negativa, aditiva e tem um extremo no caso da equiprobabilidade. A diferença fundamental entre estas entropias consiste na não conservação de côncavidade, propriedade esta que depende do valor do parâmetro α.

Na figura seguinte está representado o gráfico da função entropia de Rényi para diferentes valores de α, no caso de distribuições completas com dois resultados possíveis.



Entropia de Rényi para distibuições completas.
Gráfico de $S_a^R(p_1, p_2)$ para diferentes valores de a.

De seguida, listam–se algumas das principais propriedades da entropia de Rényi definida por (4.8), algumas das quais estão enunciadas no original de Rényi [18] e no artigo de Curado e Tsallis [9]. Uma vez que se está a considerar o caso particular (4.8) da entropia de Rényi, o domínio desta função é um subconjunto de Δ_n, nomeadamente

$$\delta_n = \left\{ (p_1, \ldots, p_n) : 0 \leq p_i \leq 1 \wedge \sum_{i=1}^{n} p_i = 1 \right\}$$

exceptuando o caso de $\alpha < 0$ para o qual os p_i têm de ser não nulos. No que se segue, $P = (p_1, \ldots, p_n)$ denota uma distribuição de probabilidade completa.

R1. $S^R_\alpha(P) \geq 0$ verificando–se a igualdade no caso da certeza e de $\alpha > 0$, isto é,

$$\alpha > 0 \quad \wedge \quad \exists|\, i \in \{1, \ldots, n\} \, : \, p_i = 1 \, .$$

R2. $S_\alpha(p_1, \ldots, p_n)$ é contínua no seu domínio.

R3. $S^R_\alpha(P)$ é uma função simétrica nos seus argumentos, isto é, se θ for uma permutação de 1 até n, tem–se

$$S^R_\alpha(p_1, \ldots, p_n) = S^R_\alpha(p_{\theta(1)}, \ldots, p_{\theta(n)}) \, .$$

R4. Se $\alpha > 0$, então $S^R_\alpha(p_1, \ldots, p_n, 0) = S^R_\alpha(p_1, \ldots, p_n)$.

R5. Se $P = (\frac{1}{n}, \ldots, \frac{1}{n})$, então a função é monótona crescente em n e é definida por

$$S^R_\alpha(P) = \ln n \, .$$

R6. Quando $0 < \alpha \leq 1$, a função $S^R_\alpha(P)$ é côncava na variável P.

R7. A função $S^R_\alpha(P)$ tem um extremo no caso de equiprobabilidade. E verifica–se que,

$$S^R_\alpha(p_1, \ldots, p_n) \leq \ln n \, , \text{ para } \alpha \in]0, 1]$$

$$S^R_\alpha(p_1, \ldots, p_n) \geq \ln n \, , \text{ para } \alpha \notin]0, 1] \, .$$

R8. Sejam P e Q funções probabilidade e $P * Q$ o seu produto directo. Então,

$$S^R_\alpha(P * Q) = S^R_\alpha(P) + S^R_\alpha(Q) \, .$$

R9. Sejam P e Q as funções probabilidade marginais de duas variáveis aleatórias e C a sua distribuição conjunta. Então,

$$S^R_\alpha(C) \leq S^R_\alpha(P) + S^R_\alpha(Q) \;\; \Leftrightarrow \;\; \alpha = 1 \, .$$

Para quem está familiarizado com a entropia de Shannon, a propriedade (R9.) é uma grande surpresa. De facto, a desigualdade é válida[9] para a entropia de Shannon. A surpreendente implicação contrária encontra–se demonstrada em [18]. De seguida procede–se à demonstração das propriedades (R6.) e (R7.).

[9]Cf. propriedade (S11.) da entropia de Shannon, enunciada na página 33.

R6. Seja $0 < \alpha < 1$. Considerem–se duas funções probabilidade $(p_1, \ldots, p_n)$ e $(q_1, \ldots, q_n)$. Para que a função S_α^R seja côncava é necessário que, para todo o $\lambda \in [0,1]$, se verifique o seguinte:

$$S_\alpha^R(\{\lambda p_i + (1-\lambda) q_i\}_i) \geq \lambda S_\alpha^R(\{p_i\}_i) + (1-\lambda) S_\alpha^R(\{q_i\}_i)\,.$$

A função $\phi(x) = x^\alpha$ é côncava em [0,1] para $0 < \alpha < 1$, portanto

$$(\,\lambda\, p_i + (1-\lambda)\, q_i\,)^{\,\alpha} \;\geq\; \lambda\, p_i^\alpha + (1-\lambda)\, q_i^\alpha \qquad \text{para} \quad i = 1, \ldots, n$$

logo,

$$\sum_{i=1}^{n} (\lambda\, p_i + (1-\lambda)\, q_i)^{\,\alpha} \geq \lambda \sum_{i=1}^{n} p_i^\alpha + (1-\lambda) \sum_{i=1}^{n} q_i^\alpha\,. \tag{4.10}$$

A função $\varphi(x) = \ln x$ é côncava em $\mathbb{R}^+$, portanto

$$\ln\left(\lambda \sum_{i=1}^{n} p_i^\alpha + (1-\lambda) \sum_{i=1}^{n} q_i^\alpha\right) \geq \lambda\, \ln\left(\sum_{i=1}^{n} p_i^\alpha\right) + (1-\lambda)\, \ln\left(\sum_{i=1}^{n} q_i^\alpha\right)\,. \tag{4.11}$$

Uma vez que $0 < \alpha < 1$, então $\frac{1}{1-\alpha}$ é positivo e pode–se escrever as seguintes desigualdades

$$\begin{aligned}
&S_\alpha^R(\{\lambda p_i + (1-\lambda) q_i\}) \\
&= \frac{1}{1-\alpha}\, \ln\left(\sum_{i=1}^{n} (\lambda\, p_i + (1-\lambda)\, q_i)^{\,\alpha}\right) \\
&\geq \frac{1}{1-\alpha}\, \ln\left(\lambda \sum_{i=1}^{n} p_i^\alpha + (1-\lambda) \sum_{i=1}^{n} q_i^\alpha\right) && \text{por (4.10)} \\
&\geq \frac{1}{1-\alpha}\, \left(\lambda\, \ln\left(\sum_{i=1}^{n} p_i^\alpha\right) + (1-\lambda)\, \ln\left(\sum_{i=1}^{n} q_i^\alpha\right)\right) && \text{por (4.11)} \\
&= \lambda\, S_\alpha^R(\{p_i\}) + (1-\lambda)\, S_\alpha^R(\{q_i\})\,.
\end{aligned}$$

Portanto, se $0 < \alpha < 1$ a função S_α^R é côncava.

No caso limite $\alpha = 1$, a entropia de Rényi coincide com a entropia de Shannon, que é uma função côncava.

No caso de $\alpha > 1$, é possível concluir graficamente que a entropia de Rényi nem sempre mantém o sentido de concavidade. Veja–se o exemplo de $\alpha = 20$ no gráfico anterior. Contudo, através da análise gráfica da segunda derivada da função S_α^R, conjecturamos que a entropia de Rényi é ainda côncava para $1 < \alpha < 2$.

Na literatura consultada, não se encontrou nenhuma referência sobre a concavidade da função no caso de $\alpha < 0$. Graficamente, conjectura–se que a função seja convexa, no entanto não foi possível demonstrá–lo. A análise deste aspecto fica destinada a um estudo posterior.

R7. Para valores de $\alpha > 0$, a função $S_\alpha^R(p_1, \ldots, p_n)$ está definida em δ_n que é um subconjunto fechado e limitado de $\mathbb{R}^n$, pelo que é garantida a existência de mínimo e máximo da função em δ_n.

Aplicando o método dos multiplicadores de Lagrange, os candidatos a extremantes da função S_α^R, sujeita à condição de $\sum_{i=1}^n p_i = 1$, são tais que

$$\begin{cases} \nabla S_\alpha^R(p_1, \ldots, p_n) = \lambda \nabla g(p_1, \ldots, p_n) \\ g(p_1, \ldots, p_n) = 1 \end{cases} \tag{4.12}$$

desde que as derivadas parciais existam, sendo $g(p_1, \ldots, p_n) = \sum_{i=1}^n p_i$. Do sistema (4.12) obtém–se

$$\begin{cases} \dfrac{1}{1-\alpha} \dfrac{\alpha\, p_i^{\alpha-1}}{\sum p_i^\alpha} = \lambda \\ \\ \sum p_i = 1 \end{cases}$$

portanto as derivadas parciais existem sempre em δ_n. Resolvendo este sistema obtém–se

$$p_1 = \ldots = p_n = \frac{1}{n}\,.$$

Portanto, um extremo da função S_α^R é

$$S_\alpha^R\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \ln n\,.$$

Quando $\alpha \in]0, 1]$ a função é côncava, pelo que se pode concluir que este extremo é um máximo. No caso limite de $\alpha = 1$ a entropia de Rényi reduz–se à entropia de Shannon, que é máxima no caso da equiprobabilidade[10]. Por fim, no caso de $\alpha > 1$ tem–se $S_\alpha^R(1, 0, \ldots, 0) = 0 \leq \ln n$, pelo que o extremante determinado não pode ser um minimizante. Consequentemente é um maximizante.

Para valores de α negativos, o domínio da função já não é um conjunto compacto, pelo que não garante a existência de máximo e mínimo. Contudo, verifica–se que se um dos p_i tender para zero ou um, a função diverge para valores infinitamente grandes (veja a propriedade (R11.) e (R12.) mais adiante). E uma vez que a função é contínua no seu domínio e o ponto $(\frac{1}{n}, \ldots, \frac{1}{n})$ pertence ao seu interior, pode–se concluir que o extremo determinado pelo método dos multiplicadores de Lagrange é um mínimo.

Rényi apercebeu–se que, se $\alpha \leq 0$, a quantidade S_α^R verifica determinadas propriedades, que são inconvenientes para uma medida de informação. Assim, considerou que a quantidade S_α^R deve ser interpretada como uma medida de informação, apenas quando α é positivo. O autor explica esta advertência através das seguintes propriedades:

R10. $S_0^R(P)$ é constante, quaisquer que sejam os termos positivos de $P = (p_1, \ldots, p_n)$,

$$S_0^R(P) = \ln n\,.$$

[10] Cf. propriedade (S6.) enunciada na página 28.

R11. Se $\alpha < 0$, então
$$\lim_{p_i \to 0} S_\alpha^R(p_1, \ldots, p_n) = +\infty$$
sendo i um número natural entre 1 e n.

R12. Se $\alpha < 0$, então
$$\lim_{p_i \to 1} S_\alpha^R(p_1, \ldots, p_n) = +\infty$$
sendo i um número natural entre 1 e n.

