Teoria da Informação

Prof. Leonardo Araújo

https://github.com/leolca/lectures/raw/master/ti/teoria-da-informacao.pdf

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

Teoria da Informação e Codificação

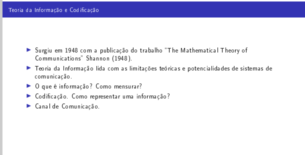
- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.



- A distinguibilidade entre as mensagens é fator importante para caracterizar informação. Pela definição de Shannon, "Informação é a habilidade de distinguir de forma confiável dentre um rol de alternativas possíveis".
- Shannon cunhou o termo 'auto-informação' de um evento ou mensagem aleatória definindo-o como "menos o logaritmo da probabilidade do evento aleatório". A 'entropia' da fonte estocástica que gera os eventos é o valor esperado da auto-informação.
- Shannon mostrou que a entropia de uma fonte estocástica possui um significado físico: em média, é o menor número de bits necessários para representar ou comunicar de forma fidedigna eventos gerados por uma fonte estocástica.
- O problema central em Teoria da Informação é a transmissão eficiente e confiável de dados, do transmissor a um receptor, através de um canal de comunicação.
- Eficiência (usar o mínimo de recursos possível).
- Confiabilidade (evitar erros, ser capaz de detectá-los e corrigi-los).

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

- A informação é um conceito paradoxal. Por um lado necessita de uma representação física, por outro lado é abstrata. Uma mesma informação pode ser representada em papel, em um meio magnético ou ótico, pode ser representada por ondas mecânicas ou elétricas.
- Linguagem. Comunicação falada e escrita. Código faz associação entre símbolo e mensagem e é ‘arbitrário’.

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

A área da ciência criada por Shannon ampliou-se ao longo do tempo e influência diversas outras áreas. Por exemplo: teoria da comunicação, criptografia, ciência da computação, física (mecânica estatística), matemática (probabilidade e estatística), filosofia da ciência, linguística e processamento de linguagem natural, reconhecimento de fala, reconhecimento de padrões e aprendizado de máquina, compressão de dados, economia, biologia e genética, psicologia, etc.

Shannon entrou para o Bell Labs para trabalhar com sistemas de controle de disparo e criptografia durante a Segunda Guerra Mundial, sob um contrato com o Comitê Nacional de Pesquisa para Defesa Em 1945, Shannon elaborou um memorando sigiloso, que posteriormente foi publicado sob o título "Communication Theory of Secrecy Systems". Este incorporava muitos dos conceitos e formulações matemáticas do artigo mais consagrado "A Mathematical Theory of Communication" (1948).

Teoria da Informação e Codificação

- Surgiu em 1948 com a publicação do trabalho "The Mathematical Theory of Communications" Shannon (1948).
- Teoria da Informação lida com as limitações teóricas e potencialidades de sistemas de comunicação.
- O que é informação? Como mensurar?
- Codificação. Como representar uma informação?
- Canal de Comunicação.

- Codificação - criar códigos com algoritmos práticos para codificação e decodificação para serem utilizados na comunicação no mundo real em canais ruidosos.
- Exemplos de codificações conhecidas para representar informação: código Morse, código ASCII, etc.

Código Morse

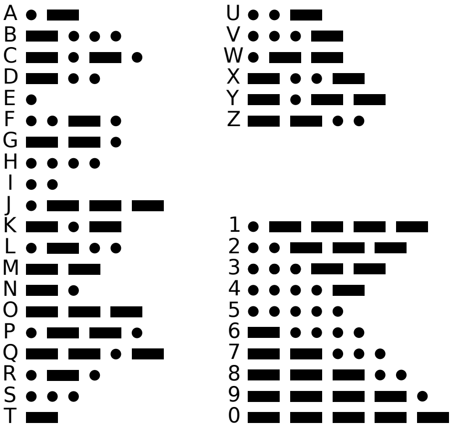


Figura 1: Código Morse internacional (Wikipedia (2020d)). Letras do alfabeto ordenadas por frequência de ocorrência no inglês: etaoin shrdlu cmfwyp vbgkjq xz (Wikipedia (2020a,c)).

