Engineering Tripos Part IIA THIRD YEAR

3F1: Signals and Systems

INFORMATION THEORY

Examples Paper Solutions

1. **Let the joint probability mass function of two binary random variables X and Y be given in the following table:**

x, y	$P_{XY}(x, y)$
0,0	**0.2**
0,1	**0.3**
1,0	**0.1**
1,1	**0.4**

Express the following quantities in terms of the binary entropy function

$$\mathbf{h}(\mathbf{x}) = -\mathbf{x}\log\mathbf{x} - (\mathbf{1}-\mathbf{x})\log(\mathbf{1}-\mathbf{x}).$$

(a) $\mathbf{H}(\mathbf{X})$

We marginalise $P_X(0) = P_{XY}(0,0) + P_{XY}(0,1) = 1/2$ and hence

$$H(X) = h(1/2) = 1 \text{ [bits]}.$$

(b) $\mathbf{H}(\mathbf{Y})$

Similarly, $P_Y(0) = P_{XY}(0,0) + P_{XY}(1,0) = .3$ and hence

$$H(Y) = h(.3).$$

(c) $\mathbf{H}(\mathbf{X}|\mathbf{Y})$

We compute

$$P_{X|Y}(0|0) = \frac{P_{XY}(0,0)}{P_Y(0)} = \frac{.2}{.3} = \frac{2}{3}$$

and

$$P_{X|Y}(0|1) = \frac{P_{XY}(0,1)}{P_Y(1)} = \frac{.3}{.7} = \frac{3}{7}$$

and hence

$$H(X|Y) = P_Y(0)H(X|Y=0) + P_Y(1)H(X|Y=1) = .3h(2/3) + .7h(3/7).$$

(d) **H(Y|X)**

We compute

$$P_{Y|X}(0|0) = \frac{P_{XY}(0,0)}{P_X(0)} = \frac{.2}{.5} = \frac{2}{5}$$

and

$$P_{Y|X}(0|1) = \frac{P_{XY}(1,0)}{P_X(1)} = \frac{.1}{.5} = \frac{1}{5}$$

and hence

$$H(Y|X) = P_X(0)H(Y|X=0) + P_X(1)H(Y|X=1) = .5h(2/5) + .5h(1/5).$$

(e) **H(XY)**

Using the chain rule, we can compute either

$$H(XY) = H(X) + H(Y|X) = 1 + .5h(2/5) + .5h(1/5) = 1.8464 \text{ [bits]}$$

or

$$H(XY) = H(Y) + H(X|Y) = h(.3) + .3h(2/3) + .7h(3/7) = 1..8464 \text{ [bits]}$$

which, as can be observed, yields the same result.

(f) **I(X;Y)**

We can write either

$$I(X;Y) = H(X) - H(X|Y) = 1 - .3h(2/3) - .7h(3/7) = 0.0349 \text{ [bits]}$$

or

$$I(X;Y) = H(Y) - H(Y|X) = h(.3) - .5h(2/5) - .5h(1/5) = 0.0349 \text{ [bits]}$$

which, as can again be observed, yields the same result.

2. **Let an N-ary random variable X be distributed as follows:**

$$\begin{cases} P_X(1) & = 1-p \\ P_X(k) & = \frac{p}{N-1} \text{ for } k = 2,3,\ldots,N. \end{cases}$$

