Channel Coding and Hypothesis Testing

Shashank Vatedka

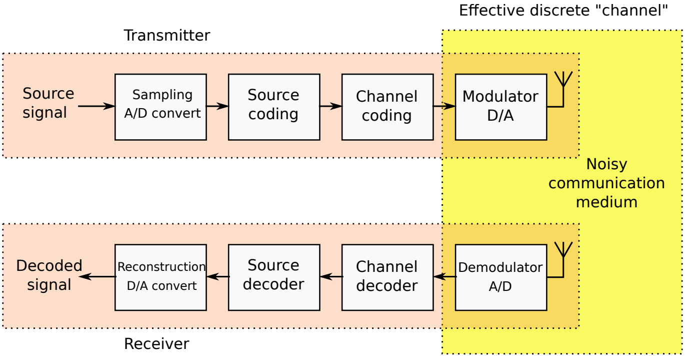
Recap

- Memoryless sources
- Data compression: rate, probability of error
- Source coding theorem
- Entropy:

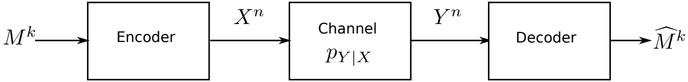
$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x).$$

Channel coding

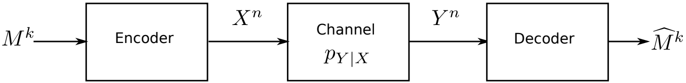
Digital Communication system



Discrete memoryless channel



Discrete memoryless channel



- $M^k \sim$ iid Unif$(\{0,1\}^k)$
- Memoryless channel:

$$p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} p_{Y|X}(y_i|x_i)$$

Common channels

Binary symmetric channel: BSC(p)

$\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and

$$p_{Y|X}(y|x) = \begin{cases} 1 - p & \text{if } x = y \\ p & \text{if } x \neq y. \end{cases}$$

Common channels

Binary erasure channel: BEC(p)

$\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, e\}$

$$p_{Y|X}(y|x) = \begin{cases} p & \text{if } y = e \\ 1 - p & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

Common channels

Additive white Gaussian noise (AWGN) channel

$\mathcal{X} = \mathcal{Y} = \mathbb{R}$

$$Y_i = x_i + Z_i, \qquad i = 1, 2, \ldots, n$$

where $(Z_1, \ldots, Z_n)$ are iid with $\mathcal{N}(0, \sigma^2)$ components.

Power constraint:

$$\|x^n\|^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i^2 \leqslant nP$$

Common channels

Complex slow/quasi-static fading channel

$\mathcal{X} = \mathcal{Y} = \mathbb{C}$

$$Y_i = hX_i + Z_i,$$

Common channels

Fast fading channel

$$Y_i = h_i X_i + Z_i,$$

Common channels

Multiple antenna/multi-input multi-output (MIMO) channels

$\mathcal{X} = \mathbb{R}^{t_s}$, $\mathcal{Y} = \mathbb{R}^{t_r}$.

$$\underline{Y}_i = \mathsf{H}_i \underline{X}_i + \underline{Z}_i, \qquad i = 1, \dots, n$$

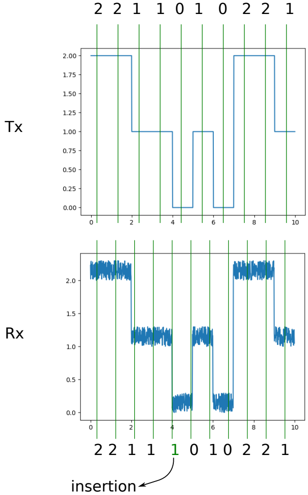
Common channels

A simple channel with memory

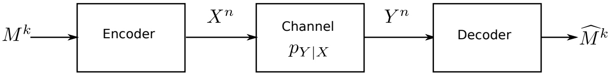
$$Y_i = a_0 X_i + a_1 X_{i-1} + \ldots + a_k X_{i-k} + Z_i$$

Common channels

Insertion/deletion channels



Channel codes



- Encoder: $f : \{0,1\}^k \to \mathcal{X}^n$
- Decoder: $g : \mathcal{Y}^n \to \{0,1\}^k$
- Rate:

$$R = \frac{k}{n}$$

- Probability of error:

$$P_e = \Pr[\widehat{M}^k \neq M^k]$$

Mutual information

$$I(X;Y) \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{XY}(x,y)\log_2 \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)},$$

Mutual information

$$I(X;Y) \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{XY}(x,y)\log_2\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)},$$

- Mutual information is symmetric
- Measures the information that X gives about Y, or Y gives about X.
- What happens if X and Y are independent?

