Section 1

Telecommunications – Section 2 (Lecture 1) 12/11/09

Information Theory: Channel Capacity

Goal: Answer the question: "What is the capacity of a communications channel to carry information?"

Recall from 2nd Year:

The information content of a symbol:

$$
\begin{aligned}
I(x_i) &= \log_2\left[\frac{1}{P(x_i)}\right] \\
&= -\log_2\left[P(x_i)\right] \\
\therefore I(x_i) &= 0 \text{ for } P(x_i)=1 \\
I(x_i) &\geq 0
\end{aligned}
$$

The information content I of a symbol i from source x is greater than that of a symbol j if the probability of occurrence of i is less than that of j, i.e.:

$$I(x_i) > I(x_j) \quad \text{when} \quad P(x_i) < P(x_j)$$

The combined information content of two symbols i and j is equal to the sum of the information content of each element, i.e.:

$$I(x_i x_j) = I(x_i) + I(x_j), \text{ provided } i \text{ and } j \text{ are independent.}$$

Average Information or Entropy:

The entropy of a data source x is equal to the expected value of the information content of x, and is measured in bits/symbol. (In this case, 'bits' refers to Shannon units of information rather than digital bits.) i.e.:

Number of symbols in source x

$$
\begin{aligned}
H(x) &= E\left[I(x_i)\right] \\
&= \sum_{i=1}^{m} P(x_i) I(x_i) \\
H(x) &= \sum_{i=1}^{m} P(x_i)\log_2\left[P(x_i)\right]
\end{aligned}
$$

It can be shown that $H(x)$, the entropy of source x, satisfies:-

Occurs in the equiprobable case: $P(x_i) = 1/m$

$$0 \le H(x) \le \log_2[m]$$

Information Rate:

If the source emits *r* symbols/s then the information rate R of the source is given by:

$$R = rH(x) \; b/s$$

Discrete Memoryless Channels

Channel Representation:

A communications channel is the path or medium through which the symbols flow to the receiver. A discrete memoryless channel (DMC) is a statistical model with an input *X* and an output *Y*. During each signalling interval the channel accepts an input signal from *X* and, in response, generates an output signal from *Y*. The channel is discrete when the alphabets of *X* and *Y* are finite. It is memoryless when the current output depends only on the current input and not on any previous inputs.

$$\left.\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{matrix}\right\} P(y_j \mid x_i) \left\{\begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{matrix}\right.$$

Where $P(y_j \mid x_i)$ represents the conditional probability of receiving y_j having sent x_i.

We assume the "a-priori" probabilities of x_i are available to us. Each possible input-output path is indicated with a conditional probability $P(y_j \mid x_i)$, also known as the "channel transmission probability", given x_i.

Channel Matrix:

For our purposes, a channel is completely specified by the complete set of transition probabilities. Accordingly, a matrix representation is often used, denoted [*P(Y|X)*], given by:

$$[P(Y \mid X)] = \begin{bmatrix} P(y_1 \mid x_1) & \cdots & P(y_n \mid x_1) \\ \vdots & \ddots & \vdots \\ P(y_1 \mid x_m) & \cdots & P(y_n \mid x_m) \end{bmatrix}$$

Channel Matrix (continued...)

$\left[P(Y|X)\right]$ is called the "Channel Matrix". Since each input to the channel results in some output. If that is the case, each row of the channel matrix must sum to unity, i.e.

$$\sum_{j=1}^{n} P\left(y_j | x_i\right) = 1 \text{ for all } i$$

Now, if the input probabilities P(x) are represented by the row matrix:

$$\left[P(x)\right] = \left[P(x_1)\, P(x_2) \dots P(x_m)\right]$$

and the output probabilities P(y) represented by the row matrix:

$$\left[P(y)\right] = \left[P(y_1)\, P(y_2) \dots P(y_n)\right]$$

then:-

$$\left[P(y)\right] = \left[P(x)\right]\left[P(Y|X)\right]$$

Alternatively, if P(x) is represented as a diagonal matrix:-

$$\left[P(x)\right]_d = \begin{bmatrix} P(x_1) & 0 & \dots & 0 \\ 0 & P(x_2) & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & P(x_m) \end{bmatrix}$$