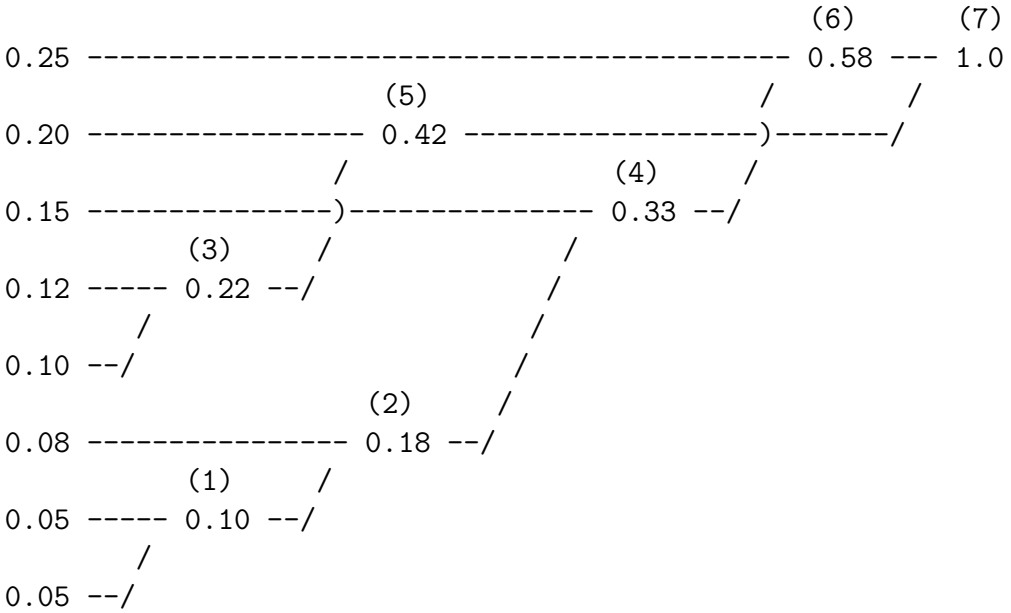
Express the entropy of X in terms of the binary entropy function $h(.)$.

$$\begin{aligned} H(X) &= -(1-p)\log(1-p) - (N-1)\frac{p}{N-1}\log\frac{p}{N-1} \\ &= -(1-p)\log(1-p) - p\log\frac{p}{N-1} \\ &= -(1-p)\log(1-p) - p\log p + p\log(N-1) \\ &= h(p) + p\log(N-1). \end{aligned}$$

3. **A discrete memoryless source has an alphabet of eight letters, $x_i, i = 1, 2, \cdots, 8$ with probabilities 0.25, 0.20, 0.15, 0.12, 0.10, 0.08, 0.05 and 0.05.**

   (a) **Use the Huffman encoding to determine a binary code for the source output.**

   The diagram below shows a the working to produce the Huffman code. The order in which the probabilities are merged is shown in brackets above the probabilities.



   Labelling the upper branch 0 and the lower branch 1 gives the code:

Symbol	Code
x_1	00
x_2	10
x_3	010
x_4	110
x_5	111
x_6	0110
x_7	01110
x_8	01111

   Note that other labellings are possible and also there is a choice at the second merge since there are two probabilities of 0.1 at that time.

   (b) **Determine the average codeword length L**

   Average codeword length:

   $L = 2 \,.\, 0.25 + 2 \,.\, 0.2 + 3 \,.\, 0.15 + 3 \,.\, 0.12 + 3 \,.\, 0.1 + 4 \,.\, 0.08 + 5 \,.\, 0.05 + 5 \,.\, 0.05$ $= 2.83$ bits.

(c) **Determine the entropy of the source and hence its efficiency**
$H(S) = -\frac{1}{\ln(2)}(0.25\ .\ \ln(0.25) + 0.2\ .\ \ln(0.2) + 0.15\ .\ \ln(0.15) + 0.12\ .\ \ln(0.12) + 0.1\ .\ \ln(0.1) + 0.05\ .\ \ln(0.05) + 0.05\ .\ \ln(0.05)) = 2.798$ bits.
The average codeword length is greater than the entropy (as expected).
Efficiency = 2.798 / 2.83 = 98.9%.

