

A MEDIDA DE INFORMAÇÃO DE SHANNON: ENTROPIA

José Carlos Magossi

Pedro Henrique Camargo de Abreu

Antônio César da Costa Barros

José Renato Paviotti

Universidade Estadual de Campinas – UNICAMP – Brasil
Faculdade de Tecnologia – FT

(aceito para publicação em outubro de 2021)

Resumo

Logo após Claude E. Shannon, em 1948, ter publicado o artigo *A Mathematical Theory of Communication*, diversas áreas se valeram de seus escritos, principalmente por ele ter desenvolvido uma fórmula para “medir informação” em seu modelo matemático de comunicação, denominando-a *entropia*. Shannon optou pela justificativa operacional da existência de sua fórmula de entropia. Por conseguinte, houve uma expansão das áreas de investigações matemáticas sobre as possíveis caracterizações de *medidas de informação*. Neste texto o objetivo é focar nas estruturas matemáticas que fundamentam o conceito de medidas de informação. Estima-se com isso, no sentido didático, que haja esclarecimentos com relação às múltiplas leituras do conceito de entropia.

Palavras-chave: entropia, medidas de informação, axiomas, equação funcional, teoria de Shannon.

[SHANNON'S INFORMATION MEASURE: ENTROPY]

Abstract

Right after Claude E. Shannon had published the article *A Mathematical Theory of Communication* in 1948, several areas used his writings, mainly because he developed a formula to “measure information” in his mathematical model of communication, naming it

as entropy. Shannon opted for the operational justification of the existence of his entropy's formula. Therefore, there was an expansion of the areas of mathematical investigations on the possible characterizations of information measures. In this text, the objective is to focus on the mathematical structures that support the concept of information measure. It is expected that, in the didactic sense, there will be clarifications regarding the multiple readings of the concept of entropy.

Keywords: *entropy, information measures, axioms, functional equation, Shannon's theory.*

Introdução

No desenvolvimento histórico da matemática há períodos em que a intuição, geométrica ou não, funciona como motor propulsor para o surgimento de novas ferramentas. Há também períodos em que a abstração e o formalismo acabam por ser “os protagonistas do espetáculo”. Oscilação entre intuição e formalismo encontra-se, por exemplo, na interconexão entre o Cálculo, de I. Newton (1643–1727) e G. W. Leibniz (1646–1716), século XVIII, os Fundamentos da Análise Matemática (HAIRER e WANNER, 1996; JAHNKE, 2003; LÜTZEN, 2003; SCHUBRING, 2005) e Séries de Fourier, séculos XIX e XX (GRATTAN-GUINNESS, 2007; VAN VLECK, 1914; BRESSOUD, 2007). Com base nisso, estima-se que algo semelhante possa ter ocorrido com a Teoria da Informação, quando Claude E. Shannon (1916-2001), em seu artigo de 1948 (SHANNON, 1948), desenvolve um modelo matemático com vistas à operacionalização dos sistemas de comunicação e à intuição tecnológica, e não se debruça nos formalismos matemáticos (SHANNON, 1956), subjacentes à sua fórmula de entropia: “É dada, principalmente, para conceder uma certa plausibilidade a algumas de nossas definições posteriores. A real justificativa dessas definições, entretanto, residirá nas suas implicações” (SHANNON, 1948, p. 393, tradução nossa).

Não há dúvidas de que a plausibilidade de seus desenvolvimentos e de sua intuição em seus teoremas é relevante, além de ser de uma elegância matemática magistral. Não há dúvidas também dos imensos impactos que sua teoria acarretou na era da informação. No entanto, é interessante a discussão sobre “intuição operacional” e axiomas, voltada a processos de comunicação, tal como ocorre na literatura (ACZÉL et al., 1974; ACZÉL e DARÓCZY, 1975; ACZÉL, 1984a; ACZÉL, 1984b; ACZÉL, 1986) sobre o quanto se tem de “natural” e “essencial” nas medidas de informação e em suas propriedades egressas. Isso traz à memória as discussões sobre os fundamentos da Análise Matemática no século XIX e proporciona o seguinte questionamento: seriam as caracterizações de medidas de informação balizas para novos desenvolvimentos em teoria da comunicação? Entende-se que uma possível resposta se encontra na relação entre o problema filosófico acerca do *conceito de informação* e o problema técnico acerca de *como medir uma informação*.

Não há como negar que a palavra “informação” se conecta com tudo aquilo que transforma, de alguma maneira, o modo de pensar. Na matemática o conceito de infinitésimos (e o conceito de *continuum*) embasou as discussões sobre intuição geométrica e rigor matemático. É possível estimar que na Teoria da Informação (Teoria de Shannon) o conceito

que possibilitaria essa propensa discussão entre intuição operacional e rigor matemático seja o conceito de informação. Uma investigação sobre uma definição precisa do conceito de informação pode muito bem revelar ser este um problema aberto na ciência. Seriam então todos os problemas abertos, correlatos ao conceito de informação, os precursores de uma hipotética “futura” era do rigor na Teoria da Informação, tal como ocorreu com os Fundamentos da Análise Matemática? Nota-se que a palavra “informação” tem um papel significativo nesse contexto, haja vista que a teoria de Shannon surgiu, conforme escrito por John Pierce em seu livro *An Introduction to Information Theory, Symbols, Signals and Noise*, com objetivo de melhorias na área de telegrafia, e expandiu-se para outros domínios da ciência que incluem o conceito de medida de informação (PIERCE, 1980). Isso acaba por gerar dúvidas com relação à palavra “informação” presente em sua teoria.

Como um exemplo, pode-se citar David. I. Spivak em seu livro *Category Theory for the Sciences*:

“Não se presta atenção ao conteúdo das mensagens – o que elas significam – como Shannon diz especificamente em seu artigo seminal: “Frequentemente as mensagens têm significado; isto é, elas referem-se ou estão correlacionadas de acordo com algum sistema com certas entidades físicas ou conceituais. Esses aspectos semânticos da comunicação são irrelevantes ao problema de engenharia”. Portanto, eu penso que este assunto está mal nomeado. Deveria ser chamado teoria da compressão ou teoria da redundância. A informação é inerentemente significativa – esse é o seu propósito – portanto uma teoria despreocupada com significados não está realmente estudando informação propriamente dita.” (SPIVAK, 2014, p.266, tradução nossa)

No presente artigo, não se entra no mérito, nem na pesquisa, dos questionamentos acerca do conceito de informação, seja com intuição, rigor, filosofia ou epistemologia. O artigo limita-se ao escopo das medidas de informação com viés matemático, sem levar em conta aspectos semânticos ou interpretativos dessas medidas.

Investiga-se, neste texto, uma trajetória histórica referente aos sistemas axiomáticos já existentes, propostos com o intuito de deduzir a fórmula de entropia de Shannon como a única possível, diante de um conjunto específico de axiomas. Questiona-se, de modo análogo aos fundamentos da Análise Matemática, qual seria então a importância do rigor matemático na caracterização de medidas de informação. Essa vertente, de certa forma, já foi avalizada por Shannon, que escreve sobre o caminho da Teoria da Informação nos anos que se seguiriam:

“De fato, o núcleo central da teoria da informação é, essencialmente, um ramo da matemática, um sistema estritamente dedutivo. Um entendimento rigoroso da fundação matemática e sua aplicação em comunicação é certamente um pré-requisito para outras aplicações. (...) mas o estabelecimento de tais aplicações não é uma simples questão de traduzir palavras para um novo domínio, mas sim o lento e tedioso processo de

hipótese e verificação experimental.” (SHANNON, 1956, p.3, tradução nossa)

Ainda não se completaram 100 anos da promulgação da Teoria de Shannon e ainda há muita discussão sobre as implicações matemáticas subjacentes à sua teoria, que está alicerçada, a princípio, numa teoria de probabilidades (nota-se, por exemplo, que há discussões sobre o fato de a teoria das probabilidades poder ser considerada um ramo da teoria da informação (INGARDEN e URBANIK, 1962)).

São clássicos os impactos da Teoria de Shannon em diversas áreas, além daquelas voltadas aos processos de comunicação (KANNAPPAN, 2009, p.404; REZA, 1961; COVER e THOMAS, 2006; VERDÚ e McLAUGHLIN, 2000). Pode-se, como exemplo, citar Sergio Verdú, quando diz das medidas de informação com Estimção Empírica, fundamentadas nos conceitos de entropia, informação mútua e entropia relativa: “A importância de medidas de informação transcende a teoria da informação” (VERDÚ, 2019, p.2, tradução nossa).

É clássica também a busca pela “correção” da matemática presente no artigo de Shannon¹, principalmente com relação à sua própria asserção, em seu artigo original, de que a intuição deveria prevalecer: “Esse teorema, e as hipóteses requeridas para sua demonstração, não é, de forma alguma, necessário para a presente teoria” (SHANNON, 1948, p. 393, tradução nossa).

Muitos trabalhos foram desenvolvidos com o intuito de dar prosseguimento a uma caracterização matemática, precisa, da fórmula de entropia

$$H(X) = - \sum_{i=1}^n p_i \log p_i,$$

e de seus possíveis desdobramentos matemáticos.

A. I. Khinchin (1894–1959), em seu artigo, escrito em russo, de 1953, faz comentários sobre a matemática “incompleta” presente nos trabalhos de Shannon². O objetivo de Khinchin, em seu livro, era o de fornecer uma matemática precisa e “completa”³, fundamentada em axiomas e em equações funcionais⁴, cujo norte é a dedução da fórmula de entropia de Shannon, além, é claro, da demonstração de seus teoremas fundamentais. É o início, diga-se, de investigações matemáticas sobre a dedução da fórmula de entropia presente

¹ Podem-se citar, por exemplo, as críticas do matemático Joseph L. Doob (1910–2004) com relação à falta de rigor matemático e ausência de demonstrações suficientes na Teoria de Shannon (DOOB, 1949, p.133).

