6.441 Transmission of Information April 6, 2006

Lecture 14

Lecturer: Madhu Sudan *Scribe: Adam McCaughan*

Today

- Feedback Capacity
- Joint Source Channel Coding
- Start Continuous Channels

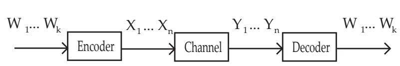
Admin

- PS3 due Thursday (04/12)
- Tuesday 4:15pm in 32-155 – Venkat Guruswami "'Channel Coding..."'

Feedback Capacity

- Recall basic model of a channel

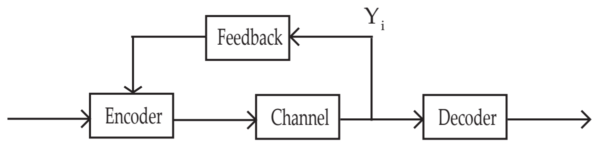
- In order to ask how well the channel performs, we apply an encoder and decoder



- Which more or less pins down exactly how the channel performs

$$C = p_x^{\max} I(X;Y)$$

How much capacity do we get with feedback?



- In other words, given $Y_1, \ldots, Y_n$ what is the maximal R such that the receiver can compute $W_1, \ldots, W_{k=Rn}$ (where $w_i \in \{0, 1\}$) with a $P_e rror \to 0$
- Denote this maximal R to be the 'feedback channel capacity', C_{FB}
- It's obvious that if you just construct an encoder with zero feedback, you're able achieve at least C, ie $C_{FB} \geq C$
- Now the question remains: Is it possible to improve capacity with feedback? Short answer: No. Proving this shows the strength of Shannon's coding theorem.

Lemma 1 **($C_{FB} \leq C$)**

- $H(W) = Rn$ - The entropy of W is fairly large
- $H(W|Y^n) \leq 1 + P_{error}Rn$. Fano's Inequality
- If $H(W|Y^n)$ wasn't small we wouldn't be able to calculate W, given Y
- These two points imply that Y^n contains a lot of information

$$
\begin{aligned}
I(W; Y^n) &= H(W) - H(W|Y^n) \\
&\geq R_n - 1 - P_e rror Rn
\end{aligned}
$$

- $P_{error}Rn$ is vanishingly small

Question: Is $I(W; Y^n) \leq nC$ **?**

$$
\begin{aligned}
I(W; Y^n) &= H(Y^n) - H(Y^n|W) && (1) \\
&\leq \sum_{i=1}^{n} H(Y_i) - H(Y^n|W) && (2)
\end{aligned}
$$

$Y_1 \ldots Y_{i-1}, W$ is enough to fully determine X_i:

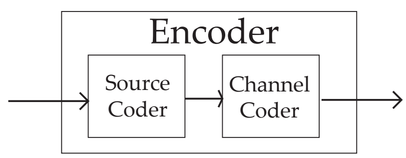
$$
\begin{aligned}
H(Y^n|W) &= \sum_{i=1}^{n} H(Y_i|Y_1 \ldots Y_{i-1}, W) \\
&= \sum_{i=1}^{n} H(Y_i|Y_1 \ldots Y_{i-1}, W, X_i) \\
&= \sum_{i=1}^{n} H(Y_i|X_i)
\end{aligned}
$$

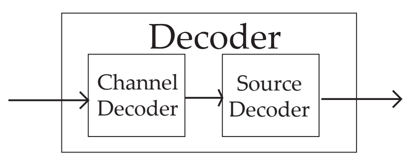
Then from (2)

$$
\begin{aligned}
I(W;Y^n) &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&= \sum_{i=1}^{n} I(Y_i;X_i) \\
&= nC
\end{aligned}
$$

Conclusion: Feedback doesn't contribute to capacity

- Previously we always looked at either
  - Uniform distributions on the source with a noisy channel
  - Clean channels with non-uniform sources
- We have now learned enough to combine non-uniform source with a noisy channel
- Simply need to look at the rate of the source, and the capacity of the channel. Then compare R and C
  - Apply compression algorithm to the source
  - Apply channel coding algorithm





Joint Source-Channel Coding Theorem

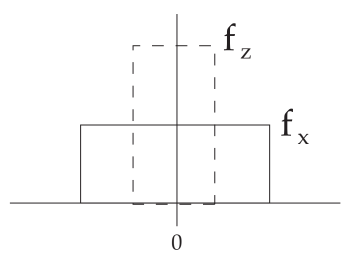
- If $W_1 \ldots W_k$ is produced by source $\mathscr{W}$ with entropy rate $H(\mathscr{W}) \rightarrow$ (source satisfies AEP)
- and if it's on a DMC with capacity C
  - Then communcation is possible with $P_{error} \rightarrow 0$ iff $H(\mathscr{W}) < C$
- After n steps of the source, it's producing on a uniform distribution of size $2^{H(\mathscr{W})n}$
- This concludes our discussion of the DMC

