Question Paper

Govt. Mahila Engineering College, Ajmer

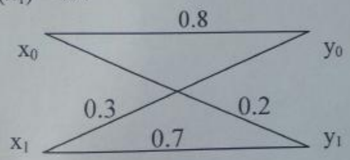
Ist Mid Term Examination-2017-18 (Even Sem.)

Information Theory and Coding (2MDC2)

Max. Marks: 10 | Max Time: 01 Hour

Note: - All Questions are Compulsory and carry equal marks.

1. State and prove Shannon's channel capacity theorem. What is its significance?
2. Prove the relation $0 \leq H(X) \leq \log_2 M$. Where M is size of the alphabet of X.
3. Describe the rate distortion theory and show that it may be viewed as a natural extension of Shannon coding theorems.
4. Find the mutual information and channel capacity of channel shown below:
$P(x_0) = 0.6$, $P(x_1) = 0.4$



ANS 1

Capacity of a Gaussian Channel :- Shannon - Hartley theorem :-

The noise char^r of channels encountered in practice is generally gaussian. If particular encoder - decoder is used with a Gaussian channel, giving an error probability P_e, then, with a non - Gaussian Channel, another encoder - decoder can be designed for which the error probability will be less than P_e.

for a Gaussian channel $\quad P(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{x^2}{2\sigma^2}\right] \quad$ P.D.F. $\quad$ (1) $\quad x(t)$

Hence $H(x) = -\int_{-\infty}^{\infty} P(x) \log P(x) dx$

$$-\log P(x) = \log\sqrt{2\pi\sigma^2} + \log e^{x^2/2\sigma^2}$$

Hence $H(x) = \int_{-\infty}^{\infty} P(x) \log\sqrt{2\pi\sigma^2}\, dx + \int_{-\infty}^{\infty} P(x) \log e^{x^2/2\sigma^2} dx$

$$H(x) = \log\sqrt{2\pi e\sigma^2} \quad \text{bits/message} \qquad (2)$$

Now, if the signal is band limited to B Hz, then it may be uniquely specified by taking 2B samples per second. Hence, the rate of infor transmission is $R(x) = 2B\,H(x) = 2B \log\left(\sqrt{2\pi e\sigma^2}\right)$

$$= B \log\left(\sqrt{2\pi e\sigma^2}\right)^2$$

$$\boxed{R(x) = B\log(2\pi e\sigma^2)} \qquad (3)$$

If $P(x)$ is a band limited Gaussian noise with an average noise power N, then we have $R(n) = \cancel{R(x)} = B\log(2\pi e N) \qquad (\because \sigma^2 = N)$

Now Consider a Continuous source transmitting infor over noisy Channel. If the received signal is Composed of a transmitted signal x and a noise n, then the joint entropy (bit/sec) of the Source and noise is given by

$$R(x,n) = R(x) + R(n/x)$$

Assuming that the transmitted signal and noise are independent

$$R(x,n) = R(x) + R(n) \qquad (4)$$

Since the received signal y is the sum of the transmitted signal x and the noise n, we may equate $\quad H(x,y) = H(x,n)$

or $\quad H(y) + H(x/y) = H(x) + H(n)$

or $\quad R(y) + R(x/y) = R(x) + R(n) \qquad (5)$

The rate at which the information is received from a noisy channel is (6)

$$R = R(x) - R(x/y)$$

and $R = R(y) - R(n)$ bit/sec   Using eq (5)

The channel capacity is bit/sec is $C = \max[(R)]$ bits/sec

$$C = \max[R(y) - R(n)] \text{ bits/sec} \quad \text{—— (6)}$$

Since $R(n)$ is assumed to be independent of $x(t)$, maximing R requires maximizing $R(y)$.

Let a transmitted signal be limited to an average power S and the noise on the channel be white Gaussian with an average power N within the bandwidth B of the channel. The received signal will now have an average power $(S+N)$. $R(y)$ is maximum when $y(t)$ is also a Gaussian random process because noise is assume to be Gaussian.

The entropy $\therefore R(y) = B \log[2\pi e(S+N)]$ bits/sec

while the entropy of the noise is given by $R(n) = B \log[2\pi eN]$ bit/sec.