Then:-

$$\left[P(y)\right] = \left[P(x)\right]_d \left[P(Y|X)\right]$$

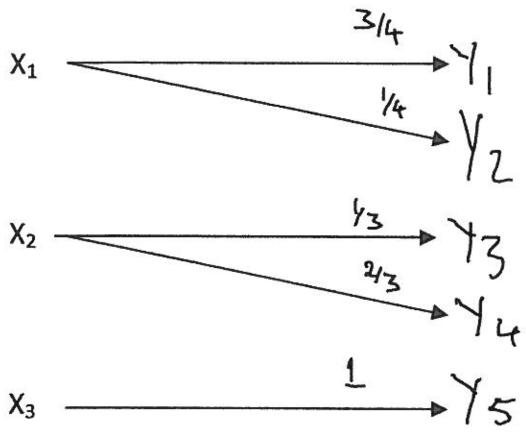
where the *(i,j)* element of *[P(X,Y)]* has the form $P(x_i, y_j)$. (i.e. probability of sending x_i and subsequently receiving y_j. The matrix *[P(X,Y)]* is known as the "**Joint Probability Matrix**".

Special Channels

Lossless Channels

Lossless channels have only one non-zero element in each column. For example: -

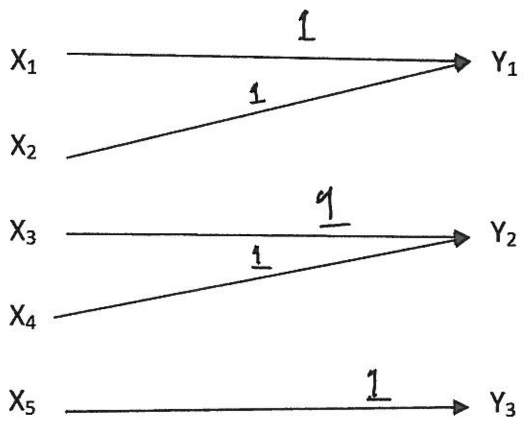
$$[P(X|Y)]=\begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Deterministic Channels

A channel whose channel matrix has only one non-zero element in each row is called a **"Deterministic Channel".** For example, consider:

$$[P(X|Y)]=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Noiseless Channel

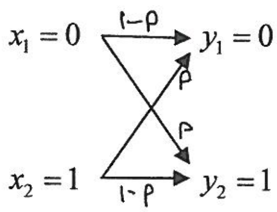
This channel is both lossless and deterministic. For example,

$$\begin{array}{ccc} x_1 & \longrightarrow & y_1 \\ x_2 & \longrightarrow & y_2 \\ \vdots & & \vdots \\ x_m & \longrightarrow & y_n \end{array}$$

Binary Symmetric Channel

The binary symmetric channel is defined by: -

$$\left[P(Y|X)\right]=\begin{bmatrix}1-p & p \\ p & 1-p\end{bmatrix}$$



The transition probability *p* denotes the probability of error, given a particular input symbol. Note that the channel is symmetric since p applies in both error types (i.e. send 0, receive 1 and send 1, receive 0).

Mutual Information

Conditional and Joint Entropies

Using the input probabilities *P(x*$_i$*)*, the transition probabilities *P(y*$_j$*,x*$_i$*)* and joint probabilities *P(x*$_i$*,u*$_j$*)* we can define the following entropy functions for a channel with *m* inputs and *n* outputs: -

$$H(X)=-\sum_{i=1}^{m}P(x_i)\log_2\left[P(x_i)\right]$$

$$H(Y)=-\sum_{j=1}^{n}P(y_j)\log_2\left[P(y_j)\right]$$

$$H(X|Y)=-\sum_{j=1}^{n}\sum_{i=1}^{m}P(x_i,y_j)\log_2\left[P(x_i|y_j)\right]$$

Equivocation of x w.r.t. y.

$$H(Y|X) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2\left[P(y_j | x_i)\right]$$

Prevarication of y w.r.t. x.

(Generally easiest to use)

$$H(X,Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2\left[P(x_i, y_j)\right]$$

H(x) is the average uncertainty of the channel input. *H(y)* is the average uncertainty of the channel output. *H(X/Y)* is a measure of the average uncertainty about the channel input after the channel output has been observed (*Equivocation*). *H(Y/X)* is the measure of the average uncertainty about the channel output given the channel input *(Prevarication – relates to average uncertainty from the transmitter's viewpoint)*.

