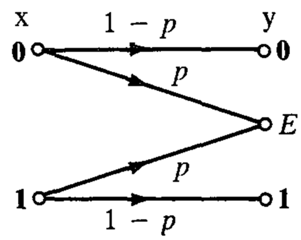
WEEK 10

Binary Erasure Channel

In data communication using error detection code, as soon as an error is detected, an automatic request for retransmission, enables retransmission of the data in error. In such a channel the data in error is erased. Hence, there is an erase probability ***p*,** but the probability of error is zero. Such a channel, known as a **binary erasure channel** (BEC), can be modeled as shown in Fig.



Note that for the BEC, the probability of "bit error" is zero. In other words, the following conditional probabilities hold for any BEC model:

$$P(Y = \text{"erasure"} \mid X = 0) = P$$

$$P(Y = \text{"erasure"} \mid X = 1) = P$$

$$P(Y = 0 \mid X = 0) = 1 - P$$

$$P(Y = 1 \mid X = 1) = 1 - P$$

$$P(Y = 0 \mid X = 1) = 0$$

$$P(Y = 1 \mid X = 0) = 0$$

Example :

The joint probability of a system is given by:

$$P(X,Y) = \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.125 \\ 0.0625 & 0.0625 \end{bmatrix}$$

Where, $p(x_j) = [0.75, 0.125, 0.125]$, and $p(y_k) = [0.5625, 0.4375]$

Find:

1- Marginal entropies. 2- Joint entropy

3- Conditional entropies. 4- The mutual information between x_1 and y_2.

5- The transinformation.

Sol:

1- $P(X) = \begin{bmatrix} x_1 & x_2 & x_3 \\ 0.75 & 0.125 & 0.125 \end{bmatrix}$ $P(Y) = \begin{bmatrix} y_1 & y_2 \\ 0.5625 & 0.4375 \end{bmatrix}$

$$H(X) = -[0.75\ln(0.75) + 2 \times 0.125\ln(0.125)]/ln2 = 1.06127\ bits/symbol$$

$$H(Y) = -[0.5625\ln(0.5625) + 0.4375\ln(0.4375)]/ln2$$
$$= 0.9887\ bits/symbol$$

2-

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

$$H(X,Y) = -\frac{[0.5\ln(0.5) + 0.25\ln(0.25) + 0.125\ln(0.125) + 2 \times 0.0625\ln(0.0625)]}{ln2}$$
$$= 1.875 \quad bits/symbol$$

3- $H(Y \mid X) = H(X,Y) - H(X) = 1.875 - 1.06127 = 0.813\ \frac{bits}{symbol}$

$$H(X \mid Y) = H(X,Y) - H(Y) = 1.875 - 0.9887 = 0.886\ bits/symbol$$

4- $I(x_1,y_2) = \log_2\left(\frac{P(x_1|y_2)}{P(x_1)}\right)$, $but\ P(x_1 \mid y_2) = P(x_1,y_2)/P(y_2)$

$$I(x_1,y_2) = \log_2\left(\frac{P(x_1,y_2)}{P(x_1)P(y_2)}\right) = \log_2\frac{0.25}{0.75\times0.4375} = -0.3923\ \ bits$$

-ve values for I means y_2 gives ambiguity about x_1

5- $I(X,Y) = H(X) - H(X \mid Y) = 1.06127 - 0.8863 = 0.17497\ bits/symbol.$

Channel Capacity per Second

The channel capacity gives the maximum possible information transmitted when one symbol (digit) is transmitted. If K symbols are being transmitted per second, then the maximum rate of transmission of information per second is K C_S. This is the channel capacity in information units per seconds and will be denoted by C (in bits per second):

$$C = K\,C_S \text{ bit/s}$$

The channel capacity C_s is the maximum value of $H(x)$ — $H(x|y)$; clearly, $C_s < H(x)$ [because H(x|y) > 0]. But $H(x)$ is the average information per input symbol. Hence, C_s is always less than (or equal to) the average information per input symbol. If we use binary symbols at the input, the maximum value of $H(x)$ is 1 bit, occurring when $P(x_1) = P(x_2) = 0.5$. Hence, for a binary channel, $C_s \le 1$ bit per binary digit. If we use r-ary symbols, the maximum value of $H_r(x)$ is 1 r-ary unit. Hence, $C_s \le 1$ r-ary unit per symbol.

CHANNEL CAPACITY OF A CONTINUOUS CHANNEL

For a discrete random variable x taking on values $x_1, x_2, \ldots x_n$ with probabilities $P(x_1), P(x_2), \ldots, P(x_n)$, the entropy $H(x)$ was defined as:

$$H(x) = \sum_{i=0}^{n} \mathrm{P(x_i)} \log \frac{1}{\mathrm{P(x_i)}}$$

For analog data, we have to deal with continuous random variables. Therefore, we must extend the definition of entropy to continuous random variables. For continuous random variables, Entropy is obtained by using the integral instead of discrete summation in above Eq. :

$$\mathrm{H(x)} = \int_{-\infty}^{\infty} \mathrm{p(x)} \log \frac{1}{\mathrm{p(x)}}\, \mathrm{dx}$$