4. **Show that for statistically independent events**

$$H(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} H(X_i)$$

First expand out $H(X_1, X_2, \cdots, X_n)$

$$H(X_1, X_2, \cdots, X_n) = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_n=1}^{N_n} P(x_{i_1}, x_{i_2}, \cdots x_{i_n}) \log_2(P(x_{i_1}, x_{i_2}, \cdots x_{i_n}))$$

The probabilities are independent so this is

$$= \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_n=1}^{N_n} \left(\prod_{j=1}^{n} P(x_{i_j}) \right) \log_2 \left(\prod_{j=1}^{n} P(x_{i_j}) \right)$$

The last summation only indexes i_n so we can move all the other terms in the first product to the left of it. We can also replace the log of a product with the sum of the logs.

$$= \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_{n-1}=1}^{N_{n-1}} \left(\prod_{j=1}^{n-1} P(x_{i_j}) \right) \sum_{i_n=1}^{N_n} P(x_{i_n}) \left(\sum_{j=1}^{n} \log_2(P(x_{i_j})) \right)$$

All terms in the right hand sum (except the last one) do not depend on i_n and since $\sum_{i_n=1}^{N_n} P(x_{i_n}) = 1$ we can transform

$$\sum_{i_n=1}^{N_n} P(x_{i_n}) \left(\sum_{j=1}^{n} \log_2(P(x_{i_j})) \right) = \sum_{j=1}^{n-1} \log_2(P(x_{i_j})) + \sum_{i_n=1}^{N_n} P(x_{i_n}) \log_2(P x_{i_n})$$

Now the right hand term is independent of $i_1 \cdots i_{n-1}$ and

$$\sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_{n-1}=1}^{N_{n-1}} \left(\prod_{j=1}^{n-1} P(x_{i_j}) \right) = 1$$

so we have

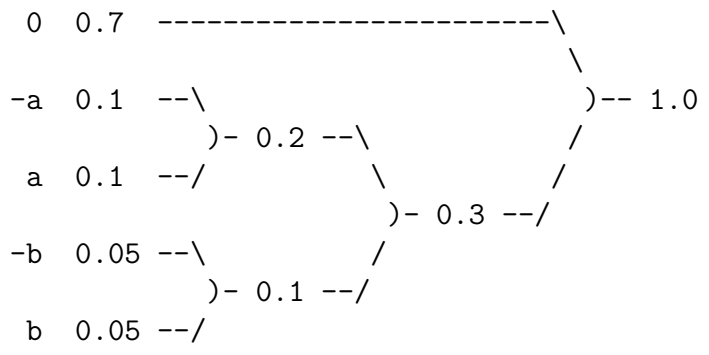
$$
\begin{aligned}
&H(X_1, X_2, \cdots, X_n) \\
&\quad = \sum_{i_n=1}^{N_n} P(x_{i_n}) \log_2(Px_{i_n}) + \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_{n-1}=1}^{N_{n-1}} \left(\prod_{j=1}^{n-1} P(x_{i_j}) \right) \sum_{j=1}^{n-1} \log_2(P(x_{i_j})) \\
&\quad = H(X_n) + H(X_1, X_2, \cdots, X_{n-1})
\end{aligned}
$$

by induction this gives

$$
= \sum_{i=1}^{n} H(X_i)
$$

5. **A five-level non-uniform quantizer for a zero-mean signal results in the 5 levels $-b, -a, 0, a, b$ with corresponding probabilities of occurrence $p_{-b} = p_b = 0.05$, $p_{-a} = p_a = 0.1$ and $p_0 = 0.7$.**

   (a) **Design a Huffman code that encodes one signal sample at a time and determine the average bit rate per sample.**



giving code:

Symbol	Code
0	0
$-a$	100
a	101
$-b$	110
b	111

   Average bit rate $= 1 \,.\, 0.7 + 3 \,.\, (0.1 + 0.1 + 0.05 + 0.05) = 1.6$ bits.