The joint entropy *H(X,Y)* is the average uncertainty of the communications channel as a whole. Two useful relationships amongst the various entropies are:-

$$\text{Chain Rules for Entropy}\begin{cases} H(x,y)=H(x\mid y)+H(y) \\ H(x,y)=H(y\mid x)+H(x) \end{cases}$$

<u>Aside:</u>

Prove the equivocation formula:

$$H(X\mid Y)=-\sum_{j=1}^{n}\sum_{i=1}^{m}P(x_i,y_j)\log_2\left[P(x_i\mid y_j)\right]$$

Solution:

$$H(X\mid Y=y_j)=\sum_{i=1}^{m}P(x_i\mid y_j)\log_2\left(P(x_i\mid y_j)\right)$$

Now we average with respect to Y

$$\Rightarrow H(X\mid Y)=\sum_{j=1}^{n}P(y_j)H(x\mid y=y_j)$$

$$=-\sum_{j=1}^{n}\sum_{i=1}^{m}\underbrace{P(y_i)P(x_i\mid y_j)}_{P(x_i,y_j)}\log_2\left(P(x_i\mid y_j)\right)$$

Mutual Information

The guiding principle is "Information is always a measure of the decrease of uncertainty at a receiver". With this in mind, the mutual information *I(x;y)* of a channel is defined by:-

$$I(x;y)=H(X)-H(X\mid Y)\ \textit{(bits/sym)}$$

Since H(x) represents the average uncertainty about the channel input before the channel output is observed and *H(x|y)* (equivocation) represents the average uncertainty about the channel input after the output has been observed, *I(x;y)* represents the uncertainty about the channel input that is resolved by observing the channel output.

Properties of *I(x;y)*

$$
\begin{aligned}
&\textit{Symmetric:} && I(x;y)=I(y;x)\\
& && I(x;y)\geq 0\\
&\textit{Normally, we use these} && \begin{cases} I(x;y)=H(Y)-H(Y\mid X)\\ I(x;y)=H(X)-H(X\mid Y)\end{cases}
\end{aligned}
$$

Channel Capacity

Channel capacity per symbol Cs:

The channel capacity per symbol of a discrete memoryless channel is defined by:

$$C_s=\underbrace{Max}_{\{P(x_i)\}}\left[I(x;y)\right]$$

Where the maximisation is over all possible input probability distributions $\{P(x_i)\}$ on X. Note that C_s is a function only of the channel transition probabilities.

Channel Capacity per Second (=C)

If r symbols per second are being transmitted (on average), then the channel capacity per second is:-

$$C=rC_s \qquad (bits/s)$$

This is the max rate of transmission of information per second.

Channel Coding Theorem

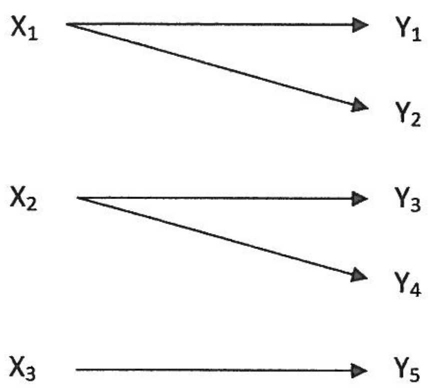
Given a discrete memoryless source with entropy *H(X)* bits/symbol and a discrete memoryless channel with capacity C_s bits/symbol, **"if $H(X)\leq C_s$, there exists a coding scheme for which the source output can be transmitted over the channel with an arbitrarily small probability of error"**. Conversely, if $H(X)>C_s$, it is impossible to transmit information over the channel with an arbitrarily small probability of error → unrecoverable errors are guaranteed.

Telecommunications – Section 2 (Lecture 4) 07/01/10

Capacities of Typical Channels

Lossless Channel

E.g. recall:



It is clear which *x* was sent for a given *y*.

What is the capacity of a lossless channel?

When we observe output Y_i in a lossless channel, it is clear which *x* was transmitted.

Therefore:

$$P(x_i \mid y_j) = 0 \; or \; 1$$

Now,

$$
\begin{aligned}
H(x \mid y) &= -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i \mid y_j)\log_2\left[P(x_i \mid y_j)\right] \\
&= -\sum_{j=1}^{n}\sum_{i=1}^{m} P(y_i)P(x_i \mid y_j)\log_2 P\left[(x_i \mid y_j)\right] \\
&= -\sum_{j=1}^{n} P(y_i)\sum_{i=1}^{m} \underbrace{P(x_i \mid y_j)\log_2 P\left[(x_i \mid y_j)\right]}_{0.\log_2(0) \quad or \quad 1\log_2(1)} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
&Now, \\
&I(x;y) = H(X) - H(X \mid Y) \\
&= H(X) \\
&\text{since } H(X \mid Y) = 0
\end{aligned}
$$

