Lecture 1

Introductions

course web site:

markwilde.com/teaching

review syllabus

overview of book & homeworks

office hours

rearrange time/days of class to be MW

What is quantum information theory?

Main goal is to address the following question:

"What are the fundamental limits of communication?"

Question asked & then addressed by Shannon in 1948 w/ a breakthrough paper.

- Shannon single-handedly introduced information theory to address this question
- he did not incorporate quantum mechanics, even though it was invented around 1925
- this came much later

- Shannon used entropy & mutual information to quantify uncertainty & correlations, respectively
- he also introduced concepts like typical sequences & random coding, as mathematical methods to address the fundamental question.

Let's briefly review Shannon's contributions.

1) data compression
2) channel coding

Data compression

Suppose that an information source randomly emits a symbol from $\{a, b, \ldots, g, h\}$ according to the histogram / probability dist.

What should we do for compression to ensure that we can represent the source faithfully such that error prob. in recovering is $\leq \varepsilon$?

if we use 3 bits to encode source, then we can recover it <u>perfectly</u>.

However if we use 2 bits, then we can recover w/ error prob. $\leq \varepsilon$.

<u>Strategy</u>: keep only the likely or typical symbols.

This basic idea generalizes beyond this simple example.

Now consider an i.i.d. information source modeled by a prob. distribution $p_X(x)$

Source emits n independent samples, modeled by random variables

$X_1, \ldots, X_n$ w/ prob. dist.

$$P_{X^n}(x^n) = \prod_{i=1}^{n} P_X(x_i)$$

$$\left(\text{shorthand: } \begin{array}{l} X^n \equiv X_1 \cdots X_n \\ x^n \equiv x_1 \cdots x_n \end{array} \right)$$

Recall the law of large numbers:

denote the expectation of $f(X)$ by

$$\mathbb{E}[f(X)] = \sum_x p(x) f(x)$$

& sample mean by

$$\overline{f}(X^n) = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

$\forall \varepsilon \in (0,1)$, $\delta > 0$ &
sufficiently large n,

$$\Pr\left[\left| \bar{f}(x^n) - \mathbb{E}[f(x)] \right| \le \delta \right] \ge 1-\varepsilon$$

Define typical set by picking

$$f(x) = \underbrace{-\log_2 P_X(x)}_{\text{called \underline{surprisal} of } x}$$

$$T_\delta^n \equiv \left\{ x^n \in \mathcal{X}^n : \left| -\frac{\log P_{X^n}(x^n)}{n} - H(X) \right| \le \delta \right\}$$

where entropy $H(X) = -\sum_x P_X(x) \log P_X(x)$

By LLN,

$$\Pr\left[X^n \in T_\delta^n \right] \ge 1-\varepsilon$$

$\forall$ sufficiently large n.

Size of typical set

$$|T_\delta^n| \le 2^{n[H(X)+\delta]}$$

follows from

$$1 = \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \ge \sum_{x^n \in T_\delta^n} p_{X^n}(x^n)$$

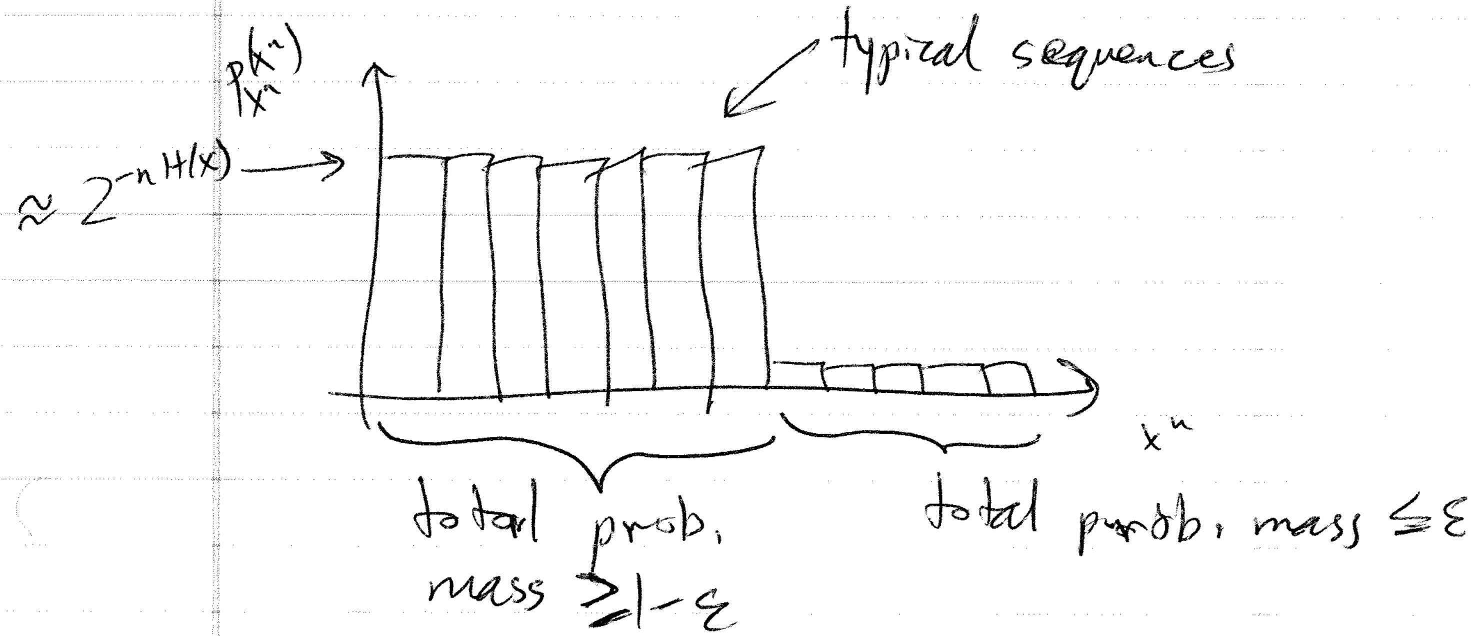
$$\ge \sum_{x^n \in T_\delta^n} 2^{-n[H(X)+\delta]}$$