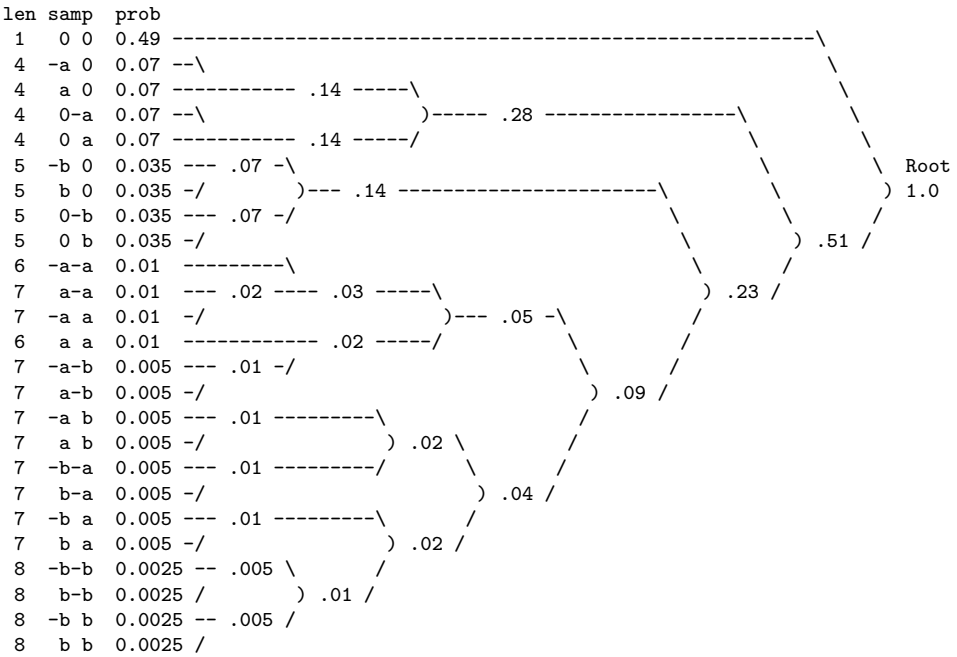
   Another version of the code would give 2 bits to $-a$ (or to a) and 4 bits each to $-b$ and b. It has the same mean length.

   (b) **Design a Huffman code that encodes two output samples at a time and determine the average bit rate per sample.**

   We combine the 5 states into 25 pairs of states, in descending order of probability and form a Huffman code tree as follows.

Note that there are many different possible versions of this tree as many of the nodes have equal probabilities and so can be taken in arbitrary order. However the mean length (and hence performance) of the code should be the same in all cases.

The code lengths are shown on the left, but should be filled in last, according to the number of branch-points in the tree that are passed, going from the root to each code state.



Summing probabilities for codewords of the same length:

Av. codeword length $= 1 \,.\, 0.49 + 4 \,.\, 0.28 + 5 \,.\, 0.14 + 6 \,.\, 0.02 + 7 \,.\, 0.06 + 8 \,.\, 0.01$ $= 2.93$ bits per pair of samples.

Hence code rate $= 2.93/2 = 1.465$ bit / sample

(c) **What are the efficiencies of these two codes?**

Entropy of the message source $= 0.7 \log_2(0.7) + 2 \,.\, 0.1 \log_2(0.1) + 2 \,.\, 0.05 \log_2(0.05)$ $= 1.4568$ bit / sample

Hence:

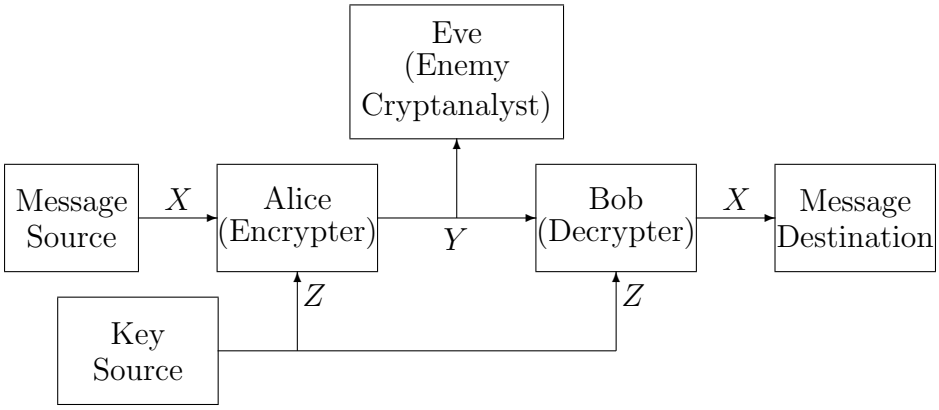
$$\text{Efficiency of code 1} = \frac{1.4568}{1.6} = 91.05\%$$

$$\text{Efficiency of code 2} = \frac{1.4568}{1.465} = 99.44\%$$

Code 2 represents a significant improvement, because it eliminates the 'zero' state of code 1 which has a probability well above 0.5 .