² Esse artigo aparece no livro da Dover Publications (tradução do russo), *Mathematical Foundations of Information Theory* (KHINCHIN, 1957)

³ A primeira caracterização axiomática de uma medida de informação deve-se a Shannon (1948). Como Shannon enfatizou fortemente a intuição em sua caracterização, Khinchin, em 1953 (exposto em Khinchin (1957)), fornece um tratamento mais sofisticado, mas, segundo Aczél e Daróczy (1975), ainda com alguns vieses de intuição operacional. Já em 1956, Faddeev desenvolve uma axiomática menos restritiva que Khinchin (REZA, 1961).

⁴ Sobre equações funcionais ver Kannappan (2009), 2009Aczél (1966) e Aczél e Dhombres (1989).

no trabalho de Shannon⁵. Concomitante aos trabalhos de Khinchin, surgem também fortes desenvolvimentos agregados ao conceito de entropia, sejam sob a ótica da física (BRILLOUIN, 1956; WEHRL, 1978; WEHRL, 1991), sejam sob a ótica semântica (BAR-HILLEL e CARNAP, 1953), aspectos não abordados por Shannon em seu artigo, haja vista não estarem inseridos no propósito de melhorias dos processos de “comunicação elétrica”. Interessante é a observação de que o artigo de Shannon fomentou a diversificação de leituras associadas ao termo “entropia”, mesmo que não estejam associadas ao conceito de entropia desenvolvido por Rudolf J. E. Clausius (1822–1888) em trabalhos sobre termodinâmica (CLAUSIUS, 1865, 1867), nem ao conceito de entropia visto sob a ótica da mecânica estatística, conforme indicado em Pierce (1980).

No presente artigo, analisam-se, sob a forma de compilação de resultados já obtidos, haja vista serem resultados clássicos, algumas caracterizações de medidas de informação, sendo que a entropia de Shannon se configura como uma delas. Esta exposição tem como meta principal esclarecer a contraparte matemática da entropia de Shannon, e com isso, balizar as possíveis investigações sobre a diferença conceitual entre o clássico conceito de entropia da Física, da mecânica estatística, e o de Shannon, mesmo tendo eles o mesmo nome⁶.

O sentido matemático dessas caracterizações é o de analisar quais são os axiomas que fundamentam a fórmula de entropia de Shannon, tornando, de certa forma, os teoremas da Teoria de Shannon deduções em um sistema axiomático. Com isso, é possível demonstrar que a fórmula de entropia de Shannon se revela como única possível, levando-se em conta um determinado sistema de axiomas. Mais ainda, estima-se que com isso se abram espaços não somente para possíveis inovações tecnológicas, mas também para colocar os conceitos associados à entropia em patamares matemáticos sólidos. Nesse caso, o foco na busca por axiomas é o foco também na busca por equações funcionais, que fundamentem e determinem de forma única a fórmula de entropia proposta por Shannon em seu trabalho de 1948. Com viés didático, opta-se por expor uma demonstração clássica da fórmula

$$h(p) = -\log_2 p \quad p \in (0,1]$$

para um sistema simples de informação de um evento. Desse modo, entende-se, facilita-se a interpretação da relação entre “informação” e “incerteza”, presentes na teoria matemática de comunicação de Shannon. Indicam-se também os axiomas utilizados por Shannon em seu artigo, muito bem apresentados no livro de Robert B. Ash (1935–2015) (ASH, 1965), além da exposição de outras axiomáticas que caracterizam a medida de informação de Shannon. Expõe-se ainda uma breve discussão sobre intuição operacional relacionada às equações funcionais presentes nesses sistemas axiomáticos.

⁵ Nota-se que todo o trabalho de Shannon é “palco” de investigações matemáticas, sejam elas na fórmula de entropia, sejam em seus teoremas. Isso significa que há uma riqueza de desenvolvimentos matemáticos, físicos, tecnológicos etc., egressos do seu trabalho fundamental de 1948 (SHANNON, 1948).

⁶ Sobre a escolha do nome entropia feita por Shannon ver, por exemplo, Magossi e Paviotti (2019).

Considerações sobre medidas de informação

Neste trabalho discute-se a relação que existe entre o conceito de *medida de informação*, no sentido geral (DARÓCZY, 1970; ACZÉL e DARÓCZY, 1975; ACZÉL, 1984a; EBANKS et al. 1998; ACZÉL, 1986), e a Teoria Matemática de Comunicação de Shannon. Em Shannon (1948), a incerteza (entropia para Shannon) presente *a priori* e a quantidade de informação, obtida *a posteriori* como resultado da eliminação da incerteza, fundamentam-se numa certa definição de probabilidade. Enquanto, por um lado, é intuitiva a utilização da incerteza de Shannon (entropia) em seu modelo matemático de comunicação, por outro, ainda há investigação no sentido da matemática subjacente e também no sentido de quais propriedades são essenciais, “naturais”, na dedução da fórmula da entropia de Shannon fundamentada num sistema de axiomas (ACZÉL, 1984a, 1984b, 1986). Ou seja, quais propriedades “decorrem normalmente de uma ordem regular das coisas” e quais são indispensáveis, seja para a consistência da matemática utilizada, seja para a realização dos modelos em sistemas de comunicação entre “máquinas”.

É evidente que essa área, por mais investigada que tenha sido, pode ainda fornecer bons resultados, levando-se em conta a avalanche de desenvolvimentos tecnológicos (inteligência artificial, ciência dos dados, segurança de sistemas, teoria de questionários etc.) imersos numa necessidade básica, qual seja, “medir informação”. Mas a situação pode se tornar ainda mais complexa ao considerar a palavra “entropia” encontrada em diversas ramificações da ciência, tais como na Física, advinda dos trabalhos de R. Clausius (CLAUSIUS, 1965, 1967), na mecânica estatística (WEHRL, 1978, 1991), na matemática, nas engenharias⁷ e em outras áreas da ciência (COVER e THOMAS, 2006). Tal como exposto no artigo *Incerteza em Entropia* (MAGOSSO e PAVIOTTI, 2019), pode-se dizer que o excesso de áreas que se valem do termo “Teoria da Informação” se relaciona diretamente com as facilidades de atribuição de significados aos termos “incerteza”, “informação”, “comunicação” e “entropia”. A ideia é a de que uma exposição da entropia de Shannon como sendo determinada com base em axiomas possa possibilitar novas leituras de teoremas, desigualdades etc., com base num estudo unicamente matemático. Estima-se que possa possibilitar também que cenários sejam criados, no sentido matemático ou no sentido prático-tecnológico, com fins de provar a existência (ou não) de propriedades, desejáveis para uma intuição operacional. Vale citar I. Csiszár:

“Este autor tem conhecimento de uma ocasião [31], quando uma abordagem axiomática levou a uma nova medida de informação de interesse prático, e de outra [48], quando tal abordagem iniciou pesquisa que teve sucesso em encontrar significados operacionais de uma medida de informação previamente insignificante.” (CSISZÁR, 2008, p. 269, tradução nossa)

⁷ Por exemplo, em Agrell et al. (AGRELL et al., 2016), discutem-se as aplicações da Teoria da Informação em fibras óticas. Nesse caso o problema de determinar a capacidade de canal, no sentido de Shannon, para fibra óticas, é um problema bem interessante.

Na citação acima as referências citadas, [31] e [48], correspondem neste texto, respectivamente, às referências (HAVRDA e CHARVÁT, 1967; RÉNYI, 1961). Nesse sentido, tanto a ideia de probabilidade, como um ramo da teoria da medida no sentido estritamente formal, quanto a ideia de uma possível medida de informação, dependente de um conjunto de axiomas, estão abertas a novos desenvolvimentos matemático-tecnológicos que podem impactar na Teoria da Informação, na Teoria da Medida e na Teoria de Equações Funcionais, para dizer apenas de algumas áreas de pesquisa.

Neste trabalho o viés é muito mais exibir a matemática de uma medida de informação do que justificar sua implicação tecnológica, operacional ou pragmática. O objetivo é pensar e refletir sobre uma possível era do rigor no quesito “medida de informação”, que poderia fomentar, quiçá, a solução de problemas abertos correlatos ao conceito de informação.

Seriam então a intuição, a plausibilidade e a operacionalização na Teoria da Informação suficientes para “medir informação”, ou seria necessário, tal como no surgimento da era do rigor na matemática, que intuição fosse vista com cautela e se debruçasse sobre a busca pelo rigor matemático subjacente ao desenvolvimento pragmático, intuitivo? Essa ideia é clássica na matemática, conforme pode ser observado no prefácio da tradução do livro de Cauchy, *Cours d’analyse*, por Robert E. Bradley e C. Edward Sandifer.

