INFORMATION THEORY & CODING

Channel Code - 1

Dr. Rui Wang

Department of Electrical and Electronic Engineering
Southern Univ. of Science and Technology (SUSTech)

Email: wang.r@sustech.edu.cn

November 29, 2022

Review Summary

- **Entropy rate**. Two definitions of entropy rate for a stochastic process are

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n),$$
$$H^{'}(\mathcal{X}) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2} \ldots, X_1).$$

For a **stationary** stochastic process, $H(\mathcal{X}) = H^{'}(\mathcal{X})$.

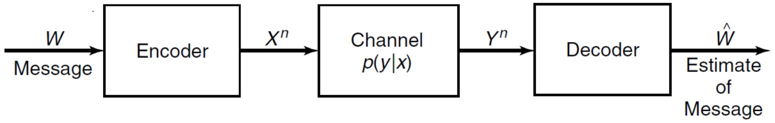
- Entropy rate of a stationary Markov chain.

$$H(\mathcal{X}) = -\sum_{i,j} \mu_i P_{ij} \log P_{ij}.$$

Outline

- **Channel model**: conditional distribution
- **Channel capacity**: defined in a pure way of information theory, not operational
- **Channel coding & data rate**: operational indicator of channel

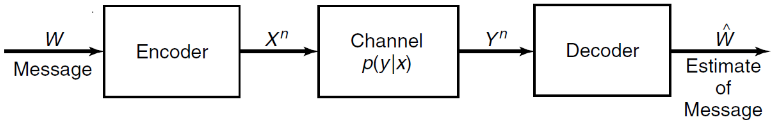
Communication System Model



- $X^n = [X_1, X_2, \ldots, X_n]$
- $Y^n = [Y_1, Y_2, \ldots, Y_n]$
- Channel $p(y^n|x^n)$: probability of observing y^n given input input sequence x^n

  $\Pr[Y^n = y^n | X^n = x^n]$

Discrete memoryless channel (DMC) 每次发射都不会对未来造成影响.

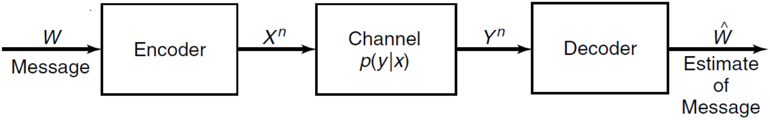


Definition

A discrete channel consists of an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and a probability transition matrix $p(y^n|x^n)$ that expresses the probability of observing the output sequence y^n given that we send the sequence x^n.

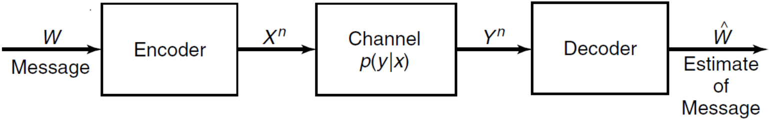
Definition

The channel is called memoryless if $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$.



- $X^n = [X_1, X_2, \ldots, X_n] \in \mathcal{X}^n$, $Y^n = [Y_1, Y_2, \ldots, Y_n] \in \mathcal{Y}^n$
  Channel $p(y^n|x^n)$: probability of observing y^n given input symbol x^n
  Memoryless: $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$
- Messages are mapped into some sequence of the channel symbols. Output sequence is random but has a distribution that depends on the input sequences. Each possible input sequence may induce several possbile outputs, and hence inputs are confusable. Can we choose a *non-confusable* subset of input sequences?

Duality



- Data compression: we remove all the redundancy in the data to form the most compressed version possible.

- Data transmission: we add redundancy in a controlled manner to combat errors in the channel.

"Survivor"

- You were deserted on a small island. You met a native and asked about the weather.

- True weather is a random variable X

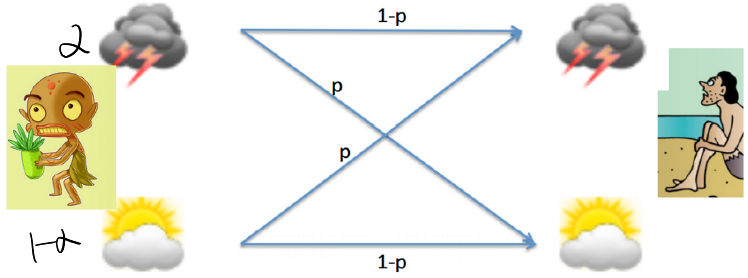
$$X = \begin{cases} \text{rain} & \text{w.p. } \alpha, \\ \text{sunny} & \text{w.p. } 1-\alpha, \end{cases}$$

- Native knows tomorrow's weather perfectly, but only tells truth with probability $1-p$.

- Native's answer is a random variable $Y \in \{\text{rain, sunny}\}$.

