$$R = I(X\,;\,Y) = H(Y) - H(Y/X)$$

$$= -(p_1' \log p_1' + p_2' \log p_2') + p_1(p_{11}Q_1 + p_{12}Q_2) + p_2(p_{21}Q_1 + p_{22}Q_2)$$

$$= -(p_1' \log p_1' + p_2' \log p_2') + p_1' Q_1 + p_2' Q_2 \qquad (2.61)$$

where,

$$p_1' = p_1\, p_{11} + p_2\, p_{21}$$

$$p_2' = p_1\, p_{12} + p_2\, p_{22}$$

To maximize $I(X\,;\,Y)$, one uses the calculus of variations (method of Lagrangian Multipliers) and maximize the function

$$F = I(X\,;\,Y) + \lambda\,(p_1' + p_2')$$

through a proper selection of λ. Differentiating with respect to p_1' and p_2', the condition to be satisfied is :

$$\frac{\partial F}{\partial p_1'} = -(\log_2 e + \log p_1') + Q_1 + \lambda = 0$$

$$\frac{\partial F}{\partial p_2'} = -(\log_2 e + \log p_2') + Q_2 + \lambda = 0$$

Solving the equations,

$$\lambda = -Q_1 + (\log_2 e + \log p_1') = -Q_2 + (\log_2 e + \log p_2')$$

and $$Q_1 = Q_2 + \log p_1' - \log p_2'$$

Using these values in Equation (2.61)

$$C = R_{max} = -p_1' \log p_1' - p_2' \log p_2' + p_1' Q_2' + p_1' \log p_1' - p_1' \log p_2' + p_2' Q_2$$

$$= Q_2 - \log p_2' \quad [\text{since } (p_1' + p_2') = 1]$$

$$= Q_1 - \log p_1'$$

where Q_1 and Q_2 are obtained by solving equation (2.60)

Further,

$$p_1' = \exp(Q_1 - C) = \frac{e^{Q_1}}{e^{C}}$$

$$p_2' = \exp(Q_2 - C) = \frac{e^{Q_2}}{e^C}$$

and summing $p_1'$and p_2',

$$\frac{e^{Q_1} + e^{Q_2}}{e^C} = 1$$

Therefore,

$$C = \log\left[e^{Q_1} + e^{Q_2}\right] \text{ nats/symbol}$$

$$C = \log\left[2^{Q_1} + 2^{Q_2}\right] \text{ nats/symbol} \qquad (2.62)$$

f. BSC and BEC

The simplest type of source alphabet to be considered is binary $\{0, 1\}$. We assume that the output of such a source is transmitted through a binary symmetric channel or a binary erasure (BEC) channel. Fig. 2.7 shows a BSC and Fig. 2.8. shows a BEC.

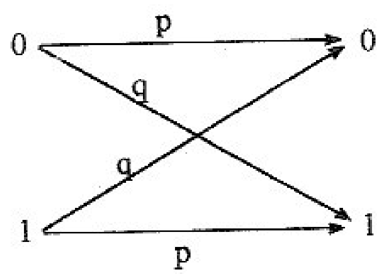

Fig. 2.7 Binary Symmetric channel

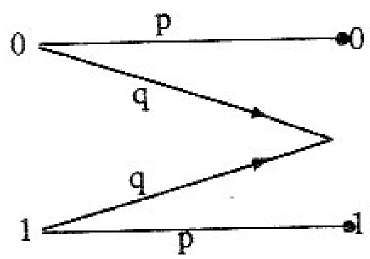

Fig. 2.8 Binary Erasure channel

For BSC, let

$$p(0) = \alpha \qquad p(1) = 1 - \alpha$$
$$p(0/0) = p(1/1) = p$$
$$p(0/1) = p(1/0) = q$$

Then

$$H(X) = H(\alpha, 1 - \alpha) = -\alpha\log\alpha - (1 - \alpha)\log(1 - \alpha)$$

$$H(Y/X) = -(p\log p + q\log q)$$

$$I(X ; Y) = H(Y) + p\log p + q\log q$$

$$C = \max I(X ; Y) = 1 + p\log p + q\log q \qquad (2.63)$$

For BEC, the channel has two input {0, 1} and three output symbols {0, y, 1}. The letter y indicates the fact that the output is erased and no deterministic decision can be made as to whether the transmitted letter was 0 or 1.

Let $p(0) = \alpha \qquad p(1) = 1 - \alpha$

$$p(0/0) = p(1/1) = p$$

$$p(y/0) = p(y/1) = q$$

Then

$$H(X) = H(\alpha, 1 - \alpha)$$

$$H(X/Y) = (1 - p)\, H(X)$$

$$I(X ; Y) = p\, H(X)$$

$$C = p \tag{2.64}$$

Equations (2.63) and (2.64) give the channel capacities of BSC and BEC respectively.

2.5 Decoding Schemes

Let us now consider the problem of reliable transmission of messages through a noisy communication channel. In order to achieve reliability, we must be able to determine the input message with a high degree of accuracy after seeing the received sequence of symbols. Therefore we are looking for the *best decoding scheme* or the best way of finding the correct input.

Suppose a channel has an input alphabet $x_1, x_2, \ldots, x_M$, an output alphabet $y_1, y_2, \ldots, y_N$ and a channel matrix $[p(y_j/x_i)]$. For simplicity let us consider first the special case in which a single symbol, chosen at random according to a known input description $p(x_i)$, is transmitted through the channel.

A *decoder* or *decision scheme* is an assignment to every output symbol y_j of an input symbol x_j^* from the alphabet $x_1, x_2, \ldots, x_M$.

The meaning here is that if y_j is received, it will be decoded as x_j^*. The decoder may be thought of as a deterministic channel with input alphabet $y_1, \ldots, y_N$ and output alphabet $x_1, x_2, \ldots, x_M$. If Z is the output of the decoder, then we may express Z as a function of Y, say Z = g(Y). (See Fig. 2.9). Equivalently, we may think of the decoders as partitioning the values of Y into disjoint subsets $B_1, \ldots, B_M$, such that every y in B_i is decoded as x_i.

X		Y		Z = g(Y)
x_1		y_1		x_1
x_2	channel	y_2	decoder	x_2
x_M		y_N		x_M

Fig. 2.9 Representation of a decoder

As an example consider the channel and decoder combination of Fig.2.10. The symbol y_1 is decoded as x_1 and the symbols y_2 and y_3 as x_3. The probability of error in this case is just the probability that x_2 is chosen, since x_1 and x_3 are always decoded perfectly.

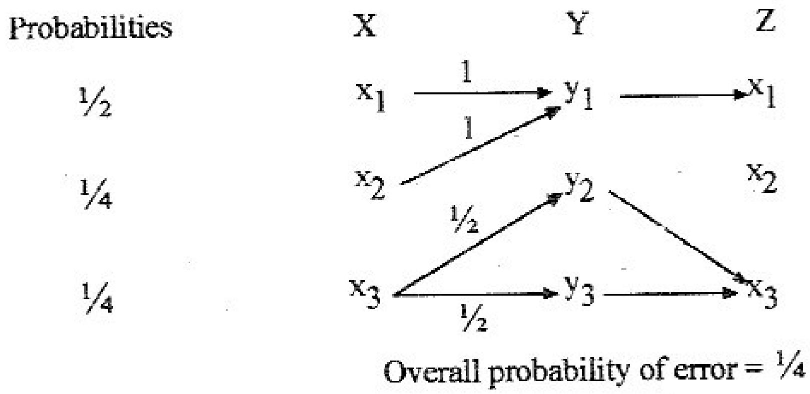

Fig. 2.10 Example of channel and decoder

We now propose the following problem. For a given input distribution $p(x_i)$, construct the decision scheme that minimizes the overall probability of error. Such a decision scheme is called *ideal observer*. To find out the required decoder, assume that each output symbol y_j is associated with an input symbol x_j^* ($j = 1, 2, \ldots N$) (see fig. 2.11). Let p(e) be the overall probability of error and p(e′) the overall probability of correct transmission. Given that y_j is received, the probability of correct transmission is the probability that the actual input is x_j^*.