*“Muitas vezes na história da matemática, vê-se um padrão em que as ideias e aplicações ocorrem primeiro e o rigor vem depois. Isso aconteceu nos tempos antigos, quando a geometria prática dos povos da Mesopotâmia e do Egito evoluíram para os esforços dos gregos. Isso aconteceu de novo com o Cálculo. O Cálculo foi descoberto, alguns dizem inventado, quase independentemente por Isaac Newton (1642–1727) em 1666 e por Gottfried Wilhelm von Leibniz (1646–1716) quase 10 anos mais tarde, mas as suas fundações rigorosas não foram estabelecidas, apesar de várias tentativas, por mais de 150 anos. (...) Em 1821, Augustin-Louis Cauchy (1789–1857) publicou um livro texto, o *Cours d’analyse*, (...) não somente Cauchy forneceu uma definição viável de limites e um meio de os transformar na base de uma teoria rigorosa do cálculo, mas também revitalizou a ideia de que toda a matemática poderia ser colocada sobre tais fundações rigorosas. Hoje, a qualidade de um trabalho matemático é avaliada em parte pela qualidade do seu rigor.”* (BRADLEY e SANDIFER, 2009, p.vii, tradução nossa)

É evidente que alguns problemas que a intuição geométrica não dava conta de responder, no século XVIII, impulsionaram a revisão dos fundamentos da Análise Matemática, revendo conceitos como os de funções, continuidade, convergência etc. (KLINE, 1975; BOTTAZZINI, 1986).

Seria pertinente, nessa mesma esteira, questionar sobre quais problemas abertos na Teoria da Informação poderiam ser vistos como motores propulsores para uma propensa justificativa do fortalecimento do rigor em detrimento de uma visão única voltada ao sentido operacional, pragmático, no quesito “medida de informação”? Há evidentemente muitos

problemas abertos em Teoria da Informação. Qual seria o impacto das análises das caracterizações axiomáticas de medidas de informação nos problemas de canais de difusão, conhecidos como *broadcast channel* (COVER, 1972; COVER e THOMAS, 2006), ou no problema da inserção de aspectos semânticos na Teoria da Informação (BAR-HILLEL e CARNAP, 1953), ou ainda em todos os diversos desenvolvimentos tecnológicos (robótica, *big data*, economia, criptografia, comunicação 5G, comunicação 6G, canais de comunicação, teoria de questionários etc.) que se valem de uma noção de “medida de informação”?

Talvez seja então possível intuir, diante da plasticidade matemática subjacente à fórmula de entropia de Shannon, e de suas extensões como caracterizações de medidas de informação, uma possível resposta à pergunta clássica sobre o porquê de Shannon ter escolhido o nome entropia para sua fórmula (TRIBUS e McIRVINE, 1983; ELLERSICK, 1984; MAGOSSO e PAVIOTTI, 2019). Mais ainda, inferir que há muito mais a ser discutido sobre medidas de informação do que a sua visualização apenas como uma fórmula dentro da Teoria da Informação, principalmente se a atenção se voltar para o conceito de entropia revelado na termodinâmica por R. Clausius, na mecânica estatística, e com as inúmeras associações entre informação⁸, energia e medidas (CLAUSIUS, 1965, 1967; BRILLOUIN, 1956; TRIBUS e McIRVINE, 1983). Na sequência, neste artigo, com foco nos sistemas axiomáticos e na possibilidade de múltiplas interpretações⁹, expõe-se a demonstração da fórmula de entropia para um simples evento (ACZÉL e DARÓCZY, 1975) e em seguida mostram-se os axiomas de Shannon, de 1948 (SHANNON, 1948), tal como também exposto no livro *Information Theory* (ASH, 1965). Indicam-se referências às axiomáticas de A. Khinchin, de 1953 (KHINCHIN, 1957), e de outros sistemas de axiomas que revelam a unicidade da fórmula da entropia de Shannon condicionada a axiomas predefinidos.

Entropia de um simples evento

O objetivo da exposição de uma fórmula para a entropia, de Shannon, de um simples evento, é lançar as bases, didáticas e simples, para a fórmula “geral” de entropia de Shannon¹⁰, e das inúmeras caracterizações de medida de informação (ACZÉL e DARÓCZY, 1975; ACZÉL, 1984a, 1986; EBANKS et al. 1998; DARÓCZY, 1970). Com base nisso, entende-se que haverá facilidades na compreensão não só da fórmula de entropia de Shannon como também das inúmeras caracterizações de medidas de informação. Na apresentação seguinte leva-se em conta, tal como exposto no livro *On measures of information and their characterizations* (ACZÉL e DARÓCZY, 1975) que se tem um simples evento estocástico (variável aleatória X) com probabilidade $P(X) \neq 0$. Para essa demonstração, a equação funcional de Cauchy (ACZÉL, 1966) é parte fundamental. Os teoremas a seguir podem ser encontrados

⁸ Por exemplo, em Ben-Naim (BEM-NAIM, 2010), o autor indica que a sigla *SMI* (*Shannon's Measure of Information*) é utilizada para representar a medida de informação de Shannon.

⁹ Tal como ocorrido com a geometria de Euclides e com os trabalhos de D. Hilbert (1862-1943) em seu livro *Foundations of Geometry* de 1903.

¹⁰ Uma exposição, com pretensões didáticas, da demonstração da fórmula $H(X)$ de entropia de Shannon pode ser encontrada em Magossi e Barros (2021).

no livro *Functional equations and how to solve them* (SMALL, 2007) e também no livro *On measures of information and their characterizations* (ACZÉL e DARÓCZY, 1975).

Teorema 3. (SMALL, 2007, pp.31-32) *Seja $f: \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua de números reais que satisfaz a equação de Cauchy $f(x + y) = f(x) + f(y)$, para todo $x, y \in \mathbb{R}$. Então, para quaisquer $x_i \in \mathbb{R}$, com $i = 1, 2, \dots, n$, tem-se que $f(x_1 + x_2 + \dots + x_n) = f(x_1) + f(x_2) + \dots + f(x_n)$.*

Teorema 3.2 (SMALL, 2007, pp.31-32) *Seja $f: \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua de números reais que satisfaz a equação de Cauchy $f(x + y) = f(x) + f(y)$, para todo $x, y \in \mathbb{R}$. Então, existe um número $a \in \mathbb{R}$, tal que $f(q) = aq, \forall q \in \mathbb{Q}$.*

Teorema 3.3 (SMALL, 2007, p.33) *Se $f: \mathbb{R} \rightarrow \mathbb{R}$ e $g: \mathbb{R} \rightarrow \mathbb{R}$ são funções contínuas tal que $f(q) = g(q)$ para todo $q \in \mathbb{Q}$. Então, $f(x) = g(x)$ para todo $x \in \mathbb{R}$.*

Teorema 3.4. (SMALL, 2007, p.34) *Seja $f: \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua de números reais que satisfaz a equação de Cauchy $f(x + y) = f(x) + f(y)$, para todo $x, y \in \mathbb{R}$. Então, existe um número $a \in \mathbb{R}$ tal que $f(x) = ax, \forall x \in \mathbb{R}$.*

Teorema 3.5. (SMALL, 2007, p.34) *Seja $f: \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua de números reais que satisfaz a equação de Cauchy $f(x + y) = f(x) + f(y)$, para todo $x, y \in \mathbb{R}$. Então, $f(x) = ax, \forall x \in \mathbb{R}$, é não decrescente para todo $x \in [0, +\infty)$ e $a \geq 0$.*

Claude E. Shannon, em seu clássico artigo de 1948, ao caracterizar, via axiomas, sua medida *entropia*, assumiu que ela dependia unicamente das probabilidades acerca dos eventos considerados. Sua interpretação levou em conta a realização de eventos probabilísticos. Antes de um experimento há uma incerteza da ocorrência de um evento; após a realização do experimento, uma quantidade de incerteza é eliminada, e tem-se, segundo Shannon, informação. Ou seja, incerteza *a priori* pode ser transformada em quantidade de informação *a posteriori*. Shannon indicou que algumas propriedades podem ser consideradas intuitivas, ou seja, adaptáveis ao senso comum em processos de comunicação (inseridos em seu modelo matemático de comunicação). Assim, ao considerar um simples evento, a quantidade

$$h(p) = -\log_2 p \quad p \in (0, 1] \quad (1)$$

representa uma medida de informação.

Seguindo as ideias de Shannon (não propriamente seus axiomas), é possível verificar que os seguintes axiomas são satisfeitos ao assumir a fórmula (1) (ACZÉL e DARÓCZY, 1975).

Axioma 3.1 *h é uma função não negativa: $h(p) \geq 0, \forall p \in (0, 1]$,*

Axioma 3.2 h é uma função aditiva: $h(pq) = h(p) + h(q)$, $\forall p, q \in (0, 1]$,

Axioma 3.3 h é normalizado: $h\left(\frac{1}{2}\right) = 1$.

Em termos intuitivos, o axioma 3.1 indica que, por mais que se tenha incerteza sobre um experimento, após sua realização a incerteza tende a diminuir e a quantidade de informação é positiva. Ou seja, sempre se ganha alguma informação após a realização de um experimento. Ainda em relação à intuição do axioma 3.1, pode-se dizer que $h(p) = 0$ representa o evento certo, aquele cuja probabilidade é 1. Nesse caso, a informação é nula.

À parte das interpretações intuitivas dos outros axiomas, é possível demonstrar que, ao assumir a fórmula (1), os axiomas 3.1, 3.2 e 3.3 se tornam verdadeiros. Na sequência demonstra-se o oposto, isto é, que, ao se assumirem como hipótese os axiomas 3.1, 3.2 e 3.3, deduz-se a fórmula (1).

Teorema 3.6 (SMALL, 2007; ACZÉL, DARÓCZY, 1975). A fórmula $h(p)$, para $p \in (0, 1]$, é a única dedução possível com base nos axiomas 3.1, 3.2 e 3.3. Assim, $h(p)$ indica a medida de informação mais geral, não negativa e aditiva produzida por um simples evento.