A propriedade (R10.) evidencia que a entropia $S_0^R(P)$ é independente da distribuição de probabilidade P, dependendo apenas da sua dimensão (número de resultados possíveis). Por sua vez, a propriedade (R11.) revela que, depois de adicionar um novo evento de probabilidade quase nula, o ganho de informação é infinito[11]. Ambas as propriedades são impróprias para uma medida de informação. A propriedade (R12.) é uma consequência de (R11.).

A quantidade S_α^R deve ser interpretada como uma medida de informação, apenas quando α é positivo. Contudo, em certas situações é desejável admitir a possibilidade de α ser negativo, pois existem sistemas caóticos em que o parâmetro α toma valores negativos [27].

No próximo capítulo apresenta–se uma outra forma de entropia paramétrica, a entropia de Tsallis. Esta entropia é mais recente que a de Rényi e tem merecido grande atenção por parte da comunidade científica. Durante a última década muitos trabalhos têm revelado a sua aplicabilidade em domínios tão complicados como o do estudo da turbulência.

[11] Em oposição ao que acontece no caso de $\alpha > 0$. Relembre–se a propriedade (R4.) que afirma que a entropia não se altera quando se adiciona um evento de probabilidade nula

Capítulo 5

Entropia de Tsallis

> "Tive a ideia de propor o uso de uma generalização da fórmula de Boltzmann em 1985. Levei três anos para estudar se esta equação fazia sentido fisicamente."
>
> Constantino Tsallis, [35]

A Mecânica estuda as interacções entre a matéria e as forças que nela actuam. A Mecânica Estatística[1] é um ramo da Física onde os métodos estatísticos são aplicados sobre as componentes microscópicas de um sistema, por forma a permitir a previsão de propriedades macroscópicas. A Mecânica Estatística e Termodinâmica não extensivas extendem o domínio de aplicabilidade dos procedimentos da Mecânica Estatística usual, a sistemas onde a estatística de Boltzmann–Gibbs apresenta sérias dificuldades matemáticas, ou até mesmo falha. De entre uma longa lista, dão–se os exemplos de sistemas que envolvam interacções de longo alcance, memória microscópica de longo alcance o os buracos negros [23]. Esta nova corrente que atravessa os fundamentos da Mecânica Estatística iniciou–se, em 1988, com a proposta de Tsallis [21] de uma entropia não extensiva, a entropia de Tsallis ou q–entropia.

Neste capítulo propomos uma viagem sobre a entropia dos subíndices q. Na primeira secção, apresentamos algumas funções generalizadas necessárias à abordagem da Mecânica Estatística Não Extensiva, nomeadamente q–exponencial, q–logaritmo e q–esperança. Na secção 2, definimos a q–entropia e na secção seguinte são enunciadas e demonstradas algumas das principais propriedades desta entropia. Na secção 4, apresentamos a generalização do conceito de ganho de informação de Kullback–Leibler. Por fim, na secção 5, efectuamos uma análise crítica ao artigo [19] de Santos, que propõe uma definição axiomática da q–entropia.

[1] "Mechanics" e "Statistical Mechanics"– A Dictionary of Physics. Ed. Alan Isaacs. Oxford University Press, 2000. Oxford Reference Online.

5.1 Definições

A definição das funções q-exponencial e q–logaritmo encontra–se num artigo de Tsallis publicado em 1994 (Quimica Nova, vol.17, pp.468). Não sendo possível consultar esta fonte, o que aqui se apresenta baseia–se num apêndice de um trabalho posterior do mesmo autor, [23].

Função q–Exponencial

A generalização da função exponencial, denonimada função q–exponencial, é dada por

$$e_q^x \equiv [1+(1-q)x]^{\frac{1}{1-q}} \qquad (q \in \mathbb{R}) \tag{5.1}$$

e verifica–se que tem as seguintes características:

1. Para $q > 1$, a função é contínua e monótona crescente em $]-\infty, \frac{1}{q-1}[$. O gráfico da função tem uma assímptota horizontal $x = 0$ quando $x \to -\infty$ e uma assímptota vertical em $x = \frac{1}{q-1}$. O contradomínio da função é $\mathbb{R}^+$. A função e_q^x é convexa.
2. Para $q < 1$, a função é contínua e monótona crescente em $]\frac{1}{q-1}, +\infty[$ variando de 0 a $+\infty$. Para valores de x inferiores a $\frac{1}{q-1}$, a função é definida como identicamente nula. A função e_q^x é convexa para $0 < q < 1$ e côncava para $q < 0$.
3. No caso limite $q = 1$ verifica–se que e_q^x coincide com a função exponencial usual. De facto, tendo em conta que $\lim\limits_{n\to\infty}\left(1+\dfrac{x}{n}\right)^n = e^x$, tem-se

$$\lim_{q\to 1} e_q^x = \lim_{q\to 1}(1+(1-q)x)^{\frac{1}{1-q}} = \lim_{q\to 1}\left(1+\frac{x}{\frac{1}{1-q}}\right)^{\frac{1}{1-q}} = e^x$$

4. Quando $x \to 0$ tem-se $e_q^x \sim 1 + x$. Esta aproximação resulta da expansão em série de Taylor em torno de $x = 0$,

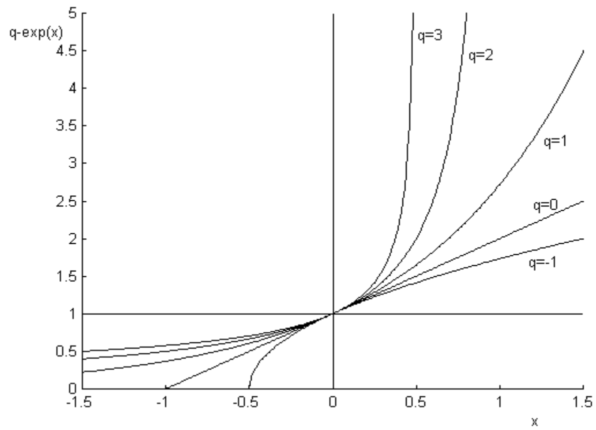
$$\begin{aligned} e_q^x = \; & 1 + x + \tfrac{q}{2}x^2 + \tfrac{q(2q-1)}{6}x^3 + \tfrac{q(2q-1)(3q-2)}{24}x^4 + \dots \\ & + \tfrac{1}{n!}\left(\prod_{i=1}^{n-1}(iq-i+1)\right)x^n + \dots \qquad (\forall q;\ x \to 0) \end{aligned}$$

   Graficamente, esta aproximação significa que o gráfico da função q–exponencial se aproxima da recta $y = 1 + x$ quando $x \to 0$.

Verifica–se ainda que

$$e_q^0 = 1 \qquad \text{e} \qquad e_q^x\, e_q^y = e_q^{x+y+(1-q)xy}\,.$$

Apresenta–se de seguida o gráfico da função q–logaritmo, para alguns valores de q



Função q–exponencial para valores típicos de q.

Em 1995 foi publicada uma abordagem diferente à definição de q–exponencial, baseada em séries[2] [5].

Função q–Logaritmo

A função inversa da q–exponencial é a chamada função q–logaritmo, que se define por

$$\ln_q x \equiv \frac{x^{1-q}-1}{1-q} \qquad (x \in \mathbb{R}^+;\, q \in \mathbb{R}) \tag{5.2}$$

e verifica o seguinte:

1. No caso $q = 1$, a função q–logaritmo reduz-se ao usual logaritmo neperiano. De facto, tendo em conta que $\lim\limits_{n\to 0} \dfrac{x^n - 1}{n} = \ln x$, resulta que $\lim\limits_{q\to 1} \ln_q x = \ln x$.
2. Para $q > 1$, a função é monótona crescente em $\mathbb{R}^+$. O gráfico da função tem uma assímptota horizontal de equação $y = \frac{1}{q-1}$ quando $x \to +\infty$, e uma assímptota vertical em $x = 0$. O contradomínio da função é $]-\infty, \frac{1}{q-1}[$. A função $\ln_q$ é côncava.
3. Para $q < 1$, a função é monótona crescente em $\mathbb{R}^+$. A função converge para $\frac{1}{q-1}$ quando $x \to 0^+$ e diverge quando $x \to +\infty$. A função $\ln_q x$ é côncava quando $0 < q < 1$ e convexa quando $q < 0$.
4. Quando $x \to 1$ tem-se $\ln_q x \sim x - 1$. De facto, da expansão em série de Taylor em torno de $y = 0$ obtém–se

$$\begin{aligned} \ln_q(1+y) = \quad & y - \tfrac{q}{2}y^2 + \tfrac{q(q+1)}{6}y^3 - \tfrac{q(q+1)(q+2)}{24}y^4 + \ldots \\ & + \tfrac{(-1)^{n+1}}{n!} \left(\prod_{i=0}^{n-2} (q+i) \right) y^n + \ldots \qquad (\forall q,\, y \to 0)\,. \end{aligned}$$

[2]McAnally (1995), J. Math. Phys, vol.36, pp.546–573; Kassel (1995), "Quantum Groups", New York, Springer.

Fazendo $x = 1 + y$ obtém–se que $\ln_q x \sim x - 1$ quando $x \to 1$.

Graficamente, esta aproximação significa que o gráfico da q–exponencial se aproxima da recta $y = x - 1$ quando $x \to 1$.

Verifica–se ainda que

$$\ln_q(1) = 0$$

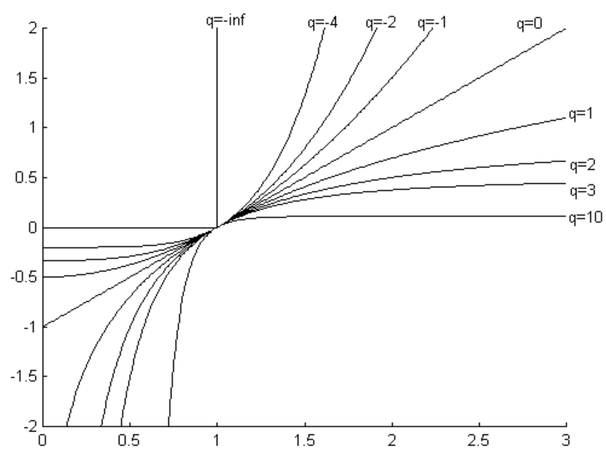
e

$$\ln_q(x\,y) \;=\; \ln_q(x) + \ln_q(y) + (1-q)\ln_q(x)\ln_q(y)\,.$$

Uma vez que as funções ln_q e e_q são inversas uma da outra, então

$$e_q \circ \ln_q = \ln_q \circ e_q = id\,.$$

Na figura que se segue ilustra–se o gráfico da função q–logaritmo, para alguns valores do parâmetro q:



Função q–logaritmo para alguns valores típicos de q.