Hence, our interpretation of *H(x)* as the average information content of a source (ee2009) is valid. However, in general, entropy *H()* measures average uncertainty. Since the mutual information (*I(x;y)*) is equal to the input source entropy, no source information is lost during transmission (hence the name). Hence, the channel capacity per symbol, C_s, is given by:-

$$
C_s = \underbrace{\max}_{\{P(x_i)\}}\left[H(X)\right] = \log_2[m]
$$

Number of symbols in x

Deterministic Channel

$$
\begin{aligned}
&H(X \mid Y) = 0 \text{ for all input distributions } P(x_i) \\
&\text{and } I(X;Y) = H(Y) \\
&\rightarrow C_s = \underbrace{\max}_{\{P(x_i)\}}\left[H(Y)\right] = \log_2[n]
\end{aligned}
$$

Noiseless Channel

Since this is both lossless and deterministic:

$$
\begin{aligned}
&I(x;y) = H(x) = H(y) \\
&\Rightarrow C_s = \log_2[m] = \log_2[n] \\
&\text{since } n = m
\end{aligned}
$$

Summer 08 – Question 4

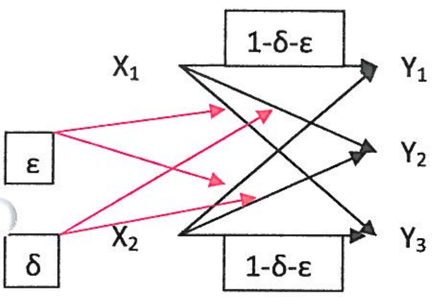
$$[P(Y|X)] = \begin{bmatrix} 1-\delta-\varepsilon & \delta & \varepsilon \\ \varepsilon & \delta & 1-\delta-\varepsilon \end{bmatrix}$$

Binary Symmetric Erasure Channel

$(BSEC)$

a) Show: $H(Y|X) = -(\delta \log_2[\delta] + \varepsilon \log_2[\varepsilon] + (1-\delta-\varepsilon)\log_2[1-\delta-\varepsilon])$

Solution:



$$Let\ P(x_1) = \alpha \quad \Rightarrow \quad P(x_2) = 1-\alpha$$

$$H(Y|X) = -\sum_{j=1}^{3}\sum_{i=1}^{2} \underbrace{P(x_i, y_j)}_{P(x_i)P(y_j|x_i)} \log_2(P(y_j|x_i))$$

$$= (1-\delta-\varepsilon)\log_2[1-\delta-\varepsilon](1-\alpha+a)$$

$$+$$

$$\log_2[\delta](\alpha+1-\alpha)$$

$$+$$

$$\log_2[\varepsilon](\alpha+1-\alpha)$$

as stated.

b) If the input symbols are equiprobable, show that:

$$C_{BSEC} = (\delta - 1)\log_2\left[\frac{1-\delta}{2}\right] + \varepsilon\log_2[\varepsilon] + (1-\delta-\varepsilon)\log_2[1-\delta-\varepsilon]$$

$$= \underbrace{Max}_{\{P(x_i)\}}\left[H(Y) - H(Y\,|\,X)\right]$$

$$H(Y) = \sum_{j=1}^{3} P(y_j)\log_2\left[P(y_j)\right]$$

$$P(y_1) = \alpha(1-\delta-\varepsilon) + (1-\alpha)(\varepsilon)$$

$$= \alpha(1-\delta) + \varepsilon(1-2\alpha) = \gamma \text{ say}$$

$$P(y_2) = \delta(\alpha + 1 - \alpha) = \delta$$

$$P(y_3) = 1 - \left(P(y_1) + P(y_2)\right)$$

$$= 1 - (\delta + \gamma)$$

In the equiprobable case, $\alpha = ½$

$$\Rightarrow \gamma = \left({}^{1-\delta}/_{2}\right)$$

$$P(y_3) = 1 - (\delta + \gamma) = 1 - \left(\frac{1+\delta}{2}\right) = \frac{1-\delta}{2} = \gamma = P(y_1)$$

Now, substitute back in to yield:-

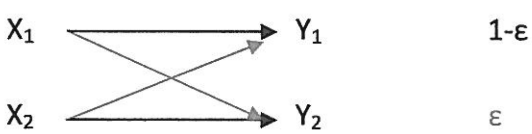
$$C_{BSEC} = \left(H(Y) - H(Y\,|\,X)\right)\big|_{\alpha=½}$$

$$-\sum_{j=1}^{3} P(y_j)\log_2\left(P(y_j)\right)$$

and the answer emerges

c) Consider the case $\delta \to 0$. Deduce the channel capacity of the resulting "binary symmetric channel".

