Universidade Federal do ABC

BC-0504
Natureza da Informação

Aulas 2

Entropia na termodinâmica e na teoria da informação

Equipe de professores de Natureza da Informação

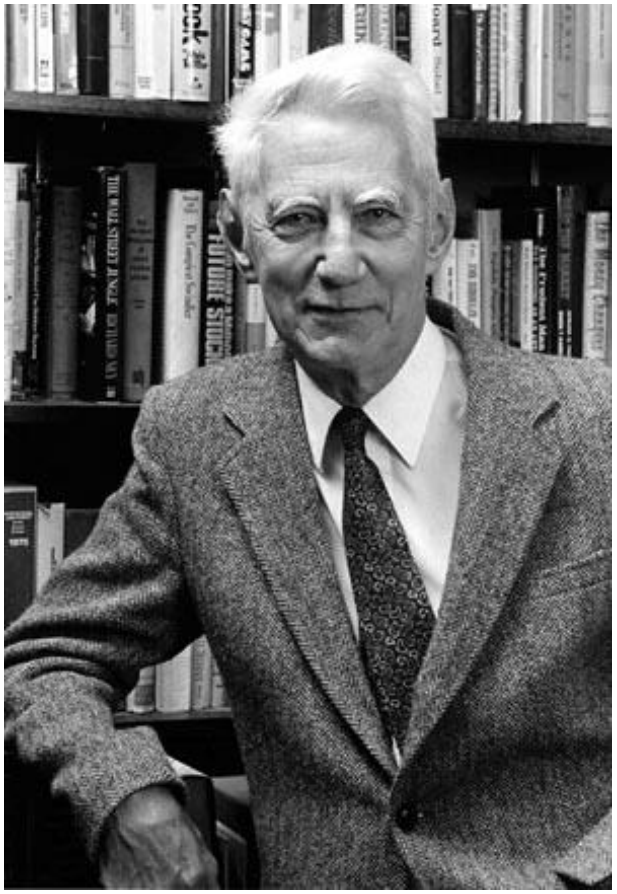

Parte 4

Os pilares da teoria da informação

- Os estudos de criptografia desenvolvidos na 2ª Guerra mundial
- Os estudos de “termodinâmica”
- **As tecnologias de transmissão de informação**

Claude Elwood Shannon (1916-2001)

- Trabalhava nos Laboratórios Bell
- Quanta informação pode passar por uma linha de telefone?
- Inventou o término bigit ou bit



Shannon’s ultimate machine

- Quando se abre a caixa, a mãozinha solta o pino e a caixa fecha de novo.

O Bit como unidade de informação

- Mede o grau de imprevisibilidade.
- 1 bit é a quantidade de informação necessária para tomar uma decisão perante duas opções igualmente prováveis
- Segundo fórmula de Boltzmann, S=k log(W) onde W é o n. de possíveis configurações para um determinado arranjo de partículas
- Segundo fórmula de Shannon.

$$S = -\sum p_i \log_2 p_i$$

A informação é medida em bits

$$S = \log_2(n)$$

- Se todos os estados tem a mesma probabilidade

$$p_i = 1/n, \forall i$$

$$S = -\sum p_i \log_2 p_i = -n\left(\frac{1}{n}\right)\log_2\left(\frac{1}{n}\right) =$$

$$= -n\left(\frac{1}{n}\right)\log_2\left(\frac{1}{n}\right) = \log_2(n)$$

A informação é medida em bits

$$S = \log_2(n)$$

$$S = \log_2(1) = 0$$

- Ele diz bom dia todo dia
  - Há apenas um estado possível
- 0 bit. Não há informação
- Entropia = 0
- Não precisamos transmitir nada
- Já sabemos que ele diz bom dia.

A informação é medida em bits

$$S = \log_2(n)$$

$$S = \log_2(2) = 1$$

- 1 bit – dois estados igualmente prováveis
- Precisamos transmitir um bit para informar sobre o estado da moeda
- Mas sabemos que só pode ser cara ou coroa. Um bit resolve.