Código Unário

Claude Mendibil utilizou o código unário (1, 01, 001, 0001, ...) para representar as letras do alfabeto ESARINTULOMDPCFBVHGJQZYXKW. Utilizando este código, Jean-Dominique Bauby ditou o livro *Le Scaphandre et le Papillon* (O Escafandro e a Borboleta).

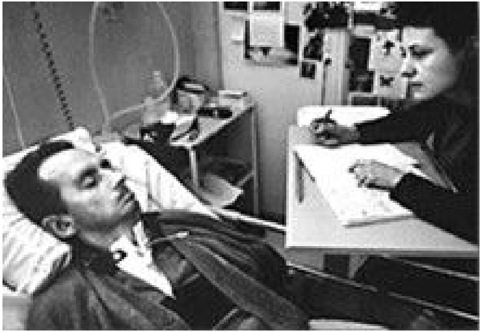


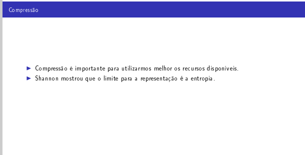
Figura 2: Foto de Bauby em 1996 ditando suas memórias para Claude Mendibil (Wikipedia (2020b)).

Compressão

- Compressão é importante para utilizarmos melhor os recursos disponíveis.
- Shannon mostrou que o limite para a representação é a entropia.

Compressão

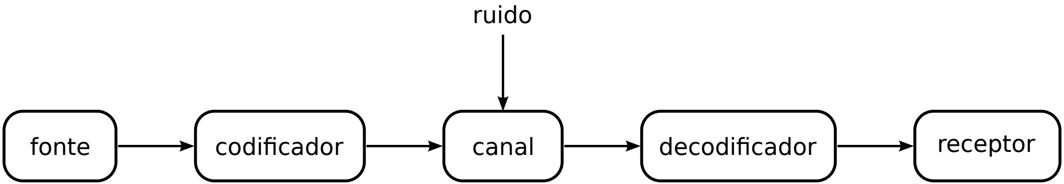
- Compressão é importante para utilizarmos melhor os recursos disponíveis.
- Shannon mostrou que o limite para a representação é a entropia.



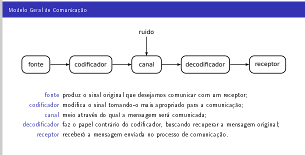
Compressão é importante para utilizarmos melhor os recursos disponíveis.

1. Armazenar mais dados em meio (disco rígido, memória, fita, etc).
2. Transmitir mais informação através de um canal (essencialmente, armazenar e transmitir são o mesmo problema).
3. Diminuir o desgaste do meio ao reduzir o número de vez que se faz leitura e escrita. Solid State Drives (SSDs) baseados em memórias flash NAND possuem um número finito de ciclos de programar/apagar. É importante reduzir a quantidade de bits que serão gravados para aumentar a vida útil dessas memórias/discos. (A compressão LZ4 vem sendo utilizada com esta finalidade, e também para que o S.O. tenha um boot mais rápido)

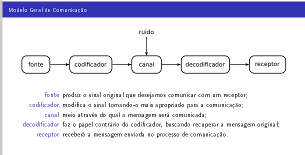
Modelo Geral de Comunicação



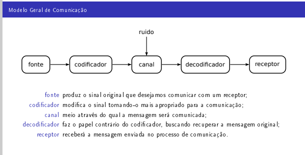
fonte	produz o sinal original que desejamos comunicar com um receptor;
codificador	modifica o sinal tornando-o mais apropriado para a comunicação;
canal	meio através do qual a mensagem será comunicada;
decodificador	faz o papel contrário do codificador, buscando recuperar a mensagem original;
receptor	receberá a mensagem enviada no processo de comunicação.



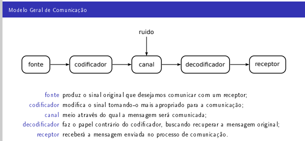
- Separação do codificador/decodificador em duas partes: codificador/decodificador de fonte e codificador/decodificador de canal.
- Remover redundância do sinal produzido pela fonte e acrescentar redundância por causa do ruído no canal de comunicação.
- Fontes e Canais de comunicação: discretos ou contínuos.