"Survivor"

- How informative is the native's answer?



Truth X R "1" S "0" $\xrightarrow{1-P}$ P Y

$$X = \begin{cases} R, & \alpha \\ S, & 1-\alpha \end{cases}$$

$I(X;Y)$

$\Downarrow$

$P(x,y)$

$P(x)$ $\downarrow$ source

$P(y|x)$ $\downarrow$ channel.

What is $I(X;Y)$?

- $I(X;Y) = H(X) - H(X|Y)$
- $H(X) = H(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$
- $H(X|Y) = H(X|Y = \text{rain})p(\text{rain}) + H(X|Y = \text{sunny})p(\text{sunny})$
- $H(X|Y = \text{rain})$ is equal to $-\sum_{i\in\{\text{rain},\text{sunny}\}} p(X = i|Y = \text{rain}) \log p(X = i|Y = \text{rain})$. Note that

  $$p(X = \text{rain}|Y = \text{rain}) = \frac{p(\overset{Y}{\cancel{X}}=\text{rain}|\overset{X}{\cancel{Y}}=\text{rain})p(X=\text{rain})}{p(Y=\text{rain})} = \frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}$$

  Thus, $H(X|Y) = \alpha H\left(\frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}\right) + (1-\alpha)H\left(\frac{p\alpha}{p\alpha+(1-p)(1-\alpha)}\right)$
- $I(X;Y) = H(\alpha) - \alpha H\left(\frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}\right) - (1-\alpha)H\left(\frac{p\alpha}{p\alpha+(1-p)(1-\alpha)}\right)$

Special Cases

- $I(X;Y) = H(\alpha) - \alpha H\left(\frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}\right) - (1-\alpha)H\left(\frac{p\alpha}{p\alpha+(1-p)(1-\alpha)}\right)$
- Always telling the truth: $p = 0$

$$I(X;Y) = H(\alpha) - \alpha H(1) - (1-\alpha)H(0) = H(\alpha) \leq 1 \text{ bit}$$

- Telling truth half of the time: $p = 1/2$

$$I(X;Y) = H(\alpha) - \alpha H(\alpha) - (1-\alpha)H(\alpha) = 0 \text{ bit}$$

- Fix p, maximize with respect to α, maximum achieved when $\alpha = 1/2$

$$\max_{\alpha} I(X;Y) = H(1/2) - \frac{1}{2}H(1-p) - \frac{1}{2}H(p) = 1 - H(P)$$

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Special Cases

- $I(X;Y) = H(\alpha) - \alpha H\left(\frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}\right) - (1-\alpha)H\left(\frac{p\alpha}{p\alpha+(1-p)(1-\alpha)}\right)$
- Always telling the truth: $p = 0$, $p=1$.

$$I(X;Y) = H(\alpha) - \alpha H(1) - (1-\alpha)H(0) = H(\alpha) \le 1 \text{ bit}$$

- Telling truth half of the time: $p = 1/2$

$$I(X;Y) = H(\alpha) - \alpha H(\alpha) - (1-\alpha)H(\alpha) = 0 \text{ bit}$$ → Y没有从X中获取任何信息

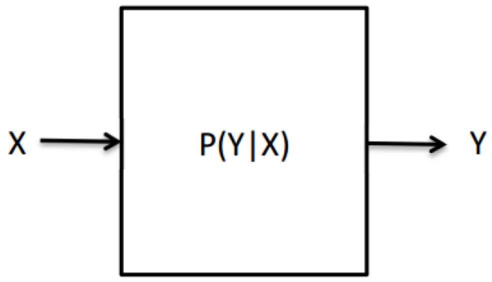
- Fix p, maximize with respect to α, maximum achieved when $\alpha = 1/2$

$$\max_{\alpha} I(X;Y) = H(1/2) - \frac{1}{2}H(1-p) - \frac{1}{2}H(p) = 1 - H(P)$$ → 与source没有关系,只与channel有关.

Definition (“Information” Channel Capacity)

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

仅由channel本身决定，X传递给Y的最大信息量



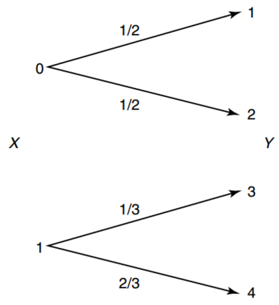
Examples

- Binary noiseless channel

$$C = \max I(X;Y) = \log 2 = 1 \text{ bits } \left(\text{with } p(x) = (\frac{1}{2}, \frac{1}{2})\right)$$

Examples

- Noisy channel with nonoverlapping outputs



$$C = \max I(X;Y) = \log 2 = 1 \text{ bits } \left(\text{with } p(x) = (\frac{1}{2}, \frac{1}{2})\right)$$