A informação é medida em bits

$$S = \log_2(6) = 2{,}58\,bits$$

- Dado: 6 estados.
- 2 bits não são suficientes
- 3 bits: “sobram” 2 estados
- O que quer dizer 2.58 bit?
  - Quanto é .58 bit?

A informação é medida em bits

$$S = \log_2(6) = 2{,}58\,bits$$

- Sobram 2 estados
- Opção: código redundante.
- Podemos ser mais eficientes?

Código	**Estado Do Dado**
000	1
001	2
010	3
011	4
100	5
101	5
110	6
111	6

A informação é medida em bits

$$S = \log_2(6) = 2{,}58\,bits$$

- Para estados 5 e 6, não transmitimos o último bit
- Média de bits transmitidos:
- (4*3+2*2)/6 = 16/6= 2.67

Código	**Estado Do Dado**
000	1
001	2
010	3
011	4
10	5
10	5
11	6
11	6

A informação é medida em bits

$$S = \log_2(6) = 2{,}58\,bits$$

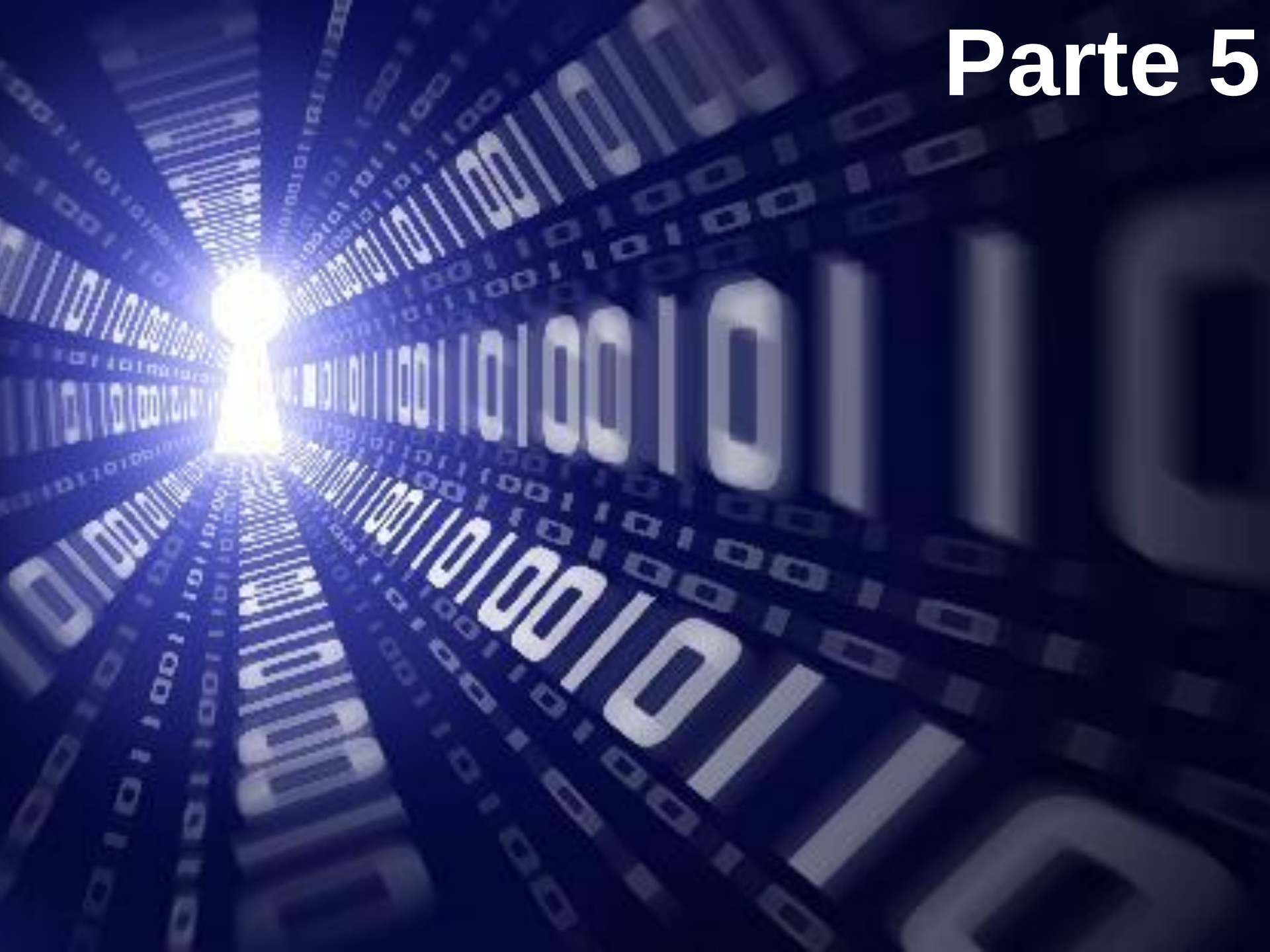
- Média de bits transmitidos:
- (4*3+2*2)/6 = 16/6= 2.67
- Muito próximo de S
- S é um ***limite inferior***

