6.441 Transmission of Information	Feb 7, 2005
Lecture 01	
Lecturer: Madhu Sudan	*Scribe: Guner Celik*

GENERAL COURSE INFORMATION

Lecturer:Madhu Sudan, madhu@mit.edu

TA: Chung Chan, chungc@mit.edu

Website: http://stellar.mit.edu/S/course/6/sp06/6.441/index.html

Text: Elements Of Information Theory by Cover and Thomas

- MIT Engineering Library
- CSAIL Reading Room
- Should you decide to buy, try the online retailers

Grading:

- 4 Problem Sets: First one out today, due approximately two weeks (encourage collaboration, but write-ups separate, mention all sources)
- 1 Midterm
- 1 Project
- 1 Script

Course Topics

- Mathematics of information transmission
- Quantify Information
- Quantify "capacity" of a channel
- How do you manipulate these quantities?

AN EXAMPLE OUTLINING THE CONCEPTS

- Satellite through space
- Sensor measuring temperature
- Transmitter beams back one bit/time (Transmission is not perfectly accurate)

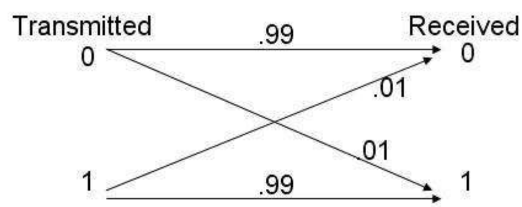
Is communication feasible?

Sensor: $X_0, X_1, X_2, ..., X_t, X_{t+1}$
X_t: Temperature at time t: integer

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

$$\begin{aligned} X_{t+1} \quad &= X_t \quad w.p.\ 7/8, \\ &= X_t + 1\ w.p.\ 7/128, \\ &= X_t - 1\ w.p.\ 7/128, \\ &\ldots \end{aligned}$$

Transmission Channel:



Each time unit is independent of past and future.

$$Y_t = X_t - X_{t-1} \Rightarrow$$

$$\begin{aligned} Y_t \quad &= \ \ 0\ w.p.\ 7/8, \Rightarrow 0 \\ &= +1\ w.p.\ 7/128 \Rightarrow 100 \\ &= -1\ w.p.\ 7/128, \Rightarrow 101 \\ &= +2\ w.p.\ ....., \Rightarrow 1100 \\ &= -2\ w.p.\ ....., \Rightarrow 1101 \\ &\ldots \end{aligned}$$

Then, $E[encoding\, length] > 1\ and\ E[capacity] < 1\, bit/time$

Idea1: Transmit Y_t each t : Didn't work
Idea2: Buffer information for 100 units of time

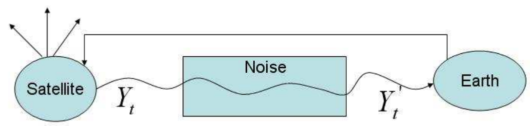
"Rate at which information is produced?"

-Expect to see 87 zeros $\Rightarrow log_2\binom{100}{87} \approx 53bits$
-11 of these are +/- 1's $\Rightarrow log_2\binom{13}{11} \approx 7bits$

-11 bits for +/- 1's.
-2 Symbols $Y_t's, |Y_t| \geq 2$
-Also Expected cost $\leq$ 3 bits per Y_t

$\Rightarrow$ Total Cost: 77 bits, seems good since we improved it from 100 bits.

Simplifying Assumption: Feedback



This feedback is assumed errorless, which is not a reasonable assumption. Since the probability of success in the channel is 0.99 we expect 1 bit error in every 100 bits (which is the same assumption as we did before, and it is also not very accurate) Therefore, there are 100 different possibilities for the earth station to transmit this error back to satellite. As a result, at least $log_2 100$ bits are required for the feedback channel. Including the 77 bits from the previous analysis, we end up with approximately 84 bits which is lower than 100 bits, showing that the transmission with buffering of 100 bits may be successful as opposed to transmitting Y_t each t.

