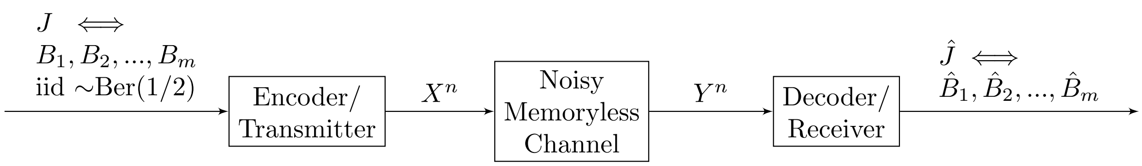
**EE376A Information Theory** Lecture 10 - 02/10/2015

# Lecture 10

*Lecturer: Tsachy Weissman* *Scribe: Daniel Miller and Gabriele Carotti-Sha*

## 1 Communication Problem (Chapter 7)

### 1.1 Recap




- rate $= \frac{m}{n} \ \frac{bits}{channel\ use}$
- $P_e = \mathbb{P}(\hat{J} \neq J)$
- R is achievable if $\forall\ \epsilon > 0,\ \ \exists$ a scheme $(m,\ n,\ encoder,\ decoder)$ with $\frac{m}{n} \geq R$ and $P_e < \epsilon$.
- Capacity: $C = \sup\{$R: R is achievable.$\}$

"Channel Coding Theorem": $C = \max_X I(X;Y)$ (sometimes written as $\max_{P_X}$)

*Note:* The Channel Coding Theorem is equally valid for analog signals, e.g., the AWGN channel. However, we must extend our definition of the various information measures such as entropy, mutual information, etc.

### 1.2 Information Measures for Continuous Random Variables

1. **Relative Entropy**
   with f, g pdfs:
   $$D(f||g) = \int f(x)\ log\frac{f(x)}{g(x)}dx, \quad x \in \mathbb{R}.$$
   Can similarly define for $x \in \mathbb{R}^n$.

2. **Mutual Information**
   with X, Y $\sim f_{X,Y}$ (X, Y are jointly continuous with distribution $f_{X,Y}$):
   $$I(X;Y) = D(f_{X,Y}||f_X \times f_Y),$$
   where $f_X \times f_Y$ is the product of the marginal distributions.

3. **Differential Entropy**
   with X $\sim f_X$:
   $$h(X) = E[-\log f_X(X)]$$
   with X, Y $\sim f_{X,Y}$:
   $$h(X,Y) = E[-\log f_{X,Y}(X,Y)] \qquad \text{(“Joint Differential Entropy”)}$$

$$h(X|Y) = E[-\log f_{X|Y}(X|Y)] \qquad \text{("Conditional Differential Entropy")}$$

Each of the above definitions is totally analogous to the discrete case.

### 1.3 In Homework

$$\begin{aligned} I(X;Y) &= h(X) + h(Y) - h(X,Y) \\ &= h(X) - h(X|Y) \\ &= h(Y) - h(Y|X) \end{aligned}$$

This is the main/only interest in differential entropy.

*Note:* Unlike discrete entropy H(X), differential entropy can be positive or negative. This is not the only way in which they differ.

$$\begin{aligned} h(X+c) &= h(X), & &\text{for constant c} \\ h(X \cdot c) &= h(X) + \log|c|, & &c \neq 0 \end{aligned}$$

### 1.4 Gaussian Distribution

**Claim:** The Gaussian distribution has maximal differential entropy, i.e.,:
If $X \sim f_X$ with $E[X^2] \leq \sigma^2$ (second moment), and $G \sim N(0, \sigma^2)$,
Then $h(X) \leq h(G)$, with equality iff $X \sim N(0, \sigma^2)$.

*Note:* If $E[X^2] \leq \sigma^2$ and $Var(X) = \sigma^2$, then necessarily $E[X] = 0$.