Demonstração: Considera-se, para que a equação de Cauchy (Teorema 3.4) seja satisfeita, que $f(x) = h(2^{-x})$. Assim, $f(x + y) = h(2^{-(x+y)}) = h(2^{(-x)+(-y)}) = h(2^{-x} \cdot 2^{-y})$, que pelo axioma 3.2 se tem que: $h(2^{-x}) + h(2^{-y}) = f(x) + f(y)$.

Portanto, com o auxílio do axioma 3.2, a fórmula de Cauchy é válida. Assim, os Teoremas 3.4 e 3.5 indicam que a função contínua, não decrescente f , que satisfaz as condições de Cauchy, para algum $a \in \mathbb{R}$, é do tipo: $f(x) = ax$.

Logo, existe $a \in \mathbb{R}$ tal que $h(2^{-x}) = ax$. Portanto, sendo $p = 2^{-x}$ e $q = 2^{-y}$, tem-se que $x = -\log_2 p$ e $y = -\log_2 q$. Assim,

$$h(p) = h(2^{-x}) = a \cdot x = a \cdot (-\log_2 p) = -a \log_2 p \quad \forall p \in (0, 1].$$

O mesmo vale para

$$h(q) = h(2^{-y}) = a \cdot y = a \cdot (-\log_2 q) = -a \log_2 q \quad \forall q \in (0, 1].$$

Nota-se que, no Teorema 3.4, $f(x) = ax$, para algum $a \in \mathbb{R}$, e para todo $x \in [0, +\infty)$. Já, em $h(p) = -\log_2 p$, o domínio limita-se a todo $p \in (0, 1]$. De fato, caso se tenha $x \rightarrow +\infty$, então $p \rightarrow 0$, enquanto se $x \rightarrow 0$, então $p \rightarrow 1$. Daí, ao assumir o axioma 3.2, tem-se que:

$$h(p \cdot q) = h(2^{-x} \cdot 2^{-y}) = h(2^{-x}) + h(2^{-y}) = -a \log_2 p - a \log_2 q = h(p) + h(q).$$

Portanto, $h(u) = -a \log_2 u$, $u \in (0, 1]$. Ao considerar o axioma 3.3, tem-se que:

$$h\left(\frac{1}{2}\right) = -a \log_2 \frac{1}{2} \Rightarrow a = 1.$$

Portanto, existe $a = 1$, e daí tem-se que, um logaritmo na base 2, $h(p) = -\log_2 p$.

Nota-se que, sem o axioma 3.3, $h(p)$ não precisaria ser um logaritmo na base 2, pois seria bem possível estabelecer que $f(x) = h(10^{-x})$ ou $f(x) = h(e^{-x})$, por exemplo. Mas, ao se colocar o axioma 3.3, tem-se a única fórmula $h(p) = -\log_2 p$.

Em Teoria da Informação as bases do logaritmo na fórmula de entropia que não sejam logaritmo na base 2 recebem outros nomes. Se o logaritmo é escrito na base 2, então a unidade de medida é *bits*; se o logaritmo é escrito na base 10, então a unidade de medida é *decits* (ou *Hartley*); e se a base é e , a unidade é *nats*.

No gráfico a seguir, em que se representa a função informação $I(p) = \log \frac{1}{p}$ ou, de modo equivalente, $I(p) = -\log p$, visualiza-se que, para $0 < p \leq 1$, quanto mais incerto (mais p tende a 0), mais se tem de aumento de informação. Por outro lado, quanto mais certo é o evento (p tende a 1), menos informação se tem. Isso justifica, pela intuição, uma das leituras do porquê da presença do sinal de “menos” na fórmula $h(p) = -\log p$. Dessa forma, $I(p)$, que é o caso particular de um único evento, representa a simples intuição de que quanto mais se tem de “liberdade de escolha”, mais se tem de “informação”. Ou seja, quanto mais provável, menos informação, quanto menos provável, mais informação. Isso, conforme exposto no parágrafo anterior, não depende da base do logaritmo.

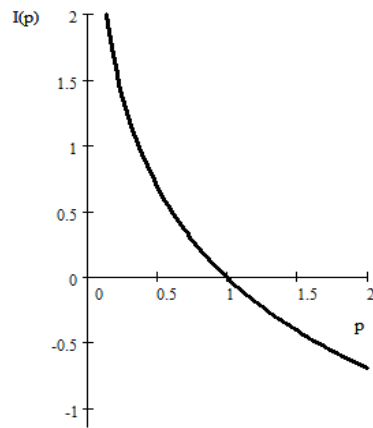


Figura 1. Informação $I(p) = -\log p$ de um único evento¹¹.

¹¹ Elaboração dos autores (2021).

Uma medida de informação: a entropia de Shannon

Em seu artigo, nas páginas 392 e 393 (SHANNON, 1948), Shannon expõe três propriedades que julga serem suficientes para deduzir sua fórmula de entropia de forma única. São elas:

- “1. H deve ser contínua em p_i .
2. Se todos os p_i forem iguais, $p_i = \frac{1}{n}$, então H deve ser uma função de n monotônica crescente. Com eventos igualmente prováveis, há mais escolha, ou incerteza, quando há mais eventos possíveis.
3. Se uma escolha for decomposta em duas escolhas sucessivas, a H original deve ser a soma ponderada dos valores individuais de H (...)” (SHANNON, 1948, p. 392-393, tradução nossa)

Logo em seguida Shannon escreve: “A única H satisfazendo as três hipóteses acima é da forma: $H = -K \sum_{i=1}^n p_i \log p_i$, onde K é uma constante positiva” (SHANNON, 1948, p. 392-393, tradução nossa).

Esse resultado é estabelecido, segundo Shannon, no apêndice 2 de seu artigo, páginas 419 e 420 (SHANNON, 1948). Logo após a exibição da função H , na página 393, Shannon faz o comentário que é *um ponto-chave* de discussão deste artigo, qual seja, sobre as demonstrações axiomáticas acerca de sua fórmula de entropia. Shannon escreve a frase: “This theorem, and the assumptions required for its proof, are in no way necessary for the present theory” (Shannon, 1948, p. 393). E completa no mesmo parágrafo: “The real justification of these definitions, however, will reside in their implications.” Entende-se que Shannon esclareceu o objetivo de seu trabalho, qual seja, melhorias nos processos de comunicação herdados da telegrafia, mas abriu espaço também para que outros caminhos, digam-se sistemas axiomáticos, pudessem ser abordados. Logo na sequência de seu artigo, no parágrafo seguinte, mesma página, ele introduz a frase: “Play a central role in information theory as measures of information.” Ocorre desse modo o nome *Information Theory*.

O artigo de Shannon causou impacto na comunidade acadêmica, sendo que seu modelo foi utilizado em muitas áreas, num amplo espectro de atuação. Isso não significa, como se mostra na literatura (AFTAB et al., 2001), que em todas as áreas se aplicavam realmente os conceitos desenvolvidos por Shannon. O problema residia nas palavras “informação” e “comunicação”, que, devido a plasticidade de leituras, favorecia aplicações em áreas diversas¹². Muitos autores entenderam com correção as ideias de Shannon e prontamente iniciaram desenvolvimentos e divulgação de textos em Teoria da Informação. É o caso, por exemplo, dos livros de Fazlollah M. Reza, *An Introduction to Information Theory* (REZA, 1961), primeira edição em 1961, e Robert B. Ash, *Information Theory* (ASH, 1965), primeira edição em 1965, que expõem a Teoria de Shannon com clareza.

¹² No modelo de Shannon, não somente as palavras “informação” e “comunicação”, presentes em seu modelo, têm amplo espectro de leituras, pode-se dizer o mesmo também de “canal”, “ruído”, “redundância”, “emissor”, “receptor” e “códigos”.

Em seu livro *Information Theory*, Robert Ash expõe a axiomática de Shannon e deduz a fórmula $H(p_1, p_2, \dots, p_M) = -C \sum_{i=1}^M p_i \log p_i$ de entropia de Shannon¹³ fundamentada nos seguintes axiomas (ASH, 1965, p. 8):

Axioma 4.1 $H\left(\frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M}\right) = f(M)$ é uma função monotonamente crescente em M ($M = 1, 2, \dots$).

Axioma 4.2 $f(ML) = f(M) + f(L)$ $M, L = (1, 2, \dots)$.

Axioma 4.3 $H(p_1, p_2, \dots, p_M) = H(p_1 + p_2 + \dots + p_r, p_{r+1} + \dots + p_M) +$

$$(p_1 + p_2 + \dots + p_r)H\left(\frac{p_1}{\sum_{i=1}^r p_i}, \dots, \frac{p_r}{\sum_{i=1}^r p_i}\right) +$$

$$(p_{r+1} + p_{r+2} + \dots + p_M)H\left(\frac{p_{r+1}}{\sum_{i=r+1}^M p_i}, \dots, \frac{p_M}{\sum_{i=r+1}^M p_i}\right).$$

Axioma 4.4 $H(p, 1 - p)$ é uma função contínua de p .

Pode-se inferir que R. B. Ash indica a fertilidade do tratamento axiomático na dedução de uma *medida de incerteza*, haja vista que o nome “entropia” é indicado numa seção de *notas e observação*: “Os axiomas para a medida da incerteza dados no texto são essencialmente aqueles de Shannon (1948). (...) A quantidade $H(X)$, a qual nos referimos como a “incerteza de X ”, também tem sido chamada de “entropia” ou “entropia de comunicação” de X ” (ASH, 1965, p.24, tradução nossa)

Entropia de Shannon: abordagem de A. Khinchin

No prefácio do livro de Khinchin, *Mathematical Foundations of Information Theory*, há um comentário em que se diz dos aspectos teóricos, matemática pura, diga-se de passagem, em oposição aos desenvolvimentos fundamentados na operacionalidade, *grosso modo*, pragmatismo: “É muito natural que Shannon e seus primeiros discípulos, cujo objetivo básico era obter resultados práticos, não foram capazes de prestar atenção suficiente a essas dificuldades matemáticas no começo” (KHINCHIN, 1957, p. 30, tradução nossa).

