# Fundamentos de Teoria da Informação

Israel Andrade Esquef [a] Márcio Portes de Albuquerque [b]
Marcelo Portes de Albuquerque [b]

[a] *Universidade Estadual do Norte Fluminense - UENF*
[b] *Centro Brasileiro de Pesquisas Físicas - CBPF*

**Resumo**

A Teoria da Informação é uma ramificação da teoria de probabilidades introduzida por Claude Shannon com a publicação do artigo "The Mathematical Theory of Communication"[1] em 1948, apresentando um novo modelo matemático para o estudo de sistemas de comunicação. Os principais objetivos de Shannon eram descobrir as leis que regulam os sistemas usados para comunicar e manipular a informação e definir medidas quantitativas para a informação e para a capacidade de determinados sistemas transmitirem, armazenarem e processarem a informação. Uma das inovações mais importantes do modelo introduzido por Shannon, foi considerar os componentes de um sistema de comunicação (fontes de informação, canais de comunicação) como elementos probabilísticos. Alguns dos problemas abordados por Shannon estão relacionados com a descoberta de melhores métodos para utilizar os sistemas de comunicação existentes e as melhores formas de separar a informação desejada (sinal) da informação desprezível (ruído). Um outro tema abordado é a definição de limites superiores para as possibilidades do meio de transporte de informação, também chamado um canal de comunicação.

Shannon propôs uma forma de medição quantitativa da informação fornecida por um evento probabilístico, baseada na tradicional expressão de entropia de Boltzmann (1896) presente na termodinâmica e física estatística. Foi Shannon quem primeiro relacionou entropia e informação. Em seu modelo de comunicação (fonte-canal-receptor)[1], a quantidade de informação transmitida em uma mensagem é função de previsibilidade da mensagem. A noção de entropia está ligada ao grau de desorganização existente na fonte de informação. Quanto maior a desordem, maior o potencial de informação desta fonte. Uma fonte que responda com uma única e mesma mensagem a toda e qualquer pergunta não transmite informação, já que não há redução de incerteza.

## 1 Informação, Incerteza e Entropia

Uma *fonte de informação* é um modelo matemático para um sistema físico que produz uma sucessão de símbolos de maneira aleatória chamados *eventos*. Os símbolos produzidos podem ser números reais como valores de voltagens provenientes de um transdutor, números binários de dados computacionais [1], etc. O espaço contendo todos os eventos possíveis é usualmente chamado de *alfabeto* da fonte de informação, ao qual é atribuído um conjunto de probabilidades de ocorrência.

[1] No caso do sistema físico ser uma imagem digital, pode-se considerar como fonte de informação os valores das voltagens emitidas por um sensor CCD (charge-coupled device), relativas à ocorrência dos fótons incidentes em cada célula.

### *1.1 Fonte Discreta de Informação*

Uma fonte discreta de informação gera símbolos de um alfabeto

$$\mathcal{A} = \{x_i, i = 1, 2, \ldots, k\} \tag{1}$$

com probabilidade de ocorrência $p_i$ tal que $\sum_{i=1}^{k} p_i = 1$.

Os símbolos gerados são estatisticamente independentes, de modo que a probabilidade de ocorrência de qualquer seqüência gerada pela fonte é dada pelo produto das probabilidades de ocorrência dos símbolos que a constituem. Se uma fonte emite uma seqüência de dois símbolos $\alpha$ e $\beta$, com probabilidades $p_\alpha$ e $p_\beta$, respectivamente, a probabilidade da seqüência gerada é definida como sendo

$$p^{\alpha \cup \beta} = \ p^{\alpha} \cdot p^{\beta} \tag{2}$$

caso os simbolos $\alpha$ e $\beta$ sejam estatisticamente independentes.

### *1.2 Medida de Informação*

Quando se descreve um processo de seleção de um objeto entre vários existentes, aparece naturalmente as noções de informação e incerteza. Se um sistema é capaz de emitir 3 símbolos distintos A, B e C e esperamos a ocorrência do primeiro evento, mantemos uma *incerteza* sobre qual símbolo aparecerá. Quando o primeiro símbolo é emitido, a incerteza desaparece e podemos considerar que houve um *ganho de informação.*

Se considerarmos a fonte de informação definida em 1, o ganho de informação associado a ocorrência de um evento é chamada de *informação própria* de cada evento $x_i$ e é definida como

$$I(x_i) = log\left(\frac{1}{p_i}\right), \tag{3}$$

representando uma forma intuitiva de medição quantitativa de informação, mesmo sendo *informação* um conceito relativamente subjetivo. Esta forma de medir informação apresenta as seguintes características:

- $I(x_i) = 0$ se $p_i = 1$
  O ganho de informação resultante da ocorrência do evento único é nulo.