Continuous Channels

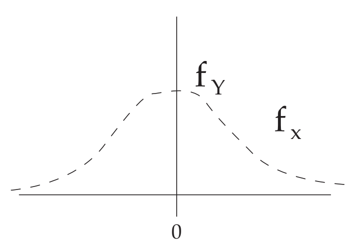
- We will begin by looking at a very simple channel
  - Input to channel X: $[-1, 1]$ (Real number)
  - Output of channel Y: Real number
  - Looking at the simplest case: noiseless channel (ie $X = Y$)
- Can't look at this in our typical manner because we can't define a finite alphabet to describe either input or output
- However, since $X = Y$, our channel capacity is apparently infinite
- What makes the channel capacity finite is the existance of noise
- **Adding noise Z to our model**
  - $Y = X + Z$
  - Z is uniform over $[-\epsilon, \epsilon]$ and independent of X
  - Divide input into intervals of 2ϵ – 'discretize it'
  - Then $C \geq \log(1/\epsilon)$
  - We will prove that the capacity is less than infinite in the future

Continuous Random Variables

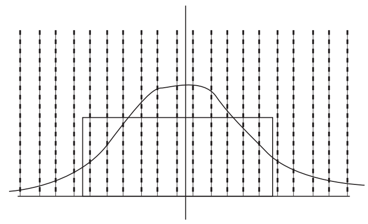
- X is a real-valued r.v.
- $f_X(x) \rightarrow$ Probability Density Function (PDF)
- $F_X(x) \rightarrow$ Cumulative Distribution Function (CDF)
  - $F_X(x) = \Pr[X \leq x] = \int_{-\infty}^{x} f_X(t)dt$
  - Monotonic, nondecreasing
- Given



- It is clear that just the pdf of the r.v. is not particularly revealing. However, comparing X and Z, one can certainly intuit that X is 'more random' than Z. How, then, do we quantitatively compute that?
- Because this is not as easy to interpret:



- X_ϵ: X discretized X by intervals of length ϵ, $Y_\epsilon \in \mathbb{Z}$



- $\lim_{\epsilon \to 0}\{H(X_\epsilon) - H(Y_\epsilon)\}$?
- Say we partition ϵ lots more: $\epsilon \to \frac{\epsilon}{2^l}$
- $\to H(\frac{X_\epsilon}{2^l}) \approx l + H(X_\epsilon)$
- and the same thing is happening to Y
- For $X+Y$, $\lim_{\epsilon \to 0}\{H(X_\epsilon) - H(Y_\epsilon)\}$ is well-behaved, but we want a quantity that only depends on X. For X alone what should we use as $H(X)$?

Differential Entropy

- $H(X) \triangleq \lim_{\epsilon \to 0}\{H(X_\epsilon) - f(\epsilon)\}$
- (We're going to be measuring against something like a baseline distribution)

$$f(\tfrac{\epsilon}{2^l}) = l + f(\epsilon) \to f(\epsilon) = \log(\tfrac{1}{\epsilon})$$
$$\text{then } H(X) = \lim_{\epsilon \to 0}\{H(X_\epsilon) + \log \epsilon\}$$

- Written in terms of the pdfs:

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log[f_X(x)]dx$$

Examples

Example 1 - Entropy of the Uniform Distribution

$$X = \text{uniform}(a, b)$$
$$f_X(x) = \tfrac{1}{b-a} \text{ if } a \leq x \leq b, \text{ 0 otherwise}$$
$$H(X) = -\int_a^b \tfrac{1}{b-a} \log(\tfrac{1}{b-a})dx = \log(b-a)$$

- Not scale invariant.

Example 2 - Entropy of a Gaussian

$$X = N(0, \sigma^2)$$
$$f_X(x) = \tfrac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{x^2}{2\sigma^2}} \text{ if } a \leq x \leq b, \text{ 0 otherwise}$$
$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x)dx$$
$$H(X) = \tfrac{1}{2} \log(2\pi \exp \sigma^2)$$

- Logarithmic in the variance

Future Lectures

- Try to understand how differential entropy behaves
- Look at AEP/LLN in this setting
- Continuous channels and how dif. entropy and mutual information play a role in determining capacity