Uma leitura que se pode fazer é que “teoria” e “prática” caminham juntas e se complementam. Desenvolvimentos teóricos propiciam novas tecnologias, que indicam novos problemas, os quais, por sua vez, retornam na busca por mais desenvolvimentos teóricos. Vale citar algumas linhas escritas por A. Khinchin:

¹³ Uma demonstração detalhada (com viés didático) da fórmula de entropia de Shannon pode ser encontrada em Magossi e Barros (2021).

“Embora o estudo de entropia tenha, de fato, evoluído num importante e interessante capítulo da teoria geral de probabilidade, até agora está faltando uma apresentação dele nesse contexto mais geral. (...) Entretanto, o tratamento de Shannon nem sempre é suficientemente completo e matematicamente correto, (...) Não há dúvidas de que, nos anos vindouros, o estudo de entropia tornar-se-á uma parte integrante da teoria de probabilidade; o trabalho que tenho feito parece-me ser um estágio necessário no desenvolvimento desse estudo.” (KHINCHIN, 1957, p.1-2, tradução nossa).

Na sequência, em seu artigo, Khinchin expõe os axiomas que considera relevantes para que a função de incerteza seja deduzida de forma única.

As seguintes “propriedades”, digam-se axiomas, encontram-se no livro de Kinchin (KHINCHIN, 1957, p.9):

Axioma K1 Para qualquer n e para $\sum_{i=1}^n p_i = 1$, a função $H(p_1, p_2, \dots, p_n)$ assume seu maior valor para $p_i = \frac{1}{n}$ para $i = 1, 2, 3, \dots, n$.

Axioma K2 $H(AB) = H(A) + H_A(B)$.

Axioma K3 $H(p_1, p_2, \dots, p_n, 0) = H(p_1, p_2, \dots, p_n)$.

Por conseguinte, Khinchin demonstra o seguinte teorema:

Teorema 5.1 (KHINCHIN, 1957, pp.9-13) *Seja $H(p_1, p_2, \dots, p_n)$ uma função definida para qualquer inteiro n e para quaisquer valores p_1, p_2, \dots, p_n , tal que $p_k \geq 0$ para $(k = 1, 2, 3, \dots, n)$ e $\sum_{k=1}^n p_k = 1$. Considera-se o caso em que H é uma função contínua com relação a seus argumentos e que satisfaz os axiomas K1, K2 e K3. Então,*

$$H(p_1, p_2, \dots, p_n) = -\lambda \sum_{k=1}^n p_k \log p_k. \quad (2)$$

No livro de Khinchin há o artigo *The Concept of Entropy in The Theory of Probability*, originalmente publicado em russo em *Uspekhi Matematicheskikh Nauk*, Vol. VII, nº 3, p. 3-20, 1953 (KHINCHIN, 1957).

Medidas de informação: caracterizações

Após a exposição no artigo de Shannon e o consequente questionamento matemático sobre a operacionalização como força propulsora (CSISZÁR, 2008; ACZÉL e DARÓCZY, 1975; EBANKS et al., 1998) e também do exposto em Khinchin (KHINCHIN, 1957), outros

desenvolvimentos axiomáticos surgiram na comunidade acadêmica. Vale citar um trecho de Pl. Kannappan (KANNAPPAN, 2009):

“Uma questão fundamental na teoria da informação é como medir a quantidade de informação, em outras palavras, como definir informação. Várias abordagens são possíveis. Uma delas é uma abordagem pragmática, que começa a partir de certo problema particular de teoria da informação e aceita como medidas da quantidade de informação as quantidades que se apresentam na solução. Outra abordagem deve ser descrita como a abordagem axiomática e postulacional, na qual se inicia com certas propriedades que uma medida de informação razoável deve possuir e, depois disso, é uma questão puramente matemática de determinar todas as expressões que possuem esses postulados.” (KANNAPPAN, 2009, p. 403, tradução nossa)

Torna-se relevante mostrar que propriedades tais como *Continuidade, Simetria, Propriedade de Valor Máximo, Aditividade*, entre outras, são satisfeitas pela medida de informação $H(p_1, p_2, \dots, p_n)$, diga-se entropia, conforme consta nos livros de Aczél (1969), Reza (1961, p.80) e Rényi (1961), para citar alguns. Vale citar o prefácio do livro *On Measures of Information and Their Characterizations*, de J. Aczél e Z. Daróczy (1975): “Nesse livro vamos lidar com medidas de informação (as mais importantes chamadas de entropias), suas propriedades e, reciprocamente, com questões a respeito de quais dessas propriedades determinam medidas de informação conhecidas, e quais são as fórmulas mais gerais que satisfazem exigências razoáveis sobre medidas de informação práticas” (ACZÉL, DAROCZY, 1975, p.xi, tradução nossa).

A essência desses desenvolvimentos, entende-se, reside nas investigações acerca das possíveis caracterizações de medidas de informação. Com esse viés, a fórmula de entropia de Shannon pode ser descrita (ACZÉL e DAROCZY, 1975) como:

Definição 6.1 *Sejam, para $n = 1, 2, 3, \dots$,*

$$\Delta_n = \left\{ (p_1, p_2, \dots, p_n) : 0 < \sum_{k=1}^n p_k \leq 1, p_k \geq 0, k = 1, 2, \dots, n \right\},$$

conjuntos de distribuições de probabilidades finitas, possivelmente incompletas. A entropia de Shannon é a sequência de funções $H_n : \Delta_n \rightarrow \mathbb{R}$, ($n = 1, 2, \dots$) definida por

$$H_n(p_1, p_2, \dots, p_n) = \frac{\sum_{k=1}^n L(p_k)}{\sum_{k=1}^n p_k},$$

em que

$$L(x) = \begin{cases} -x \log_2 x, & \text{se } x \in (0, 1] \\ 0 & \text{se } x = 0 \end{cases}.$$

De modo análogo às considerações de Khinchin sobre a alimentação recíproca entre desenvolvimentos matemáticos e desenvolvimentos tecnológicos, defende-se, no presente artigo, que as iniciativas em produzir mais e mais axiomáticas para a fórmula de Shannon são importantes, pois, de uma forma ou de outra, denotam plasticidade e completude ao conceito de *medida de informação*. Entende-se que propriedades “naturais” e “essenciais” não deixam de ser factíveis sempre que a operacionalidade, a praticidade e a consistência probabilísticas estiverem presentes. No entanto, de modo análogo à era do rigor em matemática, em que se revisaram os fundamentos, entende-se que a releitura desses fundamentos nas caracterizações de medida de informação também seja um caminho factível que pode produzir bons resultados.

Com base na vasta literatura sobre o assunto em questão, é possível *inferir* que o desenvolvimento do conceito de *entropia*, seguindo a ideia de medida de informação, assemelha-se em muito ao desenvolvimento de outros conceitos, clássicos na história da matemática, como, por exemplo, o conceito de *computabilidade* (máquinas de Turing) e o conceito de *continuum* (para o sistema de números reais).

Nos três conceitos, *continuum*, *computabilidade* e *entropia*, os desenvolvimentos iniciaram-se com intuição e sedimentaram-se com base no rigor matemático. O conceito de *continuum* veio, desde a Grécia antiga, das inúmeras discussões sobre quantidades infinitamente pequenas (LÜTZEN, 2003; JAHNKE, 2003; SCHUBRING, 2005; ALEXANDER, 2014). O conceito de computabilidade, a contraparte formal do conceito de algoritmo, emergiu após D. Hilbert ter divulgado seu *Décimo problema* no Segundo Congresso Internacional de Matemática ocorrido em 1900, em Paris, França. Hilbert, de certa forma, percebeu que a intuição plausível do conceito de algoritmo, utilizada até então, naquela época, poderia ser expandida na solução de outros problemas (REID, 1986, 2013)¹⁴ Coube a Alonzo Church e Alan Turing, em trabalhos distintos, em 1936, indicar um conceito matemático que poderia ser aceito como o conceito de algoritmo. O conceito de λ -*calculus*, de Church (CHURCH, 1936), e o conceito de *Máquinas de Turing*, de Turing (TURING, 1937), são ambos equivalentes em termos de classe de funções “computáveis” produzidas¹⁵. A tese de Church-Turing indica que não se tem ainda um *ponto final* nessa questão¹⁶. E por fim, o conceito de entropia, advindo de Clausius, investigado na mecânica estatística, discutido por Shannon sob a ótica de comunicação entre “máquinas” e expandido como *Caracterização de Medidas de Informação*. Nota-se que em todos os três conceitos ainda há desenvolvimentos a serem elaborados, mesmo que, em tese, eles estejam bem solidificados.

Uma simples busca em base de dados, assim como em sites como *Scopus*[®] (*Elsevier BV*) e *Web of Science*[™] (*Thomson Reuters*), mostrará o quanto a palavra “entropia” é utilizada. A pergunta é: todas essas utilizações remetem ao conceito de medida de

¹⁴ Os 23 problemas de Hilbert podem ser encontrados em *Mathematical problems* (HILBERT, 1902).

¹⁵ A classe de funções produzidas com base no conceito λ -*calculus* é equivalente à classe de funções produzidas por *Máquinas de Turing* (CUTLAND, 1980; DAVIS, 1985; DAVIS et al., 1994).