- $I(x_i) \geq 0$
  A ocorrência de qualquer evento produz um ganho de informação, exceto no caso de uma fonte que emite um único símbolo.
- $I(x_i) > I(x_j)$ se $p_i < p_j$
  Quanto menor a probabilidade de ocorrência de um símbolo, maior é o ganho de informação.

E se considerarmos a ocorrência simultânea de dois eventos estatisticamente independentes $x_i$ e $x_j$, o ganho de informação total é definido pela soma das informações próprias de cada um dos eventos

$$\begin{aligned} I(x_i, x_j) &= -log\,(p_i \cdot p_j) \\ &= -logp_i - logp_j \\ &= I(x_i) + I(x_j) \end{aligned} \tag{4}$$

A base da função logaritmo, presente em 3, determina a unidade de medida de informação. A rigor, a base pode ser qualquer número maior que 1 [1], sendo usual a utilização da base 2 para sistemas digitais de informação. O logaritmo de base 2 define a unidade binária (bit) como a informação própria associada a cada um dos símbolos de uma fonte binária com eventos equiprováveis:

$$I(0) = I(1) = -log_2\left(\frac{1}{2}\right) = 1\ bit \tag{5}$$

### *1.3 Entropia de uma Fonte de Informação*

A entropia de uma fonte discreta de informação é dada pela esperança matemática da informação própria dos símbolos da fonte, ou seja, o produto do ganho de informação de cada símbolo pela sua probabilidade de ocorrência. A entropia da fonte é sensível a quantidade de símbolos que a fonte é capaz de emitir. Para uma fonte discreta de informação com $k$ símbolos e probabilidade $p_i = \{p_1, p_2, \ldots, p_k\}$, a entropia é definida como sendo

$$S = E\{I(x_i)\} = \sum_{i=1}^{k} p_i \cdot log\left(\frac{1}{p_i}\right)$$

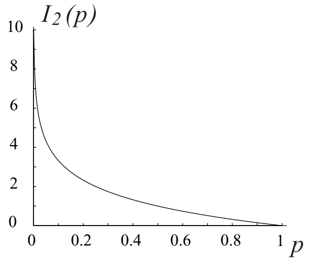
ou na forma mais utilizada na literatura

$$S = -\sum_{i=1}^{k} p_i \cdot log\ p_i \tag{6}$$

A entropia de um sistema binário, que apresenta apenas dois estados possíveis com probabilidades $p$ e $q = 1 - p$, pode ser representada como uma função de $p$,

$$S_2(p) = S_2(p, 1-p) = -p \cdot log_2\ (p)\ -\ (1-p) \cdot log_2\ (1-p) \tag{7}$$

e está ilustrada na figura 1, que também apresenta a informação própria em bits como função de $p$. Quanto menos provável é a ocorrência de um evento, maior é sua informação própria.




| $p$ | $I_2(p)$ | $S_2(p)$ |
|---|---|---|
| 0.001 | 10 | 0.011 |
| 0.01 | 6.6 | 0.081 |
| 0.1 | 3.3 | 0.47 |
| 0.2 | 2.3 | 0.72 |
| 0.5 | 1.0 | 1.0 |

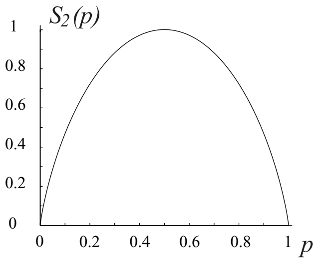



Figura 1. *Informação Própria e Entropia da fonte binária como função da probabilidade.*

Observando a figura 1, é interessante notar que:

(1) Quando $p = 0$ a entropia é nula ($S = 0$), pois $xlog\ x \rightarrow 0$ quando $x \rightarrow 0$
(2) $S = 0$ quando $p = 1$ para um único evento, definido como valor mínimo da entropia.
(3) A entropia atinge o valor máximo $S_{max} = 1\ bit/simbolo$, quando $p = q = \frac{1}{2}$ ou seja, quando os símbolos são equiprováveis.

Portanto, os valores para a entropia de uma fonte de informação com $k$ símbolos é limitada segundo a desigualdade a seguir

$$0 \leq S \leq log_2\ k \tag{8}$$

em que o $log_2$ representa a função logaritmo na base 2.

