• Notes Set 24: Sketch of information theory

- Statistical entropy. Typical sequences. Conditional distributions and conditional entropy. Mutual information. Entropy of a signal sequence.Channel capacity. Capacity of bandlimited additive Gaussian channel.

ECE ______ Notes Set 24

Information theory

* Looks at communication problem with transmitter allowed to adjust message signalling method so as to maximize communication rate & minimize errors.

* Ends up contributing both ideas & methods to probability theory itself.

* Optimum codes (modulation methods, signalling methods) exist, but corresponding receiver architecture becomes unboundedly difficult → practicality impaired.

* Intimately related to statistical Thermo.

* Intimately related to ciphers & secure communication.

— Let's get started —

We will find that

information theory,

like stat thermo

... cannot deal properly with continuous random variables.

→ message & noise must, strictly speaking

be discrete r.v.'s.

→ noise can be allowed to become

continuous by limiting arguments, but message

cannot.

→ both information theory & stat thermo

saved by quantization of physics...

Statistical Entropy

random sequence x_i : $\quad x_i = \begin{cases} 0 & p \\ 1 & q = 1-p \end{cases}$ probability.

Look at sequence of length N : $\vec{X} = (x_1, \ldots \; x_N)$

$$p(\vec{X} = \vec{x}) = p^{(\#\,zeros)} \cdot q^{(\#\,ones)}$$

$$\#\,zeros = \sum x_i$$

and: $\#\,ones = \sum (1 - x_i)$

$$P(\vec{X}) = p^{\sum_n x_i} \, q^{\sum (1 - x_i)}$$

$$= p^{n\left[\frac{1}{n} \sum_n x_i\right]} \cdot q^{n\left[\frac{1}{n} \sum (1 - x_i)\right]}$$

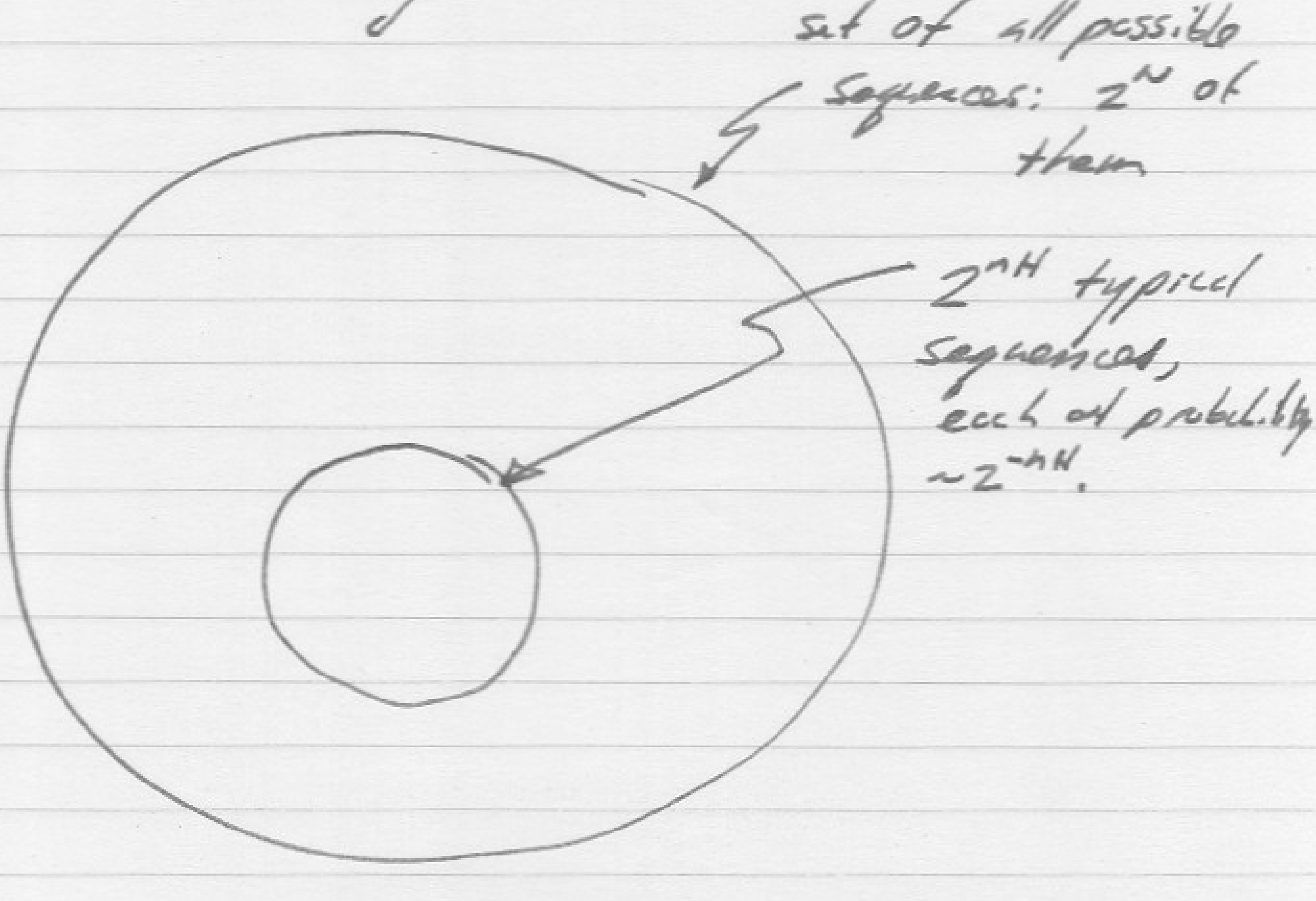
by Law of Large #'s: $\left.\begin{array}{l} \frac{1}{n} \sum_n x_i \Rightarrow p \\ \frac{1}{n} \sum_n (1 - x_i) \Rightarrow q \end{array}\right\}$ with probability one!