But is it a meaningful interpretation as uncertainty?. A random variable x takes a value in the range $(n\Delta x, (n + l)\Delta x)$ with probability $\mathrm{p}(n\Delta x)\, \Delta x$ in the limit as $\Delta x \to 0$. The error in the approximation will vanish in the limit as $\Delta x \to 0$. Hence H(x), the entropy of a continuous random variable x, is given by:

$$\mathrm{H(x)} = \int_{-\infty}^{\infty} \mathrm{p(x)} \log \frac{1}{\mathrm{p(x)}}\, \mathrm{dx} - \lim_{\Delta x \to 0} \log \Delta x$$

The limit as $\Delta x \to 0$, log $\Delta x \to \infty$. It therefore appears that the entropy of a continuous random variable is infinite. This is quite true. The magnitude of uncertainty associated with a continuous random variable is infinite. This fact is also apparent intuitively. A continuous random variable assumes a nonenumerably infinite number of values, and, hence, the uncertainty is on the order of infinity. Does this mean that there is no meaningful definition of entropy for a continuous random variable? On the contrary, we shall see that the first term in the above equation serves as a meaningful measure of the entropy (average information) of a continuous random variable x. This may be argued as follows. We can consider $\int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx$ as a relative entropy with - log Δx serving as a datum, or reference. The information transmitted over a channel is actually the difference between the two terms $H(x)$ and H(x|y). Obviously, if we have a common datum for both $H(x)$ and H(x|y), the difference $H(x)$-$H(x|y)$ will be the same as the difference between their relative entropies. We are therefore justified in considering the first term in the above equation as the **differential entropy** of x. We must, however, always remember that this is a relative entropy and not the absolute entropy. Failure to realize this point generates many apparent misunderstandings !!

Example on differential Entropy

A signal amplitude x is a random variable uniformly distributed in the range (-1, 1). This signal is passed through an amplifier of gain 2. The output y is also a random variable, uniformly distributed in the range (-2, 2). Determine the (differential) entropies H(x) and H(y).

Sol:

$$P(x) = \begin{cases} \frac{1}{2} & |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$P(y) = \begin{cases} \frac{1}{4} & |y| < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$H(\mathrm{x}) = \int_{-1}^{1} \frac{1}{2} \log 2 \, dx = 1 \text{ bit}$$

$$H(\mathrm{y}) = \int_{-2}^{2} \frac{1}{4} \log 4 \, dx = 2 \text{ bits}$$

The entropy of the random variable y is twice that of x. This result may come as a surprise, since a knowledge of x uniquely determines y, and vice versa, because y = 2x. Hence, the average uncertainty of x and y should be identical.

Amplification itself can neither add nor subtract information. Why, then, is $H(y)$ twice as large as H(x)? This becomes clear when we remember that $H(x)$ and $H(y)$ are differential (relative) entropies, and they will be equal if and only if their datum (or reference) entropies are equal. The reference entropy $R1$ for x is - log Δx, and the reference entropy R_2 for y is - log Δy (in the limit as Δx, $\Delta y \to 0$) .

$$R_1 = \lim_{\Delta x \to 0} -\log \Delta x$$

$$R_2 = \lim_{\Delta y \to 0} -\log \Delta y$$

$$R_1 - R_2 = \lim_{\Delta x\, \Delta y \to 0} \log \frac{\Delta y}{\Delta x} = \log \frac{dy}{dx} = \log 2 = 1\ bit$$

Thus, R_1 the reference entropy of x, is higher than the reference entropy R_2 for y. Hence, if x and y have equal absolute entropies, their differential (relative) entropies must differ by 1 bit.

The information capacity of a continuous channel of bandwidth B hertz, perturbed by additive white Gaussian noise of power spectral density $N_0/2$ and limited in bandwidth to B, is given by:

$$C = B\, log_2(1 + \frac{P}{N_o B}) \qquad \text{bits per second}$$

where P is the average transmitted power.

The information capacity theorem is one of the most remarkable results of information theory for, in a single formula, it highlights most vividly the interplay among three key system parameters: channel bandwidth, average transmitted power (or, equivalently, average received signal power), and noise power spectral density at the channel output. The dependence of information capacity C on channel bandwidth B is *linear*, whereas its dependence on signal-to-noise ratio P/N_0B is *logarithmic*. Accordingly, *it is easier to increase the information capacity of a communication channel by expanding its bandwidth than increasing the transmitted power for a prescribed noise variance.*

For a White gaussian noise with mean square N = N_oB:

$$C = B\, log_2(1 + \frac{S}{N})$$

Example:

A voice-grade channel of the telephone network has a bandwidth of 3.4 kHz.

(a) Calculate the information capacity of the telephone channel for a signal-to-noise ratio of 30 dB.

(b) Calculate the minimum signal-to-noise ratio required to support information transmission through the telephone channel at the rate of 9,600 b/s.

Solution:

a:) Channel bandwidth B = 3.4 kHz

Received signal-to-noise ratio SNR = 30 dB = 10^3

Hence, the channel capacity is:

$C = B \log_2(1 + SNR) = 3.4 \times 10^3 \log_2(1 + 10^3) = 33.9 \times 10^3$ bits/second

b:) $9600 = 3.4 \times 10^3 \log_2 (1 + SNR)$, solving for the unknown SNR, we get

$SNR = 1.66 = 2.2$ dB