3.1 The mutual information between X and a two valued random variable Y can be written as

$$I(X;Y) = H(Y) - H(Y|X) \leqslant H(Y) \leq \log 2$$

Now Y could be statistically independent of $X \Rightarrow H(Y|X) = H(Y)$

$\therefore$

$$0 \leqslant I(X;Y) \leqslant H(Y) \leq \log 2$$

We are interested in finding the case where $I(X;Y)$ is largest. This occurs when $H(Y|X)$ is zero and $H(Y) = \log 2$.

$\Rightarrow$ Y takes on each of two values with equal probability and Y is completely determined by X. In this case

$$I(X;Y) = H(Y) = \log 2 \longleftarrow$$

Many such specific Y could be formulated, One such case is

$$Y = \begin{cases} 1, & \text{Sum of outcomes of the two individual dice is odd} \\ 0, & \text{" " " " " " " " even.} \end{cases}$$

In this case, it is easy to see that both $H(Y|X) = 0$ and $H(Y) = \log 2$.

3.2 For this channel, we have

$P_X(0) = p$ $\quad P_{XY}(0,0) = p/2$ $\quad I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

$p_X(1) = 1-p$ $\quad P_{XY}(0,?) = p/2$

$P_Y(0) = p/2$ $\quad P_{XY}(0,1) = 0$ $\quad = \cdots = \frac{1}{4}\left[2p\log 2 - 2p\log p\right.$

$P_Y(?) = (1+p)/4$ $\quad P_{XY}(1,0) = 0$

$P_Y(1) = \frac{3}{4}(1-p)$ $\quad P_{XY}(1,?) = \frac{1}{4}(1-p)$ $\quad \left. -(1+p)\log(1+p) - 3(1-p)\log(1-p)\right]$

To find which p achieves capacity, we take the derivative w.r.t p and equate it to zero (convexity $\cap$ of $I(X;Y)$ gives unique max):

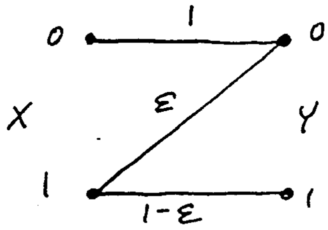
$$4\frac{d}{dp} I(X;Y) = 2\log 2 - 2\log p - \log(p+1) + 3\log(1-p) = 0$$

$$\Rightarrow \log \frac{2^2(1-p)^3}{p^2(p+1)} = 0 \Rightarrow 5p^3 - 11p^2 + 12p - 4 = 0$$

The one real root of this cubic is $p = 0.526826$

Substituting this value into $I(X;Y)$ yields $C = 0.4554$ nats $= 0.657$ bits.

3.3



$P_X(0) = p$ $\qquad P_{XY}(0,0) = p$

$P_X(1) = 1-p$ $\qquad P_{XY}(0,1) = 0$

$P_Y(0) = p + (1-p)\varepsilon$ $\qquad P_{XY}(1,0) = (1-p)\varepsilon$

$P_Y(1) = (1-p)(1-\varepsilon)$ $\qquad P_{XY}(1,1) = (1-p)(1-\varepsilon)$

$$I(X;Y) = \sum_{x,y} p(x,y) \log \left[\frac{p(x,y)}{p(x)\,p(y)} \right]$$

$$= p \log\left[\frac{p}{p[p+(1-p)\varepsilon]}\right] + (1-p)\varepsilon \log\left[\frac{(1-p)\varepsilon}{(1-p)[p+(1-p)\varepsilon]}\right]$$

$$+ (1-p)(1-\varepsilon) \log\left[\frac{(1-p)(1-\varepsilon)}{(1-p)(1-p)(1-\varepsilon)}\right]$$

$$= \cdots = \varepsilon(1-p)\log\varepsilon - (1-\varepsilon)(1-p)\log(1-p) - [\varepsilon(1-p)+p]\log[\varepsilon(1-p)+p]$$

