Lecture 12: Channel Capacity

- Definition
- Examples

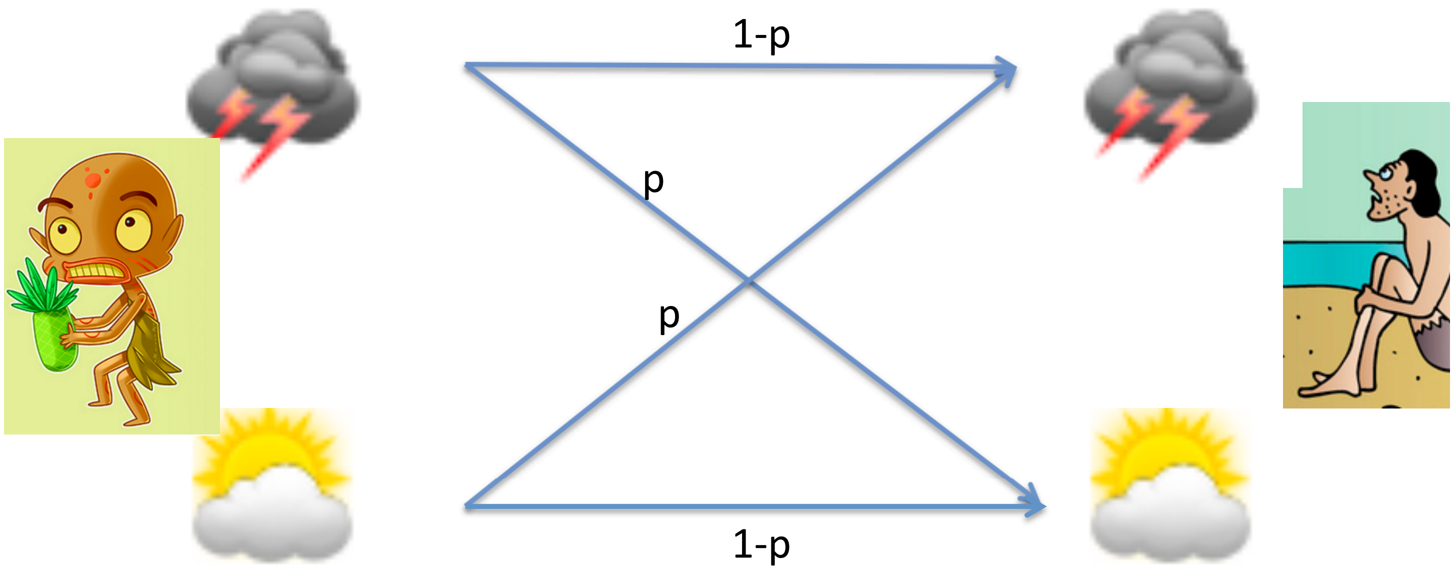
“Survivor”

- 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 $Y \in \{\text{rain}, \text{sunny}\}$

- How informative the native's answer is?



- Let us study $I(X;Y)$
- $I(X;Y) = H(X) - H(X|Y)$
- $H(X) = H(\alpha)$

$$H(p) = -p\log p - (1-p)\log(1-p)$$

- $H(X|Y) = H(X|Y = \text{rain})p(\text{rain}) + H(X|Y = \text{sunny})p(\text{sunny})$
- Find the posterior $p(x|Y = \text{rain})$ and $p(x|Y = \text{sunny})$ use Bayes' rule
- $H(X|Y) = \alpha H\left(\frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}\right) + (1-\alpha)H\left(\frac{p\alpha}{p\alpha+(1-p)(1-\alpha)}\right)$
- $I(X;Y) = H(\alpha) - \alpha H\left(\frac{(1-p)\alpha}{(1-p)\alpha+p(1-\alpha)}\right) - (1-\alpha)H\left(\frac{p\alpha}{p\alpha+(1-p)(1-\alpha)}\right)$

Special cases

- 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 $\alpha = 1/2$

$$\max_{\alpha} I(X;Y) \leq$$

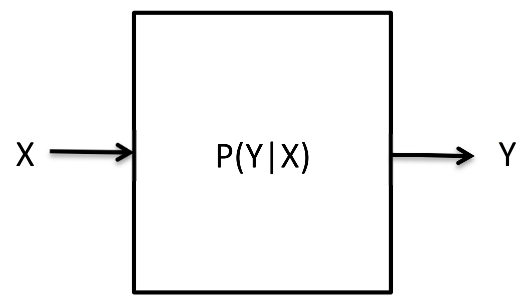
$$H(1/2) - \frac{1}{2}H(1-p) - \frac{1}{2}H(p) = 1 - H(p).$$

"Information" channel capacity

- Definition: "information channel capacity"