Thus we may write the probability of correct transmission as

$$p(e') = \sum_{j=1}^{N} p(y_j)p(e'/y_j) = \sum_{j=1}^{N} p(y_j)p(X = x_j^*/y_j) \qquad (2.65)$$

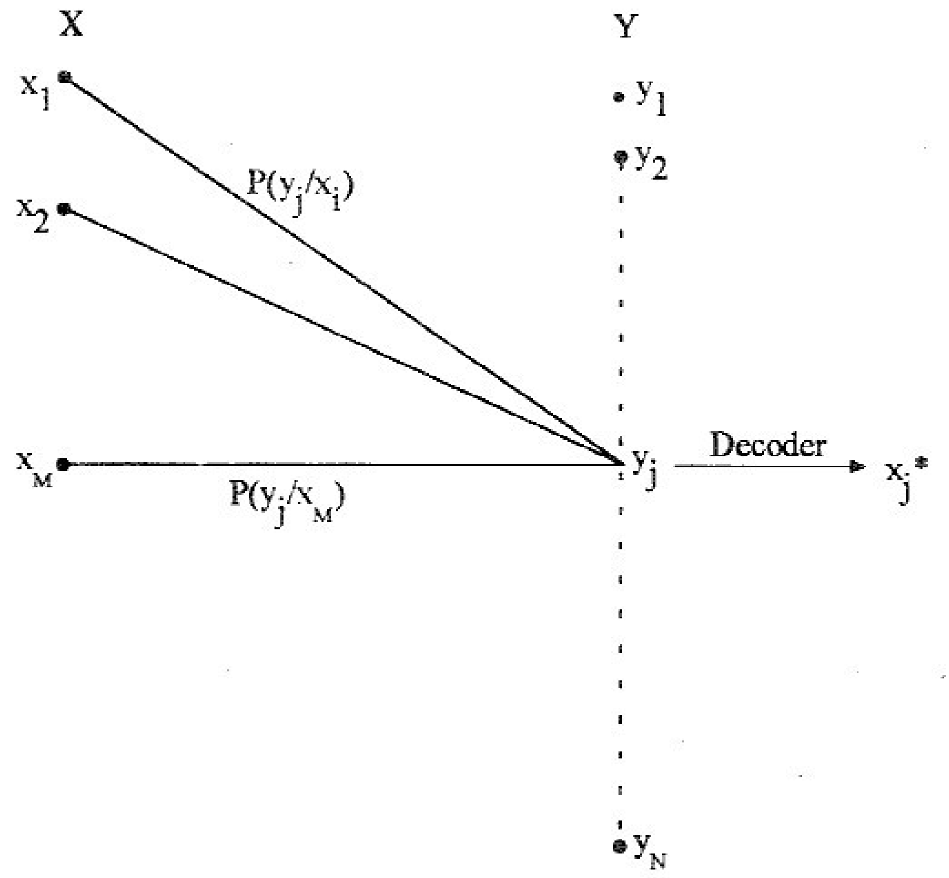

Fig. 2.11 Construction of the ideal observer

The probability $p(y_j)$ is the same for any decision scheme since $p(y_j)$ depends only on the input distribution and the channel matrix.

For each symbol y_j we are free to choose the corresponding x_j^*. It follows from (2.65) that if we choose x_j^* as that value of X which maximizes $p(x/y_j)$, we have maximized $p(e'/y_j)$ for each j, and therefore we have maximized the probability of correct transmission. To summarize, the *ideal-observer* decision scheme associates with each output symbol y_j the input symbol x that maximizes $p(x/y_j)$. If more than one input symbol yields a maximum, any one of the maximizing inputs may be chosen; the probability of error will not be affected.

Similar to the previous case, we consider a situation in which the sequence $X = (\alpha_1, \ldots, \alpha_n)$ chosen in accordance with a distribution $p(X) = p(\alpha_1, \ldots, \alpha_n)$ is transmitted through the channel. The probability that an output sequence $(\beta_1, \ldots, \beta_n)$ is produced is then given by

$$p(\beta_1, \ldots, \beta_N) = \sum_{\alpha_1 \ldots \alpha_n} p(\alpha_1, \ldots \alpha_n) p(\beta_1 \ldots \beta_n / \alpha_1 \ldots \alpha_n)$$

$$= \sum_{\alpha_1 \ldots \alpha_n} p(\alpha_1 \ldots \alpha_n) p(\beta_1 / \alpha_1) p(\beta_2 / \alpha_2) \ldots p(\beta_n / \alpha_n) \quad (2.66)$$

We may define a decoder or decision scheme as a function that assigns to each output sequence ($\beta_1, \ldots, \beta_n$) an input sequence ($\alpha_1, \ldots, \alpha_n$). The ideal observer, that is, the decision scheme minimizing the overall probability of error for the given input distribution, is just as before, the decoder that selects, for each ($\beta_1, \ldots, \beta_n$) , the input sequence ($\alpha_1, \ldots, \alpha_n$) which maximizes the conditional probability

$$p(\alpha_1, \ldots, \alpha_N / \beta_1, \ldots, \beta_N) = \frac{p(\alpha_1 \ldots \alpha_n) \underset{k=1}{\pi} p(\beta_k / \alpha_k)}{p(\beta_1 \ldots \beta_n)} \quad (2.67)$$

A special case occurs when all inputs are equally likely. If say, $x_1, \ldots x_M$ have the same probability, then

$$p(x_i / y) = \frac{p(x_i)\, p(y / x_i)}{p(y)} = \frac{1}{M\, p(y)} p(y / x_i) \quad (2.68)$$

Hence for a fixed y, maximizing $p(x_i / y)$ is equivalent to maximizing $p(y / x_i)$. Thus when all the inputs are equally likely , the ideal observer selects the input x_i for which $p(y / x_i)$ is a maximum. The resulting decoder is called *maximum Likelihood* decision scheme.

The ideal observer has several disadvantages. It is defined only for a particular input distribution ; if the input probabilities change, the decision scheme will in general also change. It may also happen that certain inputs are never received correctly. For example, in the channel of Fig 2.10 an error is always made whenever x_2 is sent. It would be more desirable to have a decision scheme with a *uniform error bound.* A decoding scheme with uniform error bound $\in$ is a decoder for which the probability of error given that x_i is sent is less than $\hat{I}$ for all i. For such a decision scheme the overall error probability is less than $\hat{I}$ for any input distribution.

2.6 Fano's Inequality

It is not possible to maintain a transmission rate R > C, while at the same time reducing the probability of error to zero. Thus we relate the probability of error to the uncertainity measure in the following theorem.

