# Introdução à Teoria da Informação

Teoria da Informação - AULA 01 (Parte 1)
Prof$^{a}$. Verusca Severo

Universidade de Pernambuco
Escola Politécnica de Pernambuco

16 de junho de 2021

# Introdução

- A TEORIA DA INFORMAÇÃO lida com:
  - Sistemas de comunicação;
  - Transmissão de dados;
  - Criptografia;
  - Codificação;
  - Compressão.

# Introdução

**O que é Teoria da Informação?**

**O que é Teoria da Informação?**

- TEORIA

**O que é Teoria da Informação?**

- TEORIA
  - É o conhecimento descritivo puramente racional.
  - O substantivo theoría significa ação de contemplar, olhar, examinar, especular. Também pode ser entendido como forma de pensar e entender algum fenômeno a partir da observação.

**O que é Teoria da Informação?**

- INFORMAÇÃO

**O que é Teoria da Informação?**

- INFORMAÇÃO
  - É a resultante do processamento, manipulação e organização de dados, de tal forma que represente uma modificação no conhecimento do sistema que a recebe.

**O que é Teoria da Informação?**

- A Teoria da Informação procura explicar as modalidades de transferência das mensagens de um emissor a um receptor.

- A Teoria da Informação contempla uma abordagem matemática voltada para o estudo do armazenamento e manipulação de informação.

A Teoria da Informação é, sobretudo, uma teoria matemática que trata de três conceitos básicos:

- A medida da informação;
- A capacidade de um canal de comunicações transferir Informação;
- A codificação, como meio de utilizar os canais com toda a sua capacidade.

# Introdução

- **Teoria da Informação** é o nome da disciplina científica criada por Claude Shannon ao publicar em 1948 um dos mais importantes artigos na história da engenharia, intitulado *A Mathematical Theory of Communication*, que foi publicado no Bell System Technical Journal.

Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

## A Mathematical Theory of Communication

By C. E. SHANNON

### INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

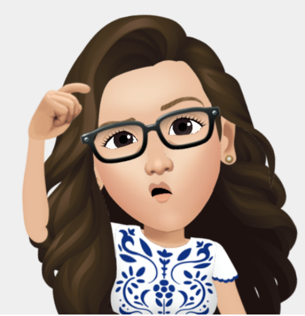
# Introdução

A Teoria da Informação tenta responder a perguntas do tipo:

- O que é informação? Como a medimos?
- Quais são os limites fundamentais à transmissão de informação?
- Quais são os limites fundamentais à extração de informação do meio ambiente?
- Como é que se deve projetar dispositivos ou equipamentos que se aproximem desses limites?
- Os dispositivos e equipamentos atuais aproximam-se, ou não, desses limites?

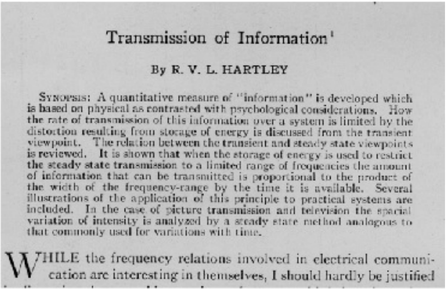
# Introdução

Como medir a **quantidade de informação** associada a uma mensagem?



# Introdução - Medida de informação de Hartley

- O único trabalho anterior ao de Shannon, do qual se tem notícia, foi desenvolvido por Hartley e intitula-se *Trasmission of Information*, Bell System Technical Journal, em 1928.



Transmission of Information[1]

By R. V. L. HARTLEY

SYNOPSIS: A quantitative measure of "information" is developed which is based on physical as contrasted with psychological considerations. How the rate of transmission of this information over a system is limited by the distortion resulting from storage of energy is discussed from the transient viewpoint. The relation between the transient and steady state viewpoints is reviewed. It is shown that when the storage of energy is used to restrict the steady state transmission to a limited range of frequencies the amount of information that can be transmitted is proportional to the product of the width of the frequency-range by the time it is available. Several illustrations of the application of this principle to practical systems are included. In the case of picture transmission and television the spacial variation of intensity is analyzed by a steady state method analogous to that commonly used for variations with time.

WHILE the frequency relations involved in electrical communication are interesting in themselves, I should hardly be justified

# Introdução - Medida de informação de Hartley

Hartley propôs uma medida de informação baseada no seguinte raciocínio:

- Considere a ocorrência de um símbolo, pertencente a um conjunto de $K$ símbolos.
- A informação provida pela ocorrência de $n$ de tais símbolos é:

| posição 1 | posição 2 | ... | posição $n$ |
|---|---|---|---|
| $K$ | $K$ | ... | $K$ |

- Logo:

$$\log(K^n) = n \log K,$$

em que a base selecionada para o logaritmo fixa o tamanho da unidade de informação, nas palavras do próprio Hartley.

