A Mathematical Theory of Communication

By C.E.Shannon

Presented by Ling Shi

Part II: The Discrete Channel with Noise

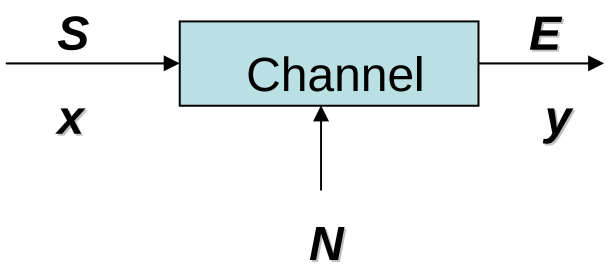
- Representation of a noisy discrete channel
- Equivocation and channel capacity
- Fundamental theorem for a discrete channel with noise
- Example of a discrete channel and its capacity

- Sets and ensembles of functions (from Part III)
- Entropy of a continuous distribution (from Part III)

2.1 Representation of a noisy discrete channel

- ***E=f(S,N)***
- ***p(β, j | α,i) => p(j|i)***

  α, β are states of channel and i is the send sym and j is the received sym



- ***H(x):*** Entropy of input

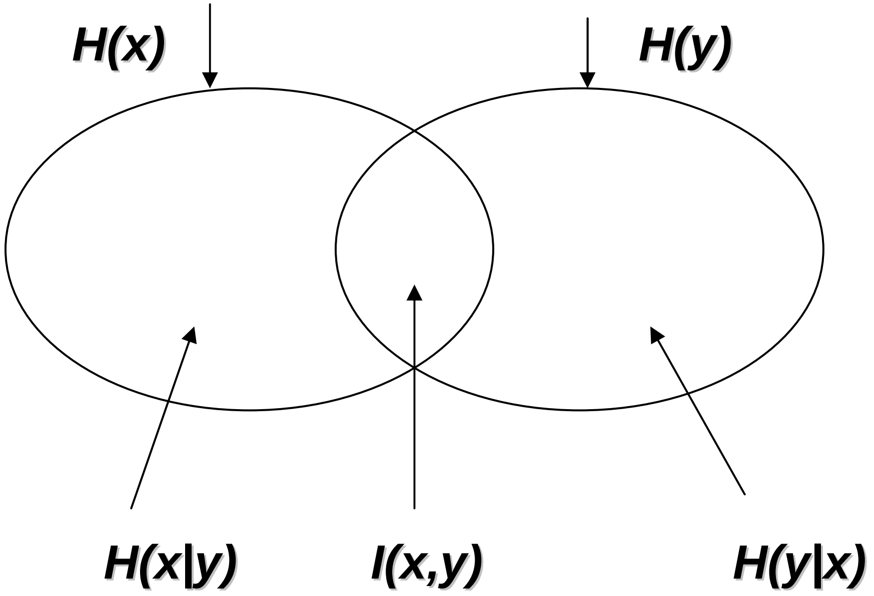
  H(y): Entropy of output

- ***H(y|x):*** Entropy of output when input is known

  H(x|y): Entropy of input when output is known

- ***H(x,y) = H(x) + H(y|x) = H(y) + H(x|y)***

  I(x, y) = H(x) – H(x|y) = H(y) – H(y|x)



2.2 Equivocation and channel capacity

(1) Objective: minimize the error caused by the noise

(2) $\boldsymbol{R = I(x,y) = H(x) - H(x|y)}$**,** where $\boldsymbol{H(x|y)}$ is named equivocation

Rate = 1000 symbols per second

$p_0 = p_1 = 1/2$

$p(1|0) = 0.01 \qquad p(1|1) = 0.99$

$p(0|0) = 0.99 \qquad p(0|1) = 0.01$

$H(x) = 1 \qquad H(x|y) = 0.081$

$R = 0.919$

Rate = 1000 symbols per second

$p_0 = p_1 = 1/2$

$p(1|0) = 1/2 \qquad p(1|1) = 1/2$

$p(0|0) = 1/2 \qquad p(0|1) = 1/2$

$H(x) = 1 \qquad H(x|y) = 1$

$R = 0$

(3) Theorem 10: *If the correction channel has a capacity equal to H(x|y) it is possible to so encode the correction data as to send it over this channel and correct all but an arbitrarily small fraction* ε *of the errors. This is not possible if the channel capacity is less than H(x|y).*