## 2 Entropia Relativa

Nesta seção será introduzido o conceito de *entropia relativa*. A entropia relativa é a medida de uma distância estatística entre duas distribuições definidas sobre um mesmo alfabeto. Em estatística, isto significa o valor esperado do logaritmo da relação entre as probabilidades. A entropia relativa é definida como sendo

$$D_{KL}(p:p') = \sum_{i=1}^{k} p_i \cdot log \frac{p_i}{p'_i} \tag{9}$$

e é também conhecida como *Distância Kullback-Leibler*, *Entropia Kullback-Leibler* ou *Divergência I*. Na definição acima, assumimos por convenção (baseada em argumentos de continuidade) que $0 \cdot log \frac{0}{p'} = 0$ e $p \cdot log \frac{p}{0} = \infty$.

A entropia relativa é sempre não negativa, satisfazendo a desigualdade de Gibbs:

$$\sum_{i=1}^{k} p_i \cdot log \frac{1}{p_i} \leq \sum_{i=1}^{k} p_i \cdot log \frac{1}{p'_i} \tag{10}$$

considerando $p_i$ uma distribuição de probabilidades qualquer e $p'_i$ uma outra distribuição que satisfaz a condição a seguir

$$\sum_{i=1}^{k} p'_i \leq 1 \tag{11}$$

Apesar da entropia relativa ser chamada de distância Kullback-Leibler, esta não pode ser considerada uma distância verdadeira entre duas distribuições. A entropia relativa não apresenta simetria,

$$\sum_{i=1}^{k} p_i \cdot log \frac{p_i}{p'_i} \neq \sum_{i=1}^{k} p'_i \cdot log \frac{p'_i}{p_i} \tag{12}$$

e portanto não pode ser considerada uma distância métrica. Mesmo assim, é usual pensar na entropia relativa como uma medida de "distância" entre duas distribuições estatísticas.

Uma versão simétrica de entropia relativa, conhecida como divergência $J$, foi introduzida por H. Jeffreys [2] e é definida como a soma das duas divergências diretas (divergências $I$):

$$\begin{aligned} D(p:p') &= D_{KL}(p:p') + D_{KL}(p':p) \\ &= \sum_{i=1}^{k} p_i \cdot log \frac{p_i}{p'_i} + \sum_{i=1}^{k} p'_i \cdot log \frac{p'_i}{p_i} \end{aligned} \tag{13}$$

é valido notar que a divergência $J$ é simétrica em relação aos argumentos, de forma que $D(p:p') = D(p':p)$.

# 3 Entropia Generalizada em Sistemas de Informação

## *3.1 Conceitos Fundamentais da Entropia Não Extensiva*

Durante aproximadamente 120 anos, o conceito de entropia tem sido descrito através de uma expressão particular, chamada *Entropia Boltzmann-Gibbs*

$$S = k \log W \tag{14}$$

onde a entropia ($S$) é o produto da constante de Boltzmann ($k$) pelo logaritmo de microestados ($W$) do sistema.

O conceito de entropia é de fundamental importância na termodinâmica, mecânica estatística e teoria da informação. Recentemente, estudos de sistemas físicos que envolvem a presença de efeitos não extensivos têm despertado um grande interesse, principalmente porque tais sistemas não são convenientemente descritos pela mecânica estatística de Boltzmann-Gibbs. A presença de características não extensivas é comum em sistemas astrofísicos, sistemas magnéticos e sistemas que apresentam evolução temporal da entropia. De um modo simplificado, podemos afirmar que a não extensividade pode ocorrer quando:
($i$) as interações são de curto alcance sendo o sistema de tamanho finito e menor do que o alcance das interações ou
($ii$) com interações de longo alcance sendo o tamanho do sistema qualquer.

Com base neste contexto, acredita-se que a mecânica estatística atual possui limitações, existindo a necessidade de uma reformulação dos conceitos de entropia e extensividade.