Theorem : (Fano's Inequality) Given an arbitrary code (s, n) consisting of words $x^{(1)}, \ldots, x^{(s)}$., let $X = (X_1, \ldots X_n)$ be a random vector that equals $x^{(i)}$ with probability $p(x^{(i)})$, $i = 1, 2, \ldots s$, where $\sum_{i=1}^{s} p(x^{(i)}) = 1$. Let $Y = (Y_1, \ldots Y_n)$ be the corresponding output sequence. If p(e) is the probability of error of the code, computed for the given input distribution, then

$$H(X/Y) \leq H\,[p(e), 1 - p(e)] + p(e) \log(s - 1) \qquad (2.69)$$

Proof : Let us consider any output sequence y. Let g(y) be the input selected by the decoder ; thus if y is recieved, an error will occur if and only if the transmitted sequence is not equal to g(y).

Now let us divide the set of values of X into two groups, one group consisting of all other code words.

From the basic axioms of Entropy we have,

$$H(X/Y = y) - H(q, 1 - q) + q\,H(1) + (1 - q)\,H(q_1, \ldots q_{s-1}).$$

where $q = p[\,X = g(y)/\,Y = y\,]$ and $q_1, \ldots q_{s-1}$ are of the form

$$\frac{p(X/Y)}{\sum\limits_{x \neq g(y)} p(X/Y)}$$

with x ranging over all the code words except g(y).

We have a theorem,

$$H(p_1, p_2, \ldots p_M) \leq \log M, \text{ with equality if and only if } p_i = \frac{1}{M}.$$

Thus by this theorem

$$H(q_1, \ldots q_{s-1}) \leq \log(s-1),$$

we obtain

$$H(X / Y = y) \leq H\,[\,p(e/y), 1 - p(e/y)\,] + p(e/y) \log(s-1) \qquad (2.70)$$

Now by the convexity of H,

$$H\,[p(e), 1 - p(e)] = H\left(\sum_y p(y)p(e/y), 1 - \sum_y p(y)p(e/y)\right)$$

Theorem : (Fano's Inequality) Given an arbitrary code (s, n) consisting of words $x^{(1)}, \ldots, x^{(s)}$., let $X = (X_1, \ldots X_n)$ be a random vector that equals $x^{(i)}$ with probability $p(x^{(i)})$, $i = 1, 2, \ldots s$, where $\sum_{i=1}^{s} p(x^{(i)}) = 1$. Let $Y = (Y_1, \ldots Y_n)$ be the corresponding output sequence. If p(e) is the probability of error of the code, computed for the given input distribution, then

$$H(X/Y) \le H\,[p(e), 1 - p(e)] + p(e)\log(s-1) \qquad (2.69)$$

Proof : Let us consider any output sequence y. Let g(y) be the input selected by the decoder ; thus if y is recieved, an error will occur if and only if the transmitted sequence is not equal to g(y).

Now let us divide the set of values of X into two groups, one group consisting of all other code words.

From the basic axioms of Entropy we have,

$$H(X/Y = y) - H(q, 1 - q) + q\,H(1) + (1 - q)\,H(q_1, \ldots q_{s-1}).$$

where $q = p[\,X = g(y)/\,Y = y\,]$ and $q_1, \ldots q_{s-1}$ are of the form

$$\frac{p(X/Y)}{\sum_{x \ne g(y)} p(X/Y)}$$

with x ranging over all the code words except g(y).

We have a theorem,

$$H(p_1, p_2, \ldots p_M) \le \log M, \text{ with equality if and only if } p_i = \frac{1}{M}.$$

Thus by this theorem

$$H(q_1, \ldots q_{s-1}) \le \log(s-1),$$

we obtain

$$H(X / Y = y) \le H\,[\,p(e/y), 1 - p(e/y)\,] + p(e/y)\log(s-1) \qquad (2.70)$$

Now by the convexity of H,

$$H\,[p(e), 1 - p(e)] = H\left(\sum_y p(y)p(e/y), 1 - \sum_y p(y)p(e/y)\right)$$

$$= H\left(\sum_y p(y)p(e/y), \sum_y p(y)[1-p(e/y)]\right)$$

$$\geq \sum_y p(y).H[p(e/y), 1-p(e/y)] \qquad (2.71)$$

Therefore, if we multiply (2.70) by p(y) and sum over all y, we find using (2.71) that

$$H(X/Y) \leq H[\,p(e), 1-p(e)\,] + p(e)\log(s-1) \qquad (2.72)$$

Hence the Fano's Inequality has been proved.

In the next section we shall discuss about the fundamental theorem of Information Theory.

2.7 Shannon's Fundamental Theorem (Noisy coding theorem)

The capacity of a channel is a fundamental property of an information channel in the sense that it is possible to transmit information through the channel at any rate less than the channel capacity with arbitrarily small probability of error.

The noisy coding theorem can be stated more precisely as follows : Consider a discrete memoryless channel with nonzero capacity C ; fix two numbers H and e such that

$$0 < H < C \qquad (2.73)$$

and

$$e > 0 \qquad (2.74)$$

Let us transmit *m* messages $u_1, u_2, \ldots, u_m$ by code words each of length n binary digits. The positive integer *n* can be chosen so that

$$m \geq 2^{nH} \qquad (2.75)$$

In addition at the destination the *m* sent messages can be associated with a set $V = \{v_1, v_2, \ldots, v_m\}$ of received messages and with a decision rule $d(v_j) = u_j$ such that

$$p[\,d(v_j)/v_j\,] \geq 1-e \qquad (2.76)$$

that is, decoding can be accomplished with a probability of error that does not exceed *e*.

2.8 Capacity of a bandlimited gaussian channel

The rate of transmission I(x ; y) for a continuous channel is defined as

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy \tag{2.77}$$

We may express I(x ; y) in all those forms in which we expressed it for the discrete channel.

Thus

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x,y)dxdy$$

$$- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)[\log p(x) + \log p(y)]dxdy$$

$$= -H(x,y) + H(x) + H(y)$$

Since

$$\int_{-\infty}^{\infty} p(x,y)dx = p(y)$$

$$\int_{-\infty}^{\infty} p(x,y)dy = p(x).$$

Also,

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x/y)}{p(x)}dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x/y)dxdy$$

$$- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x)dxdy$$

$$= -H(X/Y) + H(X)$$

$$= H(X) - H(X/Y) \tag{2.78}$$

Since $p(x_i, y) = p(x/y).p(y)$

Still another form of I(X, Y) may be obtained as

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(y/x)}{p(y)}dxdy$$

$$= -H(Y/X) + H(Y)$$

$$= H(Y) - H(Y/X) \tag{2.79}$$

We shall now show that I(x ; y) is nonnegative

$$I(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x/y)}{p(x)}\,dxdy$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\left[\frac{p(x)}{p(x/y)}\right]dxdy$$

$$\geq -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\left[\frac{p(x)}{p(x/y)} - 1\right]\log e\ \ dxdy$$

[using lemma, $\ln x \leq x - 1$]

$$\geq -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(y)\log e\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log e\,dxdy$$

$$\geq -\log e + \log e$$

$$I(X\,;\,Y) \geq 0 \tag{2.80}$$

Although we shall not prove it here, but it can be shown that I(X;Y) is unaffected by a linear transformation, although all individual entropies will be affected. Thus we conclude that I(X;Y) does not maintain its significance in continuous channels, even though the individual entropies appear to have no direct interpretation as far as we are concerned.

Now if x is the transmitted random variable (sampled message), y is the received variable, then y must contain the effect of channel noise. If the noise is additive, and the channel does not introduce any phase and attenuation, we must have

$$y = x + n \tag{2.81}$$

where n is the channel noise random variable.

If x is held constant at x_0, then

$$p[\,(y_1 < y \leq y_2)\,/\,x = x_0\,] = p[\,(n_1 < n \leq n_2)\,/\,x = x_0\,] \tag{2.82}$$

(where $y_1 = x_0 + n_1$ and $y_2 = x_0 + n_2$).