The channel capacity may now be obtained directly since $R(y)$ has been maximized

$$C = \max[R(y) - R(n)] = B \log[2\pi e[S+N]] - B \log[2\pi eN]$$

$$C = B \log\left[\frac{S+N}{N}\right] = B \log\left[1 + \frac{S}{N}\right] \text{ bits/sec.} \quad \text{—— (7)}$$

equation seven is the famous Shannon-Hartley theorm, which is Complementry to shannon's theorm applies to Gaussian noise channel.

"The channel capacity of a white bandlimited Gaussian channel is

$C = B \log\left[1 + \frac{S}{N}\right]$ bits/sec.   where B = Channel Bandwidth
S = avg. signal power
N = avg. noise power

If $\eta/2$ is two sided power spectral density of noise in watts/Hz then

$N = \eta B$ and $C = B \log\left(1 + \frac{S}{\eta B}\right)$ bits/sec. —— (8)

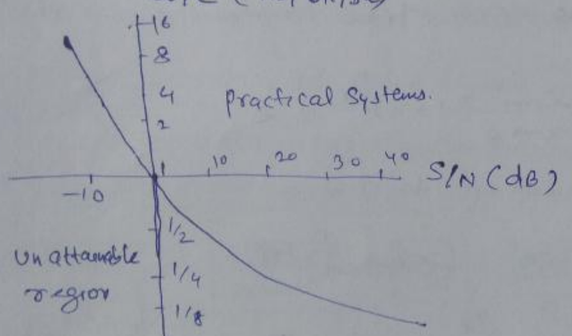
SHANNON - HARTLEY CAPACITY THEORM & Shannon bound :-

* The System Capacity of a channel Perturbed by additive white Gaussian noise (AWGN) is a function of the average received signal power S, the average noise power N, and the bandwidth W.

$$C = W \log_2\left(1+\frac{S}{N}\right) \quad \text{—①}$$

W in Hertz
Capacity C in bits/sec

* It is theoretically possible to transmit information over a such a channel at any rate R, where $R \le C$, with an arbitrarily small error probability by using a sufficiently complicated coding scheme.
* Set a limit on transmission rate, not on error probability.



Normalized channel bandwidth (W/C) versus channel (S/N).

The detected noise power is propotional to bandwidth

$$N = N_0 W \quad \text{—②}$$

so $$\frac{C}{W} = \log_2\left(1+\frac{S}{N_0 W}\right) \quad \text{—③}$$

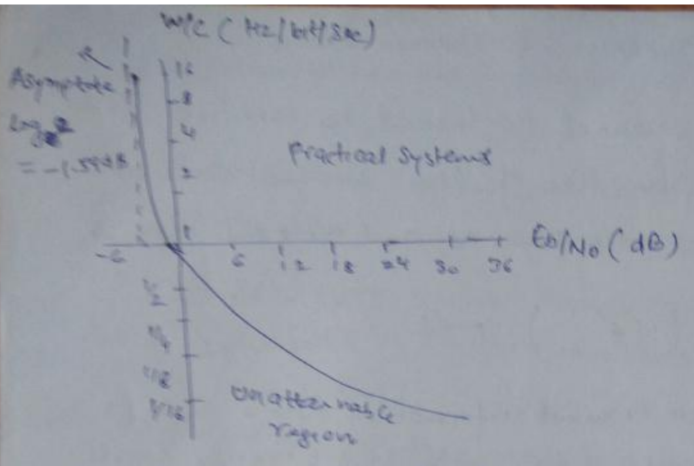
For the case where transmission bit rate is equal to channel capacity $R=C$, we can use the identity

$$\frac{E_b}{N_0} = \frac{ST_b}{N/W} = \frac{S/R}{N/W} = \frac{S}{N}\left(\frac{W}{R}\right) \quad \text{—④}$$

so $$\frac{E_b}{N_0} = \frac{S}{N_0 W}\left(\frac{W}{R}\right) = \frac{S}{N_0 R} = \frac{S}{N_0 C} \quad (R=C)$$

so $$\frac{C}{W} = \log_2\left(1+\frac{E_b}{N_0}\left(\frac{C}{W}\right)\right) \quad \text{—⑤}$$

$$2^{C/W} = 1+\frac{E_b}{N_0}\left(\frac{C}{W}\right) \quad \text{or} \quad \frac{E_b}{N_0} = \frac{W}{C}\left(2^{C/W}-1\right) \quad \text{—⑥}$$