Função q–Esperança

O conceito de esperança também é generalizado. Considerando que $\{p_i\}_i$ são as probabilidades dos resultados associados à variável aleatória X, a q-esperança[3] é definida por

$$\langle X \rangle_q \equiv \sum_i (p_i)^q\, x_i\ . \tag{5.3}$$

[3]Ou, esperança de segunda espécie

Obviamente que $\langle X \rangle_1 = \langle X \rangle$.

5.2 Função q–Entropia

A forma entrópica não extensiva definida por Tsallis é

$$S_q(p_1, \dots, p_n) = k \, \frac{1 - \sum_{i=1}^{n} (p_i)^q}{q - 1} \tag{5.4}$$

onde $q \in \mathbb{R}$, k é uma constante positiva que define a unidade em que a entropia é medida e n é o número de configurações microscópicas possíveis, cujas probabilidades são $\{p_i\}_i$, sendo $\sum_{i=1}^{n} p_i = 1$. A função S_q chama–se q–entropia ou entropia de Tsallis. No caso de $q > 0$, o domínio da função (5.4) é

$$\Delta_n = \left\{ (p_1, \dots, p_n) \in \mathbb{R}^n : 0 \leq p_i \leq 1 \;\wedge\; \sum_{i=1}^{n} p_i = 1 \right\}$$

E no caso de $q < 0$, deve impôr–se ainda que as probabilidades sejam estritamente positivas.

Na figura seguinte estão representados gráficos da q–entropia de uma distribuição com dois resultados possíveis $S_q(p_1, p_2)$, para alguns valores de q.

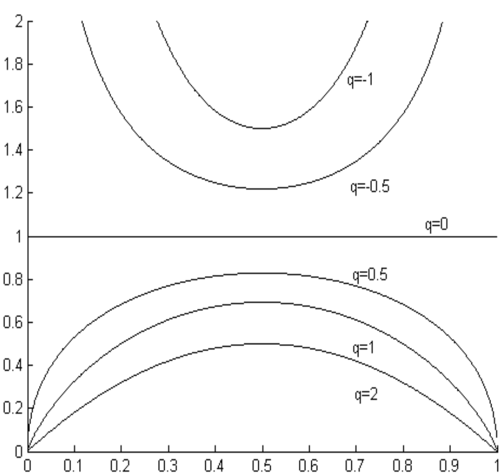


Gráfico de $S_q(p_1, p_2)$, para os valores de q indicados.

Ao tomar o caso limite $q = 1$, a q–entropia reduz-se à entropia de Boltzmann–Gibbs–Shannon, recuperando todo o formalismo usual da Mecânica Estatística. De facto, tendo em conta o limite notável,

$$\lim_{n\to 0} \frac{x^n - 1}{n} = \ln x$$

verifica–se imediatamente que,

$$\begin{aligned}
S_1(p_1, \ldots, p_n) &= \lim_{q\to 1} S_q(p_1, \ldots, p_n) && = \lim_{q\to 1} k\, \frac{1 - \sum_i p_i^q}{q-1} \\
&= \lim_{q\to 1} k\, \frac{\sum_i p_i - \sum_i p_i\, p_i^{q-1}}{q-1} && = \lim_{q\to 1} k\, \frac{\sum_i p_i (1 - p_i^{q-1})}{q-1} \\
&= k\, \textstyle\sum_i p_i\, \displaystyle\lim_{q\to 1} \frac{1 - p_i^{q-1}}{q-1} && = k\, \textstyle\sum_i p_i(-\ln p_i) \\
&= -\, k\, \textstyle\sum_i p_i \ln p_i\, .
\end{aligned}$$

No caso de uma variável contínua, com função densidade de probabilidade p, a q–entropia S_q é definida por,

$$S_q(p) = k\, \frac{1 - \displaystyle\int_{-\infty}^{+\infty} p(x)^q dx}{q-1}\, .$$

Tal como a entropia de Shannon representa a esperança matemática dos $\{-k \ln p_i\}_i$, a entropia de Tsallis pode ser interpretada como a q-esperança dos $\{-k \ln_q p_i\}_i$. De facto,

$$\begin{aligned}
S_q(p_1, \ldots, p_n) &= k\, \frac{1 - \sum_{i=1}^{\omega} p_i^q}{q-1} = k \sum_{i=1}^{\omega} p_i^{\,q}\, \frac{1 - p_i^{1-q}}{1-q} \\
&= k \sum_{i=1}^{\omega} p_i^{\,q}\, (-\ln_q p_i) = k\, \langle -\ln_q p_i \rangle_q = \langle -k\, \ln_q p_i \rangle_q\, .
\end{aligned}$$

Na entropia de Shannon os eventos com probabilidade muito alta ou muito baixa não têm grande peso no valor da entropia. Porém, na entropia de Tsallis, no caso de $q > 1$, os eventos com probabilidades mais elevadas contribuem mais para o valor da entropia do que os eventos de probabilidade baixa (e o recíproco para $q < 1$). Estabelecendo um paralelo com a entropia de Shannon, a média usual é aqui substituida por uma média de potência q. Assim, uma mudança do valor de q altera a contribuição relativa a um dado evento para a soma total. Quanto mais elevado for q, mais peso têm os eventos de probabilidade elevada, na soma total.

Citanto o próprio Tsallis[4], a origem da sua proposta para a entropia

> *...está no facto de que, em muitos fenómenos da natureza, percebe–se um viés, no sentido que alguns eventos pouco prováveis (ou então muito prováveis) parecem controlar o fenómeno. A forma mais simples que encontrei de traduzir isto matematicamente foi introduzir na entropia termos do tipo p^q.*

[4]Excerto de uma mensagem electrónica escrita por Constantino Tsallis, em Outubro de 2002.

5.3 Propriedades da q–Entropia

De seguida, apresentam–se algumas das propriedades mais importantes da entropia de Tsallis. No que se segue, considera–se que $k = 1$. No entanto, todas as propriedades permanecem válidas para qualquer valor positivo de k.

A entropia de Tsallis satisfaz algumas das propriedades da entropia de Shannon e da de Rényi. As principais diferenças entre estas encontram–se, fundamentalmente, na aditividade e na convexidade. Enquanto que a entropia de Shannon e de Rényi verificam a aditividade usual[5], a entropia de Tsallis não verifica a aditividade, mas sim a q–aditividade. A entropia de Shannon é côncava, enquanto que as de Rényi e de Tsallis nem sempre o são (dependendo tal propriedade do valor dos parâmetros α e q, respectivamente). Para além disso, a entropia de Shannon e de Tsallis não são aplicáveis a distribuições incompletas.

5.3.1 Não Negatividade

A propriedade não negativa da função usual de entropia[6] é preservada nesta generalização. De facto, ao reescrever a equação (5.4) na seguinte forma

$$S_q = \frac{1}{q-1} \sum_{i=1}^{n} p_i \left(1 - p_i^{q-1}\right)$$

e tendo em conta que $p_i \in [0, 1]$, facilmente se conclui que S_q é uma função não negativa para todo o valor de q, como se segue:

$$q \geq 1 \quad \Rightarrow \begin{cases} q - 1 \geq 0 \\ 0 \leq p_i^{q-1} \leq 1 \end{cases} \quad \Rightarrow \begin{cases} q - 1 \geq 0 \\ 1 - p_i^{q-1} \geq 0 \end{cases} \quad \Rightarrow \tfrac{1}{q-1} \textstyle\sum_i p_i (1 - p_i^{q-1}) \geq 0$$

$$q \leq 1 \quad \Rightarrow \begin{cases} q - 1 \leq 0 \\ p_i^{q-1} \geq 1 \end{cases} \quad \Rightarrow \begin{cases} q - 1 \leq 0 \\ 1 - p_i^{q-1} \leq 0 \end{cases} \quad \Rightarrow \tfrac{1}{q-1} \textstyle\sum_i p_i (1 - p_i^{q-1}) \geq 0$$

Estas implicações resultam do facto da função exponencial de base $a \in]0, 1]$ verificar a desigualdade $0 \leq a^x \leq 1$ quando $x \geq 0$ e verificar $a^x \geq 1$ quando $x \leq 0$.

A função S_q toma valor nulo apenas quando o resultdao é conhecido, ou seja:

[5] Aditividade de uma função f, no sentido de se verificar a igualdade $f(P * Q) = f(P) + f(Q)$, onde $P * Q$ representa o produto directo de P e Q.

[6] A entropia de Boltzmann–Gibbs ou de Shannon.

(i) quando existe apenas um resultado possível ($n = 1$);

(ii) se existirem vários resultados possíveis ($n > 1$), todos os eventos têm de ter probabilidade nula, excepto um (que tem probabilididade unitária).

5.3.2 Concavidade e Extremos

A entropia de Boltzmann–Gibbs é definida por uma função côncava, pelo que apresenta um único máximo. É esta a propriedade na qual se baseia a Segunda Lei da Termodinâmica e que garante a estabilidade dos sistemas (os sistemas tendem para o estado de máxima entropia, num sistema em estado de equilíbrio a entropia é máxima).

A entropia generalizada é uma função côncava para todo o $q > 0$ e convexa para todo o $q < 0$. Desta forma, a Segunda Lei da Termodinâmina deve ser reescrita como, *a entropia de um sistema em equilíbrio é um extremo* [6]. Esse extremo poderá ser um máximo (se $q > 0$, estando aqui incluído o caso usual de $q = 1$) ou pode ser um mínimo (se $q < 0$), como se mostra de seguida.

Considerem–se duas funções probabilidade $(p_1, \ldots, p_n)$ e $(q_1, \ldots, q_n)$, e defina–se

$$C_q \equiv S_q(\{\lambda p_i + (1-\lambda)q_i\}) - \lambda S_q(\{p_i\}) - (1-\lambda)S_q(\{q_i\})\,.$$

Para que a função S_q seja côncava, ou convexa, é necessário que C_q não mude de sinal para todo o $\lambda \in [0,1]$. Tem–se

$$C_q \equiv \frac{1}{1-q}\sum_{i=1}^{n}\left[\,(\lambda\, p_i + (1-\lambda)\, q_i)^q - \lambda\, {p_i}^q - (1-\lambda)\, {q_i}^q\right]\,.$$

Para estudar o sinal de C_q vai–se recorrer à função auxiliar $\varphi(x) = x^q$, definida em $[0,1]$, que é convexa quando $q < 0$ ou $q > 1$ e côncava quando $0 < q < 1$.

(i) Se $q < 0$:
Por um lado tem–se que $\frac{1}{1-q} > 0$ e por outro, como φ é convexa, tem–se

$$(\lambda\, p_i + (1-\lambda)\, q_i)^q - \lambda\, {p_i}^q - (1-\lambda)\, {q_i}^q \leq 0 \qquad \forall\, i \in \{1, \ldots, n\}$$

Portanto, $C_q \leq 0$, pelo que S_q é uma função convexa.