6. **While we cover in 3F1 and 4F5 the application of Shannon's theory to data compression and transmission, Shannon also applied the concepts**

of entropy and mutual information to the study of secrecy systems. The figure below shows a cryptographic scenario where Alice wants to transmit a secret plaintext message X to Bob and they share a secret key Z, while the enemy Eve has access to the public message Y.



(a) **Write out two conditions using conditional entropies involving X,Y and Z to enforce the deterministic encryptability and decryptability of the messages.**

Hint: **the entropy of a function given its argument is zero, e.g., for any random variable X, $\mathbf{H(f(X)|X) = 0}$.**

The conditions are

$$H(Y|XZ) = 0$$

because the ciphertext is a function of the secret plaintext message and the secret key, and

$$H(X|YZ) = 0$$

because the secret plaintext message can be inferred from the ciphertext and the key.

(b) **Shannon made the notion of an "unbreakable cryptosystem" precise by saying that a cryptosystem provides perfect secrecy if the enemy's observation is statistically independent of the plaintext, i.e., $I(X;Y) = 0$. Show that this implies Shannon's much cited bound on key size**

$$H(Z) \geq H(X),$$

i.e., perfect secrecy can only be attained if the entropy of the key (and hence its compressed length) is at least as large as the entropy of the secret plaintext.

Since $I(X;Y) = 0$, we have

$$\begin{aligned} H(X) = H(X|Y) &= H(X,Z|Y) - H(Z|X,Y) \\ &\leq H(X,Z|Y) \\ &= H(Z|Y) + H(X|Z,Y) \\ &= H(Z|Y) \\ &\leq H(Z) \end{aligned}$$

(c) **Vernam's cipher assumes a binary secret plaintext message X with any probability distribution $P_X(0) = p = 1 - P_X(1)$ and a binary secret key Z that's uniform $P_Z(0) = P_Z(1) = 1/2$ and independent of X. The encrypter simply adds the secret key to the plaintext modulo 2, and the decrypter by adding the same key to the ciphertext can recover the plaintext. Show that Vernam's cipher achieves perfect secrecy, i.e., $\mathbf{I(X;Y) = 0}$.**

We have

$$\begin{aligned} P_Y(0) &= \sum_x \sum_z P_{XYZ}(x,0,z) \\ &= \sum_x \sum_z P_{Y|XZ}(0|x,z) P_X(x) P_Z(z) \end{aligned}$$

and note that $P_{Y|XZ}(0|x,z) = 1$ if $x + z \mod 2 = 0$, and $P_{Y|XZ}(0|x,z) = 0$ otherwise, and so the expression above becomes

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

and hence

$$H(Y) = 1 \text{ [bits]}.$$

Furthermore,

$$\begin{aligned} P_{Y|X}(0|0) &= \frac{P_{XY}(0,0)}{P_X(0)} \\ &= \frac{P_{XYZ}(0,0,0) + P_{XYZ}(0,0,1)}{P_X(0)} \\ &= \frac{P_X(0)P_Z(0)P_{Y|XZ}(0|0,0) + P_X(0)P_Z(1)P_{Y|XZ}(0|0,1)}{P_X(0)} \\ &= \frac{p\frac{1}{2} + 0}{p} = \frac{1}{2} \end{aligned}$$

and similarly

$$\begin{aligned} P_{Y|X}(0|1) &= \frac{P_{XY}(1,0)}{P_X(1)} \\ &= \frac{P_{XYZ}(1,0,0) + P_{XYZ}(1,0,1)}{P_X 1} \\ &= \frac{P_X(1)P_Z(0)P_{Y|XZ}(0|1,0) + P_X(1)P_Z(1)P_{Y|XZ}(0|1,1)}{P_X(1)} \\ &= \frac{0 + (1-p)\frac{1}{2}}{1-p} = \frac{1}{2} \end{aligned}$$

and hence,

$H(Y|X) = P_X(0)H(Y|X=0)+P_X(1)H(Y|X=1) = ph(1/2)+(1-p)h(1/2) = 1$ [bits]

and so, finally,

$$I(X;Y) = H(Y) - H(Y|X) = 1 - 1 = 0 \text{ [bits]}.$$

7. **What is the entropy of the following continuous probability density functions?**