In order to find the value of p that maximizes $I(X;Y)$, take deriv. w.r.t. p and set $= 0$

$$\frac{d}{dp} I(X;Y) = \cdots = -\varepsilon \log \varepsilon + (1-\varepsilon)\log\left[\frac{1-p}{p(1-\varepsilon)+\varepsilon}\right]$$

Setting this equal to zero and solving for p, we get

$$p_{max} = \frac{1-K\varepsilon}{1+K(1-\varepsilon)}, \text{ where } K = \exp\left\{\frac{\varepsilon \log \varepsilon}{1-\varepsilon}\right\} = \varepsilon^{\frac{\varepsilon}{1-\varepsilon}}$$

(A plot of p_{max} is on following page)

If we define $U(\varepsilon)$ as $\quad U(\varepsilon) = 1 - \dfrac{1-K\varepsilon}{1+k\varepsilon} = \dfrac{\varepsilon}{\varepsilon^{1-\varepsilon}+\varepsilon^{\varepsilon}-1}$

We can write C by substituting p_{max} back into $I(X;Y)$, and we get

$$C = \varepsilon U(\varepsilon)\log\varepsilon - (1-\varepsilon)U(\varepsilon)\log U(\varepsilon) - \left[(1-U(\varepsilon)+\varepsilon U(\varepsilon)\right]\log\left[1-U(\varepsilon)\right] + \varepsilon \log U(\varepsilon)$$

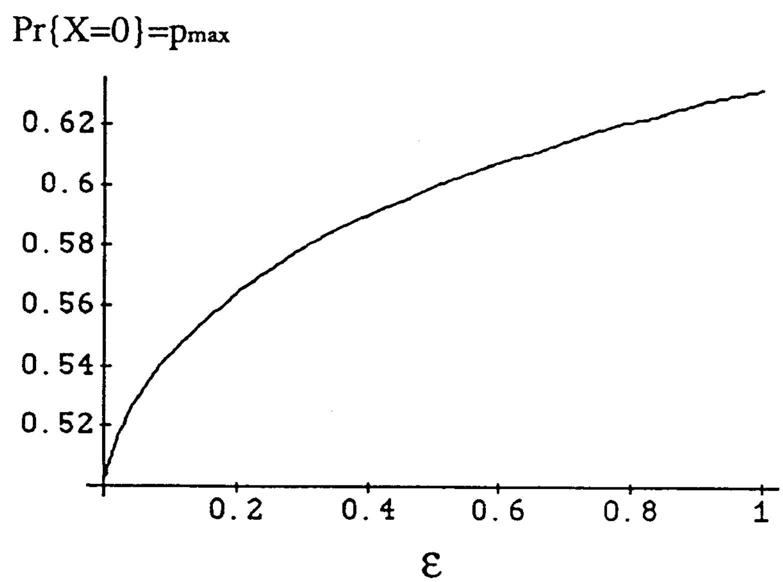
$$= \mathcal{H}(U(\varepsilon)) + \varepsilon(1+U(\varepsilon))\log U(\varepsilon) + \varepsilon U(\varepsilon)\log\left\{\frac{\varepsilon}{1-U(\varepsilon)}\right\}$$

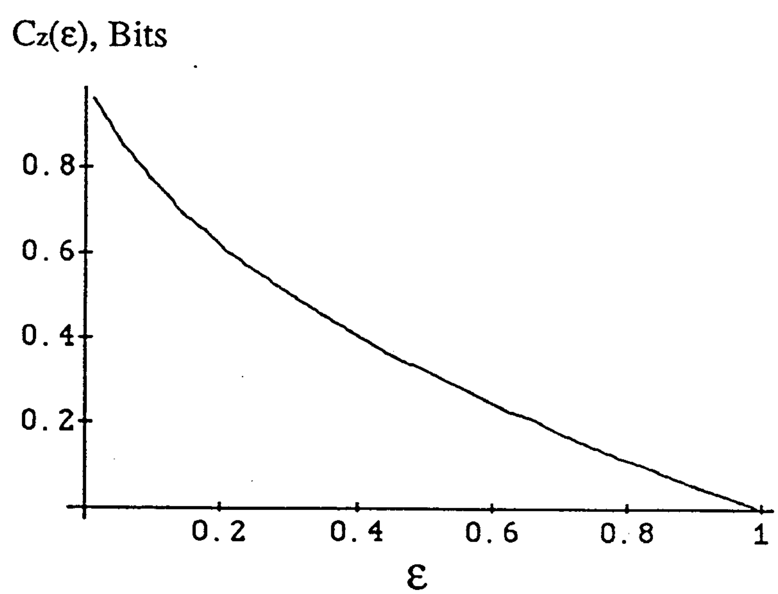
$\therefore$ The capacity of the Z-channel, as a fn. of ε is