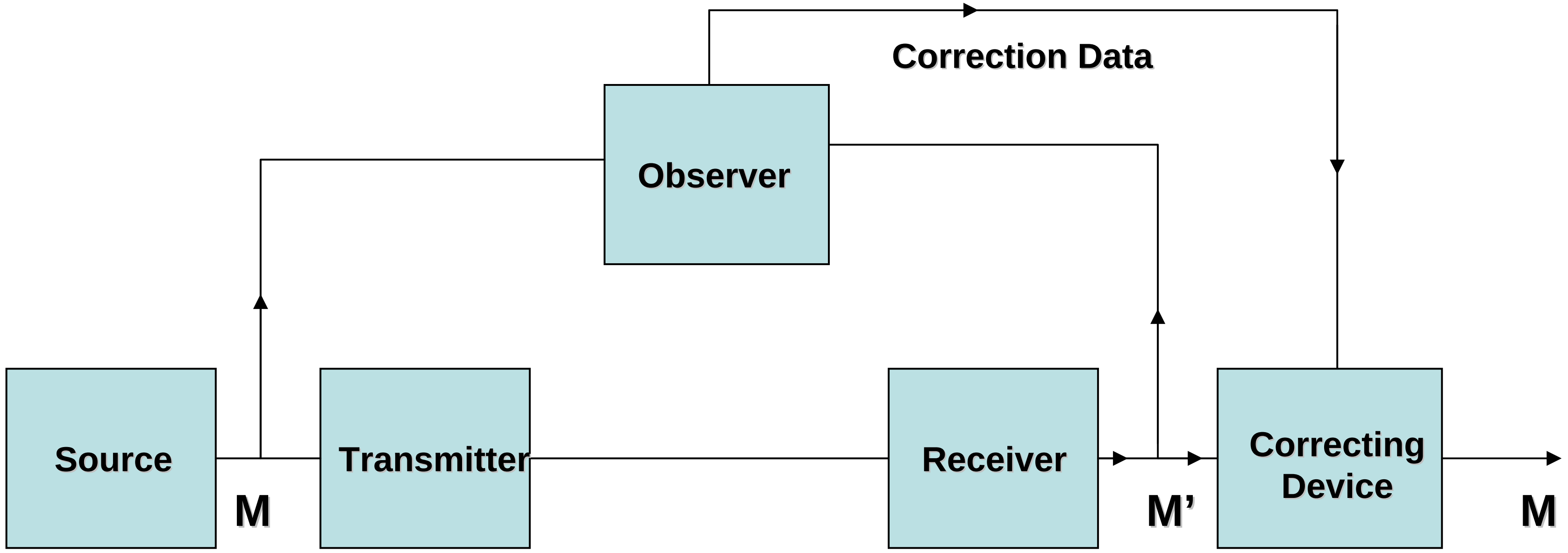
2.2 Equivocation and channel capacity (cont'd)

(4) Remarks: *$H(x|y)$ is the amount of additional information that must be supplied per second at the receiving point to correct the received message.*

(5) Definition:

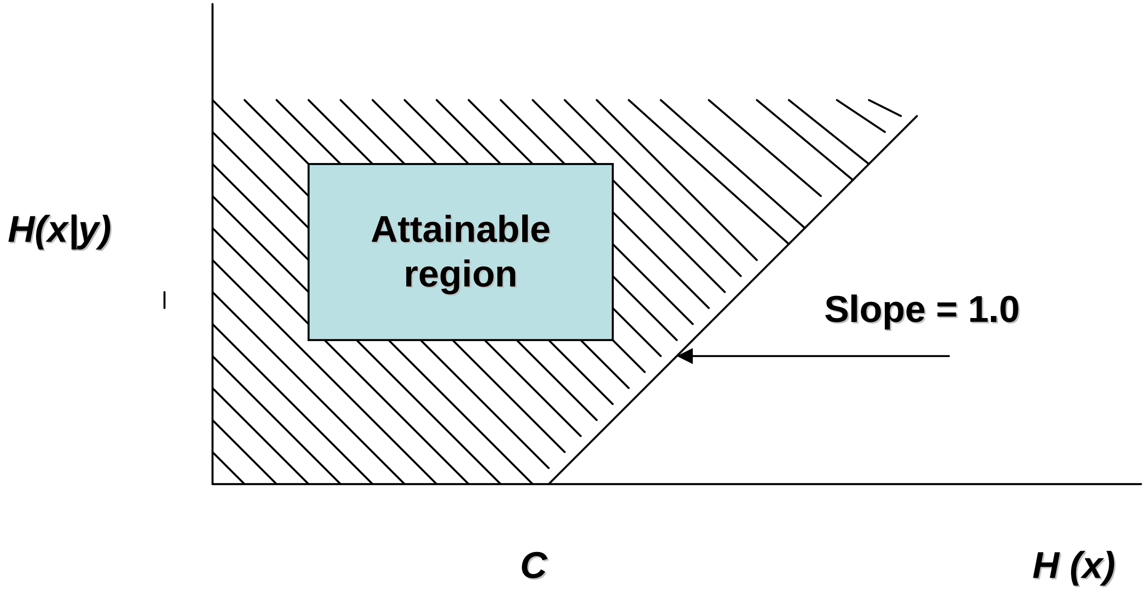
$$C = \mathrm{Max}(\, H(x) - H(x|y) \,)$$

where the max is with respect to all possible information sources used as input to the channel.