Código	**Estado Do Dado**
000	1
001	2
010	3
011	4
10	5
10	5
11	6
11	6

A informação é medida em bits

$$S = \log_2(6) = 2{,}58\,bits$$

- Supomos dado de prob. iguais
- Suponha um dado viciado com mais probabilidade para 1 e 2
- Então seria mais vantajoso codificar os estados 1 e 2 com 2 bits e os outros com 3 bits
- A entropia do dado viciado é menor


Parte 5

Verdadeira contribuição de Shannon

- Quanta informação cabe num mensagem.
- Como arranjar melhor a informação para que a mensagem seja mais enxuta.
- Exemplo: programas de compressão de arquivos (Zip)
- Outro exemplo são as sinais em jogos de cartas como o “truco”.

Algumas sinais do “truco”

- Piscar um olho = Zap (Paus)
- Subir as sobrancelhas = Sete Copas (Copas)
- Fazer um montinho na bochecha usando a língua = Espadilha (Espadas)
- Mostrar a ponta da língua = Pica Fumo (Ouros)
- Levantar um ombro = Três
- Levantar dois ombros = Um par de três
- Encher as bochechas de ar = "To cheio de manilha"
- Tocar no peito = "deixa aqui que eu faço"
- Passar a mão no pescoço como se estivesse cortando-o= "Peça Truco”

Algumas sinais do “truco”

- Cartas: 4 a 10; JQKA
- Basicamente há um sinal diferente para cada carta
- Queremos essa ineficiência para que os adversários não saibam onde estará o nosso sinal.
- Mas podemos usar o mesmo tipo de sinal para mais de uma carta? (mesmo que não seja bom para esconder dos adversários?)

Um código mais enxuto

- Pro "Pica-fumo" as cartas do baralho bem abertas, como um leque de três carta.

- Pra "Espadilha" as cartas do baralho bem abertas, como um leque de duas cartas só, vc coloca uma carta sobre uma outra, assim abre um leque somente como se fosse duas cartas só.

- Pro "Sete Copas" vc fecha o baralho como num maço de uma carta só

- Pro "ZAP" vc deita todas as cartas na mesa .

$S=\log_2(w)$

- Se W é o numero de diferentes mensagems que precisamos transmitir o código mais enxuto é aquele que tem log(w) bits.
- No caso do “truco”, se quisermos transmitir os quatro naipes com dois bits seria suficiente usando, talvez, os dois olhos.
- Para transmitir o número (de 1 para 7) poderíamos colocar algumas cartas deitadas (viradas para abaixo) representando os zeros, e algumas em pé representando os uns, formando um código binário.


Parte 6

Codificação binária (em 0 e 1)

- Permite identificar qualquer

  magnitude em termos de sim ou não

  Exemplo: Para identificar número entre 1 e n precisaríamos realizar log(n) questões (bem formuladas).
- Exemplo para identificar um átomo no meio do universo que têm 10^{80} átomos
- Precisariamos Log2(10^{80}) = 266 perguntas

Codificação binária permite responder qualquer pergunta

- Exemplo: Qual é a capital da Islândia?
- *Reykjavík.* Poderia ser escrito em ASCII que é código binário que utilizam os computadores na hora de digitar as letras.