Channel capacity

Maximum rate R for which $\lim_{n\to\infty} P_e = 0$.

Theorem (Shannon)

$$C = \max_{p_X} I(X;Y).$$

Capacity of BSC

$$C = 1 - H_2(p)$$

- Hamming distance: $d_H(x^n, y^n)$

Capacity of BSC

$$C = 1 - H_2(p)$$

- Hamming distance: $d_H(x^n, y^n)$
- Chernoff bound: $d_H(X^n, Y^n) \leqslant np(1 + \epsilon)$ with high probability

Capacity of BSC

$$C = 1 - H_2(p)$$

- Hamming distance: $d_H(x^n, y^n)$
- Chernoff bound: $d_H(X^n, Y^n) \leqslant np(1+\epsilon)$ with high probability
- Channel coding as a packing problem

Relation between entropy and mutual information

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Classification

Back to classifying spam

Model:

$$\text{Input email } X_1, \ldots, X_n \text{ iid} \sim p_X$$

$$p_X = \begin{cases} p_s & \text{if spam} \\ p_g & \text{not spam (good)} \end{cases}$$

Want

$$Pr[\text{declare not spam}|\text{spam}]$$

to be as small as possible subject to

$$Pr[\text{declare spam}|\text{not spam}] \leqslant \epsilon$$

Optimal test

$$\text{Output} \begin{cases} \text{spam} & \text{if } \log_2 \frac{p_s(x^n)}{p_g(x^n)} > \alpha \\ \text{not spam} & \text{if } \log_2 \frac{p_s(x^n)}{p_g(x^n)} \leqslant \alpha \end{cases}$$

α chosen to satisfy

$$\Pr\left[\log_2 \frac{p_s(X^n)}{p_g(X^n)} > \alpha \middle| \text{not spam}\right] = \epsilon$$

Performance of optimal test

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Pr}[\text{declare not spam}|\text{email is spam}] = -D(p_s\|p_g),$$

where

$$D(p_s\|p_g) \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} p_s(x)\log_2 \frac{p_s(x)}{p_g(x)}$$

Kullback-Liebler (KL) divergence (or the relative entropy)

KL divergence

- "distance" between distributions
- Not symmetric:
$$D(p\|q) \neq D(q\|p)$$
- $$I(X;Y) = D(p_{XY}\|p_X p_Y).$$

Continuous random variables

Differential entropy

$$h(X) \stackrel{\text{def}}{=} -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x) dx.$$

Conditional differential entropy

$$h(X|Y) \stackrel{\text{def}}{=} -\int_{x,y} f_{XY}(x,y) \log_2 f_{X|Y}(x|y) dxdy.$$

Mutual information

$$\begin{aligned} I(X;Y) &= h(X) - h(X|Y) = h(Y) - h(Y|X) \\ &= \int_{x,y} f_{XY}(x,y) \log_2 \frac{f_{XY}(x,y)}{f_X(x) f_Y(y)} dxdy \end{aligned}$$

Caution

Differential entropy is not a measure of the information content of a system

- Differential entropy can be negative
- Differential entropy is not invariant to invertible transformations

Mutual information for continuous rvs

However, mutual information is more well behaved

- For a continuous channel $f_{Y|X}$, the capacity is

$$C = \max_{f_X} I(X; Y)$$

- Even for continuous rvs, $I(X; Y) \geqslant 0$.

Gaussian random variables

- The differential entropy of $\mathcal{N}(\mu, \sigma^2)$ random variable is

$$h(X) = \frac{1}{2} \log_2(2\pi e \sigma^2)$$

- The capacity of an AWGN channel is

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2}\right).$$