Solution:

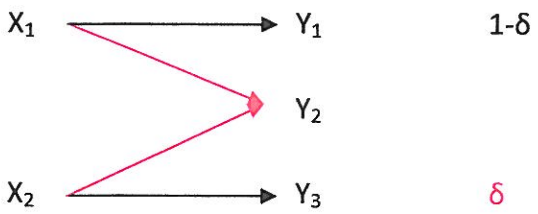


Let $\delta \to 0$ in part (b)

$C_{bsc} = 1 + \varepsilon\log_2[\varepsilon] + (1-\varepsilon)\log_2[1-\varepsilon]$

d) Consider the case $\varepsilon \to 0$. Deduce the channel capacity of the "binary erasure channel" (BEC).

Solution:



Let $\varepsilon \to 0$

$$\therefore C_{BEC} = (1-\delta)\log_2[1-\delta] + (\delta - 1)\underbrace{\log_2\left[\tfrac{1-\delta}{2}\right]}_{\log_2[1-\delta]-\log_2[2]}$$

Example: Summer 2007 Q.4

$$P(Y_1 \mid X) = \begin{bmatrix} 1-\rho & \alpha & \alpha \\ \alpha & 1-\rho & \alpha \\ \alpha & \alpha & 1-\rho \end{bmatrix}$$
$$= \underbrace{\begin{bmatrix} -1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} 1-{}^{3\rho}/_{2} & 0 & 0 \\ 0 & 1-{}^{3\rho}/_{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 1 & 1 & 1 \end{bmatrix} \frac{1}{3}}_{F^{-1}}$$

Also, $\alpha = {}^{\rho}/_{2}$ and the columns of F are eigenvectors of $P[Y_1|X]$. If n such that "3-ary uniform channels" are connected in series,

a) Show that the composite channel and matrix

$$\left[P(Y_n \mid X)\right] = \begin{bmatrix} 1+2q & 1-q & 1-q \\ 1-q & 1+2q & 1-q \\ 1-q & 1-q & 1+2q \end{bmatrix}$$
$$where\ q = \left(1 - {}^{3\rho}/_{2}{}^{n}\right)$$

Solution:

Recall:

$$\left[P(Y)\right] = \left[P(X)\right]\left[P(Y \mid X)\right]$$
$$\Rightarrow \left[P(Y_1)\right] = \left[P(X)\right]\left[P(Y_1 \mid X)\right]$$
$$\Rightarrow \left[P(Y_2)\right] = \left[P(Y_1)\right]\left[P(Y_2 \mid X)\right] = \left[P(X)\right]\left[P(Y_1 \mid X)\right]^2$$
$$\vdots$$
$$\left[P(Y_n)\right] = \left[P(X)\right]\underbrace{\left[P(Y_1 \mid X)\right]^n}_{Composite\ Channel}$$

①

14/01/10 Telecoms Q4 '07

c)

If p is sufficiently small so that its square and higher powers can be neglected show that.

$C_s^c \simeq$ that of a single 3-ary uniform channel with probability of error-free transmition $= 1 - n{*}p$

note: $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$ and $\binom{n}{i} = \frac{n!}{i!(n-i)!}$ [4]

Solution:

Let $n = 1 \Rightarrow C_s^c = \log_2 \left[3 (1-p)^{1-p} \left(\frac{p}{2}\right)^p\right]$

So, in our system with n channels in series, we have:-

$$q = \left(1 - \frac{3p}{2}\right)^n = \sum_{i=n-1}^{n} \binom{n}{i} x^i y^{n-i}$$

where $x = 1$ and $y = -\frac{3p}{2}$

$$\Rightarrow q \simeq n\left(\frac{-3p}{2}\right) + 1$$

$$= q \simeq 1 - \frac{3(np)}{2}$$

now substitute this approximation into.

$$C_s^c \Rightarrow C_s^c \simeq \log_2 \left[3(1-np)^{1-np} \left(\frac{np}{2}\right)^{np}\right]$$

Recall with $n=1$: $C_s^c = \log_2 \left[3(1-p)^{1-p}\left(\frac{p}{2}\right)^p\right]$

$$\begin{bmatrix} 1-p & \alpha & \alpha \\ \alpha & 1-p & \alpha \\ \alpha & \alpha & 1-p \end{bmatrix}$$