Benefícios codificação binária

- Favorece a transmissão e armazenagem da informação em forma de níveis de voltagem.
- Existem códigos corretores de erro que evitam perdas da informação (exemplo: código de controle dos cartões VISA ou número de ISBN)
- Permite conversão A-D e D-A.
- A codificação binária permite facilmente as operações aritméticas e as operações booleanas (lógicas)


Parte 7

Informação e incerteza

- Uma emissor que fornece sempre a mesmo mensagem, fornece zero bits de informação.
- Enquanto que o conteúdo informativo de uma mensagem pouco previsível é grande

$$S = \log_2(1) = 0$$

- Como é um evento único, sabemos quando e onde foi tirada a foto.
- Se mostrasse o Dunga almoçando, a foto não passaria essa informação.

Uma interpretação da fórmula da Entropia: informação = -log(probabilidade)

- Numa seqüência binária, se todos são um, não há informação. Ex: 1111111111
- Mas se o número 1 aparece 10% das vezes, o número 1 possui $-\log_2(1/10)$ de informação ou segundo a fórmula de Shannon

  $-\log_2(p)=-\log_2(0,1)=\log_2(10)=3,3219$ bits
- Enquanto que o número 0 possui segundo a fórmula de Shannon

  $-\log_2(p)=-\log_2(0,9)=\log_2(10/9)=0,152$bits

$$S = -\sum p_i \log_2 p_i$$

- Esta fórmula de entropia representa uma media ponderada das informações de cada um dos eventos possíveis.
- Exemplo: 1000 lançamento de uma moeda.

  Informação em cada cara $I_{cara} = -\log_2(0,5) = \log(2) = 1$

  Informação em cada coroa $I_{coroa} = -\log_2(0,5) = \log(2) = 1$

$$I_{média} = \frac{500 I_{cara} + 500 I_{coroa}}{1000} =$$

$$= 0,5 \log_2(2) + 0,5 \log_2(2) =$$

$$= -0,5 \log_2(0,5) - 0,5 \log_2(0,5) =$$

$$= -p_{cara} \log_2(p_{cara}) - p_{coroa} \log_2(p_{coroa})$$


Parte 8

- Exemplo: Informação contida na jogada de uma moeda alterada para cair 60% das vezes em cara e 40% do tempo em coroa.

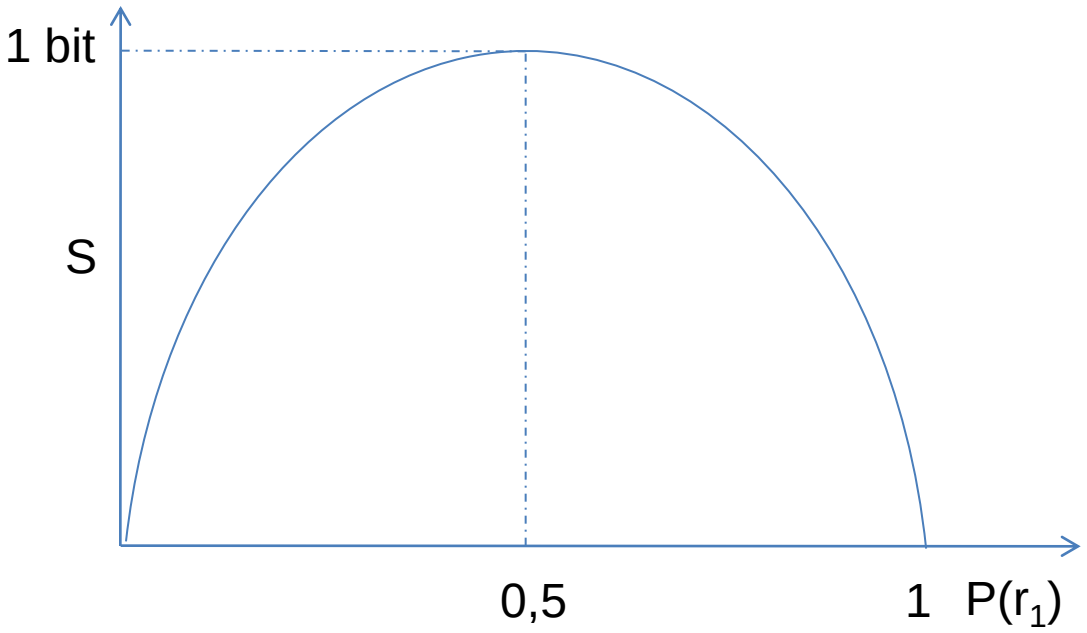
$$S = -0{,}6\log_2(0{,}6) - (0{,}4)\log_2(0{,}4) = 0{,}97\ bits$$

