Digital Communication Theory

Ky-Anh Tran

December 2021

Here is a collected notes on Digital Communications.

Contents

§1 Resources

- *Thomas Cover, Elements of Information Theory*: Anything entropy, shannon capacity, refer to this.
- *Cioffi notes*: Each chapter covers the equivalent of a small book. However lots of practical information nowhere else to be found (especially on equalization).
- *Bane, Vasic*: Surprisingly good chapters on partial response channel capacity, coding (both RS and LDPC), and timing recovery.

§2 Foundations

§2.1 Information Theory

Entropy, facts and definitions:

- Denote the **entropy** of random variable X with probability distribution $p(x)$ to be:

$$H(X) = -\langle \log(p(x)) \rangle$$

- For a continuous random variable X with density $f(x)$, the **differential entropy** $h(X)$ is denoted to be:

$$h(X) = -\int_S f(x) \log f(x) dx$$

  where S is the support of X.
- Similarly denote the **joint entropy** for 2 random variables X, Y with distributions $p(x, y)$ to be:

$$H(X, Y) = -\sum_{x,y} p(x, y) \log(p(x, y)) = -\langle \log(p(x, y)) \rangle$$

- Denote the **conditional entropy** of $H(Y|X)$ to be:

$$H(Y|X) = \langle \log(p(y|x)) \rangle$$

$$\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x) \log(p(y|x))$$

$$\sum_{x,y} p(x, y) \log(p(y|x))$$

- **Chain rule**:

$$H(X, Y) = H(X) + H(Y|X)$$

- The **kullback leibler** distance or *relative entropy* between 2 probability mass distributions p(x) and q(x) is defined as:

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = \langle \log \frac{p(x)}{q(x)} \rangle$$

  It is always non-negative, but it is not symmetric.

- The **mutual information** I(X, Y) is the relative entropy between the joint and product distribution:

$$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Example 2.1

- The differential entropy for uniform $x \in [0,a]$ is:

$$h(X) - \int_0^a dx \frac{1}{a} \log(\frac{1}{a}) = \log(a)$$

- The differential entropy for gaussian variable with $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ is:

$$h(X) = \frac{1}{2} \log(2\pi e \sigma^2)$$

Lemma 2.2

The gaussian distribution between 2 random variable x, y maximizes the differential entropy for any given positive semi-definite autocorrelation $R_{x,y}$:

$$p_{x,y} = \frac{1}{\sqrt{\pi \det(R_{xy})}} \exp\left(-\mathbf{u}^T R_{xy}^{-1} \underbrace{\mathbf{u}}_{\equiv(x,y)}\right)$$

Theorem 2.4.1 (*Mutual information and entropy*):

$$I(X;Y) = H(X) - H(X|Y)\,, \tag{2.43}$$

$$I(X;Y) = H(Y) - H(Y|X)\,, \tag{2.44}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)\,, \tag{2.45}$$

$$I(X;Y) = I(Y;X)\,, \tag{2.46}$$

$$I(X;X) = H(X)\,. \tag{2.47}$$

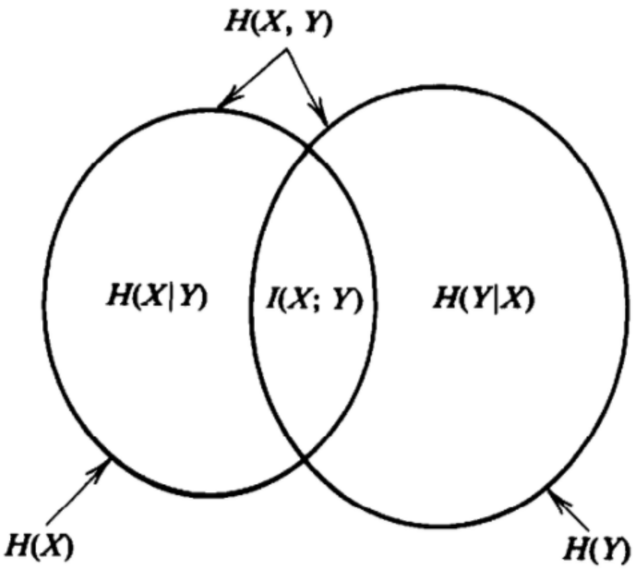

Figure 2.2. Relationship between entropy and mutual information.