(a) $P(x) = \begin{cases} 0 & x < -2 \\ 0.25 & -2 < x < 2 \\ 0 & x > 2 \end{cases}$

$$\begin{aligned} H(X) &= -\int_{-2}^{2} 0.25 \log_2(0.25)\, dx \\ &= -\log_2(0.25) \\ &= \log_2(4) = 2 \end{aligned}$$

(b) $P(x) = \frac{\lambda}{2} e^{-\lambda|x|}$

$$\begin{aligned} H(X) &= -\frac{1}{\ln(2)} \int_{-\infty}^{\infty} \frac{\lambda}{2} e^{-\lambda|x|} \ln\left(\frac{\lambda}{2} e^{-\lambda|x|}\right) dx \\ &= -\frac{2}{\ln(2)} \int_{0}^{\infty} \frac{\lambda}{2} e^{-\lambda x} \ln\left(\frac{\lambda}{2} e^{-\lambda x}\right) dx \\ &= -\frac{2}{\ln(2)} \int_{0}^{\infty} \frac{\lambda}{2} e^{-\lambda x} \left(\ln\left(\frac{\lambda}{2}\right) - \lambda x\right) dx \\ &= -\frac{\lambda \ln(\lambda/2)}{\ln(2)} \int_{0}^{\infty} e^{-\lambda x} dx + \frac{\lambda^2}{\ln(2)} \int_{0}^{\infty} x e^{-\lambda x} dx \\ &= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{\lambda}{\ln(2)} \int_{0}^{\infty} \lambda x e^{-\lambda x} dx \end{aligned}$$

integrating by parts with $u = x$ and $v' = \lambda e^{-\lambda x}$ (so $v = -e^{-\lambda x}$):

$$\begin{aligned}
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{\lambda}{\ln(2)} \left(\left[-xe^{-\lambda x} \right]_0^\infty - \int_0^\infty -e^{-\lambda x} dx \right) \\
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{\lambda}{\ln(2)} \left(0 + \frac{1}{\lambda} \right) \\
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{1}{\ln(2)} \\
&= \frac{1 - \ln(\lambda/2)}{\ln(2)}
\end{aligned}$$

(c) $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$

$$\begin{aligned}
H(X) &= -\frac{1}{\ln(2)} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \right) dx \\
&= -\frac{1}{\ln(2)} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \left(\ln\left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{x^2}{2\sigma^2} \right) \\
&= \frac{\ln(\sigma\sqrt{2\pi})}{\ln(2)} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx + \frac{1}{2\sigma\sqrt{2\pi}\ln(2)} \int_{-\infty}^{\infty} \frac{x^2}{\sigma^2} e^{-x^2/2\sigma^2} dx \\
&= \frac{\ln(\sigma\sqrt{2\pi})}{\ln(2)} + \frac{1}{2\sigma\sqrt{2\pi}\ln(2)} \int_{-\infty}^{\infty} x \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx
\end{aligned}$$

integrating by parts with $u = x$ and $v' = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$ giving $v = -e^{-x^2/2\sigma^2}$:

$$\begin{aligned}
&= \frac{\ln(\sigma\sqrt{2\pi})}{\ln(2)} + \frac{1}{2\sigma\sqrt{2\pi}\ln(2)} \left(\left[-xe^{-x^2/2\sigma^2} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-x^2/2\sigma^2} dx \right) \\
&= \frac{\ln(\sigma\sqrt{2\pi})}{\ln(2)} + \frac{1}{2\sigma\sqrt{2\pi}\ln(2)} \left(0 + \sigma\sqrt{2\pi} \right) \\
&= \frac{\ln(\sigma\sqrt{2\pi})}{\ln(2)} + \frac{1}{2\ln(2)} \\
&= \log_2(\sigma\sqrt{2\pi}) + \frac{\log_2(e)}{2} = \log_2(\sigma\sqrt{2\pi e})
\end{aligned}$$

