Introduction to Information Theory

Dr. Yoann Altmann

B39AX – Fall 2023
Heriot-Watt University

Plan

- Motivations and introduction
- Source model (DMS)
- Information content
- Entropy
- Information rate

- Based on the book:
  “Information Theory,. **Inference**, and **Learning Algorithms”**. David J.C. MacKay

General concerns

- Reducing error rate
- Keeping ”short” codes

- ”Naive” approaches seem to work
  - How good is our approach?
  - Can it be improved and how?
- Mathematical tools needed…

Information content (DMS)

- X takes values from the alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ with probabilities $p_n = p(X = a_n)$
- Definition: Information content of a_n

$$I(a_n) = \log_2\left(\frac{1}{p_n}\right) = -\log_2(p_n)$$

- “Measure of surprise” associated with a_n

Information content (DMS)

- Properties
  - If $p_n = 1$, then $I(a_n) = 0$ (no surprise)
  - $I(a_n) \geq 0$
  - If $p_i < p_n$, then $I(a_i) > I(a_n)$
  - If a_i and a_n are independent,
$$I(a_i, a_n) = I(a_i) + I\,(a_n)$$
- Rq: no surprise $\rightarrow$ no additional information $\rightarrow$ "is not worth coding/transmitting"

Entropy: a measure of information

- A measure of uncertainty associated with a random variable
- Mean value of information content of X
- Average amount of “surprise” of X
- Mean value of information content per source symbol (communications)

Entropy (discrete source)

- X takes values from the alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ with probabilities $p_n = \mathbb{p}(X = a_n)$

$$H(X) = \mathbb{E}[I(a_n)] = \sum_{n=1}^{N} p_n I(a_n) = -\sum_{n=1}^{N} p_n \log_2 p_n$$

- Measured in b/symbol

Entropy (discrete source)

$$H(X) = -\sum_{n=1}^{N} p_n \log_2 p_n$$

- Property

$$0 \leq H(X) \leq \log_2 N$$

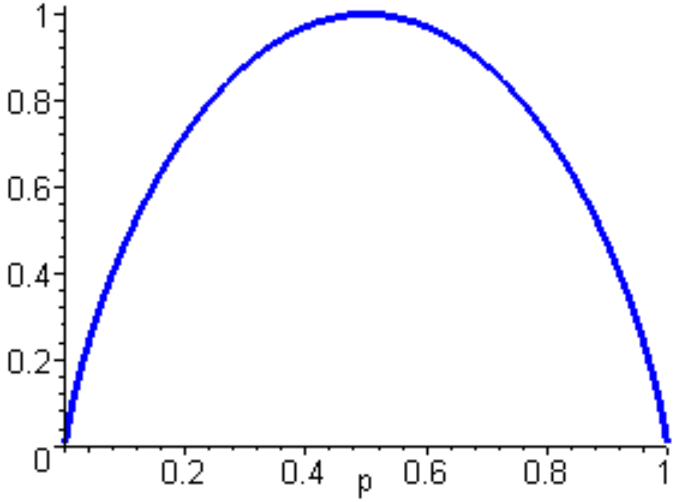
- Upper bound achieved with uniform distribution

Entropy (discrete source)

- Let X be a binary source that generates independent symbols 0 and 1 with equal probability ($\mathbb{p}(X = 0) = 0.5$).

$$H(X) = -\log_2 \frac{1}{2} = 1\text{b/symbol}$$

- In a more general case
- $\mathbb{p}(X = 0) = p$
- $H(X) = -p\log_2(p) - (1-p)\log_2(1-p)$



Entropy (discrete source)

- Consider a source X that produces five symbols with probabilities 1/2, 1/4, 1/8, 1/16, and 1/16. Determine the source entropy.

$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{2}{16}\log_2\left(\frac{1}{16}\right)$$

= 1.875 b\symbol

English text: entropy of letters

- Calculate the average information content in the English language, assuming that each of the 26 characters in the alphabet occurs with equal probability.

$$H(X) = -\log_2 \frac{1}{26} = 4.7 \text{ b/character}$$

- The actual entropy (accounting for frequencies)

$$H(X) = 2.62 \text{ b/character}$$

Average information content: “surprise” contains new information. Entropy is the average value of “surprise”. Accounting for letter frequencies makes the source more predictable

Information rate

- $H(X)$: entropy (b/symbol)
- r: time rate at which the source X emits symbols (symbols/s)
- Definition: Information rate

$$R = rH(X) \text{ b/s}$$

Information rate: example 1

A high-resolution black-and-white TV picture consists of about 10^6 picture elements and 32 different brightness levels. Pictures are repeated at the rate of 64 per second. All picture elements are assumed to be independent, and all levels have equal likelihood of occurrence.

Compute the average rate of information conveyed by this TV picture source.

Number of images possible: $N_i = 32^{10^6}$

- $H(X) = -\sum_{n=1}^{N_i} \frac{1}{N_i} \log_2 \frac{1}{N_i} = 5\times10^6$ b/image
- $r = 64$ images/s
- $R = rH(X) = 320$ Mb/s

Information rate: example 1

A high-resolution black-and-white TV picture consists of about 10^6 picture elements and 32 different brightness levels. Pictures are repeated at the rate of 64 per second. All picture elements are assumed to be independent, and all levels have equal likelihood of occurrence.

Compute the average rate of information conveyed by this TV picture source.

Alternative solution: $X = [X_1, \dots, X_{10^6}]$ concatenation of independent pixels

- $H(X) = H(X_1, \dots, X_{10^6}) = H(X_1) + \cdots + H(X_{10^6})$
- $H(X_1) = -\sum_{n=1}^{32} \frac{1}{32} \log_2 \frac{1}{32} = 5$ b/pixel
- $r = 64 \times 10^6$ pixels/s
- $R = rH(X_1) = 320$ Mb/s

Information rate: example 2

Consider a telegraph source having two symbols, dot and dash. The dot duration is 0.2 s. The dash duration is 3 times the dot duration. The probability of the dot's occurring is twice that of the dash, and the time between symbols is 0.2 s.

Compute the information rate of the telegraph source.

Entropy: $\mathbb{p}(dot) = 2\,\mathbb{p}(dash) = 2/3$

$H(X) = -\mathbb{p}(dot)\log_2 \mathbb{p}(dot) - \mathbb{p}(dash)\log_2 \mathbb{p}(dash) = 0.92$b/symbol

Symbol rate: $\mathrm{T}_s = \mathbb{p}(dot)\mathrm{t}_{\mathrm{dot}} + \mathbb{p}(dash)t_{dash} + t_{delay} = 0.5333$s/symbol

$$t_{dot} = 0.2s, \mathrm{t}_{\mathrm{dash}} = 0.6s, \mathrm{t}_{\mathrm{delay}} = 0.2s$$

$R = rH(X) = \frac{H(X)}{T_s} = 1.725$ b/s

Summary

- Information theory required to understand and build efficient transmission systems
- Source/channel coding
- Discrete memoryless source (DMS)
- Information content
- Entropy
- Information rate

Next chapter: source coding