$$P(\vec{X}) = p^{np} \cdot q^{nq}$$

$$= \exp\left[\left[p \ln p + q \ln q\right] \cdot n\right]$$

$$= \exp\left[-nH\right]$$

[where $H = -p \ln p - q \ln q$
is the entropy.]



☆ The "smaller" set of 2^{nH} vs 2^N elements contains ~100% of the probable outcomes

A "few" "typical" sequences dominate all likely possibilities.

Recognize that this is the same argument we made about atoms in a gas in stat thermo.

Entropy

A discrete r. v. X which can take on values x_i with probability $p(x_i)$ has entropy:

$$H(X) = -\sum_{x_i} p(x_i) \log(p(x_i))$$

note that the base of the logarithm is up to us: we usually will assume $\log_2(\sim)$ so entropy is in "bits"

If we use $\log_e(\sim)$, entropy is in "nits"

Note entropy is <u>not</u> a function of $\underline{x}$

but of X's distribution:

ex: $x = \begin{cases} 1 & p = 1/2 \\ 0 & q = 1/2 \end{cases}$ $H_x = -(1/2 \log 1/2 + 1/2 \log 1/2)$

$= \underline{\underline{1 \quad \text{bit}}}$

$y = \begin{cases} 1 & p = 1/4 \\ 0 & p = 3/4 \end{cases} = H_y = -(1/4 \log 1/4 + 3/4 \log 3/4)$

$= 0.81$ bits.

$\underline{\underline{y}}$ is less random than $\underline{\underline{x}}$

Note that H is the only measure of "randomness"

Ex: v = gaussian, zero mean, $\sigma = 1$

w = gaussian, zero mean, $\sigma = 2$

is w "more random" than v ?

Yes? Would you say yes if I told you that $w = 2v$??

→ H is <u>only</u> measure of randomness

→ H only so for discrete r.v.'s.