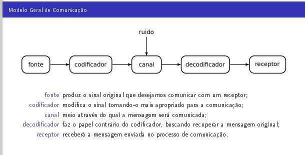
No contexto de Teoria da Informação, fonte é qualquer coisa que produza uma mensagem, um sinal que carregue informação. Podemos considerar uma fonte que produz mensagens como: voz, sons, palavras, imagens, vídeo, sequência de bits de um programa de computador, etc.



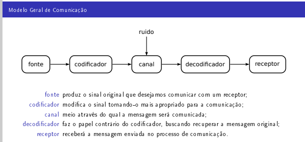
Canal é o meio através do qual o sinal produzido pela fonte será transmitido/propagado/armazenado. Por exemplo: espaço aberto (ar), linha telefônica, link de rádio, link em uma comunicação espacial, disco rígido, CD, DVD, fita magnética (armazenamento - transmissão no tempo ao invés de espaço pode sofrer deterioração ao longo do tempo); DNA de seres vivos ao longo de gerações, envio de mensagens por estímulos elétricos ou químicos em um organismo biológico.



Receptor é aquele a quem é destinada a mensagem transmitida. Exemplos: computador ou equipamento, uma pessoa, rádio, tv, sistema de áudio, etc.



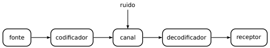
Ruído é qualquer sinal que interfere com aquele que está sendo transmitido. Exemplos: ruído térmico, ruído impulsivo, cross-talk, outro sinal qualquer indesejado. Ruído representa a nossa compreensão imperfeita do universo. Desta forma, tratamos ruído como algo aleatório e que usualmente obedece certas regras, tais como uma determinada distribuição probabilística.



O codificador processa o sinal antes de inseri-lo no canal de comunicação.

- Redução dos dados, removendo redundância do sinal.
- Inserção de redundâncias de acordo com as características do canal de comunicação, para garantir integridade aos dados transmitidos.
- Codificação para representar as informações de um sinal sob a forma de outro sinal.

Modelo Geral de Comunicação



- **fonte** produz o sinal original que desejamos comunicar com um receptor;
- **codificador** modifica o sinal tornando-o mais apropriado para a comunicação;
- **canal** meio através do qual a mensagem será comunicada;
- **decodificador** faz o papel contrário do codificador, buscando recuperar a mensagem original;
- **receptor** receberá a mensagem enviada no processo de comunicação.

O decodificador faz o papel inverso do codificador.

- Remove os erros de transmissão.
- Recupera a informação original enviada pela fonte.

Notação

X é uma variável aleatória

x é uma valor que a v.a. assume

$\mathcal{X}$ é o alfabeto de tamanho $|\mathcal{X}| = K$ dentro do qual a v.a. assume valores, $\mathcal{X} = \{a_1, \ldots, a_K\}$

$\mathcal{P}_X$ é o conjunto de probabilidades associadas aos valores, $\mathcal{P}_X = \{p_1, \ldots, p_K\}$, tais que $p_i \geq 0$ e $\sum_{i=1}^{K} p_i = 1$

p_i é a probabilidade da v.a. assumir um determinado valor, $p_i = \Pr(X = a_i)$

$\mathcal{P}_X = p$ é a distribuição da v.a., $\sum_{x \in \mathcal{X}} \Pr(X = x) = 1$

$X \sim p$, a v.a. X possui distribuição p

Informação

O conceito de informação é amplo, sendo difícil ser contemplado em sua plenitude por qualquer definição.

Shannon (1948) propôs a definição de *entropia* que possui muitas propriedades em comum o senso comum do que deve ser informação.

A informação fornecida por uma mensagem corresponde com o quão improvável é esta mensagem.

O conceito de informação é amplo, sendo difícil ser contemplado em sua plenitude por qualquer definição.

Shannon (1948) propôs a definição de *entropia* que possui muitas propriedades em comum o senso comum do que deve ser informação.

A informação fornecida por uma mensagem corresponde com o quão improvável é esta mensagem.

O que é previsível fornece pouca ou nenhuma informação.
Quanto mais incerto, mais informação há.

Informação

Hartley (1928) propõem uma medida de informação para uma variável aleatória X:

$$I(X) = \log_b L, \tag{1}$$

onde L é o número de possíveis valores que X pode assumir. Se $b = 2$, a informação será medida em 'bits' (nome sugerido por J.W. Tukey).