- Pode-se, portanto, expressar a medida de Hartley da quantidade de informação provida pela observação de uma variável aleatória discreta $X$ como:

$$I(X) = \log_b K,$$

em que $K$ denota o número de possíveis valores de $X$.

# Introdução - Medida de informação de Hartley

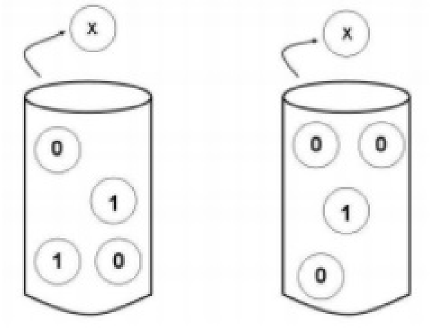
A medida de informação de Hartley permite interpretar corretamente vários problemas técnicos, por exemplo:

- Uma central telefônica com 10.000 assinantes, cada assinante requer 4 dígitos decimais de informação para identificação.
- De forma análoga, são necessários 16 dígitos binários de endereço para acessar uma determinada posição de memória, numa memória com 65.536 posições. Assim o endereço fornece 16 dígitos binários (*bits*) de informação.

Será que medida de informação de Hartley estava correta?

# Introdução - Medida de informação de Hartley

- A fim de perceber que há algo errado, é considerado o experimento aleatório no qual $X$ assume o valor 0 ou o valor 1 marcado na bola retirada aleatoriamente de uma urna contendo bolas marcadas com 0, ou marcadas com 1, idênticas exceto pelo número marcado.



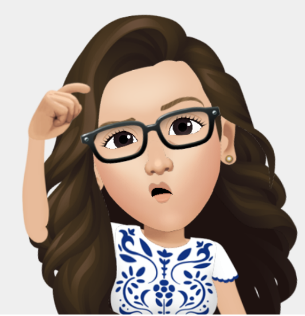
- Segundo Hartley, como $K = 2$ (números 0 ou 1 marcados nas bolas)

$$\log_2 2 = 1 \ bit.$$

- A medida de Hartley diz que a observação da ocorrência de um número fornece um dígito binário de informação, independente da proporção existente entre número de bolas 0 e bolas 1 dentro da urna.

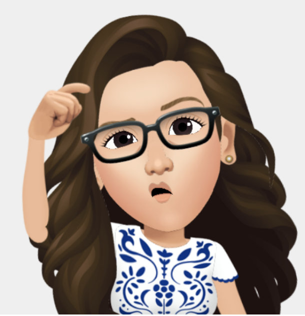
# Introdução - Medida de informação de Hartley

O que há de errado com a medida de **informação** de Hartley?



# Introdução - Medida de informação de Hartley

O que há de errado com a medida de **informação** de Hartley?



- Para ajudar na busca da resposta para essa pergunta, vamos brincar do jogo da forca... =)

# Introdução - Medida de informação

O que há de errado com a medida de informação de Hartley?

- O seguinte comentário pode auxiliar nossa compreensão:
  **"A transmissão de uma mensagem perfeitamente conhecida a priori pelo destinatário é completamente inútil".**

- Não há necessidade de transmissor e receptor: mesmo desligando-os, sabe-se tudo a respeito da mensagem e o processo de transmissão de informação é irrelevante.

O que há de errado com a medida de **informação** de Hartley?

- Isso conduz naturalmente a:
    - 1) tratar uma fonte de informação como uma série de eventos aleatórios que constituem a mensagem emitida, e
    - 2) definir a quantidade de informação associada a essa mensagem como uma medida da sua imprevisibilidade.

# Introdução - Medida de informação

- Intuitivamente, uma mensagem do tipo E = {o Sol nascerá amanhã} carrega um conteúdo de informação muito pequeno.
- Por outro lado, um evento como E = {Haverá um tremor de terra amanhã} carrega um maior conteúdo de informação.
- Imagine-se ouvindo uma sinfonia pela 1ª vez. A quantidade de informação transmitida é grande. Porém, ao ouvi-la várias vezes, a quantidade de informação transmitida em cada audição torna-se cada vez menor. No limite, ao conhecê-la completamente (a ponto de prever deterministicamente os próximos acordes), nenhuma informação é transmitida.

# Introdução

- As palavras **"incerteza"**, **"surpresa"** e **"informação"** estão relacionadas.