EX//

$$Y = \begin{cases} 4 & p = 1/4 \\ 3 & p = 1/4 \\ 2 & p = 1/4 \\ 1 & p = 1/4 \end{cases}$$

$\rightarrow H = \_\_\_\_\_ = 2$ bits!

$$Y = \begin{bmatrix} 4 & p = 0.49 \\ 3 & p = 0.49 \\ 2 & p = 0.01 \\ 1 & p = 0.01 \end{bmatrix}$$

$$H = -\left[0.49 \log 0.49 + 0.49 \log 0.49 + 0.01 \log 0.01 + 0.01 \log 0.01\right]$$

$H = 1.14$ bits

Slightly more than one bit due to slight chance of taking on 3rd & 4th values.

Joint distributions:

$$X, Y \rightarrow p(x,y)$$

discrete space

$$p(y) = \sum_x p(x,y), \quad p(x) = \sum_y p(x,y)$$

~~ase~~

as before: $p(y|x) = p(x,y)/p(x)$

Conditional Entropy for specific value of $\underline{X}$

$$H(Y|X=x) = -\sum_y p(y|x) \log(p(y|x))$$

entropy for a particular value of x.

Conditional Entropy for a General Value of X:

$$H(Y/X) = \underset{x}{E}\left[H(Y/X=x) \right]$$

$$= -\sum_{x,y} p(x,y) \log p(y/x)$$

Joint Entropy

$$H(X,Y) = -\sum_{xy} p(x,y) \log(p(x,y))$$

"Bayes' rules" for Entropy:

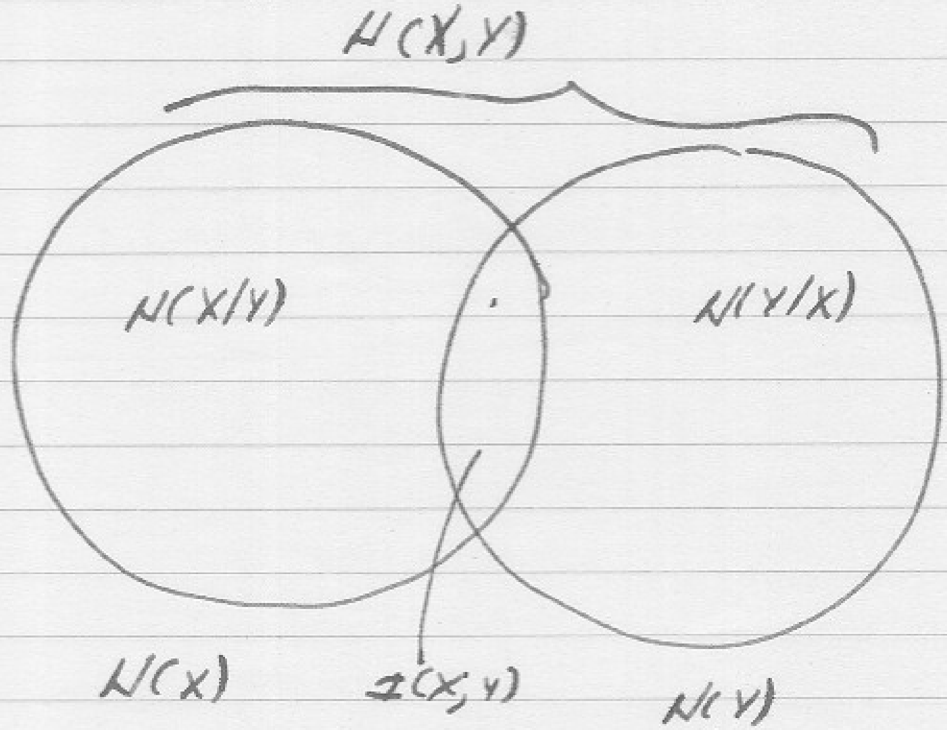
$$H(X,Y) = H(X) + H(Y/X)$$

Mutual Information

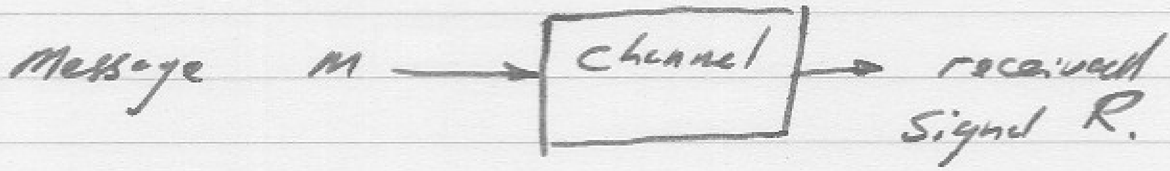
$$I(X;Y) = H(X) - H(X|Y)$$

$$= H(Y) - H(Y|X)$$

$$= H(X) + H(Y) - H(X,Y)$$



What is mutual information?