Informação

Hartley (1928) propõem uma medida de informação para uma variável aleatória X:

$$I(X) = \log_b L, \tag{1}$$

onde L é o número de possíveis valores que X pode assumir. Se $b = 2$, a informação será medida em 'bits' (nome sugerido por J.W. Tukey).

A definição de Hartley é condizente com as seguintes intuições sobre informação:

- Dois cartões de memória devem possuir o dobro da capacidade de um cartão para armazenamento de informação.
- Dois canais de comunicação idênticos devem possuir o dobro da capacidade de transmitir informação que um único canal.
- Um dispositivo com duas posições estáveis, como um relé ou um flip-flop, armazena um bit de informação. N dispositivos deste tipo podem armazenar N bits de informação, já que o número total de estados é 2^N e $\log_2 2^N = N$.

Entretanto, isto é válido apenas quando as mensagens/eventos são equiprováveis. No caso extremo, note que se o cartão de memória armazena apenas zeros, ele não é capaz de armazenar informação alguma.

Entropia

Suponha que existam eventos E_k com probabilidade de ocorrência p_k.

- Shannon: informação associada ao evento E_k é dada por $I(E_k) = \log(1/p_k)$.
  - Se $p_k = 1 \rightarrow$ não há surpresa na ocorrência do evento E_k.
  - Se $p_k = 0 \rightarrow$ surpresa infinita, afinal o evento E_k é impossível.
  - $I(E_k) = -\log p(E_k)$ é a auto-informação do evento ou mensagem E_k.
- **Sempre** utilizaremos a base 2 para o cálculo do logaritmo, desta forma $\log \equiv \log_2$, a menos que seja especificado o contrário.
- $\ln$ é o logaritmo na base natural e.

Entropia

- Notação: $p(x) = P_X(X = x)$, a probabilidade do evento $\{X = x\}$, da v.a. X assumir o valor x.
- Valor esperado da v.a. X: $E[X] = EX = \sum_x xp(x)$.
- Dada uma função $g : \mathcal{X} \to \mathbb{R}$, o valor esperado da v.a. $g(X)$ é $Eg(X) = \sum_x g(x)p(x)$.
- Considere $g(x) = \log(1/p(x))$. Então $g(x)$ é a imprevisão (surpresa) de encontrar o evento $X = x$. Tomando o valor esperado de g teremos

$$\sum_x p(x) \log \frac{1}{p(x)}, \tag{2}$$

ou seja, a esperança da surpresa, ou o valor esperado da imprevisão na variável aleatória X. Esta é a definição de entropia.

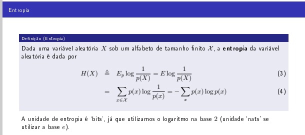
Entropia

Definição (Entropia)

Dada uma variável aleatória X sob um alfabeto de tamanho finito $\mathcal{X}$, a **entropia** da variável aleatória é dada por

$$
\begin{aligned}
H(X) &\triangleq E_p \log \frac{1}{p(X)} = E \log \frac{1}{p(X)} && (3)\\
&= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} = -\sum_{x} p(x) \log p(x) && (4)
\end{aligned}
$$

A unidade de entropia é 'bits', já que utilizamos o logaritmo na base 2 (unidade 'nats' se utilizar a base e).


Entropia

Definição (Entropia)

Dada uma variável aleatória X sob um alfabeto de tamanho finito $\mathcal{X}$, a **entropia** da variável aleatória é dada por

$$H(X) \triangleq E_p \log \frac{1}{p(X)} = E \log \frac{1}{p(X)} \quad (3)$$

$$= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} = -\sum_{x} p(x) \log p(x) \quad (4)$$

A unidade de entropia é 'bits', já que utilizamos o logaritmo na base 2 (unidade 'nats' se utilizar a base e).

- Entropia mede a grau de incerteza associado a uma distribuição.
- Entropia mede a desordem ou o espalhamento de uma distribuição.
- Entropia mede a 'escolha' que a fonte tem na escolha de símbolos de acordo com uma densidade (maior entropia implica em mais escolha).
- Vamos utilizar a seguinte convenção: $0 \log 0 = 0$.