$$= |T_\delta^n| \, 2^{-n[H(X)+\delta]}$$

Also if $x^n \in T_\delta^n$, then

$$2^{-n[H(X)+\delta]} \le p_{X^n}(x^n) \le 2^{-n[H(X)-\delta]}$$

So, for large n, histogram looks like this

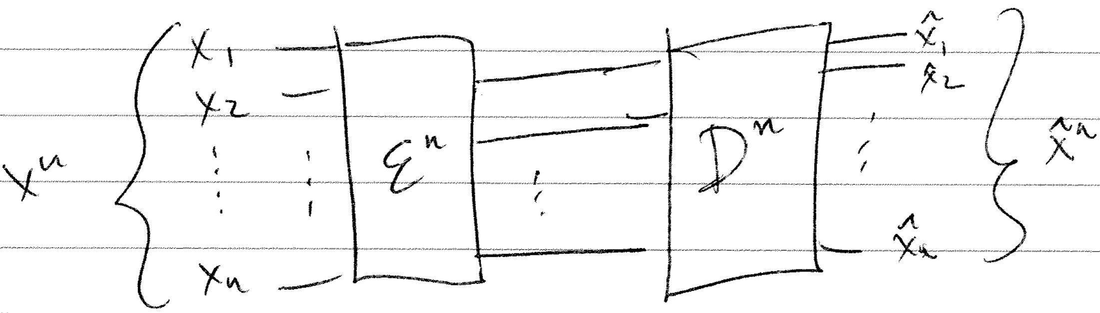


+ total # $\leq 2^{nH(X)}$

Shannon's idea for compression:

keep the typical sequences + discard the rest!!

More formally, scheme looks like



$$\mathcal{E}^n : \mathcal{X}^n \to \underbrace{\{0,1\}^{nR}}_{\text{size } 2^{nR}}$$

rate is $R = \frac{\#\text{bits}}{\text{symbol}}$

$$\mathcal{D}^n : \{0,1\}^{nR} \to \mathcal{X}^n$$

encoder is then just

receive x^n. if typical, compress to nR bits w/

$$R \approx H(X) + \delta$$

if not, set to all zeros

decoder: map from encoding of typical set back to T_δ^n.

then guaranteed that

$$\Pr\left[(D^n \circ E^n)(X^n) \neq X^n \right] \leq \varepsilon$$

$\forall \varepsilon \in (0,1)$, $\delta > 0$ + sufficiently large n.

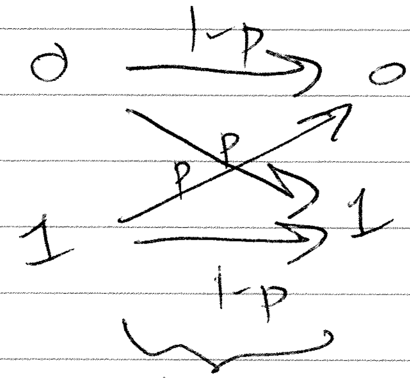
implies that entropy $H(X)$ is an achievable rate for data compression.

can also prove optimality of entropy rate

What about channel coding?

model a classical comm. channel by a conditional prob. distribution $p_{Y|X}(y|x)$

Simple example:
binary symmetric channel



transition prob. matrix

Simple idea: to reduce error,
use a repetition code
(repeating yourself to avoid mistakes in communication)

encode 0 as 000
& 1 as 111

rate is 1/3 # bits / channel use

Suppose usage of channel is i.i.d.

suppose 000 is input

then the channel output is described by the table

output	probability
000	$(1-p)^3$
001, 010, 100	$p(1-p)^2$
011, 110, 101	$p^2(1-p)$
111	p^3

Use majority vote decoder

What is error probability?

1st two rows are decoded correctly

last two are decoded incorrectly

error prob. when transmitting zero is then

$$3p^2(1-p) + p^3$$

Similar result when sending 1 due to symmetry of channel

total error probability is

$$\Pr\{e\} = \Pr\{e|0\}\Pr\{0\} + \Pr\{e|1\}\Pr\{1\}$$

$$= 3p^2(1-p) + p^3$$

$$= 3p^2 - 2p^3$$

when does coding help?

when error prob. is lower than without coding, i.e., when

$$3p^2 - 2p^3 < p$$

same as $0 < 2p^3 - 3p^2 + p$

$\Leftrightarrow$ $0 < p(2p-1)(p-1)$

$= p(1-2p)(1-p)$

i.e. when $p \in (0, 1/2)$