Thus we have the above equation written as

$$\int_{y_1}^{y_2} p(y/x)dy = \int_{n_1}^{n_2} p(n/x).dn$$

If n and x are assumed to be statistically independent as is usually the case, then

$$p(n/x) = p(n)$$

and so
$$\int_{y_1}^{y_2} p(y/x)dy = \int_{n_1}^{n_2} p(n).dn$$

Now $n = y - x_0$; $dn = dy$

We have,
$$\int_{y_1}^{y_2} p(y/x)dy = \int_{y_1}^{y_2} p(y - x_0).dy$$

$$p(y/x) = p(y - x_0) = p(n) \tag{2.83}$$

$$I(x, y) = H(y) - H(y/x) \tag{2.84}$$

$$= H(y) - H(n)$$

$$= H(y) + \int_{-\infty}^{\infty} p(n) \log p(n)\, dn \tag{2.85}$$

Let us now consider the calculation of channel capacity in the presence of additive gaussian noise with specified transmitted and noise powers.

This means that the noise n considered above has a p(n) given by

$$p(n) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-n^2/2\sigma_n^2}$$

where the noise power $N = \sigma_n^2$ is given.

(Note that σ_n^2 is the variance of noise)

Since the transmitter power S is specified, we must treat

$S = \langle x^2 \rangle$ as given,

where $$S = \langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 p(x)dx = \sigma_x^2$$

where σ_x^2 is the variance of x. We have assumed here noise and signal both as having zero mean values.

Now, $$C = \text{Max } I(X ; Y) = \text{Max } [H(y) - H(n)].$$

$$H(n) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-n^2/2\sigma_n^2} \ln\left[\frac{1}{\sqrt{2\pi}\sigma_n} e^{-n^2/2\sigma_n^2}\right] dx \quad (2.86)$$

$$= \ln \sqrt{2\pi e}.\sigma_n \text{ nats/sample} \quad (2.87)$$

(Note that H(n) is not a function of the transmitter probability density and hence is a constant as far as the maximization process is concerned.)

Thus $$C = H(y)_{max} - \ln \sqrt{2\pi e}.\sigma_n \quad (2.88)$$

Since $y = x + n$,

$$E(y) = E(x + n) = 0$$

Variance σ_y^2 of y is

$$\sigma_y^2 = \sigma_x^2 + \sigma_n^2 \quad (2.89)$$

since x and y are statistically independent.

Thus y is a random variable with specified variance and hence H(y) is maximum when y has a normal distribution (Gaussian), and so

$$H(y)_{max} = \ln \sqrt{2\pi e}\,.\sigma_y \quad (2.90)$$

Thus $$C = \ln\left(\sqrt{2\pi e\sigma_y^2}\right) - \left(\sqrt{2\pi e\sigma_n^2}\right)$$

$$= \ln\left(\frac{\sigma_y^2}{\sigma_n^2}\right)^{\frac{1}{2}}$$

$$= \ln\left(\frac{\sigma_x^2}{\sigma_n^2} + 1\right)^{\frac{1}{2}}$$

$$= \ln \left(1+\frac{S}{N}\right)^{\frac{1}{2}}$$

$$= \frac{1}{2}\ln\left(1+\frac{S}{N}\right) \text{ nats/sample} \qquad (2.91)$$

Since for the band-limited gaussian channel, we have 2W samples per second.

Hence C may be expressed as

$$C = 2W \ \frac{1}{2}\ln\left(1+\frac{S}{N}\right) \ \log_2 e \text{ bits/sec}$$

$$C = W \log_2\left(1+\frac{S}{N}\right) \text{ bits/sec} \qquad (2.92)$$

This is a well known result of communication theory (*Shannon - Hartley Law*).

Problems

Example 2.1

Calculate the capacity of lowpass channel with a usable bandwidth of 3000 Hz, and S/N = 10^3 at the channel output. Assume the channel noise to be Gaussian and white.

Solution

Channel capacity $C = W \log_2 \left(1+\frac{S}{N}\right)$

$$= (3000) \log_2 (1 + 1000)$$

$$= 30{,}000 \text{ bits/sec}$$

Example 2.2

Consider a channel with two inputs x_1, x_2 and three outputs y_1, y_2, y_3 and the noise matrix of the channel is given below. Calculate I(X, Y) with $p(x_1) = p(x_2) = 0.5$

$$P(Y/X) = \begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & 3/4 & 1/4 & 0 \\ x_2 & 0 & 1/2 & 1/2 \end{matrix}$$

Solution

We have $$P(X, Y) = \begin{bmatrix} 3/8 & 1/8 & 0 \\ 0 & 1/4 & 1/4 \end{bmatrix}$$

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

where $p(y_1) = 3/8$, $p(y_2) = 3/8$, $p(y_3) = 1/4$

Then $$H(Y) = -\frac{3}{8}\log\frac{3}{8} - \frac{3}{8}\log\frac{3}{8} - \frac{1}{4}\log\frac{1}{4}$$

$$= 1.56 \text{ bits/symbol}$$

$$H(X) = 1 \text{ bits/symbol}$$

$$H(Y/X) = 0.90 \text{ bits/symbol}$$

$$H(X/Y) = 0.4 \text{ bits/symbol}$$

Therefore,

$$I(X, Y) = H(Y) - H(Y/X) = 1.56 - 0.9 = 0.66 \text{ bits/symbol}$$

or $$I(X, Y) = H(X) - H(X/Y) = 1 - 0.34 = 0.66 \text{ bits/symbol}$$

Example 2.3

A discrete memoryless channel is characterized by the matrix

$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & 1/2 & 1/3 & 1/6 \\ x_2 & 1/6 & 1/2 & 1/3 \\ x_3 & 1/3 & 1/6 & 1/2 \end{matrix}$$

If $p(x_1) = 1/2$, $p(x_2) = p(x_3) = 1/4$, find the ideal-observer decision scheme and calculate the probability of error.

Solution

Since
$$p(y_j) = \sum_{i=1}^{3} p(x_i).p(y_j / x_i)$$

and
$$p(x_i/y_j) = \frac{p(x_i).p(y_j / x_i)}{p(y_j)}$$

we compute

$$p(x_1/y_1) = 2/3\ ,\ p(x_2/y_1) = 1/9\ ,\ p(x_3/y_1) = 2/9\ ,$$
$$p(x_1/y_2) = 1/2,\ p(x_2/y_2) = 3/8\ ,\ p(x_3/y_2) = 1/8\ ,$$
$$p(x_1/y_3) = 2/7,\ p(x_2/y_3) = 2/7\ ,\ p(x_3/y_3) = 3/7.$$

Thus the ideal observer is $g(y_1) = x_1$

$$g(y_2) = x_1$$
$$g(y_3) = x_3.$$

The probability of error is

$$\sum_{j=1}^{3} p(y_j).p(e / y_j) = \sum_{j=1}^{3} p(y_j).p\left[X \neq g(y_j) / y_j\right]$$

$$p(e) = \frac{3}{8}\left(\frac{1}{9} + \frac{2}{9}\right) + \frac{1}{3}\left(\frac{3}{8} + \frac{1}{8}\right) + \frac{7}{24}\left(\frac{2}{7} + \frac{2}{7}\right) = \frac{11}{24}$$

Example 2.4

Given a discrete memoryless channel with input alphabet $x_1, \ldots, x_M$ output alphabet $y_1, \ldots y_N$ and channel matrix $[p(y_j/x_i)]$, a randomized decision scheme may be constructed by assuming that if the channel output is y_j ($j = 1, \ldots, N$), the decoder will select x_i with probability q_{ji} ($i = 1, \ldots, M$). For a given input distribution show that no randomized decision scheme has a lower probability of error than the ideal observer.