Definition 2.3. A statistic t(X) of some random variable X is **sufficient** *for underlying parameter* θ if the conditional probability distribution of the data X given t(X) does not depend on θ. It means $I(X,\theta) = I(t(X),\theta)$

Definition 2.4.

Theorem 2.5

Consider a signal $Y(t) = X(t) + N(t)$ where $X \subseteq \mathcal{L}_2$ with an orthonormal basis $\mathcal{S}\{\phi_k\}$. Denote N(t) to be a white gaussian noise process with respect to $\mathcal{S}$. Then the set of measurements $\langle\phi_k|X\rangle$ form a set of sufficient statistics for detection of X(t) from Y(t).

Proof. See 6.451 notes, section 2.4 □

§2.2 Channel Capacity

cioffic ch. 2

- The **channel capacity** in bits/subsymbol for a channel described by $p(y|x)$ is defined by

$$\mathcal{C} = \max_{p(x)} I(x,y) \ bits/subsymbol$$

- A slightly more fancy way to describe capacity is to describe the maximum mutual information with respect to a transmit and receive sequence of subsymbols: $\mathbf{x^n} \equiv (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})$ and $\mathbf{y^n} = (\mathbf{y_1}, ..., \mathbf{y_n})$.

$$C = \lim_{n\to\infty} \frac{1}{n} \max_{p(\mathbf{x^n})} I(\mathbf{x^n}, \mathbf{y^n})$$

  The maximization $p(\mathbf{x^n})$ is taken over all probability density functions $p(\mathbf{x^n})$ which satisfy the symbol energy constraint given by

$$\langle x_k^2 \rangle \leq E_s$$

- An **Additive White Gaussian Noise** channel is a channel where

$$y(t) = x(t) + n(t)$$

  where $n(t)$ is white.

Theorem 2.6
Given a channel with capacity $\mathcal{C}$, then there exists a code with bitrate $b < \mathcal{C}$ such that $P_e \leq \delta$ for any $\delta > 0$. Furthermore, if $b > \mathcal{C}$, then $P_e \geq$ positive constant, which is typically large even for b slightly greater than $\mathcal{C}$.

Theorem 2.7
Given an AWGN channel, the channel capacity is

$$C = \frac{W}{2} \log_2 (1 + SNR)$$

It's important to notice a few things about this formula:

- The channel capacity obviously depends on the constraint on the transmit probability distribution. The probability $p(x)$ distribution that maximizes capacity and gives the formula $\frac{W}{2} \log_2 (1 + SNR)$ is gaussian. The capacity when constrained to be different (PAM-2 symbols etc...) is in general a difficult optimization problem with no closed form solution.

§3 Detection and Estimation

§3.1 Biased SNR

Consider a signal processing system where we would like to slice an output

$$y = \underbrace{\alpha}_{\text{adaptive gain}} (\underbrace{x}_{\text{symbol}} + \underbrace{n}_{\text{uncorrelated noise}})$$

Maximizing the SNR, or minimizing the MSE will lead to a **biased** gain factor α which is slightly less than 1. To show this, we just need to minimize the error, defined as:

$$\text{MSE} = \langle (y-x)^2 \rangle = (\alpha - 1)^2 \epsilon_x + \alpha^2 \epsilon_n$$

$$\frac{\partial}{\partial \alpha}\text{MSE} = 0 \rightarrow \alpha = \frac{\epsilon_x}{\epsilon_x + \epsilon_n} = \frac{SNR}{1 + SNR}$$

It is straightforward to make the decisions unbiased by scaling α by $\frac{SNR+1}{SNR}$, which leads to a relation between **biased** and **unbiased** SNR.

$$SNR = SNR_U + 1$$

The reason why one would care is that a biased decision rule, while maximizing SNR, may not optimize BER. We will further make rather trivial comments

- The distinction between biased and unbiased decreases as the SNR improves. This is why for SERDES links, people rarely care about the distinction.
- While the analysis was done for a simple gain adaptation loop, all the conclusion remains for FFE adaptation. In that case, if one uses a MMSE algorithm for adaptation, one will end up with a **biased** decision rule, while if one uses a ZF algorithm, one will end up with a **unbiased** adaptation.

§4 Math