Entropia

Se uma v.a. $X \sim p(x)$, então o valor esperado de uma função desta v.a., $g(X)$, é dada por

$$E[g(X)] = \sum_{x \in \mathcal{X}} p(x)g(x). \tag{5}$$

A entropia de X pode ser interpretada como o valor esperado da v.a. $\log \frac{1}{p(X)}$, onde X é descrita pela função massa de probabilidade $p(x)$.

$$H(X) = E\left[\log \frac{1}{p(X)}\right]. \tag{6}$$

Entropia

Se uma v.a. $X \sim p(x)$, então o valor esperado de uma função desta v.a., $g(X)$, é dada por

$$E[g(X)] = \sum_{x \in \mathcal{X}} p(x)g(x). \tag{5}$$

A entropia de X pode ser interpretada como o valor esperado da v.a. $\log \frac{1}{p(X)}$, onde X é descrita pela função massa de probabilidade $p(x)$.

$$H(X) = E\left[\log \frac{1}{p(X)}\right]. \tag{6}$$

- Entropia é uma medida da real 'incerteza' média, o que é uma medida sobre toda a distribuição.
- Entropia mede o grau de incerteza médio ou esperado do resultado de uma distribuição de probabilidade.
- É uma medida de desordem ou espalhamento. Distribuições com alta entropia devem ser planas, mais uniformes, enquanto distribuições com baixa entropia devem possuir poucas modas (unimodal, bimodal).

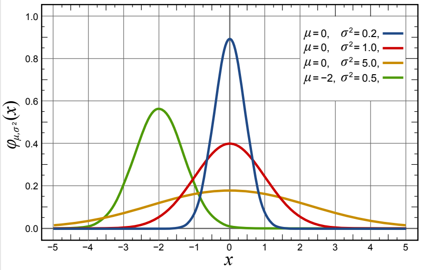
Entropia

Se uma v.a. $X \sim p(x)$, então o valor esperado de uma função desta v.a., $g(X)$, é dada por

$$E[g(X)] = \sum_{x \in \mathcal{X}} p(x) g(x). \tag{5}$$

A entropia de X pode ser interpretada como o valor esperado da v.a. $\log \frac{1}{p(X)}$, onde X é descrita pela função massa de probabilidade $p(x)$.

$$H(X) = E\left[\log \frac{1}{p(X)}\right]. \tag{6}$$



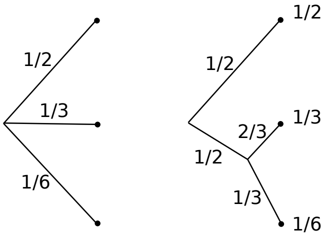
- mais concentrado: menor entropia
- mais espalhado: maior entropia
- os valores em x não importam, apenas os valores das probabilidades associadas $p(x)$ importam no cálculo da entropia

Escolha, Incerteza e Entropia I

Suponha um conjunto de eventos cujas probabilidades de ocorrências sejam dadas por $p_1, p_2, \ldots, p_n$. É possível encontrar uma medida de quanta 'escolha' está envolvida na seleção de um evento ou quão incertos estamos da saída?
Para tal medida $H(p_1, p_2, \ldots, p_n)$, é razoável requerermos as seguintes propriedades:

1) H deve ser contínuo em p_i;
2) Se todos os p_i são iguais, $p_i = \frac{1}{n}$, então H deve ser uma função monotonicamente crescente de n (quando temos eventos equiprováveis, teremos mais incerteza quão maior for o número de eventos possíveis);
3) Se for possível quebrar uma escolha em uma sequência de escolhas sucessivas, a medida H original deve ser a soma ponderada dos valores individuais das medidas H_i após a quebra.

Escolha, Incerteza e Entropia II



$$H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2} H(\frac{2}{3}, \frac{1}{3}) \tag{7}$$

Escolha, Incerteza e Entropia III

A única função H que satisfaz às suposições acima é da forma Shannon (1948):

$$H = -K \sum_{i=1}^{k} p(i) \log p(i) \ , \tag{8}$$

onde K é uma constante positiva.

Demonstração da Equação (8) I

Nesta secção iremos apresentar a demonstração de $H = -\sum p_i \log p_i$ (conforme Apêndice 2 de Shannon (1948)).
Vamos definir

$$A(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right). \tag{9}$$

Desejamos que uma escolha dentre s^m opções igualmente prováveis possa ser decomposta como uma sequência de m escolhas que se subdividem em s possibilidades igualmente prováveis.