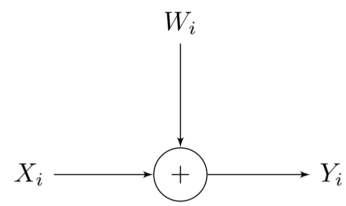
**Proof of Claim:**

$$f_G(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}X^2}, \quad -\log f_G(X) = \log\sqrt{2\pi\sigma^2} + \frac{\frac{1}{2\sigma^2}X^2}{\ln 2}$$

$$\begin{aligned} 0 \leq D(f_X||f_G) &= E\left[\log\frac{f_X(X)}{f_G(X)}\right] \\ &= -h(X) + E\left[\log\frac{1}{f_G(X)}\right] \\ &= -h(X) + E\left[\log\sqrt{2\pi\sigma^2} + \frac{\frac{1}{2\sigma^2}X^2}{\ln 2}\right] \\ &\leq -h(X) + E\left[\log\sqrt{2\pi\sigma^2} + \frac{\frac{1}{2\sigma^2}G^2}{\ln 2}\right] \\ &= -h(X) + E\left[\log\frac{1}{f_G(G)}\right] \\ &= -h(X) + h(G) \end{aligned}$$

$$\therefore h(X) \leq h(G), \text{ with equality iff } D(f_X||f_G) = 0, \ i.e., \ X \sim G$$

# 2 Example III: AWGN Channel (Additive White Gaussian Noise)




*Note:* The AWGN channel is memoryless.

- Transmission is restricted to power $P$:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P, \quad \forall n \in \mathbb{N}$$

- $R$ is achievable with power $P$ if: $\forall\ \epsilon > 0,\ \ \exists$ scheme restricted to power $P$ and with rate $\frac{m}{n} \geq R$ and probability of error $P_e < \epsilon$.
- Channel Capacity: $C(P) = \sup\{\text{R: R is achievable with power P}\}$

## 2.1 Channel Coding Theorem for this Setting

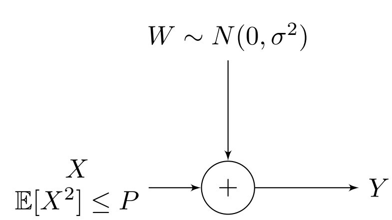
$$C(P) = \max_{\mathbb{E}[X^2] \leq P} I(X;Y)$$

*Note:* We could instead have considered the restriction that $\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i^2] \leq P$. This constitutes a relaxed constraint. However, it turns out that even with the relaxation, you cannot perform any better in terms of the fundamental limit.

## 2.2 An Aside: Cost Constraint

More generally, we can consider an arbitrary cost function constraint on X, rather than the above power constraint. We can denote this cost function by $\phi(X_i)$. The cost constraint is then $\frac{1}{n}\sum_{i=1}^{n} \phi(X_i) \leq \alpha$. This means that the average cost cannot exceed the core parameter $\alpha$, so we consider $C(\alpha)$. In this case, the coding theorem becomes $C(\alpha) = \max_{\mathbb{E}[\phi(X)] \leq \alpha} I(X;Y)$.

## 2.3 The Example



$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Y|X) \\
&= h(Y) - h(Y-X|X) && \text{(given X, X is a constant, so we can use invariance of differential entropy to constant shifts)} \\
&= h(Y) - h(W|X) \\
&= h(Y) - h(W) && \text{(since W and X are independent)} \\
&\leq h(N(0, P+\sigma^2)) - h(N(0,\sigma^2)) && (Var(Y) = Var(X+W) = Var(X) + Var(W) \leq P + \sigma^2) \\
&= \frac{1}{2}\log 2\pi e(P+\sigma^2) - \frac{1}{2} 2\pi e\sigma^2 \\
&= \frac{1}{2}\log\frac{P+\sigma^2}{\sigma^2} \\
&= \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right)
\end{aligned}
$$

So in conclusion,

$$I(X;Y) \leq \frac{1}{2} log \left(1 + \frac{P}{\sigma^2}\right)$$

with equality

$$
\begin{aligned}
&\iff \quad Y \sim N(0, P+\sigma^2) \\
&\iff \quad X \sim N(0, P)
\end{aligned}
$$

Therefore, equality is achievable. So,

$$C(P) = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right)$$

(i.e., the capacity of the AWGN channel.)