Solution

If y_j is received, the probability of correct transmission is

$$p(e'/y_j) = \sum_{i=1}^{M} q_{ji} P\left[X = (x_i / y_j)\right].$$

Necessarily $q_{ji} > 0$ for all i, j and $\sum_{i=1}^{M} q_{ji} = 1$, (j = 1, ...N).

If $p(x_{i_0} / y_j) = \max_{1 \le i \le M} p(x_i/y_j)$, then

$$p(e'/y_j) \le \sum_{i=1}^{M} q_{ji} P\left(x_{i_0} / y_j\right) = p(x_{i_0} / y_j);$$

equality can be achieved by taking $q_{ji_0} = 1$, $q_{ji} = 0$, $i \ne i_0$.

Thus $p(e'/y_j)$ is maximized, by the above choice of q_{ji}; however, this choice is precisely the ideal observer.

Example 2.5

The output of a discrete memoryless channel k_1 is connected to the input of another discrete memoryless channel k_2. Show that the capacity of the cascade can never exceed the capacity of k_i, $i = 1, 2$. or show that "information cannot be increased by data processing".

Solution

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

$$I(X ; Z) = H(X) - H(X/Z).$$

Now, $$p(x/y, z) = \frac{p(x,y,z)}{p(y,z)} = \frac{p(x,y).p(z/x,y)}{p(y)p(z/y)}$$

The essential consequence of the cascading process is that p(z/x, y) = p(z/y) for all x, y, z, in fact, this condition must be used as the definition of a cascade combination. It follows that p(x/y, z) = p(x/y), hence H(X/Y, Z) = H(X/Y). Thus

$$I(X ; Y) - I(X ; Z) = H(X/Z) - H(X/Y, Z) \ge 0$$

by problem 1.9.

Since I(X ; Y) ≥ I(X ; Z) for any input distribution, the capacity of k_1 ≥ the capacity of the cascade.

Now we write

$$I(Y ; Z) = H(Z) - H(Z/Y)$$

$$I(X ; Z) = H(Z) - H(Z/X).$$

Since $\qquad p(z/y) = p(z/x, y), H(Z/Y) = H(Z/X, Y).$

Thus $\qquad I(Y, Z) - I(X ; Z) = H(Z/X) - H(Z/X, Y) \geq 0$

so that the capacity of k_2 ≥ the capacity of cascade.

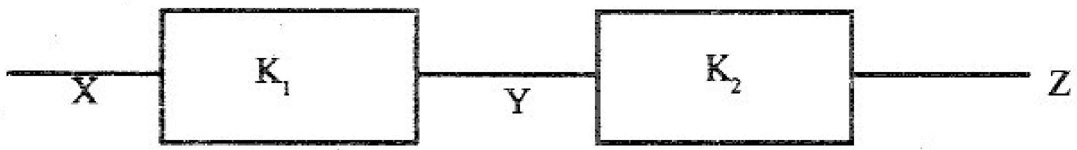

Fig. 2.12 Cascaded channels

Example 2.6

Let X and Y be discrete random variables, with X taking on the values $x_1, \ldots x_M$ and Y the values $y_1, \ldots y_N$. Let g be an arbitrary function with domain $(x_1, \ldots x_M)$. Define Z = g(X). Show that H(Y/Z) ≥ H(Y/X).

Solution

We may regard Y as the input and X the output of a discrete memoryless channel ; the output X may be thought of as being applied as the input of a deterministic channel with output Z.

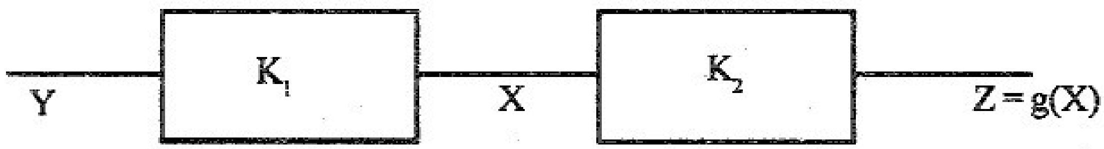

Fig. 2.13 Illustration for problem 2.6

The conditions for cascading are satisfied, that is, p(z/y, x) = p(z/x), the result then follows from the argument of problem 2.5. Alternately,

$$H(Y/Z) = E[-\log p(Y/Z)] = E[-\log p(Y/g(X))]$$

$$= -\sum_{i,j} p(x_i, y_j) \log P\left[Y = y_j / Z = g(x_i)\right]$$

$$= \sum_i p(x_i)\left[-\sum_j p(y_j / x_i) \log P\left[Y = y_j / z = g(x_i)\right]\right]$$

By problem 1.4, we have,

$$H(Y/Z) \geq \sum_i p(x_i)\left[-\sum_j p(y_j / x_i) \log P(y_j / x_i)\right]$$

$$\geq H(Y/X)$$

Example 2.7

Consider that two sources S_1 *and* S_2 *emit messages* x_1, x_2, x_3 *and* y_1, y_2, y_3 *with joint probability p(X, Y) as shown in figure. Calculate H(X), H(Y), H(X/Y), H(Y/X), and I(X ; Y) given that*

$$p(X, Y) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 3/40 & 1/40 & 1/40 \\ 1/20 & 3/20 & 1/20 \\ 1/8 & 1/8 & 3/8 \end{bmatrix} \end{array}$$

Solution

We have, $p(x_1) = 1/8$, $p(x_2) = 1/4$, $p(x_3) = 5/8$,
$p(y_1) = 1/4$, $p(y_2) = 3/10$, $p(y_3) = 9/20$,

We know, $$p(x_i, y_j) = p(x_i).p(y_j/x_i)$$

The matrices P(Y/X) and P(X/Y) are calculated by using the above equation and they are:

$$P(Y/X) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 3/5 & 1/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \\ 1/5 & 1/5 & 3/5 \end{bmatrix} \end{array}$$

$$P(X/Y) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 3/10 & 1/12 & 1/18 \\ 1/5 & 1/2 & 1/9 \\ 1/2 & 5/12 & 5/6 \end{bmatrix} \end{array}$$

$$H(X) = \frac{1}{8}\log 8 + \frac{1}{4}\log 4 + \frac{5}{8}\log\frac{8}{5} = 1.3 \text{ bits/symbol}$$

$$H(Y) = \frac{1}{4}\log 4 + \frac{3}{10}\log\frac{10}{3} + \frac{9}{20}\log\frac{20}{9} = 1.54 \text{ bits/symbol}$$

$$H(Y/X) = \frac{3}{40}\log\frac{5}{3} + \frac{1}{40}\log 5 + \frac{1}{40}\log 5 + \frac{1}{20}\log 5$$

$$+ \frac{3}{20}\log\frac{5}{3} + \frac{1}{20}\log 5 + \frac{1}{8}\log 5 + \frac{1}{8}\log 5 + \frac{3}{8}\log\frac{5}{3}$$

$$= 1.37 \text{ bits/symbol}$$

$$I(X\,;Y) = H(Y) - H(Y/X) = 1.54 - 1.37 = 0.17 \text{ bits/symbol}$$

$$= H(X) - H(X/Y) = 1.3 - 1.13 = 0.17 \text{ bits/symbol.}$$

Example 2.8

Given the noise matrix of a channel

P(Y/X) and P[X] $= [1/4, 2/5, 3/20, 3/20, 1/20]$ *calculate I(X, Y).*

$$P(Y/X) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Solution

Using P(X) as given, the P(X, Y) matrix is

$$P(X, Y) = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 1/10 & 3/10 & 0 & 0 \\ 0 & 1/20 & 1/10 & 0 \\ 0 & 0 & 1/20 & 1/10 \\ 0 & 0 & 1/20 & 0 \end{bmatrix}$$

$$P(X/Y) = \begin{bmatrix} 5/7 & 0 & 0 & 0 \\ 2/7 & 6/7 & 0 & 0 \\ 0 & 1/7 & 1/2 & 0 \\ 0 & 0 & 1/4 & 1 \\ 0 & 0 & 1/4 & 0 \end{bmatrix}$$