Durante os últimos anos, uma nova expressão para entropia, proposta pelo físico Constantino Tsallis [3], tem sido considerada uma possível generalização da entropia de Boltzmann/Gibbs. Este novo formalismo, chamado de Entropia Tsallis ou Estatística Tsallis, tem sido aplicado a inúmeros sistemas, em diversas áreas da ciência, que vão desde a física do estado sólido até a teoria da informação [4] [5] [6] [7] [8] [9] [10]. A entropia Tsallis se adapta as características físicas de muitos sistemas físicos e ainda preserva as propriedades fundamentais da entropia na segunda lei da termodinâmica, ou seja, que a entropia do universo aumenta com o tempo em todos os processos físicos. Este novo formalismo para a termodinâmica e física estatística está fundamentado na expressão para a entropia sugerida por Tsallis

$$S_q = k \cdot \sum_{i=1}^{W} p_i^q \cdot \frac{(1 - p_i^{1-q})}{q - 1} \tag{15}$$

ou na forma mais utilizada na literatura,

$$S_q = k\frac{1 - \sum_{i=1}^{W} p_i^q}{q-1} \tag{16}$$

em que $k$ é uma constante positiva (a qual é atribuído o valor unitário), $q$ é um número real, $W$ é o número total de microestados e $p_i$ é o conjunto de probabilidades associado aos estados. Pode-se facilmente demonstrar que no limite em que $q \to 1$ a equação 16 retorna para a expressão 6, a entropia de Boltzmann/Gibbs/Shannon [2].

$$S = -\sum_{i=1}^{W} p_i \cdot \log p_i \tag{17}$$

### *3.2 Entropia de Shannon Generalizada*

Nas seções anteriores deste capítulo, foram definidos os conceitos fundamentais da teoria da informação. Foi dada uma ênfase ao conceito de medida de informação proposta por Shannon a partir de um auto questionamento: "Podemos definir uma quantidade que possa medir, de alguma maneira, quanta informação é 'produzida' em um processo, ou melhor, a que taxa a informação é produzida?" A partir daí, Shannon define que para uma medida deste tipo ser possível, uma função $S$ dependente de probabilidades $p_i$ para $W$ eventos deveria satisfazer as seguintes propriedades:

(1) $S$ deveria ser contínua em $p_i$.
(2) Se a probabilidade de todos os eventos forem iguais, ou seja $p_i = \frac{1}{W}$, então $S$ deveria ser uma função monotônica crescente. Com eventos equiprováveis, há maior poder de escolha, ou maior incerteza, quando existir um número grande de eventos possíveis.
(3) Para dois subsistemas estatisticamente independentes $A$ e $B$ a função $S$ para o sistema composto $A+B$ deveria apresentar a propriedade aditiva, de forma que: $S(A+B) = S_A + S_B$.

Tendo em conta estas propriedades, Shannon enuncia o seguinte teorema:

***Teorema 1*** – A única função que satisfaz as três condições anteriores é da forma:

$$S = -K\sum_{i} p_i \cdot log\ p_i \tag{18}$$

[2] Esta demosntração matemática está presente no Apendice A.

em que $K$ é uma constante positiva que, por conveniência em sistemas de informação, é considerada de valor unitário.

Recentemente, Santos [11] generalizou o conhecido teorema de Shannon, que define a equação 18 como uma medida quantitativa da informação de uma fonte. É conhecido da estatística Tsallis que $S_q$ definido anteriormente na equação 16 satisfaz as seguintes condições:

(1) $S_q$ é contínua em $p_i$, para $0 < p_i < 1$.
(2) Para um conjunto de $W$ de eventos equiprováveis, ou seja, $p_i = \frac{1}{W}$, então $S_q$ é uma função monotônica crescente.
(3) Para dois subsistemas estatisticamente independentes $A$ e $B$ a entropia generalizada $S_q$ do sistema composto $A + B$ satisfaz a relação de pseudo-aditividade

$$S_q(A + B) = S_q(A) + S_q(B) + (1 - q) \cdot S_q(A) \cdot S_q(B) \tag{19}$$

Tendo em conta as três condições anteriores, Santos define em [11] o seguinte teorema:

***Teorema 2*** – A única expressão que satisfaz de forma simultânea todas as condições acima é a entropia generalizada Tsallis:

$$S_q = k\frac{1 - \sum_{i=1}^{W} p_i^q}{q - 1} \tag{20}$$

Deste forma, fica definido que a entropia Tsallis pode ser utilizada como uma medida de informação adequada para a utilização em sistemas de informação que apresentam características não extensivas [3] .

### *3.3 Entropia Relativa Generalizada*

Inicialmente, torna-se necessário apresentar a derivação da expressão da entropia relativa pelo caso convencional. Assumindo que um conjunto de eventos W com probabilidades $p_i$, considerando o índice $i$ para os eventos $W$, a entropia de Shannon é definida como na equação 18. Considerando a quantidade $I_i = -\log p_i$ como a informação própria de cada evento, conforme definido na seção 1.2, e tomando $p_i$ e $p_i'$ como probabilidades para dois conjuntos de eventos, podemos afirmar qua a diferença de informação obtida através destas

[3] O Apêndice A apresenta a demonstração do Teorema 1, como definido por Shannon. A demonstração do Teorema 2 pode ser encontrada em [11]

duas medidas é

$$\Delta I_i = -(\log p'_i - \log p_i) \tag{21}$$

A taxa média da informação modificada pode ser obtida através da expressão a seguir

$$D(p:p') = \sum_i p_i \cdot \Delta I_i = \sum_i p_i \cdot log\, \frac{p_i}{p'_i} \tag{22}$$

Esta é a definição convencional para o ganho de informação Kullback-Leibler, também conhecido como entropia relativa.