Normalized channel bandwidth versus channel E_b/N_0

Shannon bound :- There exist a limiting value of E_b/N_0 below which there can be no error free communication at any information rate.

$$\lim_{x \to 0} (1+x)^{1/x} = e$$

The limiting value of E_b/N_0 calculated as :-

Let $x = \frac{E_b}{N_0}\left(\frac{C}{W}\right)$ Then

$$\frac{C}{W} = x \log (1+x)^{1/x} \quad \text{and} \quad \frac{C}{W} = \frac{E_b}{N_0}\left(\frac{C}{W}\right)\log_2 (1+x)^{1/x}$$

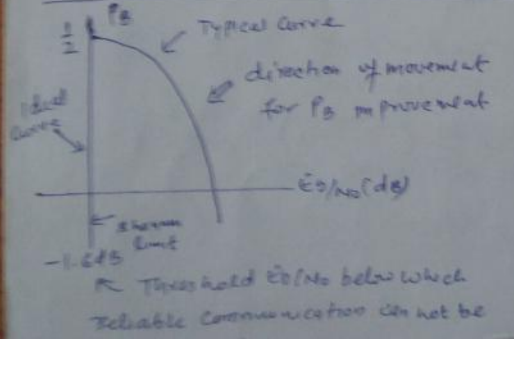
or $1 = \frac{E_b}{N_0} \log_2 (1+x)^{1/x}$

In the limit, as $C/W \to 0$ (or $W/C \to \infty$) we get $\frac{E_b}{N_0} = \frac{1}{\log_2 e} = 0.693$

or in dB $\frac{E_b}{N_0} = -1.59\,dB$

This value is a Shannon limit.

Ideal probability of bit error performance :-



* Once E_b/N_0 is reduced below the shannon limit $-1.6\,dB$, P_B degrades to the worst case $1/2$. ($P_B = 1$ is not worst case for binary signalling, since the value is just good as $P_B = 0$; if the probability of making a bit error is 100%, the bit stream simply be inverted).

ANS 2

(2) $I_t = P_1 L \log_2 \frac{1}{P_1} + P_2 L \log_2 \frac{1}{P_2} + \dots + P_m L \log_2 \frac{1}{P_m} = \sum_{k=1}^{m} P_k L \log_2 \frac{1}{P_k}$ bits — (5)

So the average information per message ~~(symbols)~~ may be obtained by dividing I_t by the number of ~~symbols in a~~ message

$$H = \frac{I_t}{L} = P_1 \log_2 \frac{1}{P_1} + P_2 \log \frac{1}{P_2} + \dots + P_m \log \frac{1}{P_m} = \sum_{k=1}^{m} P_k \log \frac{1}{P_k} = -\sum_{k=1}^{m} P_k \log P_k \quad \text{— (6) bit/symbol}$$

The ~~average~~ average information content per symbol (message) is called Source entropy.

* if there is only a single possible message $M = 1$ and $P_k = P_1 = 1$ Then
$H = P_1 \log \frac{1}{P_1} = 1 \log \frac{1}{1} = 0$ i.e. reception of that message Convey no information.

* To study the variation of H between the two extremes we consider the special case of binary system ($m = 2$) $H = P_1 \log \frac{1}{P_1} + P_2 \log \frac{1}{P_2}$ — (7)

Let the source emit two message (symbol) $P_1 = P$ and $P_2 = 1 - P_1 = 1 - P = q$ where $0 < P \leq 1$

The avg. info^{n} or entropy of the Source is $H = P \log \frac{1}{P} + (1-P) \log \frac{1}{(1-P)}$

$H = P \log \frac{1}{P} + q \log \frac{1}{q} = \Omega(P)$ — (8)
Horse shoe function

Plot of H as a function of P :-

(1) at $P = 0$ and $P = 1$, $H = 0$

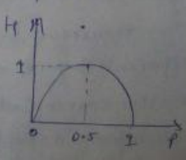
(2) The Condition for maximum entropy may be obtained by setting $\frac{dH}{dP} = 0$