Then $H(X) = 2.06$

$H(Y) = 1.86$

$H(X,Y) = 2.665$

$H(Y/X) = 0.61$

$H(X/Y) = 0.81$

Therefore, $I(X, Y) = H(Y)–H(Y/X) = 1.86–0.61 = 1.25$ bits/symbol

$= H(Y)–H(X/Y) = 2.06–0.81 = 1.25$ bits/symbol

Example 2.9

Determine the capacity of the channel (unsymmetric)

$$\begin{bmatrix} 1/3 & 2/3 & 0 \\ 2/3 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Solution

$$\begin{bmatrix} 1/3 & 2/3 & 0 \\ 2/3 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3} \\ \frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3} \\ 1 \ \log 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3}Q_1 & + & \frac{2}{3}Q_2 \\ \frac{2}{3}Q_1 & + & \frac{1}{3}Q_2 \\ & Q_3 & \end{bmatrix} = \begin{bmatrix} 0.276 \\ 0.276 \\ 0 \end{bmatrix}$$

$$Q_1 + 2Q_2 = 0.828$$

$$2Q_1 + Q_2 = 0.828$$

$$Q_3 = 0$$

Therefore $Q_1 = 0.276$

$$Q_2 = 0.276$$

and $Q_3 = 0$

$$\text{Channel capacity } C = \log_2 [2^{Q_1} + 2^{Q_2} + 2^{Q_3}]$$

$$= \log_2 [2^{0.276} + 2^{0.276} + 2^0]$$

$$= \log_2 3.42$$

$$C = 1.77 \text{ bits/symbol}$$

Example 2.10

The joint probability matrix of a channel with binary input is given below :

$$\begin{array}{c c} & \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} \end{array}$$

Find the different entropies and mutual information.

Solution

The marginal probabiolities are

$$P(x_1) = P(x_2) = 1/2$$

$$P(y_1) = P(y_2) = 1/2$$

Entropies are $H(X) = H(Y) = \left[-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} \right]$

$$H(X) = H(Y) = 1$$

$$H(X, Y) = -\left[\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4}\right]$$

$$= 2.$$

Mutual information $I(X ; Y) = H(X) + H(Y) - H(X, Y)$

$$= 1 + 1 - 2$$

$$I(X ; Y) = 0$$

Example 2.11

An ideal communication system with average power limitation and perturbed by white gaussian noise has a bandwidth of 1 MHz and a signal to noise ratio of 10.

(a) Determine the channel capacity in bits/second

(b) If the signal to noise ratio drops to 5, what Bandwidth is required for the same channel capacity?

(c) If the Bandwidth is decreased to 0.5 MHz, what S/N ratio is required to maintain the same channel capacity?

Solution

(a) Channel capacity $C = W \log_2 \left[1+\frac{S}{N}\right]$

$$= (1 \times 10^6) \log_2 [1 + 10]$$

$$C = 3.45 \times 10^6 \text{ bits/second}$$

(b)

$$W \log_2 [1 + 5] = C$$

$$W = \frac{C}{\log_2 6} = \frac{3.45 \times 10^6}{\log_2 6} = 1.33 \times 10^6$$

(c)

$$W \log_2 \left[1+\frac{S}{N}\right] = C$$

$$\left[1+\frac{S}{N}\right] = 2^{C/W}$$

$$\frac{S}{N} = 2^{C/W} - 1$$

$$= 2^{6.9} - 1$$

$$\frac{S}{N} = 118.42$$

Example 2.12

Derive the expression for channel capacity of a symmetric noise characteristic channel. From the above expression, calculate the channel capacity of the given channel noise matrix.

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} p & 1-p & 0 & 0 \\ 0 & p & 1-p & 0 \\ 0 & 1-p & p & 0 \\ 0 & 0 & 1-p & p \end{bmatrix} \end{array}$$

Solution

We have derived the expression for channel capacity of the symmetric noise characteristic channel in section 2.4.

Channel capacity $C = \log m - h$

where $h = H(Y/X)$

Here $m = 2.$

$$h = H(Y/X) = -[\,p \log p + (1-p)\log(1-p)]$$

$$C = \log_2 2 + [\,p \log p + (1-p)\log(1-p)]$$

$$C = 1 + p \log p + (1-p)\log(1-p)$$

Example 2.13

A transmitter has an alphabet consisting of five letters $(x_1, x_2, x_3, x_4, x_5)$ *and a receiver has an alphabet of four letters* (y_1, y_2, y_3, y_4). *The joint probabilities for the communication are given below.*

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.10 & 0.3 & 0 & 0 \\ 0 & 0.05 & 0.1 & 0 \\ 0 & 0 & 0.05 & 0.1 \\ 0 & 0 & 0.05 & 0 \end{bmatrix} \end{array}$$

Determine the different entropies of this channel.

Solution

$f_1(x_1) = 0.25$

$f_1(x_2) = 0.1 + 0.3 = 0.4$

$f_1(x_3) = 0.05 + 0.1 = 0.15$

$f_1(x_4) = 0.05 + 0.1 = 0.15$

$f_1(x_5) = 0.05$

$f_2(y_1) = 0.25 + 0.1 = 0.35$

$f_2(y_2) = 0.3 + 0.05 = 0.35$

$f_2(y_3) = 0.1 + 0.05 + 0.05 = 0.2$

$f_2(y_4) = 0.1$

Forward path

$$f(x_1/y_1) = \frac{f(x_1,y_1)}{f_2(y_1)} = \frac{0.25}{0.35} = \frac{5}{7}$$

$$f(x_2/y_2) = \frac{f(x_2,y_2)}{f_2(y_2)} = \frac{0.3}{0.35} = \frac{6}{7}$$

$$f(x_3/y_3) = \frac{f(x_3,y_3)}{f_2(y_3)} = \frac{0.1}{0.2} = \frac{1}{2}$$

$$f(x_3/y_4) = \frac{f(x_4,y_4)}{f_2(y_4)} = \frac{0.1}{0.1} = 1$$

$$f(x_2/y_1) = \frac{0.01}{0.35} = \frac{2}{7}$$

$$f(x_3/y_2) = \frac{0.05}{0.35} = \frac{1}{7}$$

Backward path

$$f(y_1/x_1) = \frac{f(x_1,y_1)}{f_1(x_1)} = \frac{0.25}{0.25} = 1$$

$$f(y_2/x_2) = \frac{0.3}{0.4} = \frac{3}{4}$$

$$f(y_3/x_3) = \frac{0.1}{0.15} = \frac{2}{3}$$

$$f(y_4/x_4) = \frac{0.1}{0.15} = \frac{2}{3}$$

$$f(y_1/x_2) = \frac{0.1}{0.4} = \frac{1}{4}$$

$$f(y_2/x_3) = \frac{0.05}{0.15} = \frac{1}{3}$$

$$f(x_4/y_3) = \frac{0.05}{0.02} = \frac{1}{4} \qquad f(y_3/x_4) = \frac{0.05}{0.15} = \frac{1}{3}$$