(ii) Se $0 < q < 1$:
Dado que φ é côncava, então

$$(\lambda\, p_i + (1-\lambda)\, q_i)^q - \lambda\, {p_i}^q - (1-\lambda)\, {q_i}^q \geq 0 \qquad \forall\, i \in \{1, \ldots, n\}$$

Adicionando o facto de que $\frac{1}{1-q} > 0$, resulta que $C_q \geq 0$. Portanto, S_q é uma função côncava.

(iii) Se $q = 1$:
Neste caso S_q coincide com a entropia de Boltzmann–Gibbs, pelo que é uma função côncava[7].

[7] Cf. propriedade (S6.) enunciada na página 28.

(iv) Se $q > 1$:

Tem–se $\frac{1}{1-q} < 0$ e φ é convexa, pelo que $C_q \geq 0$. Logo, S_q é uma função côncava.

Determinam–se de seguida os extremos da função q–entropia:

No caso $q > 0$ a função S_q é contínua e o seu domínio é um subconjunto compacto de $\mathbb{R}^n$, pelo que é garantida a existência de extremos. Para os determinar pode–se recorrer ao método dos multiplicadores de Lagrange, pelo que os candidatos a extremantes da função S_q, sujeita à restrição de $\sum_{i=1}^{n} p_i = 1$ são tais que

$$\begin{cases} \nabla S_q(p_1, \ldots, p_n) = \lambda \nabla g(p_1, \ldots, p_n) \\ g(p_1, \ldots, p_n) = 1 \end{cases}$$

sendo $g(p_1, \ldots, p_n) = \sum_{i=1}^{n} p_i$. Ou seja,

$$\begin{cases} \dfrac{-q\, {p_i}^{q-1}}{q-1} = \lambda \,,\; i = 1, \ldots, n \\ \displaystyle\sum_{i=1}^{n} p_i = 1 \end{cases}$$

Resolvendo o sistema, obtém–se a solução $p_1 = \ldots = p_n = \frac{1}{n}$. Uma vez que a função S_q é côncava quando $q > 0$, então o ponto $(\frac{1}{n}, \ldots, \frac{1}{n})$ é um maximizante da função.Pelo facto de S_q ser uma função não negativa, acresce ainda que

$$0 \leq S_q(p_1, \ldots, p_n) \leq \frac{n^{1-q} - 1}{q - 1} \,.$$

No caso de $q < 0$, o domínio da função não é compacto, pelo que não garante a existência de extremos. No entanto, a função é contínua e convexa, pelo que o ponto crítico é um minimizante. Assim,

$$S_q(p_1, \ldots, p_n) \geq \frac{n^{1-q} - 1}{q - 1} > 0 \,.$$

Em relação à interpretação da q–entropia como uma medida da desordem atómica, Borges [6] considera que, quando q é uma constante positiva, pode–se *manter a interpretação de entropia como medida de desordem do sistema* (quanto maior o valor da entropia, maior é a desordem); quando q é uma constante negativa, *a associação entre estes dois conceitos continua válida, mas agora o estado de completa ordem é* $S_q = +\infty$.

5.3.3 Equiprobabilidade

Como se acabou de demonstrar, é no caso da equiprobabilidade que a entropia atinge o seu valor extremo. A distribuição uniforme discreta (caso de equiprobabilidade) reveste–se da maior importância precisamente pelo facto de assumir o papel de distribuição de máxima ou de mínima entropia. Se essa distribuição tiver ω pontos, a entropia é dada por

$$S_q(\omega) = k\,\frac{\omega^{1-q}-1}{\omega-1} = k\,\ln_q(\omega)$$

(sendo k uma constante real positiva). No caso limite $q = 1$, obtém–se a conhecida fórmula de Boltzmann–Gibbs,

$$S_1(\omega) = k\,\ln\omega\,.$$

A função $S_q(\omega)$ é estritamente monótona crescente na variável ω. No caso de $q > 1$, esta aproxima–se assimptoticamente de um valor limite (assímptota $y = \frac{k}{q-1}$) e se $q < 1$ a função é divergente, como se verifica de seguida.

(i.) $S_q(\omega)$ é uma função monótona crescente na variável ω:

$$\begin{aligned}\frac{d}{d\omega}\left(k\,\frac{1-\omega^{1-q}}{q-1}\right) &= \frac{d}{d\omega}\left(\frac{k}{q-1}+\frac{k}{1-q}\omega^{1-q}\right) &= \frac{k}{1-q}(1-q)\omega^{1-q-1}\\ &= k\,\omega^{-q} > 0 &, \quad \text{pois } k>0 \text{ e } \omega \geq 0\end{aligned}$$

(ii.) Se $q > 1$, a função $S_q(\omega)$ tem assímptota horizontal $y = \frac{k}{q-1}$:

$$\lim_{\omega\to+\infty} k\,\frac{1-\omega^{1-q}}{q-1} = \frac{k}{q-1}\left(1-\frac{k}{q-1}\lim_{\omega\to+\infty}\omega^{1-q}\right) = \frac{k}{q-1} \quad , \text{ pois } 1-q<0\,.$$

(iii.) Se $q < 1$, a função $S_q(\omega)$ diverge:

$$\lim_{\omega\to+\infty} k\,\frac{1-\omega^{1-q}}{q-1} = +\infty \quad , \text{ pois } 1-q>0\,.$$

5.3.4 q–Aditividade

A generalização de uma teoria puderá pressupor a violação de um dos seus postulados. No caso da entropia generalizada proposta por Tsallis, é quebrada a propriedade aditiva da entropia usual[8].

Considerem–se duas variáveis aleatórias X e Y, com funções probabilidade marginais $P = (p_1, \ldots, p_n)$ e $Q = (q_1, \ldots, q_m)$, respectivamente. Se X e Y forem independentes, no contexto da Teoria das Probabilidades, a entropia generalizada da distribuição composta[9] verifica a chamada regra da q–aditividade, ou seja,

$$S_q(\{p_i q_j\}_{i,j}) \;=\; S_q(\{p_i\}_i) \;+\; S_q(\{q_j\}_j) \;+\; (1-q)\,S_q(\{p_i\}_i)\,S_q(\{q_j\}_j)$$

Como se passa a verificar:

[8] No capítulo 2, viu–se que Shannon [20] define a entropia de forma axiomática, sendo a aditividade não um dos axiomas, mas uma sua consequência. Em Khinchin [15], a aditividade é considerada como axioma.

[9] Interprete–se a distribuição composta como a resultante do produto directo entre P e Q.

$$S_q(\{p_i\}_i) + S_q(\{q_j\}_j) + (1-q)\, S_q(\{p_i\}_i)\, S_q(\{q_j\}_j)$$

$$= \frac{1-\sum_{i=1}^{n}(p_i)^q}{q-1} + \frac{1-\sum_{j=1}^{m}(q_j)^q}{q-1} + (1-q)\frac{1-\sum_{i=1}^{n}(p_i)^q}{q-1}\,\frac{1-\sum_{j=1}^{m}(q_j)^q}{q-1}$$

$$= \frac{1-\sum_{i=1}^{n}(p_i)^q + 1 - \sum_{j=1}^{m}(q_j)^q - 1 + \sum_{j=1}^{m}(q_j)^q + \sum_{i=1}^{n}(p_i)^q - \sum_{i=1}^{n}(p_i)^q \sum_{j=1}^{m}(q_j)^q}{q-1}$$

$$= \frac{1-\sum_{i,j}(p_i\, q_j)^q}{q-1} \;=\; S_q(\{p_i q_j\}_{i,j})$$

É a esta propriedade que se deve o nome de *mecânica estatística não extensiva.* O valor $(1-q)$ caracteriza o grau de não–extensividade[10] de um sistema. O caso $1-q>0$ $(q<1)$ corresponde à *superextensividade* e o caso $1-q<0$ $(q>1)$ corresponde à *subextensividade* de S_q.

5.3.5 q–axioma de grupo

A q–entropia permite a generalização de uma propriedade da entropia usual conhecida por axioma de grupo[11].

Considere–se um conjunto com ω configurações microscópicas de probabilidades

$$p_1, p_2, \ldots, p_\omega\,.$$

Divida–se este conjunto em dois subconjuntos L e M cada um contendo, respectivamente, ω_L e ω_M configurações. Então,

$$\omega_L + \omega_M = \omega\,.$$

Sejam $\{p_1, \ldots, p_{\omega_L}\}$ e $\{p_{\omega_L+1}, \ldots, p_\omega\}$ as probabilidades associadas a L e a M. Definam–se p_L e p_M como as probabilidades de uma configuração pertencer a cada um dos conjuntos L e M, isto é,

$$p_L \equiv \sum_{i=1}^{\omega_L} p_i \quad \text{e} \quad p_M \equiv \sum_{i=\omega_L+1}^{\omega} p_i$$

então,

$$S_q(p_1,\ldots,p_\omega) = S_q(p_L,p_M) \;+\; (p_L)^q\, S_q\left(\frac{p_1}{p_L},\ldots,\frac{p_{\omega_L}}{p_L}\right) \;+\; (p_M)^q\, S_q\left(\frac{p_{\omega_L+1}}{p_M},\ldots,\frac{p_\omega}{p_M}\right) \quad (5.5)$$

[10] Os conceitos de extensividade e aditividade são diferentes, contudo, em grande parte dos casos estudados na física, a aditividade implica a extensividade [32].

[11] A relação definida por (2.4), na página 23.

onde $\{p_i/p_L\}$ e $\{p_i/p_M\}$ são as probabilidades condicionadas.

A igualdade anterior só se diferencia do axioma de grupo da entropia de Shannon (2.4) pelo facto de apresentar as probabilidades p_L e p_M elevadas a q, isto é, surge a q–esperança (5.3) das entropias condicionadas, em vez da esperança usual.