M is random: Entropy $H(M)$

R is random Entropy $H(R)$

when we receive R, we learn something about M

hence M becomes less random:

$$H(M|R) = H(M) - I(M;R)$$

our knowledge of M is improved by

an amount equal to the mutual information

Entropy of a signal sequence

Each pixel on a TV picture is quite correlated with

- past values
- neighbors vertically
- neighbors horizontally.

Sequence $\{X_i\} \rightarrow \vec{X}$ of infinite length...

sequence entropy: $H(\vec{X}) = \lim_{n \to \infty} H(X_{n+1} / X_n, \ldots X_1)$

$\Rightarrow$ Entropy of n^{th} bit conditioned on knowledge of all prior bits, for n large.

- The Entropy of a sequence will be much smaller than the entropy of a single sample, for sequences with <u>any</u> degree of pattern dependence or memory:

  Entropy of a sequence: entropy "per bit"

First assertion:

* a sequence with rate B bits/sec
  with sequence entropy H

* can be communicated (with probability 1)
  with a bit sequence of rate (BH)

This is the rate coding theorem at the heart of

≡ hard disc compressors.

≡ video & voice digital compressors
for low binary rates...

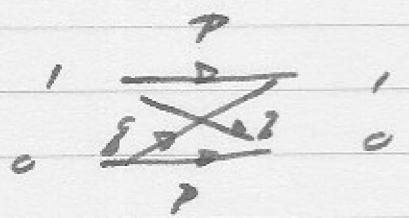
≡ TIFF image compression, etc.

Channel Capacity

$$x \longrightarrow \boxed{\text{ch.}} \longrightarrow y$$

$$p(y|x)$$

Ex:



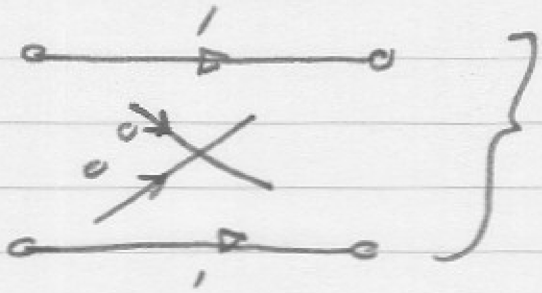
binary symmetric channel.

Channel Capacity

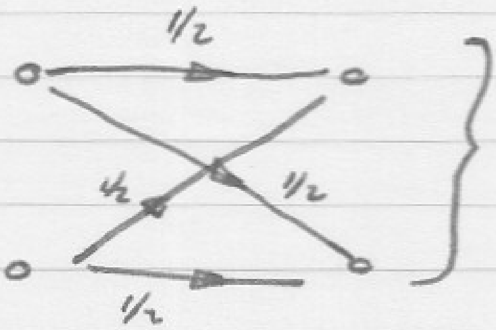
$$\boxed{C = \max_{p(x)} \left[I(x:y) \right]}$$

we pick the transmitted symbol probabilities $p(x)$ to maximize $I(x;y)$.

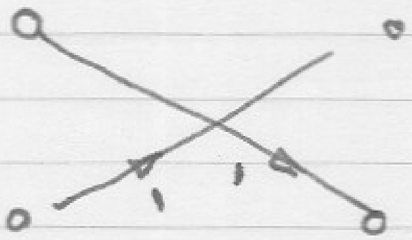
Ex:



$C = 1$ bit per transmissio

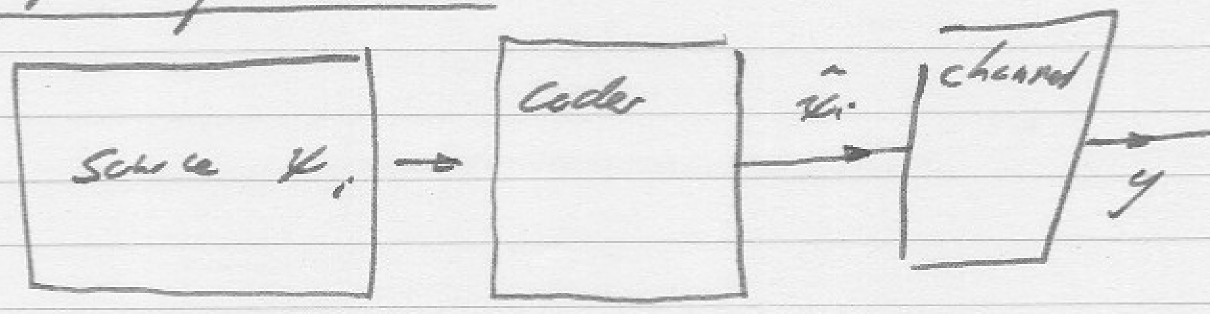


$C = 0$ bits



$C = 1$ bits!

Capacity theorem



with a source of per-bit entropy H,

codes exists to transmit messages to

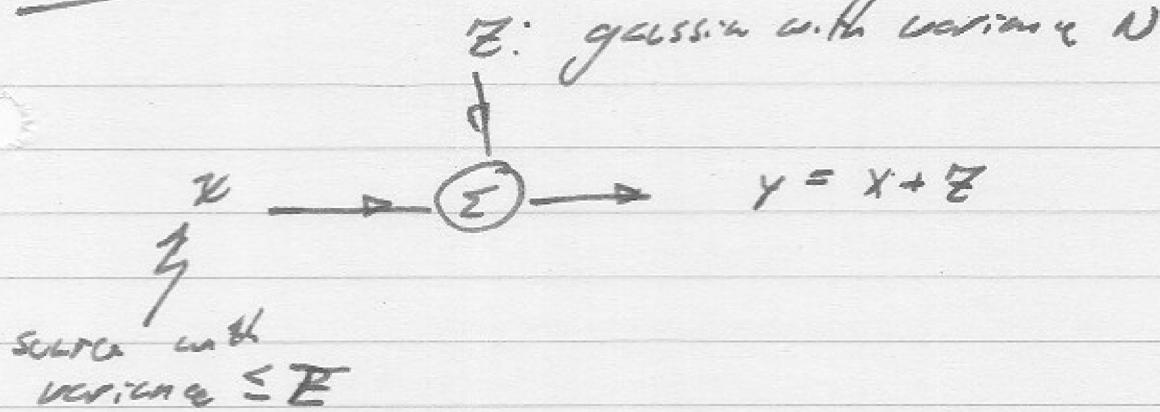
the receiver

— with arbitrarily small error for long code length —

— If the source entropy is

below the channel capacity.

Gaussian channel



$$C = \frac{1}{2}\log_2\left[1 + E/N\right]$$ bits/transmission

$$= \frac{1}{2}\log_2\left[1 + E_b/N_0\right]$$ Bits/degree of freedom

Now lets increase transmission rate:

~~E~~ symbol rate $= R$ = spectral Bandwidth used

= ~~H~~ Hz/sec.

$$E = P/R$$

& total capacity per sec

$$\tilde{C} = CR = \frac{R}{2} \log_2\left[1 + \frac{P}{\mathcal{N}_0 \cdot R}\right]$$

make $R \to \infty$

$$\tilde{C} = \frac{R}{2} \frac{1}{\ln 2} \ln\left[1 + \frac{P}{\mathcal{N}_0 R}\right]$$

$$\cong \frac{R}{2} \frac{1}{\ln 2} \cdot \left[\frac{P}{\mathcal{N}_0 R}\right]$$

$$\tilde{C} = \frac{P}{N_0}\left(\frac{1}{2 \ln 2}\right)$$

Need $N_0 \cdot (2 \cdot \ln 2)$ Watts/bit.