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

- We have proved, for fixed $p(y|x)$, $I(X;Y)$ is a concave function in $p(x)$



Why channel capacity

- Look at communication systems:
  Landline Phone, Radio $\rightarrow$ TV, Cellphone $\rightarrow$ Smartphone, WiFi

- Communication is very tied to specific source

- To break this tie, Shannon propose to focus on information, then computation

- First ask the question: what is the fundamental limit

- Then ask how to achieve this limit (took 60 years to get there! but huge success)

- All communication system are designed based on the principle of IT

Shannon's secret of success

- Start with simple model, then complicated

"Stylized" Models

- Let the code length goes to infinity, then back

- Study random coding, prove the feasibility

"Asymptotic is the first term in Taylor series expansion, and theory is the first term in the Taylor series of practice."

- Tom Cover, 1990

Channel capacity: intuition

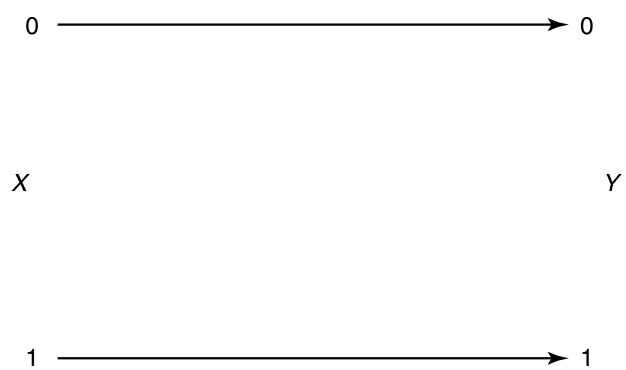
C

$= \log \#\{$ of identifiable inputs by passing through the channel with low error$\}$

Shannon's second theorem:

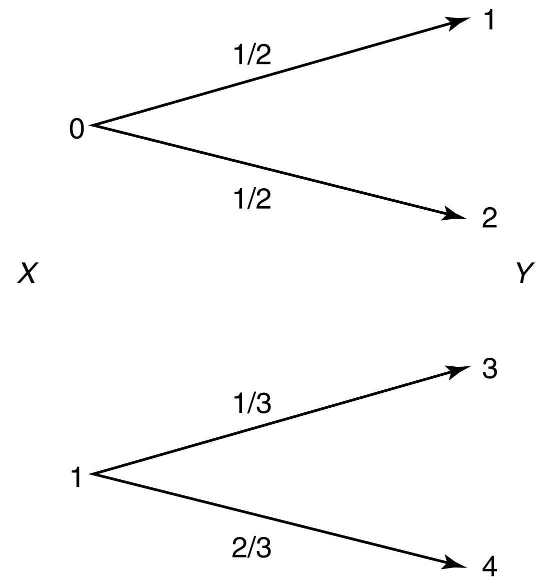
"information" channel capacity $=$ "operational" channel capacity