Facilmente se comprova a relação (5.5):

$$
\begin{aligned}
& S_q(p_L, p_M) \,+\, (p_L)^q \, S_q\left(\tfrac{p_1}{p_L}, \ldots, \tfrac{p_{\omega_L}}{p_L}\right) \,+\, (p_M)^q \, S_q\left(\tfrac{p_{\omega_L+1}}{p_M}, \ldots, \tfrac{p_\omega}{p_M}\right) \\
&= \frac{1-(p_L)^q-(p_M)^q}{q-1} + (p_L)^q \, \frac{1-\displaystyle\sum_{i=1}^{\omega_L}\left(\frac{p_i}{p_L}\right)^q}{q-1} + (p_M)^q \, \frac{1-\displaystyle\sum_{j=\omega_L+1}^{\omega}\left(\frac{p_j}{p_M}\right)^q}{q-1} \\
&= \frac{1-(p_L)^q \displaystyle\sum_{i=1}^{\omega_L}\left(\frac{p_i}{p_L}\right)^q - (p_M)^q \displaystyle\sum_{j=\omega_L+1}^{\omega}\left(\frac{p_j}{p_M}\right)^q}{q-1} \\
&= \frac{1-\displaystyle\sum_{i=1}^{\omega_L}(p_i)^q - \displaystyle\sum_{i=\omega_L+1}^{\omega}(p_i)^q}{q-1} = \frac{1-\displaystyle\sum_{i=1}^{\omega}(p_i)^q}{q-1} = S_q(p_1, \ldots, p_\omega)
\end{aligned}
$$

A equação (5.5) pode ser generalizada ao caso em que, em vez de dois, existem r subconjuntos disjuntos [1], como se passa a enunciar.

Considere r subconjuntos, cada um com ω_i elementos, sendo $\omega_1+\ldots+\omega_r=\omega$. Denotando as probabilidades associadas à ocorrência dos elementos do conjunto j por $(p_{j,1}\,,\ldots,\,p_{j,\omega_j})$ e definindo

$$
\pi_j \equiv \sum_{i=1}^{\omega_j} p_{j,i} \qquad (j=1,\ldots,r)
$$

a generalização é dada por

$$
S_q(p_1,\ldots,p_\omega) = S_q(\pi_1,\ldots,\pi_r) + \sum_{j=1}^{r} (\pi_j)^q \, S_q\left(\frac{p_{j,1}}{\pi_j}, \ldots, \frac{p_{j,\omega_j}}{\pi_j}\right) \tag{5.6}
$$

onde $\{p_i/\pi_j\}$ é o conjunto das probabilidades condicionadas associadas a ω_j. Verifica–se a igualdade (5.6):

$$S_q(\pi_1,\ldots,\pi_r) + \sum_{j=1}^{r}\left[(\pi_j)^q\, S_q\left(\frac{p_{j,1}}{\pi_j},\ldots,\frac{p_{j,\omega_j}}{\pi_j}\right)\right]$$

$$= \tfrac{1}{q-1}\left[1-\sum_{i=1}^{r}(\pi_i)^q + \sum_{j=1}^{r}(\pi_j)^q\ \left(1-\sum_{i=1}^{\omega_j}\left(\frac{p_{j,i}}{\pi_j}\right)^q\ \right)\right]$$

$$= \tfrac{1}{q-1}\left[1-\sum_{i=1}^{r}(\pi_i)^q + \sum_{j=1}^{r}(\pi_j)^q - \sum_{j=1}^{r}(\pi_j)^q\sum_{i=1}^{\omega_j}\left(\frac{p_{j,i}}{\pi_j}\right)^q\ \right]$$

$$= \tfrac{1}{q-1}\left[1-\sum_{j=1}^{r}\sum_{i=1}^{\omega_j}(p_{j,i})^q\right] = \tfrac{1}{q-1}\left[1-\sum_{k=1}^{\omega}{p_k}^q\right] = S_q(p_1,\ldots,p_\omega)$$

5.4 q–informação de Kullback–Leibler

À semelhança das generalizações efectuadas para a função exponencial, função logaritmica, esperança matemática e entropia, o ganho de informação de Kullback–Leibler também sofre uma generalização que o associa a um parâmetro real q. Em 1998, Tsallis [22] definiu essa generalização, que aqui é designada por q–informação de Kullback–Leibler. De seguida apresenta–se uma dedução dessa forma generalizada, baseada no exposto no artigo [7], da autoria de Borland, Plastino e Tsallis.

No terceiro capítulo definiu–se o ganho de informação de Kullback–Leibler como uma variação média da informação associada a duas distribuições de probabilidade. Empregando o formalismo não extensivo em vez do formalismo usual, é estabelecido um raciocínio paralelo ao apresentado naquele capítulo por forma a generalizar a medida de Kullback–Leibler.

Seja X uma variável aleatória com distribuição $P = (p_1,\ldots,p_n)$. Uma vez qua a q–entropia (5.4) é a q–esperança matemática (5.3) dos

$$(\sigma_i)_q \equiv -\frac{p_i^{1-q}-1}{1-q}$$

então $(\sigma_i)_q$ pode ser interpretado como a informação generalizada contida no resultado i.

Considerando uma nova distribuição de probabilidades $P' = (p'_1,\ldots,p'_n)$, a variação da informação (generalizada) é dada por

$$\Delta(\sigma_i)_q = \frac{1}{1-q}((1-{p_i}^{1-q}) - (1-{p'_i}^{1-q}))\ .$$

A variação média de informação é obtida efectuando a q–média dos $\Delta(\sigma_i)_q$ sobre a nova distribuição p'_i, ou seja,

$$K_q(P',P) = \sum_{i=1}^{n}{p'_i}^q\,\frac{1}{1-q}({p'_i}^{1-q} - {p_i}^{1-q})$$

que traduz o ganho de informação generalizado de Kullback–Leibler. Fazendo uso da definição de q–logaritmo (5.2), obtém-se a forma

$$K_q(P', P) = \sum_{i=1}^{n} {p'}_i \, \ln_q \frac{{p'}_i}{p_i} \tag{5.7}$$

muito semelhante à já conhecida para o ganho de informação de Kullback–Leibler (3.4).

A função generalizada do ganho de informação de Kullback–Leibler (5.7) é não negativa, anulando–se apenas quando as distribuições coincidem (ou quando $q = 0$). Por esse motivo, e apesar de não ser simétrica nem respeitar a desigualdade triangular, esta quantidade pode ser interpretada como uma espécie de "distância". Pode–se ainda provar que a função é convexa para $q > 0$ e côncava para $q < 0$, [7].

No caso de X ser uma variável aleatória contínua, a q–informação de Kullback–Leibler é definida [22] por,

$$K_q(p', p) = \int_{-\infty}^{+\infty} p'(x) \, \ln \frac{p'(x)}{p(x)} \, dx \, .$$

Neste caso, a q–informação de Kullback–Leibler apenas é não negativa quando $q > 0$, sendo identicamente nula quando $q = 0$ e não positiva quando $q < 0$. Tsallis propõe ainda a medida

$$K_q^{sim}(p, p') = \frac{1}{2} \left(\, K_q(p, p') + K_q(p', p) \, \right)$$

para o caso em que se pretende uma medida simétrica.

5.5 Definição Axiomática

A entropia de Tsallis generaliza a entropia de Boltzmann–Gibbs–Shannon e muitas das suas propriedades. Uma relação interessante entre estas duas entropias foi a encontrada por Roberto Santos [19] em 1997. A partir de quatro postulados, e seguindo o raciocínio de Shannon, provou que existe uma única função que satisfaz todos esses postulados, nomeadamente, a entropia de Tsallis.

No trabalho de Santos apenas é provada a unicidade da entropia de Tsallis para o caso $q > 1$, ficando o caso $q < 1$ por provar. De seguida apresenta–se a axiomática proposta por Santos, tendo–se considerado que, no caso de equiprobabilidade, a função de entropia é monótona crescente no sentido estrito (e não no sentido lato, como no original). Esta pequena alteração permitiu provar a unicidade da fórmula de entropia para o caso de $q > 1$, assim como para o caso de $q < 1$. Para a análise do artigo, foram pedidos alguns esclarecimentos ao próprio autor, que se mostrou sempre muito prestável.

5.5.1 Axiomática de Santos

Sejam $p_1, \ldots, p_n$ as probabilidades associadas aos resultados possíveis de uma experiência aleatória. Considerando que a função de entropia S_q associada a este espaço de resultados satisfaz as seguintes condições:

I. Para $0 < p_i < 1$, a função S_q é contínua nos $\{p_i\}$.

II. No caso da equiprobabilidade, isto é $p_i = 1/\omega$, S_q é função estritamente crescente[12] em ω.

III. (q–aditividade)
Para dois sistemas independentes A e B, no contexto da teoria das probabilidades, a entropia do sistema composto verifica

$$S_q(A \cup B) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B)$$

IV. (q–axioma de grupo)
Seja $p_1, \ldots, p_\omega$ a distribuição de probabilidades de um conjunto com ω resultados possíveis. Particione–se este conjunto nos subconjuntos L e M com ω_L e ω_M elementos, respectivamente ($\omega = \omega_L + \omega_M$). Sejam $\{p_1, \ldots, p_{\omega_L}\}$ as probabilidades associadas aos elementos de L e $\{p_{\omega_L+1}, \ldots, p_\omega\}$ as probabilidades associadas aos elementos de M. Defina–se

$$P_L = \sum_{i=1}^{\omega_L} p_i \;\; (\omega_L \text{ termos}) \qquad P_M = \sum_{i=\omega_L+1}^{\omega} p_i \;\; (\omega_M \text{ termos})$$

então

$$S_q(p_1, \ldots, p_\omega) = S_q(p_L, p_M) + (p_L)^q\, S_q\left(\frac{p_1}{p_L}, \ldots, \frac{p_{\omega_L}}{p_L}\right) + (p_M)^q\, S_q\left(\frac{p_{\omega_L+1}}{p_M}, \ldots, \frac{p_\omega}{p_M}\right)$$

Repare–se que na condição IV, os p_i/p_L e os p_i/p_M são as probabilidades condicionadas por L e M, respectivamente.

Do postulado IV resulta imediatamente que

$$S_q(1) = 0 \tag{5.8}$$

e consequentemente,

$$S_q(p_1, p_2, \ldots, p_\omega, 0) = S_q(p_1, p_2, \ldots, p_\omega). \tag{5.9}$$

[12] No trabalho de Santos [19], a condição II exige apenas que a função $S_q(\omega)$ seja monótona crescente.

De facto, para $\omega = 2$, se $\omega_L = 2$ e $\omega_M = 0$, o postulado IV garante[13] que

$$S_q(p_1, p_2) = S_q(p_1 + p_2) + (p_1 + p_2) \cdot S_q(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$$

e em particular, se $p_1 = 1$ e $p_2 = 0$, então

$$S_q(1, 0) = S_q(1) + 1^q \cdot S_q(1, 0)$$

pelo que se verifica (5.8).