- Observar que a soma das duas probabilidades é 1. 0,6+0,4=1

Exemplo de cálculo de entropia de Shannon para distintas probabilidades de cair cara ou coroa

- Caso B: Dois tipos de resposta cara e coroa, onde P(cara)+P(coroa)=1.

$$S = -P(cara)\log_2(P(cara)) - P(coroa)\log_2(P(coroa)) =$$
$$-P(cara)\log_2(P(cara)) - (1-P(cara)\log_2(1-P(cara))$$



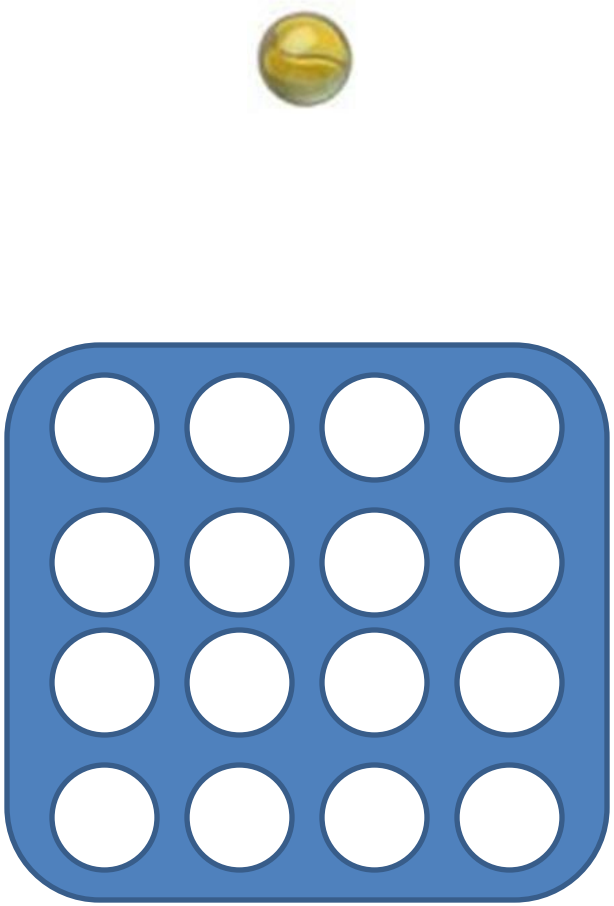
- Quando ambas possibilidades têm a mesma probabilidade de acontecer P(cara)=P(coroa)=0,5 a entropia ou imprevisibilidade é máxima, e igual a 1 bit.


Parte 9

Informação transmitida, I

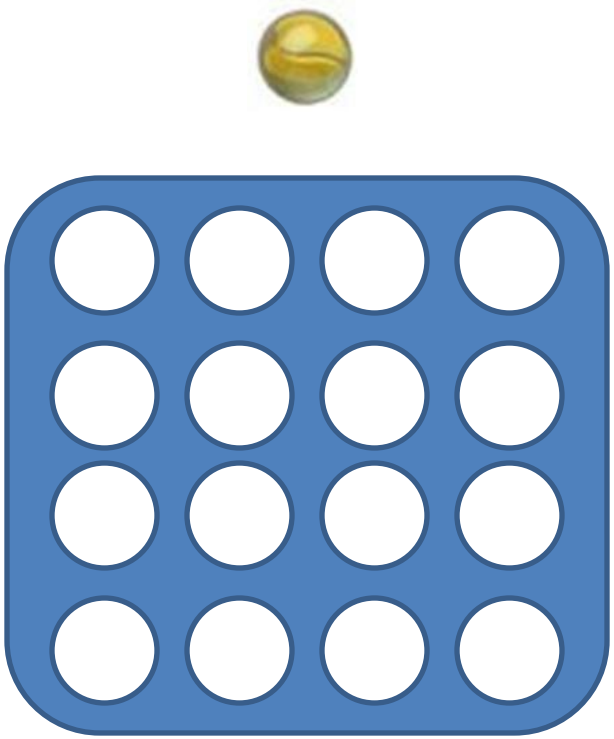
- E a diferencia entre o grau de entropia ou imprevisibilidade inicial (S) e a imprevisibilidade final S_e