¹⁶ Sobre Tese de Church-Turing, ver SIPSER (1997) e ROGERS (1967).

informação, tal como em Shannon, ou ao conceito de entropia de Clausius, ou advindos da mecânica estatística, ou são simplesmente caracterizações de medidas de informação? Estima-se que o esclarecimento sobre o que seja uma *medida de informação* e o que seja *uma extensão do conceito de entropia de Shannon* configure um ponto importante para novos desenvolvimentos e pesquisas, pois, caso contrário, pode-se entrar no mesmo círculo das publicações, ocorridas logo após o lançamento do artigo de Shannon, em que tudo que dizia respeito à comunicação, no sentido extensivo, incluía-se na Teoria da Informação (e de Comunicação) (AFTAB et al., 2001).

Outras axiomáticas

Caracterizar medidas de informação, levando em conta a existência da entropia de Shannon, é descrever axiomas e propriedades com objetivo de deduzir, analisar e investigar outras medidas de informação além da medida exposta pela fórmula

$$H(p_1, p_2, \dots, p_n) = -C \sum_{i=1}^n p_i \log p_i. \quad (3)$$

Entre elas, pode-se citar a seguinte lista, (REZA, 1961. pp.80-83; ACZÉL, 1969; ACZÉL e DAROCZY, 1975), sem pretensões de completude, não exaustiva, de propriedades (axiomas)¹⁷.

Axioma 6.1 $H^1\left(\frac{1}{2}\right) = 1$.

Axioma 6.2 $H^2\left(\frac{1}{2}, \frac{1}{2}\right) = 1$.

Axioma 6.3 $H^n(p_1, p_2, \dots, p_n) = H^n(p_{k(1)}, p_{k(2)}, \dots, p_{k(n)})$, em que $k(1), k(2), \dots, k(n)$ é qualquer permutação de $1, 2, 3, \dots, n$.

Axioma 6.4 $H^2(1, 0) = H^2(0, 1)$.

Axioma 6.5 $H^n(0, 0, 0, \dots, 0, 1) = 0$.

Em particular para $n = 1$ ou $n = 2$ tem-se:

Axioma 6.6 $H^1(1) = 0$.

¹⁷ Nessa lista (ACZÉL, 1969) $H^n(p_1, p_2, \dots, p_n)$ refere-se à entropia de n eventos mutuamente exclusivos com probabilidades p_1, p_2, \dots, p_n , para $k = 1, 2, \dots, n$, tais que $p_k \geq 0, 1 \leq k \leq n$ e $0 < p_1 + p_2 + \dots + p_n \leq 1$. No caso de $p_1 + p_2 + \dots + p_n = 1$, em um sistema completo de eventos, escreve-se $K^n(p_1, p_2, \dots, p_n)$.

Axioma 6.7 $H^2(0, 1) = 0$.

Axioma 6.8 $H^{n+1}(p_1, p_2, \dots, p_n, 0) = H^n(p_1, p_2, \dots, p_n)$.

Axioma 6.9

$$K^{mn}(p_1 q_{11}, p_1 q_{12}, \dots, p_1 q_{1n}, p_2 q_{21}, p_2 q_{22}, \dots, p_2 q_{2n}, \dots, p_m q_{m1}, p_m q_{m2}, \dots, p_m q_{mn}) = K^m(p_1, p_2, \dots, p_m) + \sum_{j=1}^m p_j K^n(q_{j1}, q_{j2}, \dots, q_{jn}).$$

Em particular ($n > m > 1, q_{jj} = 1, q_{jl} = \dots = q_{j,j-1} = q_{j,j+1} = \dots = q_{jn} = 0$ para $j = 1, 2, \dots, m-1$ e $q_{m1} = q_{m2} = \dots = q_{m,m-1} = 0$, considerando os axiomas 6.3, 6.5 e 6.8):

Axioma 6.10 $K^n(p_1, p_2, \dots, p_{m-1}, p_m q_{mm}, p_m q_{m,m+1}, \dots, p_m q_{mn}) = K^m(p_1, p_2, \dots, p_m) + p_m K^{n-m+1}(q_{mm}, q_{m,m+1}, \dots, q_{mn})$.

Ou, mais especificamente ($n = m + 1, q_{mm} = q$, em particular $n = 3$).

Axioma 6.11: ($n = m + 1, q_{mm} = q$): $K^{m+1}(p_1, p_2, \dots, p_{m-1}, p_m q, p_m(1 - q)) = K^m(p_1, p_2, \dots, p_m) + p_m K^2(q, 1 - q)$.

($n = 3$): $K^3(p_1, p_2 q, p_2(1 - q)) = K^2(p_1, p_2) + p_3 K^2(q, 1 - q)$.

Axioma 6.12 $H^n(p_1, p_2, \dots, p_n) \geq 0$, em particular para $n = 1$ tem-se:

Axioma 6.13 $H^1(p_1) \geq 0$.

Axioma 6.14 $K^n(p_1, p_2, \dots, p_n) \leq K^n\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$, que juntamente com 6.8 implica 6.15.

Axioma 6.15 $K^n\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \leq K^{n+1}\left(\frac{1}{n+1}, \frac{1}{n+1}, \dots, \frac{1}{n+1}\right)$.

Axioma 6.16 $K^n(p_1, p_2, \dots, p_n) = \sum_{k=1}^n (p_1 + p_2 + \dots + p_k) h\left(\frac{p_k}{p_1 + p_2 + \dots + p_k}\right)$.

Axioma 6.17 $K^2(p_1, p_2) = h(p_2)$. Caso particular de 6.16 quando $n = 2$.

Axioma 6.18 $K^3(p_1, p_2, p_3) = (1 - p_3)h\left(\frac{p_2}{1 - p_3}\right) + h(p_3)$. Caso particular de 6.16 quando $n = 3$.

Axioma 6.19 $p_k \rightarrow H^n(p_1, p_2, \dots, p_n)$, para $k = 1, 2, \dots, n$, são contínuas em $[0, 1]$.

Axioma 6.20 $p_2 \rightarrow h(p_2) = K^2(1 - p_2, p_2)$ é contínua.

Axioma 6.21 $\lim_{p_2 \rightarrow 0} h(p_2) = \lim_{p_2 \rightarrow 0} K^2(1 - p_2, p_2) = K^1(1) = 0$, desde que $H^2(0, 1) = 0$.

Axioma 6.22 $K^2(p_1, p_2) = h(p_2)$ é crescente em $(0, \frac{1}{2})$.

Axioma 6.23 $K^2(p_1, p_2) = h(p_2)$ é monotônica em $(0, \frac{1}{2})$.

Axioma 6.24 $K^2(p_1, p_2) = h(p_2)$ é Lebesgue integrável em $[0, 1]$.

Axioma 6.25 $K^2(p_1, p_2) = h(p_2)$ é Lebesgue mensurável em $(0, 1)$.

Conforme consta em Aczél (1969), Shannon lança mão dos axiomas 6.2, 6.3, 6.10, 6.15 e 6.19 para deduzir a fórmula

$$K^n(p_1, p_2, \dots, p_n) = - \sum_{k=1}^n p_k \log p_k. \quad (4)$$

Por sua vez, Khinchin (1957) utiliza os axiomas: 6.2, 6.3, 6.8, 6.9, 6.12, 6.14 e 6.19 para também deduzir a fórmula (4). Ainda segundo Aczél (1969), D. K. Faddeev (FADDEEV, 1956) reduz os axiomas para 6.2, 6.3, 6.11 e 6.20 para demonstrar a fórmula (4). Com base nos trabalhos de Faddeev, H. Tverberg (TVERBERG, 1958) propõe uma nova axiomática, trocando o axioma 6.20 pelo axioma 6.24; já D. G. Kendall (KENDALL, 1964) substitui o axioma 6.20 pelo axioma 6.22. P. M. Lee (LEE, 1964) substitui o axioma 6.20 pelo axioma 6.25. Conforme exposto em Aczél (1969) e também em Aczél e Daróczy (1975) e Ebanks et al (1998), há outras axiomatizações (e caracterizações de medidas de informação) em que é possível analisar, deduzir e investigar propriedades acerca da fórmula (4).

Essa breve exposição tem também como objetivo alertar para a riqueza de demonstrações da fórmula de entropia de Shannon e a importância de novas e diversificadas abordagens axiomáticas com vistas, evidentemente, ao desenvolvimento de novas tecnologias e solidificação matemática.

Intuição operacional

No que se segue, com fins didáticos, expõem-se algumas considerações acerca da motivação intuitiva, natural-essencial, sobre algumas propriedades, de algumas equações funcionais, nas caracterizações de medidas de informação.

P1: Continuidade. Pequenas alterações nas probabilidades p_k devem produzir pequenas alterações na fórmula (3) de entropia H . Se, por exemplo, num lançamento de uma moeda, o seguinte esquema se apresenta

$$\begin{pmatrix} C & K \\ 0.95 & 0.05 \end{pmatrix},$$

então a probabilidade de sair *cara* (C) é muito maior do que a probabilidade de sair *coroa* (K). Há pouca incerteza, surpresa, nesse esquema. Exigir continuidade é exigir que, caso pequenas alterações ocorram, pequenas alterações ocorrerão com o resultado do experimento, com a informação obtida como resultado de eliminação de incerteza.

P2: Simetria. A incerteza, ou surpresa sobre o resultado do experimento, não pode se alterar, caso, no exemplo do lançamento de uma moeda, troque-se *cara* por *coroa*. Ou seja, a função de incerteza deve ser independente de seus representantes. Isso está em consonância com a noção intuitiva de incerteza. Assim, $H(p_1, p_2, \dots, p_n)$ não deve se alterar quando se altera a ordem de seus representantes p_1, p_2, \dots, p_n . Como caso particular, $H(p_1, p_2, \dots, p_n) = H(p_2, p_1, \dots, p_n)$.

