ST06 : LECTURE 1

Note Title 2/6/2006

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 web site 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 : <u>Probability</u> (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?

— x —

- Need to quantify information at source
- Need to model + quantify transmission

capacity.

Modelling Sensor's information (rate)

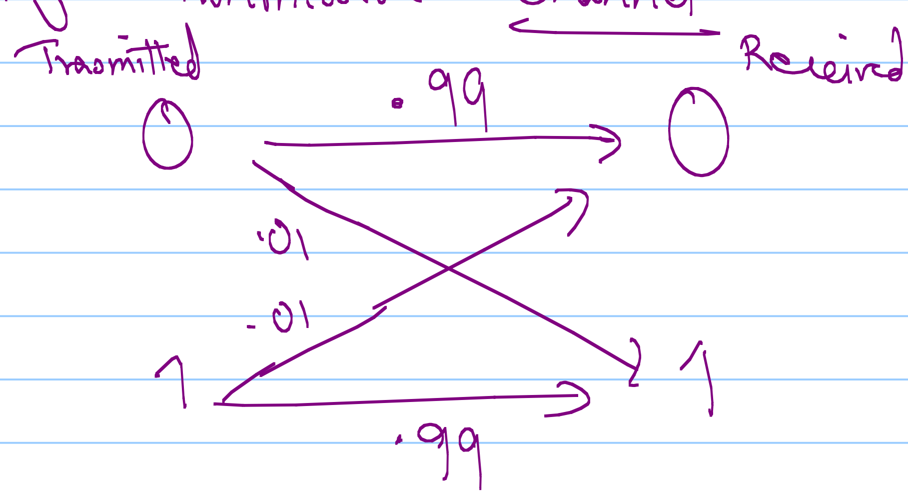
Say temperature at time $(t+1)$ X

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

$$\Pr[|X_{t+1} - X_t| \geq k] \leq 8^{-k}$$

Probability Modelling

Modelling Transmission Channel



So can we do this?

Idea 3: Compress data at source

Relevant info: at each time suffices to transmit $Y_t = X_t - X_{t-1}$.

$Y_t = 0$ w.p. $7/8$ $\Rightarrow$ 0

$= +1$ w.p. $7/128$ $\Rightarrow$ 100

$= -1$ w.p. $7/128$ $\Rightarrow$ 101

$= +2$ w.p. $7/1024$ $\Rightarrow$ 1100

$= -2$ w.p. $7/1024$ $\Rightarrow$ 1101

$\vdots$ $\vdots$

$E[\text{Encoding length}] = 1 + \dots$

On the other hand transmission rate = 1 − ... .

Conclusion:

(1) Transmission impossible

OR (2) Calculations not good enough

OR (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 $\frac{7}{64} \times 100 \approx 11 \pm 1$ 1's.

and ≤ 2 $|Y_t| \geq 2$.

$\downarrow$

Exp[encoding length] $\leq$ 3 bits

Transmission Protocol

(1) first xmit location of 0's

needs $\log_2 \binom{100}{87}$ bits $\approx$ 53 bits

(2) location of 2's and larger;

$\log \binom{13}{2}$ bits $\simeq$ 7 bits

③ Value of $(\pm 1) \sim 11$ bits.

④ Value of $(\pm 2, \pm 3 \ldots)$:

Expected ≤ 6 bits.

Sum total (modulo errors in MATLAB)

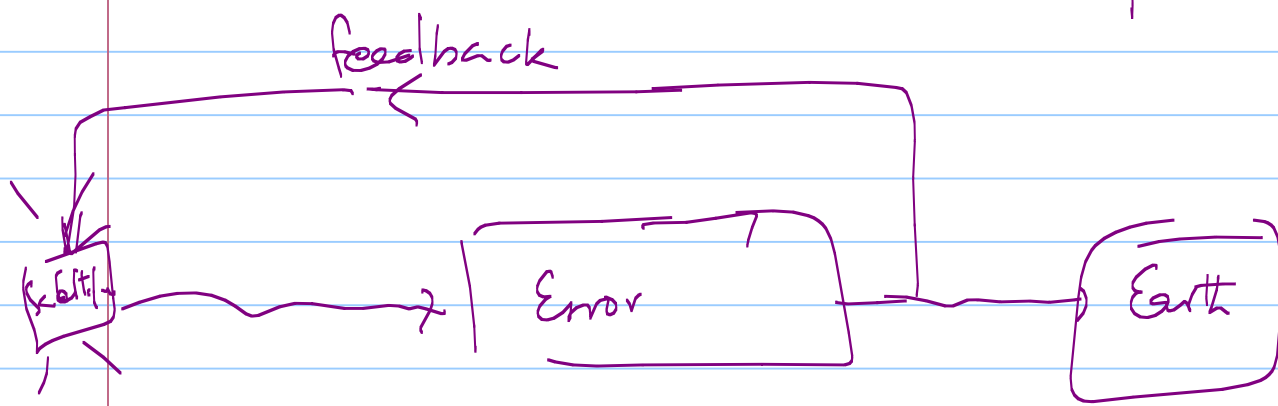
≤ 77 bits.