Exemplo:



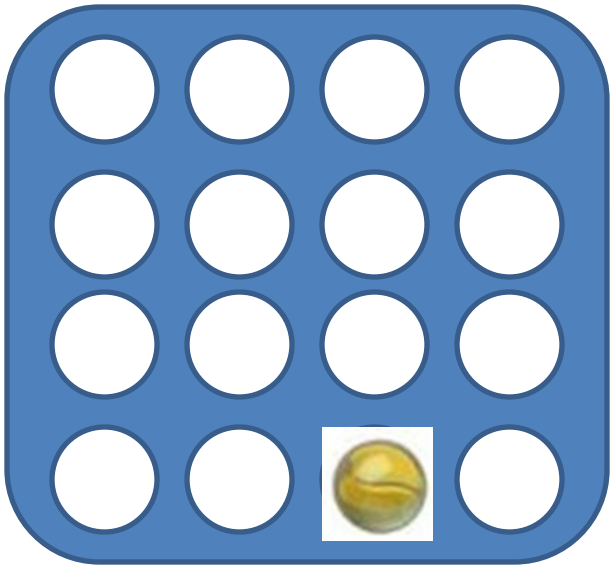
- Trata-se de jogar uma bola de numa matriz com 16 possíveis posições. Qual seria a entropia inicial e qual a entropia final, no caso de querermos lançar uma outra bola de gude.
- Qual seria a quantidade de informação fornecida pela bola?

Exemplo:



- Entropia inicial:
  $S=\log_2(16)=4$

Exemplo:



- Entropia final: $S=\log_2(15)$
- Informação (diferencia de entropias:
- I=Entropia inicial - Entropia final
- $I= \log_2(16)-\log_2(15)=$

$=\log_2(16/15)=0,09$


Parte 10

Atenção!!!

Embora a *Entropia* da Teoria da Comunicação tenha uma analogia matemática com a *Entropia* da Termodinâmica, são conceitos diferentes, e usadas para medir quantidades diferentes!

Entropia em Termodinâmica

- Entropia em Termodinâmica é uma medida da ***reversibilidade*** de um processo: se não há mudança no valor da entropia do sistema, o sistema é reversível.
- Mas a grande maioria dos fenômenos físicos são ***irreversíveis***: nestes fenômenos, a entropia do sistema envolvido aumenta.
- Uma outra interpretação em Termodinâmica nos permite dizer que quando a entropia de um sistema aumenta, diminui a ***energia disponível*** neste sistema para ser transformada em trabalho.
- Uma outra interpretação em Mecânica Estatística é que o aumento da Entropia de um sistema significa um aumento da ***desordem*** deste sistema, num sentido de imprevisibilidade.