P3: Propriedade extrema. O maior valor da função de incerteza deve ser aquele obtido quando todos os eventos são equiprováveis. Ou seja, como caso particular, no lançamento de uma moeda, a maior incerteza e, por conseguinte, maior quantidade de informação, deve advir do caso em que a moeda é “honestá”. Assim,

$$\max\{H(p_1, p_2, \dots, p_n)\} = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right).$$

P4: Grupamento. Particionar os eventos em subeventos não pode alterar a incerteza do sistema. Conclui-se então que a incerteza associada a um evento geral se mantém a mesma ao dividir o evento em subeventos.

P5: Adição de Conjunto Nulo. Adicionar um conjunto nulo ao conjunto completo de eventos não altera a incerteza do sistema.

$$H(p_1, p_2, \dots, p_n, 0) = H(p_2, p_1, \dots, p_n).$$

P6: Aditividade. Para eventos independentes a função H deve satisfazer a equação funcional $H(XY) = H(X) + H(Y)$. Ou seja, a incerteza produzida por dois eventos independentes deve ser a soma das incertezas produzidas pelos eventos de modo independente. A incerteza associada ao lançamento de um dado e uma moeda deve ser a soma das incertezas produzidas pelo lançamento de um dado e de uma moeda, independentemente.

Essas são apenas e somente algumas das muitas propriedades que revelam intuição e operacionalidade. As diferentes caracterizações baseiam-se em diferentes abordagens na escolha dos axiomas que fundamentam a demonstração da fórmula (3), tal como ocorre em Feinstein (1958), Khinchin (1957), Shannon (1948), Schutzenberger (1955)¹⁸ etc.

¹⁸ Conforme exposto em Reza (1961), Schutzenberger (1955) mostra que há outras condições para obter a fórmula $H(X)$ além daquelas expostas no modelo de Shannon-Wiener. Ele introduz uma discussão sobre os trabalhos de Fisher (1925). Indica também que a expressão funcional de Shannon-Wiener é uma das mais simples de todas as formas possíveis (WIENER, 1948)

Em Moser e Chen (2012) tem-se a apresentação da fórmula de entropia de modo um pouco diferente daquela exposta em Khinchin, mas comenta-se sobre a justificativa matemática de Shannon e do fato de ele não ter considerado a demonstração como algo importante: “Embora Shannon tenha fornecido uma justificativa matemática (...), ele não a considerou muito importante” (MOSER e CHEN, 2012, p.111, tradução nossa).

Em Csiszár e Körner (2011), há também uma outra abordagem matemática em que se explora a matemática presente na fórmula (3) em que, segundo exposto em Ash (1965), se diz que o conjunto mais fraco de axiomas é exposto em Lee (1964). Como já exposto nesse artigo, Khinchin provou, nos anos 1953 e 1956, após a publicação do artigo de Shannon em 1948, que a fórmula (3) é única, sob certas condições. F. Reza alerta sobre isso: “Tais tratamentos apareceram no trabalho de Feinstein, Khinchin, Shannon, Schutzenberger e outros.” (REZA, 1961, p.81, tradução nossa).

Ou seja, outros autores também investigaram o conceito de entropia de modo análogo ao esquema matemático exposto por A. I. Khinchin. Os livros de Aczél e Daróczy (1975) e Ebanks et al. (1998), expõem sistemas axiomáticos para caracterizações de medida de informação em abordagens muito extensivas.

Conclusões

Exceto por algumas reflexões relacionadas aos conceitos de medida de informação e entropia, e do alinhamento histórico dos fatos relativos à caracterização axiomática da fórmula de entropia de Shannon, o que se apresentou neste texto é clássico na literatura.

O objetivo é indicar uma leitura que, entende-se, possa facilitar a compreensão de um assunto não somente interessante, mas também importante para desenvolvimentos matemáticos e tecnológicos. A informação pode ser vista na Teoria da Informação como redução da incerteza, ou seja, há uma dualidade entre a medida de quantidade de informação *a posteriori* (após a observação do experimento) e incerteza *a priori* (antes da observação do experimento). Quanto maior for a incerteza, mais há para “retirar”, isto é, mais há de informação. Essa explicação remonta a Shannon, quando propôs um modelo matemático para sistematizar o tratamento de informação na transmissão de mensagens de um ponto a outro. Com isso, ele investiga, e formula, teoremas que se adaptam de modo intuitivo, e com resultados práticos, aos modelos de comunicação, e extensivamente àqueles modelos que se assemelham, em termos formais, ao processo de transmissão de mensagens por meio de uma codificação no emissor, passando por um canal, na presença de ruídos, e sendo decodificadas na recepção. É claro que sua fórmula de entropia tem um apelo intuitivo operacional forte, mas a investigação matemática, de sua, digamos, contraparte matemática, “medida de informação”, não deixa de ser importante, haja vista que além de indicar caminhos factíveis sobre a existência ou não de teoremas, abre espaço também para a intersecção com outras áreas, tais como Teoria de Equações Funcionais, Teoria de Medida, Teoria de Probabilidade, Inteligência Artificial, Ciência dos Dados etc., para citar apenas algumas delas.

Um outro ponto de suma importância é a caracterização de medidas de informação, possível, de várias formas, ao estipular axiomas fundamentais para, via dedução, obter como teorema uma determinada medida de informação. Desse modo, entende-se que a investigação matemática, via sistemas de axiomas, da entropia de Shannon, ou de quaisquer outras

medidas de informação, como proposto neste artigo, é relevante não somente para clarear a diferença entre entropia na Física, Mecânica Estatística e na Teoria da Informação, mas também para alavancar desenvolvimentos tecnológicos, tal como sucedeu com os fundamentos da Análise Matemática, na época em que A. Cauchy desenvolve o conceito de limites. Tal como sucedeu também com algoritmos e máquinas de Turing¹⁹. Segundo as ideias de Thomas Kuhn (1994), com o conceito de limites houve uma mudança de paradigmas na matemática (GRABINER, 1981); seria então factível a conjectura de que com o conceito de medida de informação, no sentido exposto nesse artigo, haveria também uma mudança de paradigmas tal como apresentado por Thomas Kuhn?

Para finalizar, defende-se, no presente artigo, que a *entropia de Shannon* e todas as axiomáticas sobre medidas de informação são relevantes não só para auxiliar na solidificação da matemática, como também para que essas estruturas constituam patamares sólidos para futuros desenvolvimentos matemáticos e tecnológicos.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Bibliografia

ACZÉL, J. **Lectures on functional equations and their applications**. New York, NY: Academic Press, 1966.

ACZÉL, J. On different characterizations of entropies. In: BEHARA, M.; KRICKEBERG, K.; WOLFOWITZ, J. (Eds.). **Probability and Information Theory**. Lecture Notes in Mathematics, v. 89. Springer, Berlin, Heidelberg, p. 1-6, 1969.

ACZÉL, J. Measuring information beyond communication theory: some probably useful and some almost certainly useless generalizations. **Information Processing & Management**, v. 20, n. 3, p. 383-395, 1984a.

ACZÉL, J. Measuring information beyond communication theory: why some generalized information measures may be useful, others not. **Aequationes Mathematicae**, v. 27, p. 1-19, 1984b.

ACZÉL, J. Characterizing information measures: approaching the end of an era. In: BOUCHON, B.; YAGER, R. R. (Eds.). **International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems**. Springer, Berlin, Heidelberg, p. 357-384, 1986.

¹⁹ Uma máquina de Turing é a contraparte formal da noção intuitiva de algoritmos (SIPSER, 1997).

ACZÉL, J.; DARÓCZY, Z. **On measures of information and their characterizations**. New York, NY: Academic Press, v. 115, 1975.

ACZÉL, J.; DHOMBRES, J. **Functional equations in several variables**. Cambridge, UK: Cambridge University Press, 1989.

ACZÉL, J.; FORTE, B.; NG, C. T. Why the Shannon and Hartley entropies are ‘natural’. **Advances in Applied Probability**, v. 6, n. 1, p. 131-146, 1974.

AFTAB, O.; CHEUNG, P.; KIM, A.; THAKKAR, S.; YEDDANAPUDI, N. **Information theory and the digital age: the structure of engineering revolutions**. Cambridge, MA: Massachusetts Institute of Technology, 2001. Disponível em: <http://web.mit.edu/6.933/www/Fall2001/Shannon2.pdf>, acesso em 22/05/2021, às 17:57.

AGRELL, E.; ALVARADO, A.; KSCHISCHANG, F. R. Implications of information theory in optical fibre communications. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, A.374: 20140438, disponível em <https://doi.org/10.1098/rsta.2014.0438>, acesso em 02/12/2020, às 10h00.

ALEXANDER, A. **Infinitesimal: a teoria matemática que mudou o mundo**. Rio de Janeiro, RJ: Jorge Zahar Editor Ltda., 2014.

ASH, R. B. **Information theory**. Dover Publications, New York, 1990. (Originalmente publicada pela Interscience Publishers, New York, 1965).

BAR-HILLEL, Y.; CARNAP, R. Semantic information. **The British Journal for the Philosophy of Science**, v. 4, n. 14, p. 147-157, 1953.

BEN-NAIM, A. **Discover entropy and the second law of thermodynamics: a playful way of discovering a law of nature**. Singapore: World Scientific Publishing Company, 2010.

BOTTAZZINI, U. **The higher calculus: a history of real and complex analysis from Euler to Weierstrass**. New York, NY: Springer-Verlag, 1986.