- Antes de um evento, há uma certa quantidade de incerteza e, depois dele, há um ganho de uma quantidade de informação.

- Se a informação é ligada a incerteza, então:
  - quanto **menos provável** um evento é, **mais informação** ele carrega

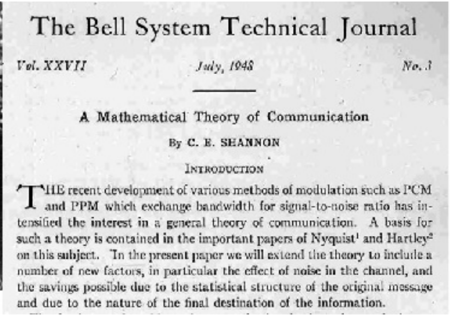
$$P \to 0 \quad \Rightarrow \quad I \to \infty$$

  - quanto **mais provável** um evento é, **menos informação** ele carrega

$$P \to 1 \quad \Rightarrow \quad I \to 0$$

# Introdução - Medida de Informação de Shannon

- Em 1948, vinte anos após a publicação do artigo de Hartley, Shannon publicou um artigo (*A Mathematical Theory of Communication*), propondo uma nova medida de informação, a qual deflagrou uma explosão de atividades fazendo uso dos conceitos de Shannon, que perdura até hoje.



The Bell System Technical Journal

Vol. XXVII — July, 1948 — No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

A definição de informação proposta por Shannon é:

- A quantidade de informação associada a um evento (mensagem) $E$, denotada $I(E)$, é expressa:

$$I(E) = \log_2 \frac{1}{P(E)}.$$

Partiremos de um exemplo para uma melhor explicação:

- Imagine nesse caso, uma moeda desonesta (enviesada), que tem probabilidade $p_1 = 0,9$ de dar cara e probabilidade $p_2 = 0,1$ de dar coroa.

- Por você estar acostumado que a jogada dessa moeda quase sempre dê **cara** , esse resultado não te surpreende. Mas um resultado **coroa** te surpreende por conta da "raridade" do evento.

- Pensando nisso, uma forma natural de se definir essa surpresa que se tem, seria como algo proporcional ao inverso da probabilidade $p$ de ocorrência do evento, de modo que quando menor essa probabilidade maior a surpresa.

# Introdução - Medida de Informação de Shannon

- Shannon definiu essa grandeza (surpresa) como:

$$h = \log_2 \left( \frac{1}{p} \right)$$

- Dessa maneira podemos calcular a surpresa da jogada da moeda retornar cara ($h_1$) ou coroa ($h_2$) de acordo com a definição anterior.

$$h_1 = \log_2 \frac{1}{0,9} = 0,152 \quad \text{e} \quad h_2 = \log_2 \frac{1}{0,1} = 3,322$$

- De fato:

$$h_1 > h_2 \quad \Rightarrow \quad p_1 < p_2$$

**Que surpresa!!!!**

# Introdução - Medida de Informação de Shannon

- Na prática, geralmente nós não estamos interessados em saber a surpresa de um valor em particular que uma variável aleatória pode assumir, e sim a **surpresa associada com todos os possíveis valores que essa variável aleatória pode ter**.

- De modo a obter a surpresa associada a todos possíveis valores que $X$ pode assumir, defini-se a **ENTROPIA**, $H(X)$, como a informação de Shannon média:

$$H(X) = p_1 h_1 + p_2 h_2 = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2}$$

$$H(X) = \sum_{i=1}^{2} p_i \log_2 \frac{1}{p_i}$$

- **Entropia:** Seja $X$ uma variável aleatória com $m$ possíveis valores, $X = \{x_1, x_2, \ldots, x_m\}$, os quais ocorrem com probabilidades $P(X = x_1), P(X = x_2), \ldots, P(X = x_m)$, respectivamente. A entropia, ou incerteza, $H(X)$, de uma variável aleatória $X$ é a quantidade:

$$H(X) = \sum_{i=1}^{m} P(X = x_i) \log_2 \frac{1}{P(X = x_i)}$$

$$H(X) = -\sum_{i=1}^{m} P(X = x_i) \log_2 P(X = x_i)$$

Obs.: Todos os logaritmos daqui em diante são na base 2 exceto quando explicitamente indicado.

- Então, voltando a questão da moeda enviesada, temos que:

$$H(X) = \sum_{i=1}^{2} p_i \log_2 \frac{1}{p_i}$$

$$H(X) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2}$$

$$H(X) = 0,9 \log_2 \frac{1}{0,9} + 0,1 \log_2 \frac{1}{0,1}$$

$$H(X) = 0,469 \ bit$$