$\frac{dH}{dP} = -\ln 2 - \log P + \ln 2 + \log(1-P) = 0$

$\log P = \log(1-P)$ or $P = 1 - P$ $P = 0.5 = \frac{1}{2}$ ~~there exist is maximum at P = 1/2~~

This concludes that there is either a maxima or minima at $P = \frac{1}{2}$. If the second derivative of H is positive, then there is a minima and if it is negative, then it is a maxima, Now

$\frac{d^2H}{dP^2} = -\frac{1}{P} - \frac{1}{1-P} < 0$ Hence H has a maxima at $P = \frac{1}{2}$



The maximum value of H can be found by putting $P = 1/2$,

$H_{max} = \frac{1}{2} \log_2 \frac{1}{1/2} + \frac{1}{2} \log_2 \frac{1}{1/2} = 1$ bit/message (Symbol)

* Similarly, it can be shown that for m-ary Case, the entropy is maximum when all the messages are equally likely. Thus $P_1 = P_2 = \dots = P_m = \frac{1}{M}$.

In this case, the maximum entropy is $H_{max} = \sum_{k=1}^{M} p_k \log_2 \frac{1}{p_k} = M\left(\frac{1}{M}\log_2 M\right)$

i.e. $\boxed{H_{max} = \log_2 M \text{ bits/message (symbol)}}$ — (9)

The value of source entropy depends upon the symbol probabilities p_k and also on the alphabet size M $\quad 0 \le H \le \log_2 M$

(1) $H = 0$ if all probabilities are zero, except for one, which must be unity. This lower limits corresponds to no uncertainity. This means that the source always emits the same symbol or message.

(2) $H = \log_2 M$ if all the probabilities are equally so that $p_k = \frac{1}{M}$ for all k's. This upper limits corresponds to maximum freedom of choice, means that all the message or symbol equally likely.

for $M = 2$	Case 1	$P_1 = 0.01$	$P_2 = 0.99$	$H = 0.08$	less uncertainity
	2	$P_1 = 0.4$	$P_2 = 0.6$	$H = 0.97$	more uncertinity
	3	$P_1 = 0.5$	$P_2 = 0.5$	$H = 1$	maximum uncertinity.

entropy is a measure of uncertainty.

Rate of information :- if a message source generates messages at the rate

ANS 3

that they have flagged a count. The remaining 6 bits in the flag/count byte will be interpreted as a 6 bit binary number for the count (from 0 to 63). This byte is then followed by the byte which represents the color. In fact, if we have a run of pixels of one of the colors with palette code even over 191, we can still code the run easily since the top two bits are not reserved in this second color code byte of a run-coding byte pair.

If a run of pixels exceeds 63 in length, we simply use this code for the first 63 pixels in the run, and that runs additional codes of that pixel until we exhaust all pixels in the run. The next question is: how do we code those remaining colors in a nearly full palette image when there is no run? We still code these as a run by simply setting the run length to 1. That means, for the case of at most 64 colors which appear as single pixels in the image and not part of runs, we expand the data by a factor of two. Luckily this rarely happens.

In the next section, we will study coding for analog sources. Recall that we ideally need infinite number of bits to accurately represent an analog source. Anything fewer will only be an approximate representation. We can choose to use fewer and fewer bits for representation at the cost of a poorer approximation of the original signal (rule of thumb: there is no free lunch). Thus, quantisation of the amplitudes of the sampled signals results in data compression. We would like to study the distortion introduced when the samples from the information source are quantised.

1.11 RATE DISTORTION FUNCTION

Although we live in an analog world, most of the communication takes place in the digital form. Since most natural sources (e.g. speech, video etc.) are analog, they are first sampled, quantised and then processed. However, the representation of an arbitrary real number requires an infinite number of bits. Thus, a finite representation of a continuous random variable can never be perfect.

Consider an analog message waveform $x(t)$ which is a sample waveform of a stochastic process $X(t)$. Assuming $X(t)$ is a bandlimited, stationary process, it can be represented by a sequence of uniform samples taken at the Nyquist rate. These samples are quantised in amplitude and encoded as a sequence of binary digits. A simple encoding strategy can be to define L levels and encode every sample using

$$R = \log_2 L \text{ bits if } L \text{ is a power of 2, or}$$

$$R = \lfloor \log_2 L \rfloor + 1 \text{ bits if } L \text{ is not a power of 2.} \qquad (1.52)$$

If all levels are not equally probable we may use entropy coding for a more efficient representation. In order to represent the analog waveform more accurately, we need more number of levels, which would imply more number of bits per sample. Theoretically we need infinite bits per sample to perfectly represent an analog source. Quantisation of amplitude results in data compression at the cost of signal distortion. It is a form of lossy data compression. Distortion implies some measure of difference between the actual source samples $\{x_k\}$ and the corresponding quantised value $\{\tilde{x}_k\}$.