Binary noiseless channel



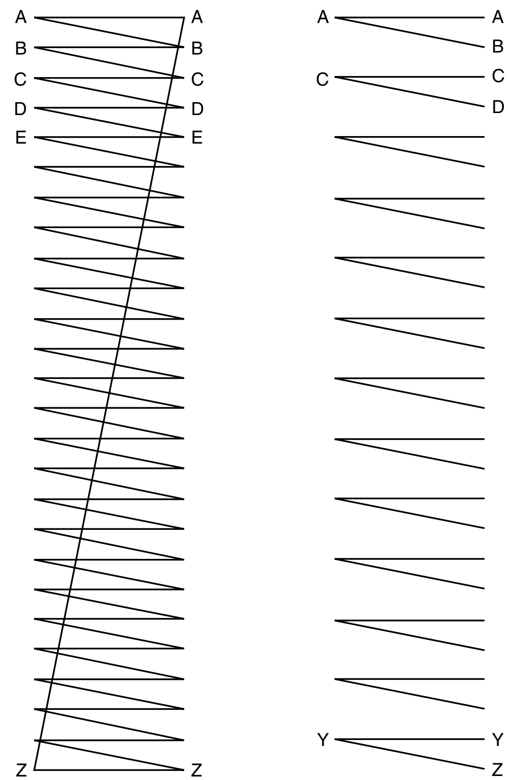
$C = \log 2 = 1$ bit

Noisy channel with non overlapping outputs



$C = \log 2 = 1$ bit

Noisy typewriter

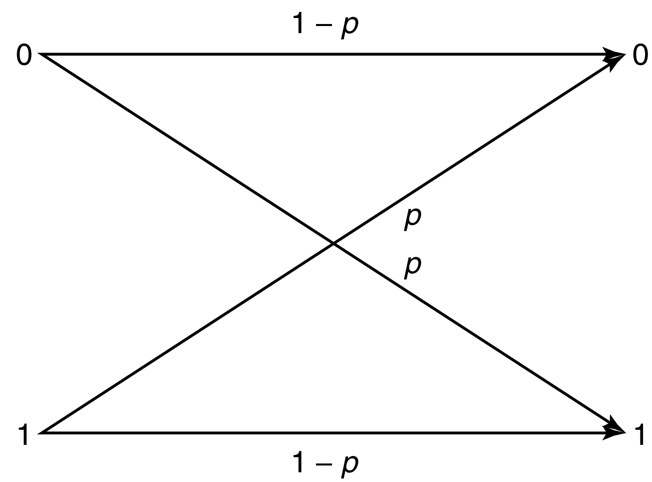


Noisy channel

Noiseless subset of inputs

$C = \log 13$ bits

Binary symmetric channel

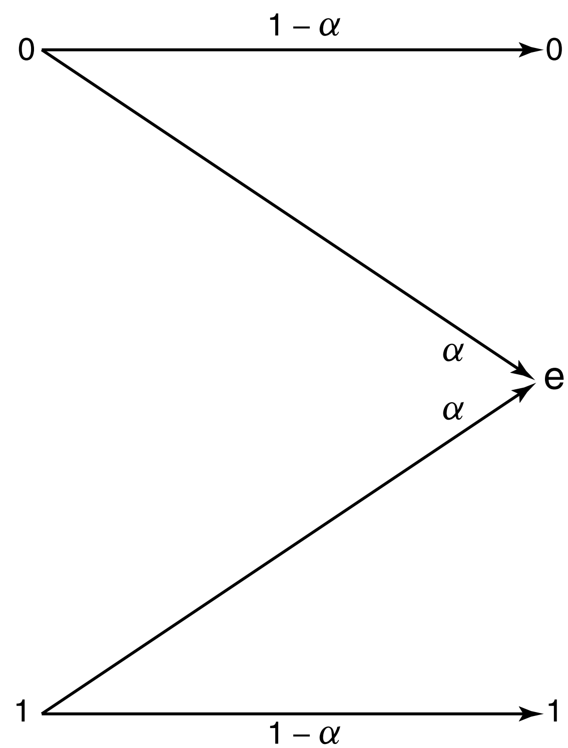


$C = 1 - H(p)$ bits.

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

CD-ROM read channel

Binary erasure channel



Some bits are lost, can be use as a model for DNA sequencing
$C = 1 - \alpha$

Symmetric channel

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

Let $\boldsymbol{r}$ be a row of the transition matrix

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(\boldsymbol{r}) \\ &\leq \log|\mathcal{Y}| - H(\boldsymbol{r}) \end{aligned}$$

with equality if $p(x) = 1/|\mathcal{X}|$:

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

Discrete Memoryless Channel (DMC)

- Discrete channel:
  - input alphabet: $\mathcal{X}$
  - output alphabet: $\mathcal{Y}$
  - probability transition matrix $p(y|x)$

- Memoryless channel:
  the probability distribution of the output depends only on the inputs at that time

Communication system model

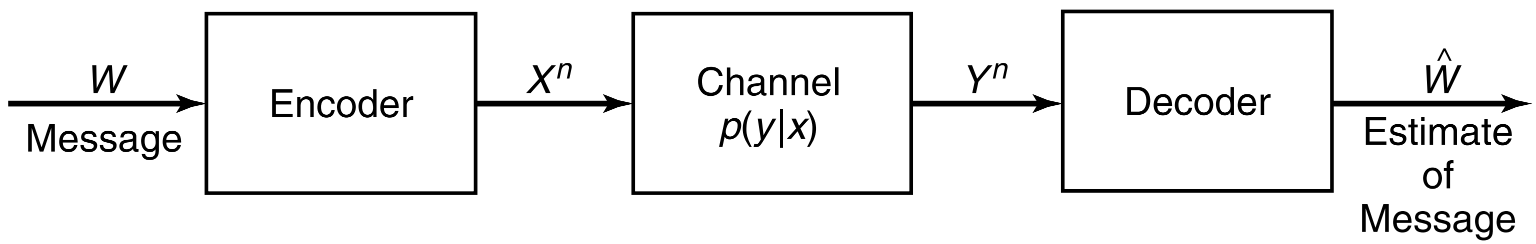


FIGURE 7.1. Communication system.

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

- Symbols from some finite alphabet are mapped into some sequence of the channel symbols

- Output sequence is random but has a distribution that depends on the input sequences

- From output sequence, we try to recover the transmitted message

- Each possible input sequences induces several possible 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

Summary

- Channel capacity:

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

intuition: $C = \log\{\#\text{of distinguishable inputs}\}$

- DMC (discrete memoryless channel)