$$C_Z(\varepsilon) = \mathcal{H}(U(\varepsilon)) + \varepsilon(1+U(\varepsilon))\log U(\varepsilon) + \varepsilon U(\varepsilon)\log\left[\frac{\varepsilon}{1-U(\varepsilon)}\right]$$

A plot is shown on the following page. Further simplification may be possible.

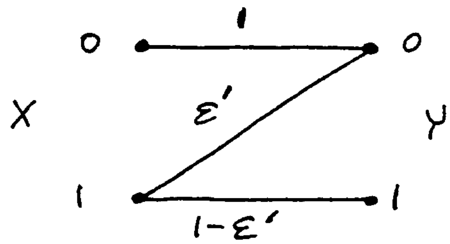
Problem 3.3 Plots for Z-Channel





3.4 Assume that the capacity of a single binary Z channel, such as the one in problem 3.3, is given by $C_Z(\varepsilon)$ as a function of the 1-to-0 cross-over probability ε. We wish to find the capacity $C_Z^{(3)}(\varepsilon)$ of a cascade of 3 such channels.

Examining the cascade of 3 channels, we note that the resulting cascade can be viewed as a single binary Z channel with crossover prob. ε'



This can be seen by noting that if a ~~zero~~ "0" enters the cascade, a "0" exits with prob. 1. However, if a "1" enters the cascade, there is some prob. ε' that it will exit the cascade as a a "0" and a prob. $1-\varepsilon'$ that it will exit the cascade a "1". It can be easily seen that

$$(1-\varepsilon)^3 = 1-\varepsilon'$$

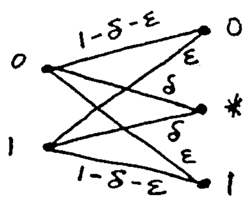
and thus we can write ε' as a function of of ε as

$$\varepsilon' = 1-(1-\varepsilon)^3 = 3\varepsilon - 3\varepsilon^2 + \varepsilon^3$$

Thus the capacity of the three channel cascade, as a function of the crossover error prob. ε of the original Z channels, is

$$C_Z^{(3)}(\varepsilon) = C_Z\left(3\varepsilon - 3\varepsilon^2 + \varepsilon^3\right)$$

3.5 (a) Channel appears as follows:



Although not a <u>symmetric channel</u> in the strict sense, there is sufficient symmetry the we might guess that $(p_0 = \frac{1}{2}, p_1 = \frac{1}{2})$ is maximizing distribution.

Checking, we see that (using theorem presented in class)

$$I(X=0;Y) = (1-\delta-\varepsilon)\log\frac{1-\delta-\varepsilon}{\left(\frac{1-\delta}{2}\right)} + \cancel{\delta \log \frac{\delta}{\delta}}^{\,0} + \varepsilon \log \frac{\varepsilon}{\left(\frac{1-\delta}{2}\right)}$$

$$= (1-\delta-\varepsilon)\log\frac{2(1-\delta-\varepsilon)}{1-\delta} + \varepsilon\log\frac{2\varepsilon}{1-\delta}$$

$$I(X=\phi;Y) = \cdots = (1-\delta-\varepsilon)\log\frac{2(1-\delta-\varepsilon)}{1-\delta} + \varepsilon\log\frac{2\varepsilon}{1-\delta}$$