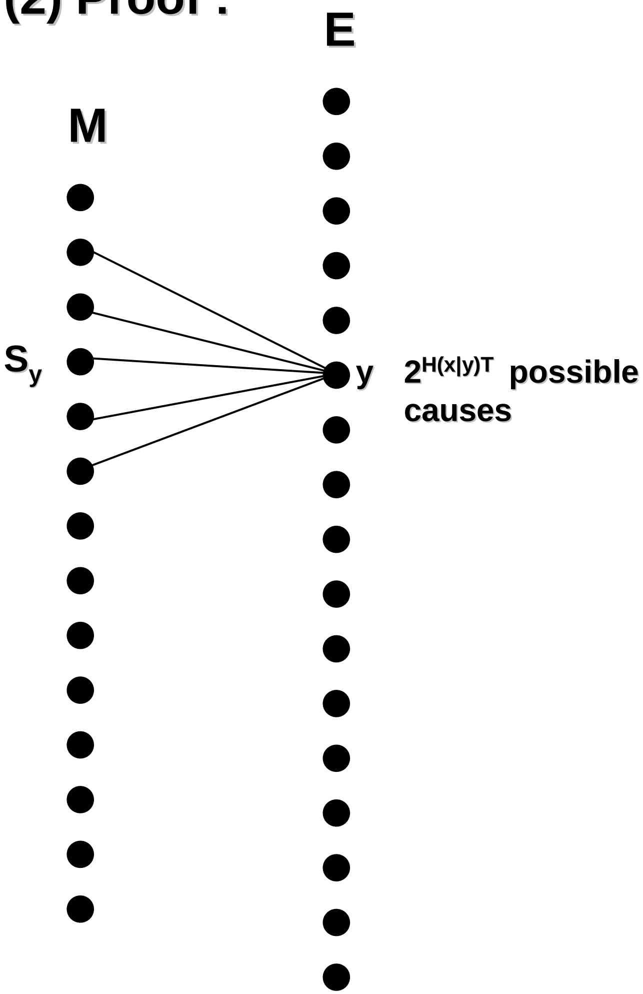
2.3 Fundamental theorem for a discrete noisy channel

(1) Theorem 11: *Let a discrete channel have the capacity C and a discrete source the entropy per second H. If H <$\leq$ C, there exists a coding system such that the output of the source can be transmitted over the channel with an aribitrarily small frequency of errors (or an arbitrarily small equivocation). If H > C, it is possible to encode the source so that the equivocation is less than H – C + ε where ε is arbitrarily small. There is no method of encoding which gives an equivocation less than H – C.*



2.3 Fundamental theorem for a discrete noisy channel (cont'd)

(2) Proof :



$2^{H(x)T}$ highly possible input messages

$2^{H(y)T}$ highly possible output messages

Assume R < C and let X be such that I(X,Y) achieves its maximum, i.e, equal to the channel capacity C. Choose 2^{RT} random sequences from the $2^{H(x)T}$ highly possible input messages. If x^i is transmitted and y is received. For y to be decoded as x^j with $j \neq i$, we must have at least one x^j in S_y with $j \neq i$, where S_y is the set of inputs associated with y. Now we have

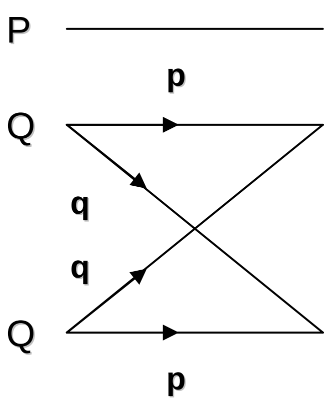
$$P(\text{at least one } x^j \text{ in } \rangle\backslash\ S_y,\ j \neq i) <= \sum_{j=1,\ j\neq i}^{2^{TR}} P(x^j \text{ in } S_y)$$