$$f(x_5/y_3) = \frac{0.05}{0.2} = \frac{1}{4} \qquad f(y_3/x_5) = \frac{0.05}{0.05} = 1$$

$$H(X;Y) = -\sum_x \sum_y f(x,y) \log f(x,y)$$

$$= -0.25 \log_2 0.25 - 0.1 \log_2 0.4 - 0.3 \log_2 0.4 - 0.05 \log_2 0.05$$

$$= -0.1 \log_2 0.1 - 0.05 \log_2 0.05 - 0.1 \log_2 0.1 - 0.05 \log_2 0.05$$

$$H(X\,;\,Y) = 2.665$$

$$H(X) = -\sum\sum f(x,y) \log f_1(x)$$

$$= 2.066$$

$$H(Y) = -\sum\sum f(x,y) \log f_2(y)$$

$$= 1.856$$

$$H(Y/X) = -\sum\sum f(x,y) \log \frac{f(x,y)}{f_1(y)} = 0.6$$

$$H(X/Y) = -\sum\sum f(x,y) \log \frac{f(x,y)}{f_2(y)} = 0.809$$

Note: $H(X, Y) = H(Y) + H(X/Y) = H(X) + H(Y/X)$.

Example 2.14

An analog signal is bandlimited to B Hz, sampled at the Nyquist rate and the samples are quantized into 4 levels. The quantization levels Q_1, Q_2, Q_3 and Q_4 (messages) are assumed independent and occur with probabilities $p_1 = p_4 = 3/8$ and $p_2 = p_3 = 1/8$. Find the information rate of the source.

Solution

If the source of the messages generator messages at the rate of *r* messages per second, then the information rate is defined to be

$$R = rH = \text{average number of bits of information/sec.}$$

where H = entropy

$r = 2f$

$$H = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + p_3 \log_2 \frac{1}{p_3} + p_4 \log_2 \frac{1}{p_4}$$

$$= \frac{1}{8} \log_2 8 + \frac{3}{8} \log_2 \frac{8}{3} + \frac{3}{8} \log_2 \frac{8}{3} + \frac{1}{8} \log_2 8$$

$H = 1.8$ bits/message

Information rate $R = rH = (2B)(1.8)$

$= 3.6$ B bits/second

Example 2.15

Consider the four messages of the previous example. Let Q_1, Q_2, Q_3 and Q_4 have probabilities 1/2, 1/4, 1/8, 1/8

(a) Calculate H

(b) Find R if r = 1 message/sec.

Solution

(a)

$$H = p_1 \log \frac{1}{p_1} + p_2 \log \frac{1}{p_2} + p_3 \log \frac{1}{p_3} + p_4 \log \frac{1}{p_4}$$

$$= \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{8} \log_2 8 + \frac{1}{8} \log_2 8$$

$$H = \frac{7}{4} \text{ bits/message}$$

(b)

$$R = rH$$

$$= (1)\ \frac{7}{4}$$

$$= \frac{7}{4} \text{ bits/sec}$$

Example 2.16

Find the capacity of the following three binary channels, given below.

(a) $p_{11} = p_{22} = 1$ (b) $p_{11} = p_{12} = p_{21} = p_{22} = \frac{1}{2}$

(c) $p_{11} = p_{12} = \frac{1}{2}$; $p_{21} = \frac{1}{4}$; $p_{22} = \frac{3}{4}$

Solution

(a)

$$p_{11} = p_{22} = 1$$

$$p_{21} = p_{21} = 0$$

$$Q_1 = \log p_{11} = \log 1 = 0$$

$$Q_2 = \log p_{22} = \log 1 = 0$$

$$\text{Channel capacity} \quad C = \log \sum_{i=1}^{n} \exp(Q_i)$$

$$C = \log_2 [2^{Q_1} + 2^{Q_2}]$$

$$= \log_2 [2^0 + 2^0]$$

$$= \log_2 2$$

$$C = 1 \text{ bit}$$

(b)

$$P_{11} = P_{12} = P_{21} = P_{22} = \frac{1}{2}$$

$$Q_1 = \log P_{11} = \log_2 \frac{1}{2} = -1$$

$$Q_2 = \log P_{22} = \log_2 \frac{1}{2} = -1$$

Channal capacity $C = \log_2 [2^{-1} + 2^{-1}]$

$= \log_2 1$

$C = 0.$

(c) $P_{11} = P_{12} = \frac{1}{2}$

$P_{21} = \frac{1}{4}\ ;\ P_{22} = \frac{3}{4}$

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} 1/2 \log 1/2 & + & 1/2 \log 1/2 \\ 1/4 \log 1/4 & + & 3/4 \log 3/4 \end{bmatrix}$$

$$= \begin{bmatrix} & -1 & \\ -2 & + & (3/4)\log 3 \end{bmatrix}$$

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -0.81 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{2}Q_1 & + & \frac{1}{2}Q_2 \\ \frac{1}{4}Q_1 & + & \frac{3}{4}Q_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -0.81 \end{bmatrix}$$

$Q_1 + Q_2 = -2$

$Q_1 + 3Q_2 = -3.24$

Therefore $Q_1 = -1.38$

$Q_2 = -0.62$

Channal capacity $C = \log_2 [2^{-1.38} + 2^{-0.62}]$

$C = 0.049$ bits

Example 2.17

Find the capacity of the channel with the noise matrix as shown below:

$$\begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/4 & 0 & 1/4 & 1/2 \end{bmatrix}$$

Solution

$$\begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/4 & 0 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4} \\ 1 \log 1 \\ 1 \log 1 \\ \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{2}\log\frac{1}{2} \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{2}Q_1 + \frac{1}{4}Q_2 + \frac{1}{4}Q_4 \\ Q_2 \\ Q_3 \\ \frac{1}{4}Q_1 + \frac{1}{4}Q_3 + \frac{1}{2}Q_4 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 0 \\ 0 \\ -1.5 \end{bmatrix}$$

$$\tfrac{1}{2}Q_1 + \tfrac{1}{4}Q_4 = -1.5$$

$$\tfrac{1}{4}Q_1 + \tfrac{1}{2}Q_4 = -1.5$$

$$Q_1 = -2 \qquad Q_2 = 0$$

and $$Q_3 = 0 \qquad Q_4 = -2$$

Channel Capacity $C = \log_2 [2^{Q_1} + 2^{Q_2} + 2^{Q_3} + 2^{Q_4}]$

$$= \log_2 [2^{-2} + 2^0 + 2^0 + 2^{-2}]$$

$$C = 1.322 \text{ bits}$$

Example 2.18

Find the channel matrix of a cascade combination in terms of the individual channel matrices.

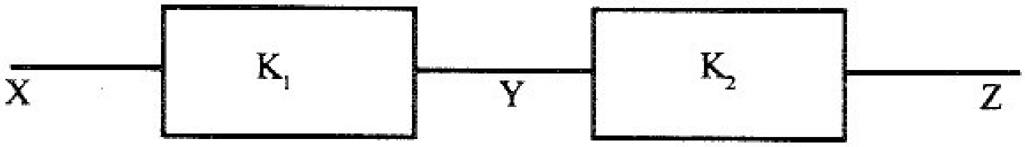

Fig. 2.14 Channel matrix of example 2.18

Solution

Let π_i = channel matrix of k_i, i =1, 2.