BRADLEY, R. E.; SANDIFER, C. E. **Cauchy’s cours d’analyse: an annotated translation**. New York, NY: Springer-Verlag, 2009.

BRESSOUD, D. M. **A radical approach to real analysis**. Washington, DC: Mathematical Association of America, v. 2, 2007.

BRILLOUIN, L. **Science and information theory**. New York, NY: Academic Press Inc., 1956.

CHURCH, A. A note on the Entscheidungs problem. **The Journal of Symbolic Logic**, v. 1, n. 1, p. 40-41, 1936.

CLAUSIUS, R. Über verschiedene für die anwendung bequeme formen der hauptgleichungen der mechanischen wärmetheorie. **Annalen der Physik**, v. 201, n. 7, p. 353-400, 1865.

CLAUSIUS, R. **Abhandlungen über die mechanische wärmetheorie**. Braunschweig, Germany: Friedrich Vieweg Und Sohn, 1867.

COVER, T. M. Broadcast channels. **IEEE Transactions on Information Theory**, v. 18, n. 1, p. 2-14, 1972.

COVER, T. M.; THOMAS, J. A. **Elements of information theory**. Hoboken, NJ: Wiley-Interscience, v. 2, 2006.

CSISZÁR, I. Axiomatic characterizations of information measures. **Entropy**, v. 10, n. 3, p. 261-273, p. 269, 2008.

CSISZÁR, I.; KÖRNER, J. **Information theory: coding theorems for discrete memoryless systems**. Cambridge, UK: Cambridge University Press, 2011.

CUTLAND, N. J. **Computability: an introduction to recursive function theory**. Cambridge, UK: Cambridge University Press, 1980.

DARÓCZY, Z. Generalized information functions. **Information and Control**, v. 16, n. 1, p. 36-51, 1970.

DAVIS, M. **Computability and unsolvability**. Mineola, NY: Dover Publications, 1985.

DAVIS, M.; SIGAL, R.; WEYUKER, E. J. **Computability, complexity, and languages: fundamentals of theoretical computer science**. Amsterdam, Netherlands: Elsevier, North Holland Publishing Co., v. 2, 1994.

DOOB, J. L. Review of C.E. Shannon's Mathematical Theory of Communication. **Mathematical Reviews**, v. 10, p. 133, 1949.

EBANKS, B.; SAHOO, P.; SANDER, W. **Characterizations of information measures**. Singapore: World Scientific Publishing Company, 1998.

ELLERSICK, F. W. A conversation with Claude Shannon. **IEEE Communications Magazine**, v. 22, n. 5, p. 123-126, 1984.

FADDEEV, D. K. On the concept of entropy of a finite probability scheme. Originalmente publicado em russo em **Uspekhi Matematicheskikh Nauk**, v. 11, n. 1, p. 227-231, 1956.

FEINSTEIN, A. **Foundations of information theory**. New York, NY: McGraw-Hill Book Company Inc., 1958.

FISHER, R. A. Theory of statistical estimation. **Proceedings of Cambridge Philosophy Society**, v. 22, n. 5, p. 700-725, 1925.

GRABINER, J. V. **The origin of Cauchy's rigorous calculus**. New York, NY: Dover Publications Inc., 1981.

GRATTAN-GUINNESS, I. Joseph Fourier and the revolution in mathematical physics. **IMA Journal of Applied Mathematics**, v. 5, n. 2, p. 230-253, 1969.

HAIRER, E.; WANNER, G. **Analysis by its history**. New York, NY: Springer-Verlag, 1996.

HAVRDA, J.; CHARVÁT, F. Quantification method of classification processes: concept of structural a -entropy. **Kybernetika**, v. 3, n. 1, p. 30-35, 1967.

HILBERT, D. Mathematical problems. **Bulletin of the American Mathematical Society**, v. 8, n. 10, p. 437-479, 1902.

INGARDEN, R.S., URBANIK, K. Information without probability. **Colloquium Mathematicum**, Vol. IX, 1962.

JAHNKE, H. N. **A History of analysis (history of mathematics)**. Providence, RI: American Mathematical Society, v. 24, 2003.

KANNAPPAN, P. **Functional equations and inequalities with applications**. New York, NY: Springer Science & Business Media, p. 404, 2009.

KENDALL, D. G. Functional equations in information theory. **Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete**, v. 2, n. 3, p. 225- 229, 1964.

KHINCHIN, A. I. **Mathematical foundations of information theory**. Translated by SILVERMAN, R. A.; FRIEDMAN, M. D. New York, NY: Dover Publications, 1957. (Originalmente publicado em russo em **Uspekhi Matematicheskikh Nauk**, v. 7, n. 3, p. 3-20, 1953 e v. 9, n. 1, p. 17-75, 1956).

KLINE, M. Les fondements des mathématiques. **La recherche**, v. 54, p. 200-208, 1975.

KUHN, T. **A estrutura das revoluções científicas**. São Paulo, SP: Editora Perspectiva, 1994.

LEE, P. M. On the axioms of information theory. **The Annals of Mathematical Statistics**, v. 35, n. 1, p. 415-418, 1964.

LÜTZEN, J. The foundation of analysis in the 19th century. In: JAHNKE, H. N. (Ed.). **A history of analysis**, American Mathematical Society, p. 155-195, 2003.

MAGOSSI, J. C.; PAVIOTTI, J. R. Incerteza em entropia. **Revista Brasileira de História da Ciência**, v. 12, n. 1, p. 84-96, 2019.

MAGOSSI, J. C.; BARROS, A. C. DA C. A entropia de Shannon: uma abordagem axiomática. **REMAT: Revista Eletrônica da Matemática**, v. 7, n. 1, p. e3013, 26 maio 2021.

MOSER, S. M.; CHEN, P. N. **A student's guide to coding and information theory**. Cambridge, UK: Cambridge University Press, 2012.

PIERCE, J. R. **An introduction to information theory-symbols signals and noise**. Dover Publications, Second Revised Edition, New York, 1980 (originally published in 1961 by Harper & Brothers).

REID, C. **Hilbert-Courant**. New York, NY: Springer-Verlag, 1986.

REID, C. **Courant**. New York, NY: Springer Science & Business Media, 2013.

RÉNYI, A. On measures of entropy and information. In: **Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics**, Berkeley, CA, USA, p. 547-561, 1961.

REZA, F. M. **An introduction to information theory**. New York, NY: McGraw-Hill, 1961.

ROGERS, H., Jr. **Theory of recursive functions and effective computability**. New York, NY: McGraw-Hill, 1967.

SCHUBRING, G. **Conflicts between generalization, rigor, and intuition: number concepts underlying the development of analysis in 17-19th century France and Germany**. New York, NY: Springer-Verlag, 2005.

SCHUTZENBERGER, M. P. On some measures of information used in statistics. In: **Proceedings of the Third Symposium on Information Theory**, London, UK, p. 18-25, 1955.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, p. 379-423 (July), 623-656 (October), 1948.

- SHANNON, C. E. The bandwagon. **IRE Transactions on Information Theory**, v. 2, n. 1, p. 3, 1956.
- SIPSER, M. **Introduction to the theory of computation**. Boston, MA: PWS Publishing Company, 1997.
- SMALL, C. G. **Functional equations and how to solve them**. New York, NY: Springer-Verlag, 2007.
- SPIVAK, D. I. **Category theory for the sciences**. Cambridge, MA: The MIT Press, 2014.
- TRIBUS, M.; McIRVINE, E. C. Energy and information. **Scientific American**, v. 225, n. 3, p. 179-190, 1971.
- TURING, A. M. On computable numbers, with an application to the Entscheidungs problem. **Proceedings of the London Mathematical Society**, v. 2, n. 1, p. 230-265, 1937.
- TVERBERG, H. A new derivation of the information function. **Mathematica Scandinavica**, v. 6, p. 297-298, 1958.
- VAN VLECK, E. B. The influence of Fourier series on the development of mathematics. **Science**, v. 39, n. 99, p. 113-124, 1914.
- VERDÚ, S. Empirical estimation of information measures: a literature guide. **Entropy**, v. 21, n. 8, 720, p. 2, 2019.
- VERDÚ, S.; McLAUGHLIN, S. W. **Information theory: 50 years of discovery**. Piscataway, NJ: IEEE Press, 2000.
- WEHRL, A. General properties of entropy. **Review of Modern Physics**, v. 50, n. 2, p. 221-260, 1978.
- WEHRL, A. The many facets of entropy. **Reports on Mathematical Physics**, v. 30, n. 1, p. 119-129, 1991.
- WIENER, N. **Cybernetics: or control and communication in the animal and the machine**. Cambridge, MA: The MIT Press, 1948.

José Carlos Magossi

Universidade Estadual de Campinas - UNICAMP, Faculdade de Tecnologia - FT, Divisão de Telecomunicações, Limeira, SP, Brasil.

e-mail: magossi@ft.unicamp.br

Pedro Henrique Camargo de Abreu

Mestre pelo Programa de Pós-Graduação em Tecnologia (PPGT) / UNICAMP / FT, Limeira, SP, Brasil.

e-mail: phcamargo1997@gmail.com

Antônio César da Costa Barros

Doutorando no Programa de Pós-Graduação em Tecnologia (PPGT) / UNICAMP / FT, Limeira, SP, Brasil

e-mail: cesar.matema@gmail.com

José Renato Paviotti

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) - BRASIL.

Mestre pelo Programa de Pós-Graduação em Tecnologia (PPGT) / UNICAMP / FT, Limeira, SP, Brasil.

e-mail: jose.renato@ifsp.edu.br