There are 2^{TR}-1 in the sum above and since

$$P(x^j \text{ in } S_y) = 2^{H(x|y)T}/2^{H(x)T} = 1/\ 2^{TC}\text{ , therefore,}$$

$$P(\text{at least one } x^j \text{ in } S_y,\ j \neq i) = 2^{T(R-C)} \to 0 \text{ as } T \to \text{infinity}$$

2.4 Example of a discrete channel and its capacity



$\alpha = p \log p + q \log q$ $\quad H(x) = -P\log P - 2Q\log Q$ $\quad H(x|y) = 2Q\alpha$

Problem: max $C = H(x) - H(x|y)$ subject to $P + 2Q = 1$

Solution: $U = -P\log P - 2Q\log Q - 2Q\alpha + \lambda(P + 2Q)$

$$\partial U / \partial P = -1 - \log P + \lambda = 0$$

$$\partial U / \partial Q = -2 - 2\log Q - 2\alpha + 2\lambda = 0$$

$P = \beta / (\beta + 2)$ $\quad Q = 1/(\beta + 2)$ $\quad C = \log (\beta + 2)/\beta$ $\quad$ Where $\beta = e^{\alpha}$

Verification:

Case I: $p=1$, $\beta = 1$, $C = \log 3$; result makes sense since now the channel is noiseless with three possible symbols

Case II: $p=0.5$, $\beta = 2$, $C = \log 2$; result also makes sense since now the second and the third symbols cannot be distinguished at all and act together like one symbol

In general cases where p has other values, C varies between log 2 and log 3

3.1 Sets and ensembles of functions

- A ***set of functions*** is merely a collections of functions, eg, $f_\theta(t) = \sin(t+\theta)$, where θ is in some set

- An ***ensemble of functions*** is a set of functions with a probability measure, eg, $P(\theta)$

- An ensemble of functions $f_\alpha(t)$ is ***stationary*** if the same ensembles results when all the functions are shifted any fixed amount in time, eg, $f_\theta(t+t_1) = \sin(t + t_1 + \theta) = \sin(t+\phi)$ with ϕ distributed uniformly from 0 to 2π if θ is uniformly distributed from 0 to 2π

- An ensemble is ***ergodic*** if it is stationary and there's no subset of the functions in the set with a probability different from 0 and 1 which is stationary.

eg, $\sin t\,(t + \theta)$ is ergodic while $a\sin(t+\theta)$ with a normally distributed and θ uniform is not ergodic

- We can obtain new ensembles from existing ones, eg, $g_\alpha(t) = Tf_\alpha(t)$ where T is an operator, eg, T=d/dt

- An operator T is called ***invariant*** if $g_\alpha(t+t_1) = Tf_\alpha(t+t_1)$ for all $f_\alpha(t)$ and t_1; Underinvariant operators, stationarity and ergodicity are preserved

- If f(t) is band limited to 0 and W, then

$$f(t) = \sum x_n \frac{\sin\pi(2Wt- n)}{\pi(2Wt-n)}$$

where $x_n=f(n/2W)$

3.2 Entropy of a continuous distribution

- **Definition:**

$H = -\iint p(x)\log p(x)dx$

$H=-\int\cdots\int p(x_1,\cdots,x_n)\log p(x_1,\cdots,x_n)dx_1\cdots dx_n$

$H(x,y)=-\iint p(x,y)\log p(x,y)dxdy$

$H(y|x)=-\iint p(x,y)\log p(x,y)/p(x)\ dxdy$

$H(x|y)=-\iint p(x,y)\log p(x,y)/p(y)\ dxdy$

- **Properties:**

1. $H(x)$ is max if x is uniform

2. *$H(x,y)\le H(x) + H(y)$ with equlity iff x & y are independent*

3. $p`(y) =\int a(x,y)p(x)dx$ *with* $\int a(x,y)dx =1$
$\int a(x,y)dy = 1,\ a(x,y) >= 0 \Rightarrow H(y)>=H(x)$

4. $H(x,y)=H(x) + H(y|x) = H(y) + H(x|y)$

5. Let $p(x)$ be a one-dim distribution. The form giving a maximum entropy subject to the condition that $\sigma(x)$ begin fixed at σ is Gaussian.

6. The entropy of a one-dim Gaussian distribution whose standard deviation is σ is given by $H(x)=\log (2\pi e)^{1/2}\sigma$

7. If x is limited to the half plane, ie

$p(x)=0$ for $x\le 0$, and $E[x]=a$ being fixed, then the maximum entropy occurs when $p(x)=1/a\ \exp(-x/a)$ and is equal to $\log ea$

8. Entropy is ***relative to coordinate system***