Welcome to Transmission of Information (6.441)

In this course we will study:
- Mathematics behind modelling and transmission of information.
- How do you quantify information?
- How do you model communication channels?
- What tools are available to study manipulation of information.

Administrivia:
- Please follow course website at http://theory.csail.mit.edu/~madhu/ST06
- Make sure you're on class's email list.

- Course Staff:
  - Madhu Sudan: madhu@mit.edu
  - Chung Chan: chungc@mit.edu

- Grading:
  - 4 Problem Sets
  - 1 Midterm
  - 1 Course Project
  - 1 Scribe work

- Background: Probability (6.041)
Motivational Problem

- Satellite flying through unknown space communicating back to Earth;
- Sensor measures some quantity (say temperature);
- Satellite has a transmitter transmitting one bit/unit time.
  Transmission is noisy.
  \[ x \]
  Can we transmit all the info we have?
  \[ \therefore \]
- Need to quantify information at source
- Need to model + quantify transmission
Modelling sensor's information (rate)

Temperature at time \( t+1 \)

\[ \mathbb{E}[X_{t+1}] = X_t. \]

\[ P_s \left[ |X_{t+1} - X_t| \geq k \right] \leq 8^{-k}. \]

Modelling transmission channel

\[
\begin{array}{c}
0 \xrightarrow{.99} 0 \\
0 \xrightarrow{.01} 7 \xrightarrow{.99} 1 \\
1 \xrightarrow{.01} 2 \xrightarrow{.99} 1
\end{array}
\]
Idea 3: Compress data at source

Relevant info: at each time, suffices to transmit \( y_t = x_t - x_{t-1} \).

\[ y_t = 0 \quad \text{w.p.} \quad \frac{7}{128} \Rightarrow 0 \]

\[ = +1 \quad \text{w.p.} \quad \frac{7}{128} \Rightarrow 100 \]

\[ = -1 \quad \text{w.p.} \quad \frac{7}{128} \Rightarrow 101 \]

\[ = +2 \quad \text{w.p.} \quad \frac{7}{1024} \Rightarrow 1100 \]

\[ = -2 \quad \text{w.p.} \quad \frac{7}{1024} \Rightarrow 1101 \]

\[ \vdots \]

\[ E[\text{Encoding length}] = 7 + \ldots \]
On the other hand transmission rate = l - 

Conclusion:

1. Transmission Impossible
2. Calculations not good enough
3. Model no good?

Turns out: Answer is 2

Better upper bound on rate 

Suppose we buffer 100 units of time & then send stuff.
- Expect to see $\frac{7}{8} \times 100 \approx 87$ 0's.
- Expect to see $\frac{7}{64} \times 100 \approx 11 \pm 1$'s.

and $\leq 2 \left| \gamma_{+} \right| \geq 2$.

$\text{Exp}[\text{encoding length}] \leq 3$ bits

**Transmission Protocol**

1. First xmit: location of 0's

needs $\log_2 \left( \frac{100}{87} \right)$ bits $\approx 5.3$ bits

2. location of 2's and longer;
\[ \log_2 (13) \text{ bits} \approx 7 \text{ bits} \]

3. Value of \((\pm 1) \approx 1\) bit.

4. Value of \((\pm 2, \pm 3, \ldots) \).

Expected \leq 6 \text{ bits}.

Sum total (module errors in MATLAB) \leq 27 \text{ bits}.

On the other hand have 100 slots.

But what about errors?

Little more complex \ldots \ldots
- Satellite gets feedback on perfect channel.
- Gets to know where error was.
- Can add info on how to correct with next block of 100.
- Comet into = Expect 1 error
  \[ \log_2 100 \approx 7 \text{ bits} \]
So adds \( 27 + 7 = 84 \) bit

next time.

Seems feasible?

Is this right?

What we need

1. More rigorous analysis
2. Better compression
3. No feedback

- Prob. Theory
  - Energy & Information
AEP: “Expectation” \[\rightarrow \]

Source Coding

Channel Coding

“Joint - Source - Channel coding”.

Eventually:

- Continuous r.v.
- Gaussian Noise
- Network Inf. Theory
- Applications: Gambling
  Stock Markets
Today: Review Probability.

Yesterday's notes:

- Random variables: can be anything
  - Value of Google stock
  - Gender of random voter
  - “Pick a city at random”

- For real-valued r.v.’s we can talk about expected value

\[ E[X] = \sum_{x \in \Omega} x \cdot P(x) \]
R.V. R.V. Events
\[ \uparrow \quad \uparrow \] Expectation vs. Probabilities
\[ \downarrow \quad \downarrow \]
Easy to manipulate \iff\ Harder to manipulate

"Meaningless?" \quad Move meaningful.

**Expectation equations:**
\[ E[ x_1 ] + E[ x_2 ] = E[ x_1 + x_2 ] \]

**Prob. Inequality**
\[ P_Y[ \bar{E}_1 \cup \bar{E}_2 ] \leq P_Y[ \bar{E}_1 ] + P_Y[ \bar{E}_2 ] \]
Tail Bounds

for $x \geq 0$, $k > 0$ [Markov's]

$$\Pr \left[ x \geq k \cdot \mathbb{E}[x] \right] \leq \frac{1}{k^2}$$

for every $x, k > 0$

$$\Pr \left[ \left| x - \mathbb{E}(x) \right|^2 \geq k^2 \left( \mathbb{E}[x^2] - \mathbb{E}[x]^2 \right) \right] \leq \frac{1}{k^2}$$

Let $\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$

(Exercise: Prove $\text{Var}[x] \geq 0$)

$\sigma[x] = \sqrt{\text{Var}[x]}$

Chebyshev bound

$$\Pr \left[ \left| x - \mathbb{E}[x] \right| \geq k \cdot \sigma[x] \right] \leq \frac{1}{k^2}$$
Conditioning

Knowing Event $E_1$ has happened, there are still some unknowns.

\[
\Pr [E_2 | E_1] \times \Pr [E_2 \cap E_1] \quad \frac{\Pr [E_1]}{\Pr [E_2]}
\]

$E_1$, $E_2$ are independent if \[ \Pr [E_1 | E_2] = \Pr [E_1] \quad \Pr [E_2] \Pr [E_2]
\]

Example: Random decreasing sequence \([A_{10}, 11, 12, 13, 14, 15] \]

$E_1 = \text{"10" appears in sequence}$

$E_2 = \text{"11" appears in sequence}$

are they independent?
Many r.v.'s

\( X, Y \) random variables with

joint distribution \( P \)

(i.e., \( P(x, y) \) is \( P[X=x, Y=y] \))

\[
P_x(x) = \sum_{y \in \mathbb{R}_y} P(x, y) \quad \leftarrow \text{Marginal Distribution}
\]

\[
P_y(y) = \sum_{x \in \mathbb{R}_x} P(x, y)
\]

\( x, y \) independent if

\(\forall y \in \mathbb{R}_y \)

\[P[X=x | Y=y] = P_X(x).
\]

Equivalent \( P(x, y) = P_x(x) \cdot P_y(y). \)
Chernoff - Hoeffding Tail Bounds

X₁, ..., Xₙ i.i.d. with mean μ

Then \[ \Pr \left[ \left| \frac{\sum X_i}{n} - \mu \right| \geq \sqrt{n} \right] \leq e^{-\frac{\varepsilon^2 n}{2}} \]

In particular, if \( \mu = \frac{1}{2} \)

\[ \Pr \left[ \sum X_i > \left( \frac{1}{2} + \varepsilon \right)n \right] \leq e^{-\frac{\varepsilon^2 n}{2}} \]...