On the other hand have 100 slots.

But what about errors?

Little more complex . . . . .

So let's make more assumptions.



- Satellite gets feedback on perfect channel.
- Gets to know where error was
- Can add info on how to correct with next block of 100.
- Correct info ≈ Expect 1 error

$\log_2 100 \approx 7$ bits

So adds $77 + 7 = 84$ bits

next time.

Seems feasible?

Is this right?

What we need

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

- Prob. Theory
  - Entropy & Information

AEP : "Expectation"

Source Coding

Channel Coding

"Joint - source - channel coding".

Eventually:

- Continuous R.V.
- Gaussian Noise
- Network Inf. Theory
- Applications : Gambling
  Stock Markets

~~2~~ 3rd

Today : Review Probability.

(Ω, P)

R.V. $X \sim P$

$\forall x \in \Omega,\ P(x) \geq 0 \qquad \sum_{x \in \Omega} P(x) = 1$

- Random variables : Can be anything
  - value of Google stock
  - Gender of random voter
  - "Pick a city at random"
- For real-valued r.v.'s can talk about expected value

$$E[X] = \sum_{x \in \Omega} x \cdot P(x)$$

R.V. R.V. Events

Expectations vs. Probabilities

Easy to manipulate $\longleftrightarrow$ Harder to manipulate

"Meaningless?" More meaningful.

Expectation equations:

$$E[X_1] + E[X_2] = E[X_1 + X_2].$$

Prob. Inequality

$$P_r[\mathcal{E}_1 \cup \mathcal{E}_2] \leq P_r[E_1] + P_r[E_2]$$

"Tail Bounds"

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

$$\left[\Pr\left[X \geq k \cdot E[X]\right] \leq \frac{1}{k} \right]$$

for every X, $k > 0$

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

Let $Var[X] = E[X^2] - E[X]^2$

(Exercise Prove $Var[X] \geq 0$]

$$\sigma[X] = \sqrt{Var[X]}$$

Chebychev bound

$$\Pr\left[|X - E[X]| \geq k\, \sigma[X] \right] \leq \frac{1}{k^2}$$

Conditioning

Knowing Event $\mathcal{E}_1$ has happened, there are still some unknowns.

$$\Pr[\mathcal{E}_2 \mid \mathcal{E}_1] \stackrel{\Delta}{=} \frac{\Pr[\mathcal{E}_2 \cap \mathcal{E}_1]}{\Pr[\mathcal{E}_1]}$$

$\mathcal{E}_1, \mathcal{E}_2$ are independent

if $\Pr[\mathcal{E}_1 \mid \mathcal{E}_2] = \Pr[\mathcal{E}_1]$.

for ind. $\mathcal{E}_1, \mathcal{E}_2$
$\Pr[\mathcal{E}_1 \cap \mathcal{E}_2]$
$= \Pr[\mathcal{E}_1] \cdot \Pr[\mathcal{E}_2]$

Example: Random decreasing sequence. [A-Vali]

$\mathcal{E}_1$ = "10" appears in sequence.

$\mathcal{E}_2$ = "11" " " .

are they independent?

Many r.v.'s

X, Y random variables with joint distribution P

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

$$P_X(x) = \sum_{y \in \Omega_Y} P(x,y)$$ ← Marginal Distribution

$$P_Y(y) = \sum_{x \in \Omega_X} P(x,y)$$

X, Y independent if

$\forall y \in \Omega_Y$

$$P[x \mid Y=y] = P_X(x).$$

Equivalent $P(x,y) = P_X(x) \cdot P_Y(y).$

Chernoff - Hoeffding Tail Bounds

r.v. R.V. $X_i \in [0,1]$

$X_1 \ldots X_n$ i.i.d with mean μ

thm $\Pr\left[\left|\frac{\sum x_i}{n} - \mu\right| \geq \lambda\sqrt{n}\right] \leq e^{-\lambda^2/2}$

in particular, if $\mu = \frac{1}{2}$

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

Next lecture: New quantities related to Prob. Spaces.