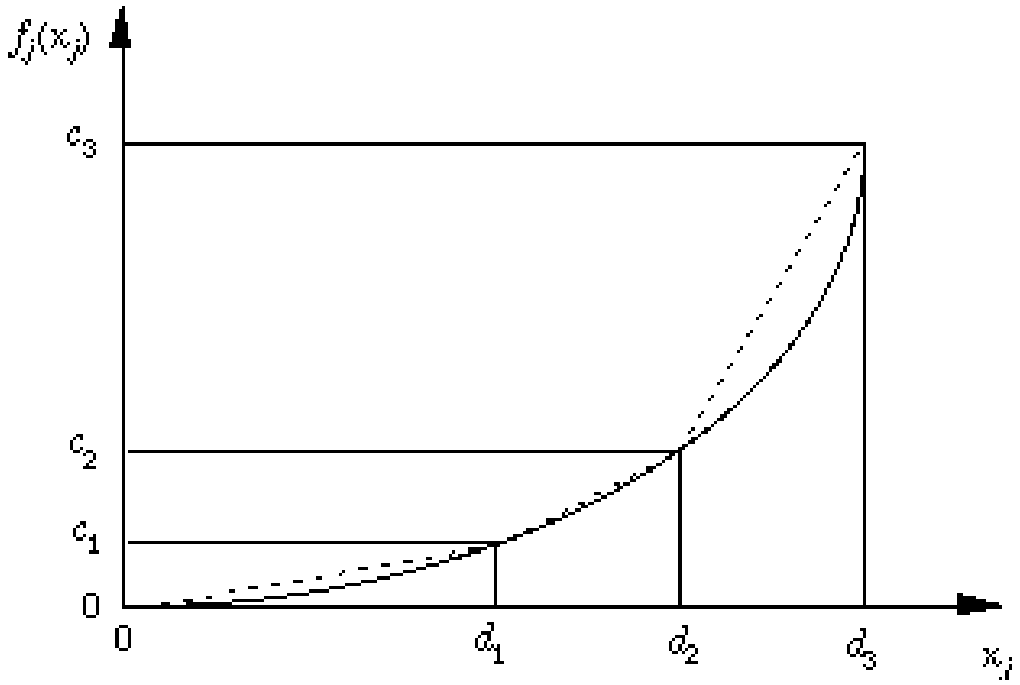
Entropia em Termodinâmica

- Existe uma relação entre matéria, energia, e informação.
  - Ainda não 100% compreendida.
- Einstein: matéria transforma-se em energia: E = mc2
- Entropia Termodinâmica:
  - Exemplo: Comprimimos um gás
  - Diminuir a entropia do gás – compressão é reversível – basta liberar o gás
  - A entropia estatística da distribuição da probabilidade de encontrar gás também aumenta
  - O que implica que ordenamos um sistema – diminui a desordem
  - Mas gastamos energia para comprimir, a entropia do universo aumentou
  - Concentramos energia nesse gás – há maior pressão
  - se liberarmos o gás, ele se expande (pode-se fazer uma bomba assim)
  - A expansão diminui a entropia termodinâmica – é irreversível - gasta a energia concentrada no gás, e torna a distribuição da probabilidade de se encontrar o gás mais uniforme.


Parte 10

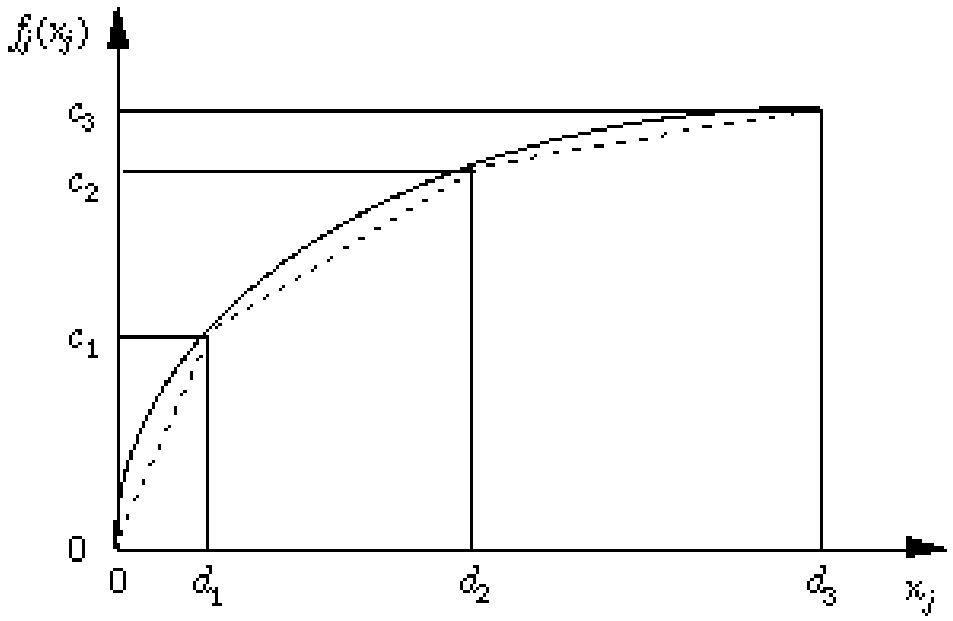
Demonstração: Distribuição uniforme é a maior entropia

- Definições necessárias
- Função concava
  - Qualquer corda está acima da função.