Por outro lado, se para a distribuição $\{p_1, \ldots, p_\omega, 0\}$ se tomar ω_L como a soma das primeiras ω probabilidades, então a partir do postulado IV e da equação (5.8), obtém–se imediatamente a equação (5.9). Pois,

$$\begin{aligned}
S_q(p_1, p_2, \ldots, p_\omega, 0) &= S_q(\sum_{i=1}^{\omega} p_i) + (\sum_{i=1}^{\omega} p_i)^q \cdot S_q\left(\frac{p_1}{\sum p_i}, \ldots, \frac{p_\omega}{\sum p_i}\right) \\
S_q(p_1, p_2, \ldots, p_\omega, 0) &= S_q(1) + (1)^q \cdot S_q(p_1, \ldots, p_\omega) \\
S_q(p_1, p_2, \ldots, p_\omega, 0) &= S_q(p_1, \ldots, p_\omega)
\end{aligned}$$

Acompanhando as linhas de [19] mostra–se de seguida que a função que satisfaz simultaneamente todas estas propriedades é a fórmula da entropia de Tsallis, definida por

$$S_q(p_1, \ldots, p_n) = \frac{1 - \sum_{i=1}^{k} p_i^q}{q - 1}\,.$$

Denote–se por $A_q(r)$ a entropia associada a um sistema com r resultados igualmente possíveis, isto é,

$$A_q(r) \equiv S_q\left(\frac{1}{r}, \ldots, \frac{1}{r}\right) \quad (r \in \mathbb{N})\,. \tag{5.10}$$

Considere–se m sistemas estatisticamente independentes $W_1, \ldots, W_m$, cada um contendo $r(\geq 2)$ resultados igualmente prováveis. Então, conforme (5.10), para cada um dos sistemas tem–se

$$S_q(W_i) = S_q\left(\frac{1}{r}, \ldots, \frac{1}{r}\right) = A_q(r) \quad (i = 1, \ldots, m)\,.$$

Como os sistemas são independentes, pode–se aplicar a propriedade III, resultando que

$$S_q(W_1 \cup \ldots \cup W_m) = \frac{[1 + (1 - q)\, A_q(r)]^m - 1}{1 - q}\,. \tag{5.11}$$

De seguida apresenta–se a dedução da equação (5.11):

[13] Convencionando que se um dos p_L ou p_M forem nulos então o termo que, no postulado IV, está multiplicado por essa probabilidade é nulo.

Seja $q \neq 1$. A condição III garante imediatamente a igualdade (5.11) para o caso $m = 2$,

$$\begin{aligned} S_q(W_1 \cup W_2) &= S_q(W_1) + S_q(W_2) + (1-q) S_q(W_1) S_q(W_2) \\ &= 2A_q(r) + (1-q)\, A_q(r)^2 \\ &= \frac{2(1-q)A_q(r) + (1-q)^2\, A_q(r)^2}{1-q} \\ &= \frac{1 + 2(1-q)A_q(r) + (1-q)^2\, A_q(r)^2 - 1}{1-q} \\ &= \frac{[1 + (1-q)\, A_q(r)]^2 - 1}{1-q}\,. \end{aligned}$$

Considerando como hipótese de indução que a equação (5.11) se verifica para $m = k$, averigua–se que também se verifica para $m = k + 1$:

$$\begin{aligned} & S_q(W_1 \cup \ldots \cup W_{k+1}) = \\ & = S_q((W_1 \cup \ldots \cup W_k) \cup W_{k+1}) \\ & = S_q(W_1 \cup \ldots \cup W_k) + S_q(W_{k+1}) + \ (1-q)\, S_q(W_1 \cup \ldots \cup W_k)\, S_q(W_{k+1}) \\ & = \frac{[\,1 + (1-q)\, A_q(r)\,]^k - 1}{1-q} + A_q(r) + (1-q)\, \frac{[1 + (1-q)\, A_q(r)]^k - 1}{1-q}\, A_q(r) \\ & = \frac{[\,1 + (1-q)\, A_q(r)\,]\,[\,1 + (1-q)\, A_q(r)\,]^k - 1}{1-q} \\ & = \frac{[\,1 + (1-q)\, A_q(r)\,]^{k+1} - 1}{1-q}\,. \end{aligned}$$

Dado que r^m representa a existência de m conjuntos (escolhas), cada um formado por r elementos equiprováveis, então existem r^m possibilidades equiprováveis. Desta forma, e de acordo com (5.10), tem–se

$$S_q(W_1 \cup \ldots \cup W_m) = A_q(r^m)$$

pelo que a equação (5.11) pode ser reescrita por

$$A_q(r^m) = \frac{[1 + (1-q)\, A_q(r)]^m - 1}{1-q}\,. \tag{5.12}$$

Sejam $r, t \in \mathbb{N} \backslash \{1\}$ fixos. Para n suficentemente elevado é possível encontrar um número natural m tal que

$$r^m < t^n < r^{m+1} \tag{5.13}$$

e dado que a condição II garante o crescimento estrito de A_q, então

$$A_q(r^m) < A_q(t^n) < A_q(r^{m+1}) \tag{5.14}$$

aplicando (5.12), resulta que

$$\frac{[1+(1-q)\,A_q(r)]^m - 1}{1-q} < \frac{[1+(1-q)\,A_q(t)]^n - 1}{1-q} < \frac{[1+(1-q)\,A_q(r)]^{m+1} - 1}{1-q} \tag{5.15}$$

pelo que,

$$\frac{m}{n} < \frac{\ln(1+(1-q)\,A_q(t))}{\ln(1+(1-q)\,A_q(r))} < \frac{m}{n} + \frac{1}{n}\,. \tag{5.16}$$

De seguida apresenta–se com maior detalhe a passagem da equação (5.15) para a equação (5.16):

Se $\mathbf{q} < \mathbf{1}$, tem–se $1-q>0$, logo da equação (5.15) resulta

$$[1+(1-q)\,A_q(r)]^m < [1+(1-q)\,A_q(t)]^n < [1+(1-q)\,A_q(r)]^{m+1}\,.$$

A desigualdade anterior permite concluir[14] que $1+(1-q)\,A_q(r) > 1$, pelo que se pode tomar o logaritmo, obtendo–se

$$m\,\ln(1+(1-q)\,A_q(r)) < n\,\ln(1+(1-q)\,A_q(t)) < (m+1)\,\ln(1+(1-q)\,A_q(r))$$

e como o argumento do logaritmo é superior à unidade, tem–se

$$\frac{m}{n} < \frac{\ln(1+(1-q)\,A_q(t))}{\ln(1+(1-q)\,A_q(r))} < \frac{m+1}{n} = \frac{m}{n} + \frac{1}{n}\,.$$

Se $\mathbf{q} > \mathbf{1}$, tem–se $1-q<0$, logo da equação (5.15) resulta

$$[1+(1-q)\,A_q(r)]^m > [1+(1-q)\,A_q(t)]^n > [1+(1-q)\,A_q(r)]^{m+1}\,.$$

A desigualdade anterior permite concluir[15] que $0 < 1+(1-q)\,A_q(r) < 1$, pelo que se pode tomar o logaritmo, obtendo–se

$$m\,\ln(1+(1-q)\,A_q(r)) > n\,\ln(1+(1-q)\,A_q(t)) > (m+1)\,\ln(1+(1-q)\,A_q(r))$$

e como o argumento do logaritmo é inferior à unidade, tem–se

$$\frac{m}{n} < \frac{\ln(1+(1-q)\,A_q(t))}{\ln(1+(1-q)\,A_q(r))} < \frac{m+1}{n} = \frac{m}{n} + \frac{1}{n}$$

o que equivale à equação (5.16), pelo que esta é válida para todo o $q \neq 1$.

[14]De facto, $a^m < a^{m+1}\ \forall m \in \mathbb{N} \Rightarrow a > 1$

[15]De facto, $a^m > a^{m+1}\ \forall m \in \mathbb{N} \Rightarrow 0 < a < 1$

A partir da inequação (5.13), obtém–se

$$m \ln r < n \ln t < (m+1) \ln r$$

e dado que $r > 1$ tem–se,

$$\frac{m}{n} < \frac{\ln t}{\ln r} < \frac{m+1}{n} = \frac{m}{n} + \frac{1}{n}\,. \tag{5.17}$$

Combinando as equações (**??**) e (5.17), resulta que

$$\left| \frac{\ln t}{\ln r} - \frac{\ln(1+(1-q)\,A_q(t))}{\ln(1+(1-q)\,A_q(r))} \right| < \frac{1}{n}$$

tomando $n \to \infty$ tem–se $\frac{1}{n} \to 0$, pelo que no limite com n se verifique

$$\frac{\ln(1+(1-q)\,A_q(r))}{\ln r} = \frac{\ln(1+(1-q)\,A_q(t))}{\ln t}\,.$$

Assim, a quantidade

$$\frac{\ln(1+(1-q)\,A_q(t)\,)}{\ln t}$$

apenas depende de q, pelo que pode ser denotada por $p(q)$, ou seja,

$$p(q) = \frac{\ln(1+(1-q)\,A_q(t)\,)}{\ln t}\,.$$

Resolvendo a equação anterior em ordem a $A_q(t)$, obtém–se

$$A_q(t) = \frac{t^{p(q)} - 1}{1-q}\,.$$

Falta agora determinar p(q). Por simplicidade de notação ir–se–á utilizar p em vez de $p(q)$, assim,

$$A_q(t) = \frac{t^p - 1}{1-q}\,. \tag{5.18}$$

Considere–se um conjunto com ω possibilidades equiprováveis (isto é cada uma com probabilidade $1/\omega$), que é particionado em k conjuntos $\omega_1, \ldots, \omega_k$, cada um deles com n_i possibilidades ($\sum_{i=1}^{k} n_i = \omega$), conforme a figura (5.1).

Para cada $i = 1, \ldots, k$, a probabilidade[16] de ocorrer ω_i é

$$p_i \equiv P(\omega_i) = \frac{n_i}{\sum_{j=1}^{k} n_j}\,. \tag{5.19}$$

[16] Segundo a definição clássica de probabilidade, a probabilidade de ocorrer um acontecimento com número finito de resultados possíveis é a razão entre o "número de casos favoráveis ao acontecimento e o número total de casos possíveis, supondo todos os casos igualmente possíveis."

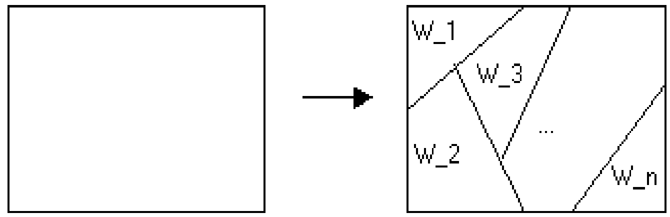


Figura 5.1: Divisão de um conjunto com ω elementos em k conjuntos, $\omega_1, \ldots, \omega_k$, cada um com n_i elementos ($\sum_{i=1}^k n_i = \omega$).