Under the light of this example, we can be more specific about the course topics.

Course Topics In More Detail

- Probability Theory Review (today)
- Entropy And Information (next few lectures)
- Asymptotic Equipartition Property (Information Theorists' Law of Large Numbers)
- Source Coding (Rate At Which A Source Is Producing Symbols)
- Channel Coding (Discrete Channels With Discrete Error)
- Continuous Channels (Gaussian Error)
- Network Information Theory (Application In Non-communication Setting; Stock Market, Gambling etc.)

REVIEW OF PROBABILITY THEORY

Probability Space: (Ω, F, P)
Underlying ground set:Ω
F: Power set of Ω
Events: $E \subseteq \Omega, E \in F$
If Ω is a finite set and X is a random variable $\sim$ (distributed according to)

P then $P(X) \geq 0 \, for\, every\, x \in \Omega$, $\sum_{x \in \Omega} P(x) = 1$

Real Valued Random Variable X$\sim$P, then expected value of X is defined as: E(X)=$\sum_{x \in \Omega} xP(x)$

Expectation of Real Valued R.V. $\Leftrightarrow$ Probability Of Events

Indicator Random Variable

E: event$\Rightarrow$

$$\begin{aligned} 1_E(X) \quad &= 1 \;\; if X \in E \\ &= 0 \, otherwise \end{aligned}$$

$\Rightarrow E[1_E] = Pr[E]$

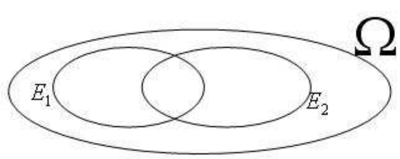
Manipulation Tools

1. $Pr(E_1 \bigcup E_2) \leq Pr(E_1) + Pr(E_2)$, $\;\; E(X_1 + E_2) = E(X_1) + E(X_2)$
2. Expectation $\Rightarrow$ Probability :$X \geq 0$, $Pr[X \geq kE(X)] \leq \frac{1}{k} \;\; Markov's\, Inequality$
3. $Pr[(X - E(x))^2 \geq k^2 E[(X - E(X))^2] \leq \frac{1}{k^2} \Rightarrow$

   By definition, $Var(X) = E(X^2) - E(X)^2$ then

   $Pr[|X - E(X)| \geq k\sqrt{Var(X)}] \leq \frac{1}{k^2} \; also \; \sqrt{Var(X)} = \sigma(X) : StandartDeviation$

Conditional Probabilities



Event E_1 has occurred, how does this change the probability space?

$$\Omega \rightarrow E_1$$

$$Pr(E_2/E_1) = \frac{Pr(E_1 \bigcap E_2)}{Pr(E_1)}$$

Independence: E_1 and E_2 are independent if $Pr(E_1/E_2) = Pr(E_1)$

Example: Random Decreasing Sequence: 1,2,3,........,100 $\rightarrow$ pick a number say 54, then 1,2,......,54 $\rightarrow$ pick a number, say 27, then 1,2,....,27

$\rightarrow$ pick a number and decrease the sequence and so on.

E_{10} = Event that the number 10 appears in this sequence; Pr(E_{10})
E_{11} = Event that the number 11 appears in this sequence; Pr(E_{11})
Question: Are these events independent?

Joint Distributions on (X,Y) One can find the marginal distributions from the joint distribution.

Chernoff-Hoefding Bound

$X_1, X_2, ..., X_n$ are iid (independent identically distributed)

$$X_i \in [0,1] \; with \; E(x) = \mu$$

$$Pr[|\frac{\sum_{k=1}^{n} X_i}{n} - \mu| \geq \epsilon] \leq e^{\frac{-\epsilon^2 n}{2}} \qquad \epsilon = \frac{1}{n^{\frac{1}{3}}}$$