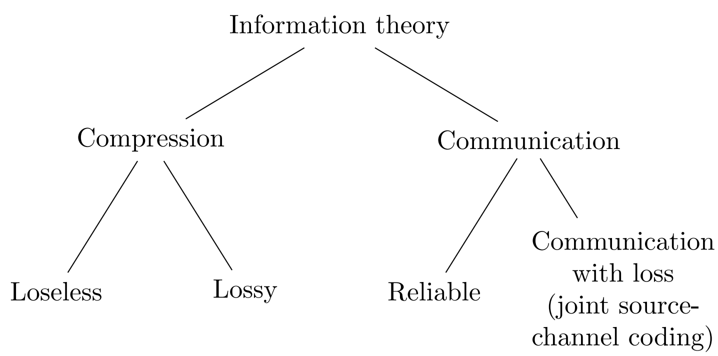
EE376A/STATS376A Information Theory Lecture 1 - 01/09/2018

Lecture 1: Introduction

Lecturer: Tsachy Weissman *Scribe: Denis Barberena, Seunghwan Lim, Suzannah Osekowsky*

1 Overview

Information theory is composed of the following topics.



After taking this course, you are expected to be familiar with the following concepts.

- measures of information: entropy, relative entropy, mutual information, chain rules.
- compression, storage, and communication of information.
- fundamental limits in information theory.
- concrete schemes for compression and communication.
- existence proofs via random construction ("random coding").
- "typical sequences" and interplay between information theory, probability and statistics

If you are more interested in information theory, you can branch out to following topics.

- *Communication*: digital, optical, wireless.
- *Coding*: algebraic coding, codes on graphs.
- *Compression & processing*: audio, image, video.
- *Statistical signal processing.*
- *Advanced information theory*: network information theory, universal schemes, wireless communication.

2 Logistics

- *Course Website*: https://web.stanford.edu/class/ee376a/
- *Lecture*: Tue & Thu, 12:00-1:20pm in Gates B1.
- *Weekly HW sets*: Handed out on Thu and due the following Thu 11:59pm on Gradescope.
- *Midterm*: Mon, Feb 12, 6-9pm.
- *Final*: Thu, March 22, 12:15-3:15pm.
- *Lecture scribes*: each student is responsible for one lecture due 48 hours after the lecture. (sign-up sheet on the website)
- *Final grade*: scribes 10%, HW 20%, midterm 30%, final 40%
- *Prerequisite*: EE178 or equivalent, and maturity & motivation to deal with abstract concepts. A second course that builds on and/or uses probability, e.g. EE278, is recommended.
- Lecture notes from the last two years are available on the website. One is more focused on core problems, the other is broader and more application-oriented.
- *Textbook*: "Elements of Information theory" by Cover & Thomas (recommended). Additional books are on the website.
- *TA*: Yanjun Han and Kedar Tatwawadi.
- *Instructor*: Tsachy Weissman. The office hour is Thursday 2:30-4pm or by appointment.
- Please join the course Piazza page.
- Course videos will be available on the website and via SCPD.

3 Examples

3.1 Lossless compression

Source is $U_1, U_2, \cdots \overset{\text{i.i.d}}{\sim} U \in \mathcal{U} = \{A, B, C\}$. The probability distribution of U is defined as $P(U = A) = 0.7$, $P(U = B) = 0.15$, $P(U = C) = 0.15$. Consider a coding that maps $u \in \mathcal{U}$ to bits. For example, we can map A to '00', B to '01', and C to '11'. This scheme requires 2 bits to store each symbol. Clearly, we should be able to do better since A occurs with such a high probability. For a better scheme, we can map

$$
\begin{aligned}
A &\rightarrow \text{`0'} \\
B &\rightarrow \text{`10'} \\
C &\rightarrow \text{`11'}
\end{aligned}
$$

Let $\bar{L}$ denote the average number of bits per symbol. For the coding above,

$$\bar{L} = 0.7 \times 1 + 0.15 \times 2 + 0.15 \times 2 = 1.3 \text{ bits/symbol}$$

Observe that allotting fewer bits to the more likely symbol gives us a lower $\bar{L}$. The coding scheme above is a *prefix code*, which means that no code word is a prefix of any other code words. This allows for linear time decoding by reading the compressed representation from left to right. As an example, one can easily check that 001101001101011 can be uniquely decode to AACABACABC.

Alternatively, we can code pairs of source symbols jointly as in Table 3.1. For this coding scheme, $\bar{L}$ is

Table 1: Coding for pairs of source symbols

pair	probability	Code word	Num. Bits Used
AA	0.49	0	1
AB	0.105	100	3
AC	0.105	111	3
BA	0.105	101	3
CA	0.105	1100	4
BB	0.0225	110100	6
BC	0.0225	110101	6
CB	0.0225	110110	6
CC	0.0225	110111	6

$$\begin{aligned}\bar{L} &= \frac{1}{2}(0.49 \times 1 + 0.105 \times 3 \times 3 + 0.105 \times 4 + 0.0225 \times 6 \times 4) \\ &= 1.1975 \text{ bits/symbol}\end{aligned}$$

We will later see that all compression schemes have $\bar{L}$ larger than or equal to source entropy of U,

$$H(U) = \sum_{u \in \mathcal{U}} p(u) \log_2 \frac{1}{p(u)} \simeq 1.1829$$

This sets a lower bound on the average number of bits per symbol. On the other hand, there is a positive result that says for any $\epsilon > 0$, there exists a scheme such that $\bar{L} \leq H(U) + \epsilon$; that is, we can come arbitrarily close to maximal efficiency, but *never actually achieve it* (we cannot achieve $\bar{L} = H(U)$ in this example; in general, there exist source distributions where the entropy can be exactly achieved). Thus the source entropy gives us a fundamental limit on compression.

3.2 Communication

Source is $U_1, U_2, \cdots \overset{\text{i.i.d}}{\sim} U \in \mathcal{U} = \{0, 1\}$ with $P(U = 0) = P(U = 1) = \frac{1}{2}$. Consider a channel that flips each bit with probability $q < \frac{1}{2}$. Let X_i denote the code word for U_i. We can write

$$Y_i = X_i \oplus W_i \text{ with } W_i \overset{\text{i.i.d}}{\sim} \mathsf{Bern}(q)$$

For example, we can let $X_i = U_i$. Then, the probability of error per source bit will be $P_e = q$. Alternatively, we can repeat each source code three times. For example, if $U = 0110\cdots$, then $X = 000111111000\cdots$.

The rate, which is ratio between the number of bits and channel input, is $\frac{1}{3}$ bits/channel use. The optimal decoding rule for this coding is taking majority for every triplet. In this case, the probability of error (per bit) is the probability that two or more bits in a given triplet are written in error; that is, $P_e = 3q^2(1-q)+q^3$, which is smaller than q.

Here we see a trade off between the rate and the probability of error: the second scheme lowers the probability of error but lowers the rate from 1 bit/channel use to $\frac{1}{3}$ bits/channel use.