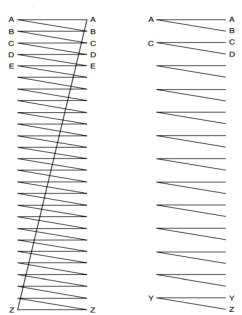
$$C = \max_{p(X)} I(X;Y)$$

$$I(X;Y) = H(X) - \underbrace{H(X|Y)}_{0} \rightarrow \text{给定}Y, X\text{唯一确定}$$

$$= H(X)$$

Examples

- Noisy typewriter



Noisy channel   Noiseless subset of inputs

$$C = \max I(X;Y) = \log \frac{26}{2} = \log 13 \text{ bits } \Big(\text{with } p(x) \text{ uniformly distributed}\Big)$$

$C = \max\limits_{p(x)} I(X;Y)$

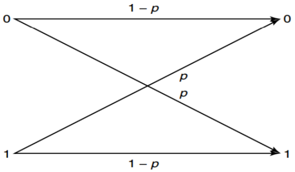
$\Downarrow$

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

$= H(Y) - \log 2.$

$C = \max\limits_{p(x)} H(Y) - \log 2 = \log 26 - \log 2 = \log 13$

- Binary symmetric channel



CD-ROM read channel

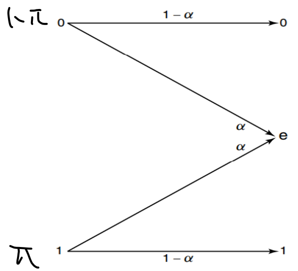
$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_{x\in\{0,1\}} p(x)H(Y|X=x)$$

$$= H(Y) - \sum_{x\in\{0,1\}} p(x)H(p) = H(Y) - H(p) \leq 1 - H(p)$$

$C = \max I(X;Y) = I - H(p)$ bits

- Binary erasure channel

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} \Big(H(Y) - H(Y|X) \Big) \\
&= \max_{p(x)} H(Y) - H(\alpha)
\end{aligned}
$$



Let $\Pr[X = 1] = \pi$, then

$$H(Y) = H\Big(\overset{Y=0}{(1-\pi)(1-\alpha)}, \overset{Y=e}{\alpha}, \overset{Y=1.}{\pi(1-\alpha)}\Big) = H(\alpha) + (1-\alpha)H(\pi)$$

? → 将前式拆开即可得.

Thus, $C = \max_{\pi}(1-\alpha)H(\pi) = 1 - \alpha$ (with $\pi = \frac{1}{2}$)

Symmetric channel

$$p(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}.$$

All the rows of the transition matrix are permutations of each other and so are the columns. Let $\mathbf{r}$ be a row of the transition matrix.

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\mathbf{r}) \le \log|\mathcal{Y}| - H(\mathbf{r})$$

with equality if $\mathcal{Y}$ is uniformly distributed. If $p(x) = \frac{1}{|\mathcal{X}|}$, Y is also uniformly distributed:

$$p(y) = \sum_{x\in\mathcal{X}} p(y|x)p(x) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} p(y|x) = \frac{c}{|\mathcal{X}|} = \frac{1}{|\mathcal{Y}|},$$

where c is the sum of the entries in one column.

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

$= H(Y) - \underbrace{H(0.2, 0.3, 0.5)}_{\parallel}$

$0.2\log\frac{1}{0.2} + 0.3\log\frac{1}{0.3} + 0.5\log\frac{1}{0.5}$

$\max I(X;Y) = \max_{p(x)} H(Y) - H(0.2, 0.3, 0.5).$

$= \log 3 - H(0.2, 0.3, 0.5)$

$(\overset{\frac{1}{3}}{u_1}, \overset{\frac{1}{3}}{u_2}, \overset{\frac{1}{3}}{u_3})\, P$

$$p(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}.$$

All the rows of the transition matrix are permutations of each other and so are the columns. Let $\mathbf{r}$ be a row of the transition matrix.

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\mathbf{r}) \le \log|\mathcal{Y}| - H(\mathbf{r})$$

with equality if $\mathcal{Y}$ is uniformly distributed. If $p(x) = \frac{1}{|\mathcal{X}|}$, Y is also uniformly distributed:

$$p(y) = \sum_{x\in\mathcal{X}} p(y|x)p(x) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} p(y|x) = \frac{c}{|\mathcal{X}|} = \frac{1}{|\mathcal{Y}|},$$

where c is the sum of the entries in one column.

Fundamental question

- How fast can we transmit information over a channel?
- Suppose a source sends r messages per second, and the entropy of a message is H bits per message, information rate is $R = rH$ bits/second.
- Intuition: as R increases, error will increase.
- Surprisingly, Shannon showed error can approach to zero, as long as

$$R < C$$