Em recente trabalho publicado, Lisa Borland [12] generalizou o ganho de informação Kullback-Leibler para a estatística não extensiva. Uma medida Kullback-Leibler generalizada se deriva naturalmente da aplicação do formalismo da entropia Tsallis no lugar do convencional de Shannon. A partir da equação 15 podemos definir portanto, a informação própria não extensiva de cada evento, ou seja, $I_i^q = -\frac{(p_i^{1-q}-1)}{1-q}$. Considerando novamente $p_i$ e $p'_i$ como as probabilidades para dois conjuntos de eventos medidos, a diferença de informação entre as medidas é

$$\Delta I_i^q = \left[\frac{1}{(1-q)}\right] \cdot \left[(1 - p_i'^{1-q}) - (1 - p_i^{1-q})\right] \tag{23}$$

A taxa média da informação modificada pode ser obtida através da expressão a seguir

$$D_q(p:p') = \sum_i \frac{p_i^q}{1-q} \cdot \left(p_i^{1-q} - p_i'^{1-q}\right) \tag{24}$$

que representa a entropia relativa generalizada. Esta expressão pode ser escrita de outra forma, considerando a função q-logarítmica, $\log_q(p) = \frac{p^{1-q}-1}{1-q}$ definida em [13], resultando em

$$D_q(p:p') = \sum_i p_i \cdot \log_q \frac{p_i}{p'_i} \tag{25}$$

e é válido ressaltar que $\log_1(p)$ retorna à expressão convencional $\log(p)$.

## Referências


[1] Shannon, C. E., "*A Mathematical Theory of Communication.*", Bell Sys. Tech. J., **27**, 379-423 e 623-656, 1948. Versão eletrônica disponivel em (http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html)

[2] Jeffreys, H., *Probability Theory*, Oxford University Press, 1961, New York.

[3] Tsallis, C.,"*Possible generalization of Boltzmann-Gibbs Statistics.*" J. Stat. Phys., **52**, 479-487, 1988.

[4] Sampaio, L. C. and Albuquerque, M. P. and Menezes, F. S., "*Nonextensivity and Tsallis statistics in magnetic systems*", Physical Review B, **55**, 9, 5611-5614, 1997.

[5] Oliveira, I. S., "*Some metallic properties in the framework of Tsallis generalized statistics.*", Eur. Phys. J. B, **14**, 1, 43-46, 2000.

[6] Martin, M. T and Plastino, A. R. and Plastino, A., "*Tsallis-like information measure and the analysis of complex signals.*", Physica A, **275**, 262-271, 1999.

[7] Tong, S. and Bezerianos, A. and Paul, J. and Zhu, Y. and Thakor, N., "*Nonextensive entropy measure of EEG following brain injury from cardiac arrest*", Physica A, **305**, 3-4, 619-628, 2002.

[8] Landsberg, P. T. and Vedral, V., "*Distributions and channel capacities in generalized statistical mechanics.*", Phys. Lett. A, **247**, 211-217, 1998.

[9] Capurro, A. and Diambra, L. and Lorenzo, D. and Macadar, O. and Martin, M. T. and Mostaccio, C. and Plastino, A. and Rofman, E. and Torres, M. E. and Velluti, J., "*Tsallis entropy and cortical dynamics: The analysis of EEG signals.*", Physica A, **257**, 149-155, 1998.

[10] Michael, F. and Johnson, M. D., "*Financial market dynamics*", Physica A, **In Press**, 2002.

[11] Santos, R. J. V., "*Generalization of Shannon's theorem for Tsallis entropy*", J. Math. Phys., **38**, 8, 4104-4107, 1997.

[12] Borland, L. and Plastino, A. R. and Tsallis, C., "*Information Gain within Nonextensive Thermostatistics.*", J. Math. Phys., **39**, 12, 1998.

[13] Borges, E. P., "*On a q-generalization of circular and hyperbolic functions.*", J. Phys. A, **31**, 23, 5281-5288, 1998.