Demonstração da Equação (8) II

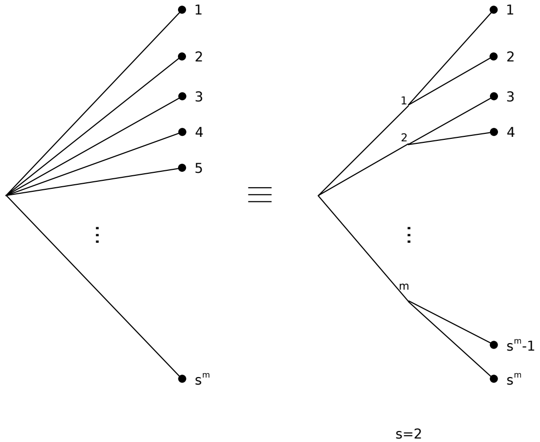


Figura 3: Exemplo de equivalência para $s = 2$.

Demonstração da Equação (8) III

Teremos então que

$$A(s^m) = mA(s). \tag{10}$$

Da mesma forma, para t e n, teremos $A(t^n) = nA(t)$. Podemos tomar n arbitrariamente grande e encontrar m que satisfaça

$$s^m \leq t^n \leq s^{(m+1)}. \tag{11}$$

Tomando o logaritmo[1] da expressão acima e dividindo por $n \log s$ todos os termos[2], teremos

$$\frac{m}{n} \leq \frac{\log t}{\log s} \leq \frac{m}{n} + \frac{1}{n}, \tag{12}$$

o que é equivalente a

$$\left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \epsilon, \tag{13}$$

Demonstração da Equação (8) IV

onde ϵ é arbitrariamente pequeno, já que n é arbitrariamente grande.
Usando agora a propriedade desejada de monotonicidade de $A(n)$, teremos

$$A(s^m) \leq A(t^n) \leq A(s^{(m+1)}) \tag{14}$$

$$mA(s) \leq nA(t) \leq (m+1)A(s) \tag{15}$$

Dividindo a expressão acima por $nA(s)$, teremos

$$\frac{m}{n} \leq \frac{A(t)}{A(s)} \leq \frac{m}{n} + \frac{1}{n}, \tag{16}$$

ou, de forma equivalente,

$$\left| \frac{m}{n} - \frac{A(t)}{A(s)} \right| < \epsilon, \tag{17}$$

Demonstração da Equação (8) V

e assim, como as duas frações ($^{\log t}/_{\log s}$ e $^{A(t)}/_{A(s)}$) estão ϵ próximas de $^{m}/_{n}$, podemos concluir que

$$\left| \frac{A(t)}{A(s)} - \frac{\log t}{\log s} \right| < 2\epsilon. \tag{18}$$

Como ϵ é arbitrariamente pequeno, no limite teremos

$$\frac{A(t)}{A(s)} = \frac{\log t}{\log s} \tag{19}$$

$$A(t) = \frac{A(s)}{\log s} \log t = K \log t, \tag{20}$$

onde K deve ser positivo, de forma que $A(n)$ seja monótona crescente.
Suponha uma escolha com n possibilidades em que as probabilidades são comensuráveis, $p_i = {^{n_i}}/{\sum n_i}$, onde n_i são inteiros. De forma equivalente, uma escolha entre $\sum n_i$ opções pode ser expressa como uma escolha dentre n opções com probabilidades $p_1, \ldots, p_n$, e para uma

Demonstração da Equação (8) VI

i-ésima dada escolha, realizar uma nova escolha dentre n_i opções igualmente prováveis. Teremos então:

$$\overbrace{K \log\left(\sum n_i\right)}^{A(\sum n_i)} = H(p_1, \ldots, p_n) + \overbrace{K \log n_i}^{A(n_i)} \tag{21}$$

$$K \underbrace{\left(\sum p_i\right)}_{=1} \log\left(\sum n_i\right) = H(p_1, \ldots, p_n) + K \underbrace{\left(\sum p_i\right)}_{=1} \log n_i. \tag{22}$$