We have
$$p(z/x) = \sum_y p(y/x)p(z/y)$$

Thus the matrix of cascade combination is $\pi_1 \pi_2$; similarly if channels k_1, k_2, . . . k_n are cascaded (in that order) the matrix of the cascade combination is the product $\pi_1 \pi_2 \ldots \pi_n$.

Example 2.19

Find the capacity of the general binary channel whose channel matrix is

$$\begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} \quad \textit{where } \alpha \neq \beta.$$

Solution

Let us state a theorem which shall be used to evaluate the channel capacity of the given binary channel.

Theorem:

Suppose that the channel matrix π of a discrete memory less channel is square and nonsingular. Let q_{ij} be the element in row i and column j of π^{-1}., i, j = 1, ...M. Suppose that for each k = 1, 2, ..., M.

$$d_k = \sum_{j=1}^{M} q_{jk} \exp_2\left[-\sum_{i=1}^{M} q_{ji} H(Y/X=x_i)\right] > 0$$

The channel capacity is given by

$$C = \log \sum_{j=1}^{M} \exp_2\left[-\sum_{i=1}^{M} q_{ji} H(Y/X=x_i)\right]$$

and a distribution that achieves capacity is given by

$$P(x_k) = 2^{-C} d_k, \quad k=1, 2, \ldots, M.$$

Let us assume here without loss of generality that $\beta > \alpha$.

We have $$\pi = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

$$\pi^{-1} = \frac{1}{\begin{vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{vmatrix}} \begin{bmatrix} 1-\beta & -(1-\alpha) \\ -\beta & \alpha \end{bmatrix}$$

$$= \frac{1}{\alpha-\beta} \begin{bmatrix} 1-\beta & \alpha-1 \\ -\beta & \alpha \end{bmatrix}$$

Let $$H_\alpha = H(\alpha, 1-\alpha) = H(Y/X=x_1)$$

$$H_\beta = H(\beta, 1-\beta) = H(Y/X=x_2).$$

We must show that the hypothesis of the theorem stated above is satisfied.

We have $d_1 = -\dfrac{1-\beta}{\beta-\alpha} \exp_2\left[\dfrac{(1-\beta)H_\alpha + (\alpha-1)H_\beta}{\beta-\alpha}\right]$

$$+\frac{\beta}{\beta-\alpha}\exp_2\left[\frac{-\beta H_\alpha+\alpha H_\beta}{\beta-\alpha}\right]$$

Thus $d_1 > 0$ if and only if

$$\log\left(\frac{1-\beta}{\beta}\right) < \frac{-\beta H_\alpha+\alpha H_\beta-(1-\beta)H_\alpha-(\alpha-1)H_\beta}{\beta-\alpha}$$

or $$\log\left(\frac{1-\beta}{\beta}\right) < \frac{H_\beta-H_\alpha}{\beta-\alpha}$$

Upon expansion, we have

$$-\alpha\log\alpha-(1-\alpha)\log(1-\alpha) < -\alpha\log\beta-(1-\alpha)\log(1-\beta)$$

which is true by the Lemma of problem (1.4). Consequently $d_1 > 0$.

A similar argument shows that $d_2 > 0$.

Thus channel capacity

$$C=\log\left\{\exp_2\left[\frac{(1-\beta)H_\alpha-(\alpha-1)H_\beta}{\beta-\alpha}\right]+\exp_2\left[\frac{-\beta H_\alpha+\alpha H_\beta}{\beta-\alpha}\right]\right\}$$

Example 2.20

Find the capacity of the channel illustrated in fig. 2.15.

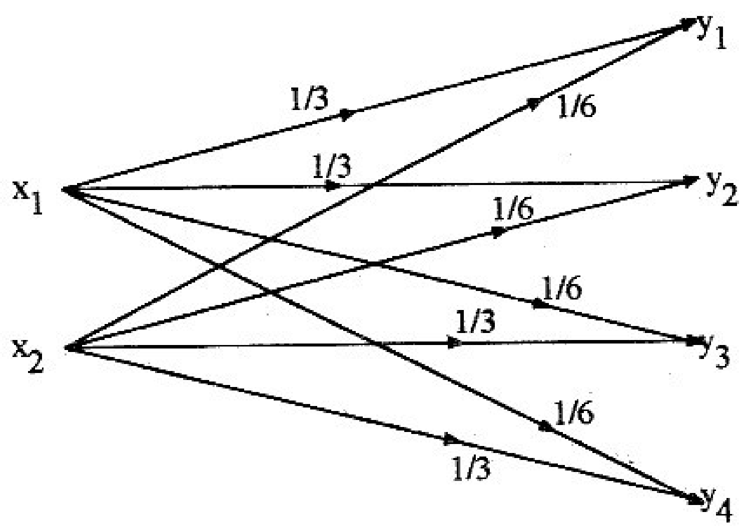

Fig. 2.15 Channel representation

Solution

For a symmetric channel,

$$C = \log_2 m - h \qquad \text{where } h = H(Y/X)$$

$$C = \log_2 4 - \frac{2}{3}\log_2 3 - \frac{1}{3}\log_2 6$$

$$C = \frac{5}{3} - \log_2 3 \text{ bits}$$

Example 2.21

A binary channel has the following noise characteristic :

$$\begin{array}{c} \\ 0 \\ \\ 1 \end{array}\!\!\begin{array}{cc} 0 & 1 \\ \left[\begin{array}{cc} 2/3 & 1/3 \\ & \\ 1/3 & 2/3 \end{array}\right] \end{array}$$

(a) If the input symbols are transmitted with respective probabilities of 3/4 and 1/4, find H(X), H(Y), H(X/Y), H(Y/X) and I(X ; Y).

(b) Find the channel capacity and the corresponding input probabilities.

Solution

(a) $H(X) = 0.81 \qquad H(Y) = 0.98$

$H(X/Y) = 0.75 \qquad H(Y/X) = 0.92 \qquad I(X ; Y) = 0.06$

(b)
$$C = 1 + \frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}$$

$$= \frac{5}{3} - \log 3$$

$$= 0.08$$

$$p(0) = p(1) = \frac{1}{2}$$

Example 2.22

Determine the capacity of a ternary channel with the stochastic matrix

$$[p] = \begin{bmatrix} \alpha & 1-\alpha & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1-\alpha & \alpha \end{bmatrix}, 0 \le \alpha \le 1$$

Solution

Since the channel matrix is a square matrix,

$$[P]\,[Q] = -\,[H]$$

$$[Q] = -\,[P]^{-1}\,[H]$$

$$\begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} = -\begin{bmatrix} \frac{1}{2\alpha} & 1 & \frac{-1}{2\alpha} \\ \frac{1}{2(1-\alpha)} & \frac{\alpha}{1-\alpha} & \frac{1}{2(1-\alpha)} \\ \frac{-1}{2\alpha} & 1 & \frac{1}{2\alpha} \end{bmatrix} \begin{bmatrix} h \\ 1 \\ h \end{bmatrix}$$

where $h = -\alpha \log \alpha - (1-\alpha) \log (1-\alpha)$

$$Q = \begin{bmatrix} -1 \\ -\left(\frac{h-\alpha}{1-\alpha}\right) \\ -1 \end{bmatrix}$$

$$C = \log_2 [2^{Q_1} + 2^{Q_2} + 2^{Q_3}]$$