Definition 1.14 The **Squared Error Distortion** is defined as

$$d(x_k, \tilde{x}_k) = (x_k - \tilde{x}_k)^2 . \tag{1.53}$$

In general a distortion measure may be represented as

$$d(x_k, \tilde{x}_k) = |x_k - \tilde{x}_k|^p . \tag{1.54}$$

Consider a sequence of n samples X_n, and the corresponding n quantized values $\tilde{X}_n$. Let $d(x_k, \tilde{x}_k)$ be the distortion measure per sample (letter). Then the distortion measure between the original sequence and the sequence of quantized values will simply be the average over the n source output samples i.e.,

$$d(X_n, \tilde{X}_n) = \frac{1}{n}\sum_{k=1}^{n} d(x_k, \tilde{x}_k) . \tag{1.55}$$

We observe that the source is a random process, hence X_n and consequently $d(X_n, \tilde{X}_n)$are random variables. We now define the distortion as follows.

Definition 1.15 The **Distortion** between a sequence of n samples X_n, and their corresponding n quantised values $\tilde{X}_n$ is defined as

$$D = E\,[d(X_n, \tilde{X}_n)] = \frac{1}{n}\sum_{k=1}^{n} E[d(x_k, \tilde{x}_k)] = E[d(x_k, \tilde{x}_k)] . \tag{1.56}$$

It has been assumed here that the random process is stationary. Next, let a memoryless source have a continuous output X and the quantised output alphabet $\tilde{X}$. Let the probability density function of this continuous amplitude be $p(x)$ and per letter distortion measure be $d(x, \tilde{x})$, where $x \in X$ and $\tilde{x} \in \tilde{X}$. We next introduce the rate distortion function, which gives us the minimum number of bits per sample required to represent the source output symbols given a prespecified allowable distortion.

Definition 1.16 The minimum rate (in bits/source output) required to represent the output X of the memoryless source with a distortion less than or equal to D is called the **Rate Distortion Function** $R(D)$, defined as

$$R(D) = \min_{p(\tilde{x}|x):E[d(X,\tilde{X})]\le D} I(X;\tilde{X}) , \tag{1.57}$$

where $I(X; \tilde{X})$ is the average mutual information between X and $\tilde{X}$.

We will now state (without proof) two theorems related to the rate distortion function.

Theorem 1.3 The minimum information rate necessary to represent the output of a discrete time, continuous amplitude memoryless Gaussian source with variance σ_x^2, based on a mean square-error distortion measure per symbol, is

$$R_g(D) = \begin{cases} \frac{1}{2}\log_2(\sigma_x^2/D) & 0 \le D \le \sigma_x^2 \\ 0 & D > \sigma_x^2 \end{cases} . \tag{1.58}$$

Consider the two cases:

(i) $D \ge \sigma_x^2$: For this case there is no need to transfer any information. For the reconstruction of the samples (with distortion greater than or equal to the variance) one can use statistically independent, zero mean Gaussian noise samples with variance $D - \sigma_x^2$.

(ii) $D < \sigma_x^2$: For this case the number of bits per output symbol decreases monotonically as D increases. The plot of the rate distortion function is given in Fig. 1.17.

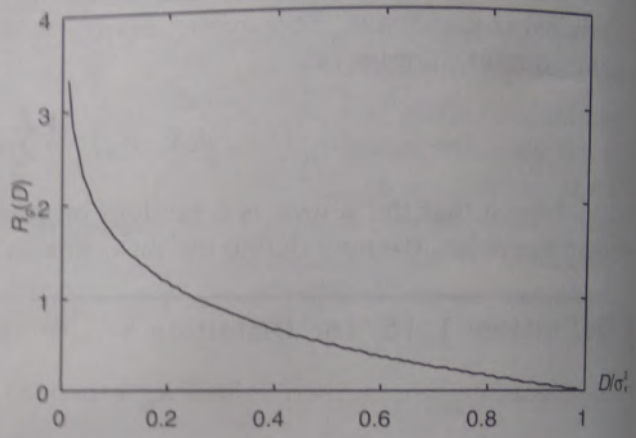


Fig. 1.17 Plot of the $R_g(D)$ versus D/σ_x^2.