Da generalização da condição IV (ver subsecção *q–axioma de grupo*) resulta que

$$\underbrace{S_q\left(\frac{1}{\sum_{i=1}^k n_i}, \ldots, \frac{1}{\sum_{i=1}^k n_i}\right)}_{\text{W argumentos}} = S_q(p_1, \ldots p_k) + \sum_{i=1}^k (p_i)^q \;\; S_q \underbrace{\left(\frac{1/\sum_{j=1}^k n_j}{p_i}, \ldots, \frac{1/\sum_{j=1}^k n_j}{p_i}\right)}_{\text{W argumentos}}$$

e tendo em consideração a equação (5.19) obtém–se

$$S_q\left(\frac{1}{\sum_{i=1}^k n_i}, \ldots, \frac{1}{\sum_{i=1}^k n_i}\right) = S_q(p_1, \ldots p_k) + \sum_{i=1}^k (p_i)^q \;\; S_q\left(\frac{1}{n_i}, \ldots, \frac{1}{n_i}\right). \qquad (5.20)$$

Na função (5.18), a variável t representa a existência de t escolhas de igual probabilidade. Dado que

$$S_q\left(\frac{1}{\sum_{i=1}^k n_i}, \ldots, \frac{1}{\sum_{i=1}^k n_i}\right) = A_q\left(\sum_{i=1}^k n_i\right)$$

e

$$S_q\left(\frac{1}{n_i}, \ldots, \frac{1}{n_i}\right) = A_q(n_i)$$

então pode–se reescrever a equação (5.20) por

$$A_q\left(\sum_{i=1}^k n_i\right) = S_q(p_1, \ldots, p_k) + \sum_{i=1}^k (p_i)^q \, A_q(n_i)\,.$$

Aplicando a igualdade (5.18), obtém–se

$$\frac{\left(\sum_{i=1}^k n_i\right)^p - 1}{1-q} = S_q(p_1, \ldots, p_k) + \sum_{i=1}^k (p_i)^q \, \frac{(n_i)^p - 1}{1-q}$$

ou seja,

$$S_q(p_1,\ldots,p_k) = \frac{1}{1-q}\left(\left(\sum_{i=1}^{k} n_i\right)^p - 1 - \sum_{i=1}^{k}(p_i)^q\,(n_i)^p + \sum_{i=1}^{k}(p_i)^q\right). \tag{5.21}$$

Da equação (5.19) tem–se

$$(n_i)^p = (p_i)^p\,\left(\sum_{j=1}^{k} n_j\right)^p$$

substituindo na equação (5.21) obtém–se

$$S_q(p_1,\ldots,p_k) = \frac{1}{1-q}\left(\left(\sum_{i=1}^{k} n_i\right)^p - 1 - \sum_{i=1}^{k}(p_i)^q\,(p_i)^p\,\left(\sum_{j=1}^{k} n_j\right)^p + \sum_{i=1}^{k}(p_i)^q\right)$$

ou seja,

$$S_q(p_1,\ldots,p_k) = \frac{1}{1-q}\left(\left(\sum_{i=1}^{k} n_i\right)^p - 1 - \sum_{i=1}^{k}(p_i)^{q+p}\,\left(\sum_{i=1}^{k} n_i\right)^p + \sum_{i=1}^{k}(p_i)^q\right).$$

Para que a condição I seja satisfeita, é necessário eliminar a dependência com os n_i. Para tal tem de se tomar $p = 1-q$,

$$\begin{aligned}
S_q(p_1,\ldots,p_k) &= \frac{1}{1-q}\left(\left(\sum_{i=1}^{k} n_i\right)^{1-q} - \sum_{i=1}^{k}(p_i)^{q+1-q}\left(\sum_{j=1}^{k} n_j\right)^{1-q} - 1 + \sum_{i=1}^{k}(p_i)^q\right)\\
&= \frac{1}{1-q}\left(\left(\sum_{i=1}^{k} n_i\right)^{1-q} - \sum_{i=1}^{k} p_i\,\left(\sum_{j=1}^{k} n_j\right)^{1-q} - 1 + \sum_{i=1}^{k}(p_i)^q\right)\\
&= \frac{1}{1-q}\left(\left(\sum_{i=1}^{k} n_i\right)^{1-q} - \left(\sum_{j=1}^{k} n_j\right)^{1-q} - 1 + \sum_{i=1}^{k}(p_i)^q\right)\quad \text{, pois } \sum_{i=1}^{k} p_i = 1\\
&= \frac{1}{1-q}\left(-1 + \sum_{i=1}^{k}(p_i)^q\right)
\end{aligned}$$

Portanto,

$$S_q(p_1,\ldots,p_k) = \frac{1 - \sum_{i=1}^{k} p_i^q}{q-1} \tag{5.22}$$

onde os p_i são todos fracionários (consequência da relação (5.19)). Devido ao axioma da continuidade, a equação (5.22) permanece válida para qualquer distribuição de probabilidade $(p_1,\ldots,p_n)$. Ficando assim demonstrado que a função que satisfaz as condições I–IV é a entropia de Tsallis.

5.5.2 Comentários

Ao considerar os axiomas, I–IV sem exigir que a função S_q seja estritamente crescente no caso da equiprobabilidade, os axiomas são suficientes para garantir que a equação (5.16) se verifica? Será necessário acrescentar alguma condição sobre a função S_q?

Ao não exigir a monotonia estrita da função A_q (função q–entropia no caso da equiprobabilidade), surge o problema da validade da aplicação do logaritmo nas desigualdades (5.15), que passamos a discutir.

Para $\mathbf{q < 1}$:

A inequação (5.14) tem de ser substituida por

$$A_q(r^m) \leq A_q(t^n) \leq A_q(r^{m+1})$$

pelo que,

$$[1 + (1-q)\, A_q(r)]^m \leq [1 + (1-q)\, A_q(t)]^n \leq [1 + (1-q)\, A_q(r)]^{m+1}$$

e esta desigualdade permite concluir[17] que

$$1 + (1-q)\, A_q(r) = 0 \quad \text{ou} \quad 1 + (1-q)\, A_q(r) \geq 1\ . \tag{5.23}$$

1. Se $1+(1-q)\, A_q(r) = 0$, resulta que $A_q(r) = \frac{-1}{1-q} < 0$ (pois, $q < 1$), surgindo o problema da impossibilidade da aplicação do logaritmo sobre aquela quantidade. Contudo, esta situação é matematicamente impossível, devido à equação (5.8),

2. No segundo caso patente em (5.23) não existe qualquer problema em se aplicar o logaritmo. Contudo, se o argumento do logaritmo for igual à unidade, não é possível proceder à divisão por $\ln(1 + (1-q)\, A_q(r))$, pelo que esta situação tem também de ser analisada em particular. Como se está a considerar $q \neq 1$, então

   $$1 + (1-q)\, A_q(r) = 1 \Leftrightarrow A_q(r) = 0$$

   A quantidade $A_q(r)$ representa a q–entropia de um sistema com r resultados equiprováveis. Dado que $r > 1$, é sabido que este resultado é fisicamente impossível, mas será matematicamente impossível? Os axiomas I–IV permitem concluir que $A_q(r) = 0$ é um absurdo? Consideramos que não. Uma forma de contornar esta situação seria a de considerar que a q–entropia apenas se anula quando um dos p_i iguala a unidade, ou seja

   $$S_q(p_1, \ldots, p_n) = 0 \Rightarrow \exists |\, i \in \{1, \ldots, n\} : p_i = 1$$

[17] De facto, $a^m \leq a^{m+1}\ \forall m \in \mathbb{N} \Rightarrow a = 0 \vee a \geq 1$

Para $\mathbf{q} > 1$:

A inequação (5.14) tem de ser substituida por

$$A_q(r^m) \geq A_q(t^n) \geq A_q(r^{m+1})$$

pelo que,

$$[1 + (1-q)\, A_q(r)]^m \geq [1 + (1-q)\, A_q(t)]^n \geq [1 + (1-q)\, A_q(r)]^{m+1}$$

e esta desigualdade permite concluir[18] que

$$0 \leq 1 + (1-q)\, A_q(r) \leq 1 \; .$$

1. Se $1 + (1-q)\, A_q(r) = 1$, não existe qualquer problema na aplicação do logaritmo. Contudo, se o argumento do logaritmo for igual à unidade, não é possível proceder à divisão por $\ln(1 + (1-q)\, A_q(r))$. Uma forma de contornar esta situação é, tal como já indicado no caso $q < 1$, considerar que a q–entropia apenas se anula quando um dos p_i iguala a unidade.

2. Se $1 + (1-q)\, A_q(r) = 0$, não se pode aplicar o logaritmo nesta quantidade. E dado que $A_q(r) = \frac{-1}{1-q} > 0$ (pois, $q > 1$), então os axiomas não permitem eliminar este caso.
   O que devemos aqui impôr para eliminar este caso? Não sabemos. Não será à toa que o artigo de Santos [19] não contempla o caso $q > 1$.

* * *

Apresentadas estas considerações sugere–se que, para contornar o caso problemático

$$1 + (1-q)\, A_q(r) = 1$$

quando $q < 1$, se acrescentem novos axiomas aos indicados em [19], reformulando a axiomática apresentada. Nomeadamente,

i. Para $0 < p_i < 1$, a função S_q é contínua nos $\{p_i\}$

ii. No caso da equiprobabilidade, isto é $p_i = 1/\omega$, S_q é função monótona crescente em ω.

iii. Para dois sistemas independentes A e B, no contexto da teoria das probabilidades, a entropia do sistema composto verifica

$$S_q(A \cup B) = S_q(A) + S_q(B) + (1-q) S_q(A) S_q(B)$$

[18] De facto, $a^m \geq a^{m+1}\ \forall m \in \mathbb{N} \Rightarrow 0 \leq a \leq 1$

iv. Nas condições referidas em IV, S_q verifica

$$S_q(p_1,\ldots,p_W) = S_q(p_L,p_M) + (p_L)^q\, S_q\left(\frac{p_1}{p_L},\ldots,\frac{p_{W_L}}{p_L}\right) + (p_M)^q\, S_q\left(\frac{p_{W_L+1}}{p_M},\ldots,\frac{p_W}{p_M}\right)$$

v. $S_q(p_1,\ldots,p_n) = 0 \Rightarrow \exists|\, i \in \{1,\ldots,n\} : p_i = 1$

Repare–se que a condição (v.) não é mais que um conceito intuitivo: se a entropia é nula, então um dos resultados possíveis é certo, pelo que não há incerteza. Mas será adequado aceitar de ânimo leve esta ideia intuitiva? Não se fazer dela um axioma?

Mesmo após o aditamento da condição (v.), é necessário ter em atenção que os axiomas permitem deduzir a equação (5.4) apenas no caso de $q < 1$. Para o caso $q < 1$, não nos foi possível chegar à mesma conclusão (tal como a Santos em [19]).