E assim,

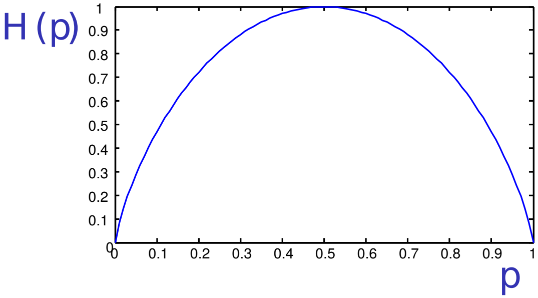
$$H(p_1, \ldots, p_n) = K\left[\left(\sum p_i\right) \log\left(\sum n_i\right) - \left(\sum p_i\right) \log n_i\right] \tag{23}$$

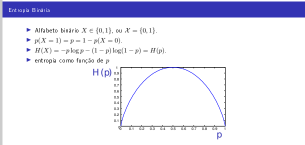
$$= -K \sum p_i \log \frac{n_i}{\sum n_i} = -K \sum p_i \log p_i. \quad \square \tag{24}$$

[1] Logaritmo é uma função monótona crescente.
[2] $n \log s$ é positivo para $n \geq 0$ e $s \geq 1$.

Entropia Binária

- Alfabeto binário $X \in \{0, 1\}$, ou $\mathcal{X} = \{0, 1\}$.
- $p(X = 1) = p = 1 - p(X = 0)$.
- $H(X) = -p \log p - (1 - p) \log(1 - p) = H(p)$.
- entropia como função de p





- maior incerteza $(H = 1)$ quando $p = 0.5$ e menor incerteza $(H = 0)$ quando $p = 0$ ou $p = 1$.
- note que a entropia $H(p)$ é concava em p.

Entropia - GNU Octave

```
function H = entropy(p,b)

  if (nargin == 0 || nargin > 2) print_usage (); endif;
  if any(p < 0) | any(p > 1) | abs(sum(p)-1) > 1E-10, error('not a
     ↪ valid pmf!'); endif;

  id = find(p!=0);
  p = p(id);
  H = sum( - p .* log2(p) );

  if nargin > 1, H *= log(2)/log(b); endif;

endfunction
```

[download do código]

Entropia - GNU Octave - demo

```
%!demo
%! p = [0.5 0.5];
%! H = entropy(p);
%! printf('The pmf p has entropy = %.2f bits.\n',H);
%! He = entropy(p,e);
%! printf('The pmf p has entropy = %.2f nats.\n',He);
%! p = [0:0.02:1];
%! for i=1:length(p), H(i) = entropy([p(i), (1-p(i))]); endfor;
%! figure; plot(p,H); xlabel('p'); ylabel('H(p) (bits)'); title('
    ↪ binary entropy');
```

Entropia - Exemplo

Suponha uma v.a. $X \in \mathcal{X} = \{a, b, c, d\}$ com distribuição dada por

$$X = \begin{cases} a, & \text{com probabilidade } \frac{1}{2}, \\ b, & \text{com probabilidade } \frac{1}{4}, \\ c, & \text{com probabilidade } \frac{1}{8}, \\ d, & \text{com probabilidade } \frac{1}{8}. \end{cases} \tag{25}$$

A entropia associada será dada por

$$H(X) = H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \tag{26}$$

$$= -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{8} \log \frac{1}{8} \tag{27}$$

$$= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} = \frac{14}{8} = \frac{7}{4} \tag{28}$$

Hartley, R. V. L. (1928). Transmission of information1. *Bell System Technical Journal*, 7(3):535–563.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423.

Wikipedia (2020a). Etaoin shrdlu. `https://en.wikipedia.org/wiki/Etaoin_shrdlu`. [Online; accessed 10-August-2020].

Wikipedia (2020b). Jean-dominique bauby. `https://en.wikipedia.org/wiki/Jean-Dominique_Bauby`. [Online; accessed 10-August-2020].

Wikipedia (2020c). Letter frequency. `https://en.wikipedia.org/wiki/Letter_frequency`. [Online; accessed 10-August-2020].

Wikipedia (2020d). Morse code. `https://en.wikipedia.org/wiki/Morse_code`. [Online; accessed 10-August-2020].