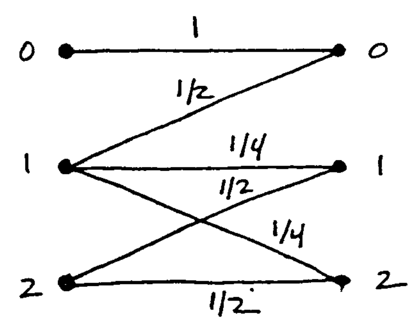
$\therefore$ $(p_0 = 1/2, p_1 = 1/2)$ is a maximizing distribution, and

$$C = (1-\delta-\varepsilon)\log\frac{2(1-\delta-\varepsilon)}{1-\delta} + \varepsilon\log\frac{2\varepsilon}{1-\delta}$$

$$= \cdots = 1-\delta + \mathcal{H}(\delta) - (1-\delta-\varepsilon)\log(1-\delta-\varepsilon) - \delta\log\delta - \varepsilon\log\varepsilon$$

$$= (1-\delta) + \mathcal{H}(\delta) + H(1-\delta-\varepsilon, \delta, \varepsilon)$$

(b) Diagramatically, the channel appears as follows:



This channel is not symmetric, and does not really look vaguely symmetric. Input-0 looks reliable, input-1 looks terrible. Suppose we only use input-0 and input-2. We could then decode output-0 as input-0, output-1 or output-2 as input-2, and it looks like we could transmit at least one-bit of information per channel use. Lets try it.

Assume $(p_0 = \frac{1}{2}, p_1 = 0, p_2 = \frac{1}{2})$

$$I(X=0;Y) = 1\log\frac{1}{1/2} = \log 2$$

$$I(X=1;Y) = \cancel{\tfrac{1}{2}\log\frac{1/2}{1/2}}^{\,0} + \cancel{\tfrac{1}{4}\log\frac{1/4}{1/4}}^{\,0} + \cancel{\tfrac{1}{4}\log\frac{1/4}{1/4}}^{\,0} = 0$$

$$I(X=2,Y) = \tfrac{1}{2}\log\frac{1/2}{1/4} + \tfrac{1}{2}\log\frac{1/2}{1/4} = \log 2$$

$\therefore$ Theorem satisfied

So $(p_0 = \frac{1}{2}, p_1 = 0, p_2 = \frac{1}{2})$ is a maximizing distribution,* and the capacity is $C = \log 2$.

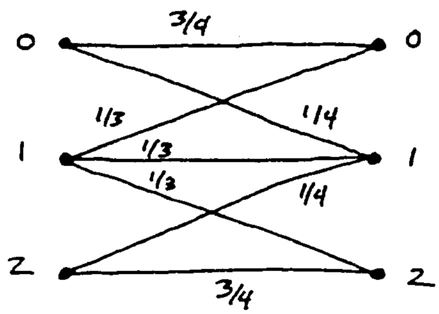
3.5 continued

(c) This is a symmetric channel (as defined in class). Thus equi-prob. probabilities on the inputs achieve capacity.

So $(p_0 = \frac{1}{3}, p_1 = \frac{1}{3}, p_2 = \frac{1}{3})$ achieves capacity. The capacity is

$$C = \varepsilon \log 3\varepsilon + (1-\varepsilon)\log 3(1-\varepsilon)$$
$$= \varepsilon[\log 3 + \log \varepsilon] + (1-\varepsilon)[\log 3 + \log(1-\varepsilon)]$$
$$= \log 3 + \varepsilon \log \varepsilon + (1-\varepsilon)\log(1-\varepsilon)$$

(d) A diagram of this channel appears as follows:



This is not a symmetric channel by definition, although there is significant symmetry. Note that input-1 looks pretty worthless. Lets split the probability between input-0 and input-2 and see what happens:

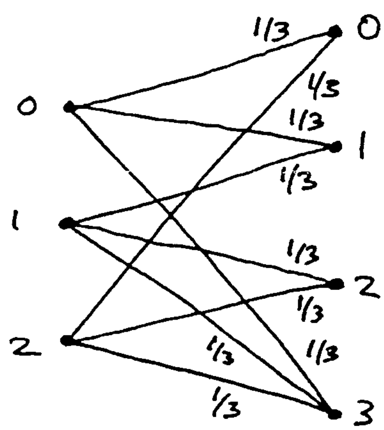
$(p_0 = \frac{1}{2}, p_1 = 0, p_2 = \frac{1}{2})$

$$I(X=0;Y) = \frac{3}{4}\log\frac{3/4}{3/8} + \cancel{\frac{1}{4}\log\frac{1/4}{1/4}} = \frac{3}{4}\log 2$$
$$I(X=1;Y) = \frac{1}{3}\log\frac{1/3}{3/8} + \frac{1}{3}\log\left(\frac{1/3}{1/4}\right) + \frac{1}{3}\log\frac{1/3}{3/8} = \frac{2}{3}\log\frac{8}{9}$$
$$I(X=2;Y) = \frac{1}{4}\log\frac{1/4}{1/4} + \frac{3}{4}\log\frac{3/4}{3/8} = \frac{3}{4}\log 2$$

$\therefore$ $(p_0 = \frac{1}{2}, p_1 = 0, p_2 = \frac{1}{2})$ achieves capacity, and

$$C = \frac{3}{4}\log 2$$

(e) A diagram of this channel appears as follows:



This is not a symmetric channel, but lets see what happens if we assign equi-probable input probabilities: $(p_0 = 1/3, p_1 = 1/3, p_2 = 1/3)$

$$I(X=0;Y) = \frac{1}{3}\log\frac{1/3}{2/9} + \frac{1}{3}\log\frac{1/3}{2/9} + \frac{1}{3}\log\frac{1/3}{2/9} = \log\frac{3}{2}$$
$$I(X=1;Y) = \frac{1}{3}\log\frac{1/3}{2/9} + \frac{1}{3}\log\frac{1/3}{2/9} + \frac{1}{3}\log\frac{1/3}{2/9} = \log\frac{3}{2}$$
$$I(X=2;Y) = \frac{1}{3}\log\frac{1/3}{2/9} + \frac{1}{3}\log\frac{1/3}{2/9} + \frac{1}{3}\log\frac{1/3}{2/9} = \log\frac{3}{2}$$

It worked!

$\therefore$ $(p_0 = \frac{1}{3}, p_1 = \frac{1}{3}, p_2 = \frac{1}{3})$ achieves capacity, and

$$C = \log\frac{3}{2}$$

3.6 (Cover and Thomas; Chapter 5, problem 25)

(a) Because $l_i = \lceil \log \frac{1}{p_i} \rceil$, we can write

$$\log \frac{1}{p_i} \le l_i < \log \frac{1}{p_i} + 1.$$

So it follows that

$$\sum_i p_i \log \frac{1}{p_i} \le \sum_i p_i l_i < \sum_i p_i \left(\log \frac{1}{p_i} + 1\right)$$

$$\Rightarrow \quad H(X) \le L(c) < H(X) + 1.$$

In order to show that this is a prefix-free code, we note that

$$2^{-l_i} \le p_i < 2^{-(l_i - 1)}.$$

Thus F_j, $j > i$ differs from F_i by at least 2^{-l_i}, and thus it will differ from F_i in at least one place in the first l_i bits of the binary expansion of F_i. Thus the codeword for F_j, which has length $l_j \ge l_i$ differs from the codeword for F_i at least once in the first l_i places. It thus follows that no codeword is a prefix of a longer codeword.

(b) We construct a code for the example distribution as follows:

Symbol	Prob.	F_i decimal	F_i binary	l_i	codeword
1	1/2	0.0	0.0	1	0
2	1/4	0.5	0.10	2	10
3	1/8	0.75	0.110	3	110
4	1/8	0.875	0.111	3	111

The Shannon code achieves the entropy bound of 1.75 bits in this case.