Theorem 1.4 There exists an encoding scheme that maps the source output into codewords such as that for any given distortion D, the minimum rate $R(D)$ bits per sample is sufficient to reconstruct the source output with an average distortion that is arbitrarily close to D.

Thus, the distortion function for any source gives the lower bound on the source rate that is possible for a given level of distortion.

Definition 1.17 The **Distortion Rate Function** for a discrete time, memoryless Gaussian source is defined as

$$D_g(R) = 2^{-2R}\sigma_x^2. \tag{1.59}$$

Example 1.21 For a discrete time, memoryless Gaussian source, the distortion (in dB) as a function of its variance can be expressed as

$$10\log_{10} D_g(R) = -6R + 10\log_{10}\sigma_x^2. \tag{1.60}$$

Thus the mean square distortion decreases at a rate of 6 dB/bit.

Ans 4

Ans $P(Y/X) = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$

channel matrix

Joint probability Matrix is obtained by multiply the rows of $P(Y|X)$ by $P(x_0)$ & $P(x_1)$, respectively

$$P[X,Y] = P[Y|X] \times P[X] = \begin{bmatrix} 0.8 \times 0.6 & 0.2 \times 0.6 \\ 0.3 \times 0.4 & 0.7 \times 0.4 \end{bmatrix} = \begin{bmatrix} 0.48 & 0.12 \\ 0.12 & 0.28 \end{bmatrix}$$

$P(Y_0)$ & $P(Y_1)$ are obtained by summing the columns of $P[X,Y]$

$P[Y_0] = 0.48 + 0.12 = 0.6$

$P[Y_1] = 0.12 + 0.28 = 0.4$

the matrix $P[X|Y]$ is obtained by dividing the columns of $P[X,Y]$ by $P[Y_0], P[Y_1]$ respectively

$$P[X|Y] = \begin{bmatrix} 0.48/0.6 & 0.12/0.4 \\ 0.12/0.6 & 0.28/0.4 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

$$H(X) = \sum_{j=0}^{1} P(x_j) \log_2 \frac{1}{P(x_j)} = 0.6 \log_2 \frac{1}{0.6} + 0.4 \log_2 \frac{1}{0.4} = 0.971 \text{ bit/message}$$

$$H(X/Y) = \sum_{j=0}^{1} \sum_{k=0}^{1} P(x_j, y_k) \log_2 \left(\frac{1}{P(x_j|y_k)}\right) = 0.48 \log_2 \frac{1}{0.8} + 0.12 \log_2 \frac{1}{0.3} + 0.12 \log_2 \frac{1}{0.2} + 0.28 \log_2 \frac{1}{0.7}$$

$$= 0.786 \text{ bit/message}$$

$$I(X;Y) = H(X) - H(X/Y) = 0.971 - 0.786 = 0.185 \text{ b/message}$$

To find the channel capacity of binary channel, the auxiliary variables Q_1 & Q_2 are defined by

$$[P][Q] = [H] \quad \text{or} \quad \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} P_{11} \log_2 \frac{1}{P_{11}} + P_{12} \log_2 \frac{1}{P_{12}} \\ P_{21} \log_2 \frac{1}{P_{21}} + P_{22} \log_2 \frac{1}{P_{22}} \end{bmatrix}$$

$[P(Y|X)]$

The channel capacity is then given by $C = \log_2 (2^{-Q_1} + 2^{-Q_2})$

$$\begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} 0.8 \log_2 \frac{1}{0.8} + 0.2 \log_2 \frac{1}{0.2} \\ 0.3 \log_2 \frac{1}{0.3} + 0.7 \log_2 \frac{1}{0.7} \end{bmatrix}$$

$Q_1 = +0.6568$

$Q_2 = +0.9764$

Hence $C = \log(2^{-0.6568} + 2^{-0.9764}) = 0.2$ bit/message