## 2.4 Rough Geometric Interpretation (Picture)

- Transmission Power Constraint: $\sqrt{\sum_{i=1}^n X_i^2} \leq \sqrt{np}$
- Noise: $\sqrt{\sum_{i=1}^n W_i^2} \approx \sqrt{n\sigma^2}$
- Channel Output Signal:

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n Y_i^2\right] &= \mathbb{E}\left[\sum_{i=1}^n (X_i + W_i)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n X_i^2 + \sum_{i=1}^n W_i^2\right] \\
&\quad \text{(independence} \Rightarrow \text{cross-terms have zero expectation)} \\
&\leq nP + n\sigma^2 \\
&= n(P+\sigma^2)
\end{aligned}
$$

See **Figure 1** for the geometric interpretation of this problem. We want the high probability output balls to not intersect. This way, we can uniquely distinguish the input sequences associated with any given output sequence.

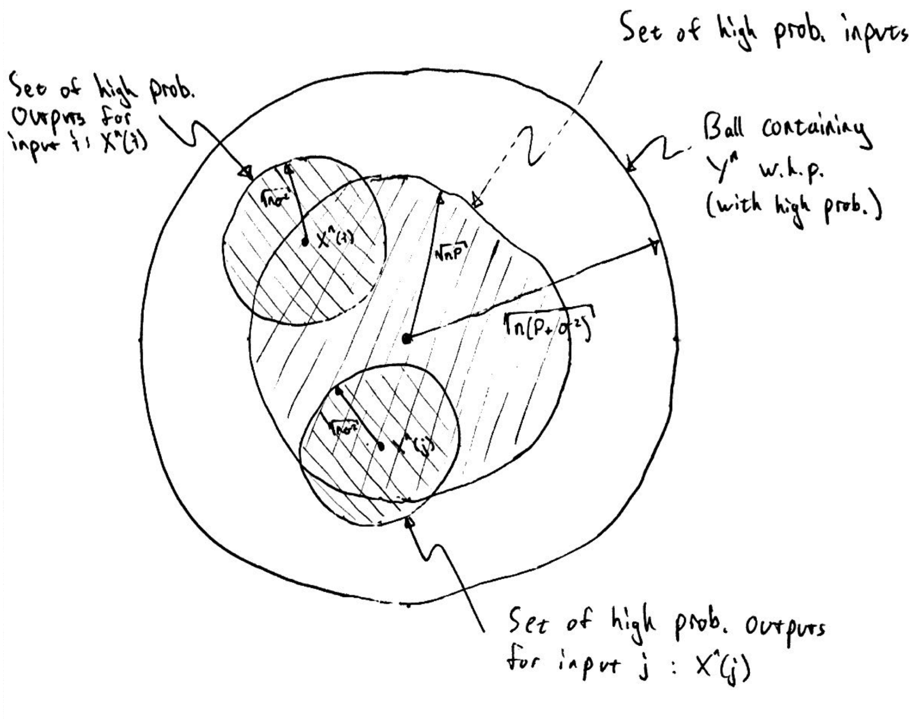



**Figure 1:** Geometrically, consider the input and output sequences as points in $\mathbb{R}^n$.

$$\# \text{ messages} \leq \frac{\text{Vol(n-dim ball of radius } \sqrt{n(P+\sigma^2)})}{\text{Vol(n-dim ball of radius } \sqrt{n\sigma^2})}$$

This inequality is due to inefficiencies in the packing ratio. Equality corresponds to perfect packing, i.e. no dead-zones. So,

$$\# \text{ of bits} = \frac{K_n(\sqrt{n(P+\sigma^2)})^n}{K_n(\sqrt{n\sigma^2})^n} = \left(1 + \frac{P}{\sigma^2}\right)^{n/2}$$

$$\Rightarrow \text{rate} = \frac{\log \# \text{ of messages}}{n} \leq \frac{1}{2}\log\left(1 + \frac{P}{Q}\right)$$

The achievability of the equality indicates that in high dimension, can pack the balls very effectively.