Ao considerar a monotonia no sentido estrito, facilmente se conclui que a entropia de Tsallis se anula apenas quando existe o resultado certo. Assim, todos os valores de q são contemplados na demonstração da unicidade da fórmula (5.4), tal como foi mostrado na subsecção (5.5.1).

Tendo ponderado todos este aspectos, emerge a vantagem de, no axioma II, se considerar a monotonia no sentido estrito.

Uma vez que o trabalho de Santos generaliza a ideia de Shannon sobre uma axiomática que define a entropia, e este não considerou a monotonia no sentido estrito, puderá parecer que a generalização se torna artificial. Contudo, tal não é o caso. Relembre–se que no capítulo 2 se fez precisamente uma observação sobre a necessidade de considerar a monotonia no sentido estrito, devido à passagem da equação (2.14) para a equação (2.15).

Conclusão

O conceito de entropia tem as suas origens no contexto da Termodinâmica, mais tarde passa pela Mecânica Estatística e Mecânica Quântica. É através de Shannon que este conceito se expande a um domínio bastante distinto dos que até aí o envolveram, nomeadamente a Teoria da Informação. Um dos campos mais recentes onde a entropia desempenha um papel fundamental é a Mecânica Estatística Não Extensiva.

Ao longo deste trabalho abordou–se a entropia do ponto de vista matemático, dando especial atenção à definição de Shannon, de Rényi e de Tsallis. Outros conceitos e definições foram aqui estudados, tais como o ganho de informação de Kullback–Leibler, a função q–exponencial e a função q–logaritmo.

O objectivo deste trabalho nunca foi o de analisar aquelas três entropias do ponto de vista da Teoria da Informação ou da Mecânica Estatística, conforme o caso, mas sim o de proceder a uma abordagem matemática. Assim, após uma breve contextualização das ideias que dão origem a cada uma das fórmulas de entropia, foram analisadas e discutidas as demonstrações de algumas das propriedades que estas verificam. Este estudo incidiu essencialmente sobre a aditividade, concavidade e existência de extremos.

É usual que a entropia total de dois sistemas coincida com a soma das entropias individuais. No entanto, existem fenómenos em que a entropia total não se resume a uma simples soma das partes, ou seja, a entropia do sistema é não aditiva. Isto ocorre quando um dos sistemas interfere substancialmente com o outro. Esta característica está patente na entropia de Tsallis, que é não aditiva (quando $q \neq 1$). A entropia de Shannon e a de Rényi são formas aditivas.

Relativamente à concavidade averiguou–se que a entropia de Tsallis tem uma concavidade bem definida para todos os valores de q, é côncava quando $q > 0$ (estando aqui incluido o caso da entropia de Shannon) e convexa quando $q < 0$. Com a entropia de Rényi não acontece o mesmo, existem valores de α para os quais a função nem é côncava nem convexa. Sobre a existência de extremos concluiu–se que as três formas de entropia atingem um extremo no caso de equiprobabilidade, mas este nem sempre traduz o máximo de entropia. Para $\alpha > 0$, na entropia de Rényi, e para $q > 0$, na entropia de Tsallis, esse extremo é um máximo. Nos restantes casos é um mínimo.

Como se pôde averiguar, existem diversas fórmulas que traduzem a entropia. Estas referem–se a diferentes interpretações, não deixando, no entanto, de envolver um mesmo conceito: uma quantidade estatística associada à diversidade de ocorrências possíveis num sistema.

A elaboração deste trabalho permitiu a aquisição de uma visão global da evolução do conceito de entropia, assim como a análise de aspectos particulares que caracterizam as várias abordagens. Como foi referido na introdução, o objectivo principal desta dissertação era o estudo da entropia de Tsallis, cuja fórmula tem sido aplicada numa grande variedade de problemas de diversas áreas. Muitos dos resultados obtidos sobre esta entropia foram encontrados e sistematizados por cientistas de diversas áreas, particularmente físicos. Alguns destes resultados carecem, entretanto, de uma fundamentação completa e de uma generalização sistemática. Do ponto de vista puramente da Matemática existe ainda um longo caminho a percorrer.

Futuramente, pretende–se fazer uso desta mesma base de conhecimentos no estudo de uma aplicação da entropia de Tsallis na Análise em Componentes Independentes e no tratamento de algum dos muitos problemas em aberto nesta área.

Referências Bibliográficas

[1] S. Abe, "Axioms and uniqueness theorem for Tsallis entropy", *Physics Letters A*, vol.271, pp.74–79, 2000.

[2] A. B. Anjo, J. M. Martins e A. H. Tavares, "Entropia – De Clausius a Jaynes", *Cadernos de Matemática do Departamento de Matemática da Universidade de Aveiro*, CM01/D–04, Aveiro, 2001.

[3] R. Ash, Information Theory, Interscience Publishers, New York, 1965.

[4] A. B. Anjo, "Têmpera Controlada pela Entropia – uma caracterização formal", Tese de Doutoramento, Departamento de Matemática da Universidade de Aveiro, 1999.

[5] E. Borges, "On a q–generalization of circular and hyperbolic functions", *J. Phys. A*, vol.31, pp.5281–5288, 1998.

[6] E. Borges, "Irreversibilidade, Desordem e Incerteza: Três Visões da Generalização do Conceito de Entropia", *Revista Brasileira de Ensino de Física*, vol.21, n.4, pp.453-463, 1999.

[7] L. Borland, A. Plastino e C. Tsallis, "Information gain within nonextensive thermostatistics", *Journal of Mathematical Physics*, vol.39, n.12, pp.6490–6501, 1998.

[8] R. Bowley, Introductory Statistical Mechanics, Clarendon Press, Oxford, 1999 (2nd ed.).

[9] E. M. Curado e C. Tsallis, "Generalized statistical mechaniscs: connection with thermodynamics", *Journal of Physics A*, vol.24, n.2, pp.L69–L72, 1991.

[10] J. Dionísio, "A definição de entropia em cálculo de probabilidades", *Gazeta de Matemática*, Ano XX, vol.74–75, pp.1–7, 1959.

[11] E. Fermi, Termodinâmica, Livraria Almedina, Coimbra, 1973.

[12] R. Graham, "Constantino Tsallis – Describing a New Entropy", *Santa Fe Institute Bulletin*, Vol.15, n.2, 2000.

[13] R. Gray, Entropy and Information Theory, Springer Verlag, New York, 1990.

[14] A. Hyvarinem, J. Karhu7nem e E. Oja, Independent Component Analysis, John Wiley and Sons, New York, 2001.

[15] A. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.

[16] S. Kullback, Information Theory and Statistics, Dover, New York, 1997. (reimporessão da edição de 1968)

[17] K. Mendelsohn, Em demanda do zero absoluto, Editorial Inova, Porto, 1962.

[18] A. Rényi, "Introdution to information theory", pp.540–616, in Rényi, Probability Theory, North–Holland Publishing Company, Amesterdam, 1970.

[19] R. Santos, "Generalization of Shannon's theorem for Tsallis entropy", *Journal Math. Phys.*, vol.38, pp.4104–4107, 1997.

[20] C. Shannon e W. Weaver, The Mathematical Theory of Communication, University Illinois Press, Urbana, pp.48–50 e 116–118, 1963. (primeira edição em 1949)

[21] C. Tsallis, "Possible Generalization of Boltzmann-Gibbs Statistics", *Journal of Statistical Physics*, vol.52, pp.479–487, 1988.

[22] C. Tsallis, "Generalized entropy–based criterion for consistent testing", *Physical Review E*, vol.58, n.2, pp.1442–1445, 1998.

[23] C. Tsallis, "Nonextensive Statistical Mechanics and Thermodynamics", pp.3–98, in Abe e Okamoto (ed.), Nonextensive Statistical Mechanics and Its Applications, Springer, Berlim, 2001.

[24] C. Tsallis, "Entropic nonextensivity: a possible measure of complexity", *Chaos, Solutions and Fractals*, n.13, pp.371-391, 2002.

[25] N. Wiener, Cibernetics - or acontrol and communication in the animal and the machine, John Wiley, New York, 1948.

[26] A. Yaglom e I. Yaglom, "Entropy and Information", pp.44–100, in Probability and Information, D. Reidel Publishing Company, Dodrecht, 1993.

Outras Referências:

[27] T. Arimitsu e P. Jizba, "The world acording to Rényi: Thermodynamics of multifractal systems" (2002), *cond–mat*/0207707
http://arxiv.org/PS_cache/cond-mat/pdf/0207/0207707.pdf, visitado em Fevereiro de 2003.

[28] D. Flamm, “History and outlook of statistical physics” (1997?), physics/9803005 v1
http://arxiv.org/PS_cache/physics/pdf/9803/9803005, visitado em Dezembro de 2002.

[29] D. Flamm, “Ludwig Boltzmann - A Pioneer of Modern Physics” (1997?), physics/9803005 v1
http://arxiv.org/PS_cache/physics/pdf/9710/9710007.pdf, visitado em Janeiro de 2003.

[30] E. Jaynes, (1994 –), Probability Theory: the logic of Science, em formato electrónico no endereço
http://omega.albany.edu:8008/JaynesBook.html, visitado em Dezembro de 2002.

[31] I. Taneja, (2001–), Generalized Information Measures and their Applications, em formato electrónico no endereço
http://jurere.mtm.ufsc.br/ taneja/book/book.html, visitado em Fevereiro de 2003.

[32] Touchette, “When is a quantity additive, and when is extensive?” (2002), *cond–mat*/0201134
http://arxiv.org/PS_cache/cond-mat/pdf/0201/0201134.pdf, visitado em Dezembro de 2002.

[33] “A Brief History of Entropy”
http://www.geocities.com/chemforum/entropyhistory.htm, visitado em Dezembro de 2002.

[34] “As atribulações de Ludwing Boltzmann”. Página do Departamento de Física da Universidade do Ceará, Brasil.
http://www.fisica.ufc.br/tintim9.htm, visitado em Dezembro de 2002.

[35] “Físico brasileiro é cotado para Nobel”, entrevista a Constantino Tsallis pelo jornal Globo de 26/02/2001
http://www.mct.gov.br/sobre/namidia/CTnamidia/26_02.htm, visitado em Janeiro de 2003.

[36] “Uma nova entropia”, Revista do CBPF (Centro Brasileiro de Pesquisas Físicas), Setembro de 2001, pp.57–59.
http://www.cbpf.br/RevistaCBPF/pdf/FisEstatist.pdf, visitado em Janeiro de 2003.

[37] A Dictionary of Scientists. Oxford Reference Online. Oxford University Press, 2003
http://oxfordreference.com, visitado em Dezembro de 2002.

[38] www.nobel.se/physics/laureates/1918/planck–bio.html, visitado em Janeiro de 2003.