Demonstração: Distribuição uniforme é a maior entropia

- Definições necessárias
- Função - convexa
  - Qualquer corda está abaixo da função.



Demonstração: Distribuição uniforme é a maior entropia

- Definições necessárias
- Função côncava: definição matemática
  - Uma função real f é **côncava** em um intervalo I se, p/ qualquer x e y em I:

$$f\left(\frac{x+y}{2}\right) \geq \frac{f(x)+f(y)}{2}$$

  - A função será **estritamente côncava** se a desigualdade for sempre verdadeira (> ao invés de >=)

Demonstração: Distribuição uniforme é a maior entropia

- Definições necessárias
- Desigualdade de Jensen
  - Suponha função real f, **estritamente côncava** e contínua em um intervalo I, e:

$$\sum_{i=1}^{n} a_i = 1; \;\; a_i > 0; 1 \le i \le n$$

  - Então: $\sum_{i=1}^{n} a_i f(x_i) \le f\left(\sum_{i=1}^{n} a_i x_i\right)$
  - Onde $x_i \in I, 1 \le i \le n$
  - Igualdade sse $x_1 = \ldots = x_n$

Demonstração: Distribuição uniforme é a maior entropia

- Teorema: Suponha $\mathbf{X}$ váriavel aleatória com distribuição de probabilidade $p_1, p_2,..., p_n$, onde $p_i,>0, 1<=i<=n$. Então $Entropia(\mathbf{X})<=log_2 n$, com igualdade sse $p_i,=1/n, 1<=i<=n$.

$$Entropia(\mathbf{X}) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

$$= \sum_{i=1}^{n} p_i \log_2 1/p_i \leq (Jensen)$$

$$\leq \log_2 \sum_{i=1}^{n} \left(p_i \times 1/p_i \right) =$$

$$= \log_2 n$$

Exercícios

Exercícios - Individual

- 1) Dado viciado: Suponha que o dado está viciado, com as seguintes probabilidades:
- 1: 0.05 ; 6:0.3 ; de 2 a 5: 0,1625
- a) qual é a entropia do dado viciado?
- b) Qual o número médio de bits transmitidos por jogada se para cada um dos códigos a seguir.

Códigos para estado do dado

Cód. a	Estado
000	1
001	2
010	3
011	4
10	5
11	6

Cód. b	Estado
000	6
001	5
010	4
011	3
10	2
11	1

Cód. c	Estado
000	1
001	2
010	3
011	4
100	5
101	5
110	6
111	6

Códigos para estado do dado

Cód. d	Estado
0000	1
000	2
001	3
010	4
011	5
1	6

- c)
  - i) Aqui utilizamos 4 bits para o ‘1’ e apenas 1 bit para o ‘6’. Conseguimos um código mais eficiente?
  - Ii) Podemos conseguir um código muito mais eficiente do que este para o dado viciado?

2) Melhor de Três

- 2 times A, e B, jogam uma melhor de três jogos. Cada jogo é independente, e os times tem igual probabilidade de ganhar.
- Defina X como a seqüência de vencedores. Por exemplo X=AA; ou X=BAB;
- Defina Y como o número de jogos jogados (pode ser 2 ou 3)
- Defina Z como o vencedor final: A ou B

Melhor de Três

- a) Qual a entropia de X?
  - Dica: liste todos os possíveis valores de X e a probabilidade de cada um.
- b) Qual a entropia de Y?

Melhor de Três

- c) Suponha que sabemos que A ganhou a melhor de três.
  - i) Qual a entropia de X? Compare com (a), quantos bits de informação foram ganhos?
  - ii) Qual a entropia de Y? Compare com (b), quantos bits de informação foram ganhos?
- d) Suponha que sabemos o valor de X. Qual a entropia de Z?