8. $*$ **Continuous variables X and Y are normally distributed with standard deviation $\sigma = 1$.**

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad\qquad P(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

A variable Z is defined by $z = x + y$. What is the mutual information of X and Z?

$$I(Z;X) = H(Z) - H(Z|X)$$

Z is a normally distributed variable with standard deviation $\sigma = \sqrt{2}$ (since it is the sum of two independent variables with standard deviation σ). Hence

$$H(Z) = \log_2(\sqrt{2}\ .\ \sqrt{2\pi}) + \frac{\log_2(e)}{2}$$

by 6(c) above.

If X is known then Z is still normally distributed, but now the mean $= x$ and the standard deviation is 1 since $z = x + y$ and Y has zero mean and standard deviation $\sigma = 1$. So:

$$H(Z|X) = \log_2(\sqrt{2\pi}) + \frac{\log_2(e)}{2}$$

So

$$\begin{aligned} I(Z;X) &= \log_2(\sqrt{2}\ .\ \sqrt{2\pi}) - \log_2(\sqrt{2\pi}) \\ &= \log_2(\frac{\sqrt{2}\ .\ \sqrt{2\pi}}{\sqrt{2\pi}}) \\ &= \log_2(\sqrt{2}) \\ &= 0.5 \text{ bit} \end{aligned}$$

9. **A symmetric binary communications channel operates with signalling levels of ± 2 volts at the detector in the receiver, and the rms noise level at the detector is 0.5 volts. The binary symbol rate is 100 kbit/s.**

   (a) **Determine the probability of error on this channel and hence, based on mutual information, calculate the theoretical capacity of this channel for error-free communication.**

   Using notation and results from sections 2.2 and 2.4 of the 3F1 Random Processes course and section 3.5 of the 3F1 Information Theory course:

$$\text{Probability of bit error, } \ p_e = Q\left(\frac{v_0}{\sigma}\right) = Q\left(\frac{2}{0.5}\right) = Q(4) = 3.17\ .\ 10^{-5}$$

   The mutual information between the channel input X and the channel output Y is then:

$$I(Y;X) = H(Y) - H(Y|X) = H(Y) - H([p_e \ \ (1-p_e)])$$

   Now, assuming equiprobable random bits at input and output of the channel:

$$H(Y) = -[0.5\log_2(0.5) + 0.5\log_2(0.5)] = 1 \text{ bit / sym}$$

and, from the above p_e:

$$\begin{aligned} H([p_e \;\; (1-p_e)]) &= -[3.17 \,.\, 10^{-5} \log_2(3.17 \,.\, 10^{-5}) \\ &\quad + \;\; (1 - 3.17 \,.\, 10^{-5}) \log_2(1 - 3.17 \,.\, 10^{-5})] \\ &= 4.739 \,.\, 10^{-4} + 4.57 \,.\, 10^{-5} \\ &= 5.196 \,.\, 10^{-4} \text{ bit / sym} \end{aligned}$$

So

$$I(Y;X) = 1 - 5.196 \,.\, 10^{-4} = 0.99948 \text{ bit / sym}$$

Hence the error-free capacity of the binary channel is

$$\text{(Symbol rate)} \times I(Y;X) = 100\text{k} \,.\, 0.99948 = 99.948 \text{ kbit/s}.$$

(b) **If the binary signalling were replaced by symbols drawn from a continuous process with a Gaussian (normal) pdf with zero mean and the same mean power at the detector, determine the theoretical capacity of this new channel, assuming the symbol rate remains at 100 ksym/s and the noise level is unchanged.**

For symbols with a Gaussian pdf, section 4.4 of the course shows that the mutual information of the channel is now given by:

$$I(Y;X) = h(Y) - h(N) = \log_2 \frac{\sigma_Y}{\sigma_N} = \tfrac{1}{2} \log_2 \left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)$$

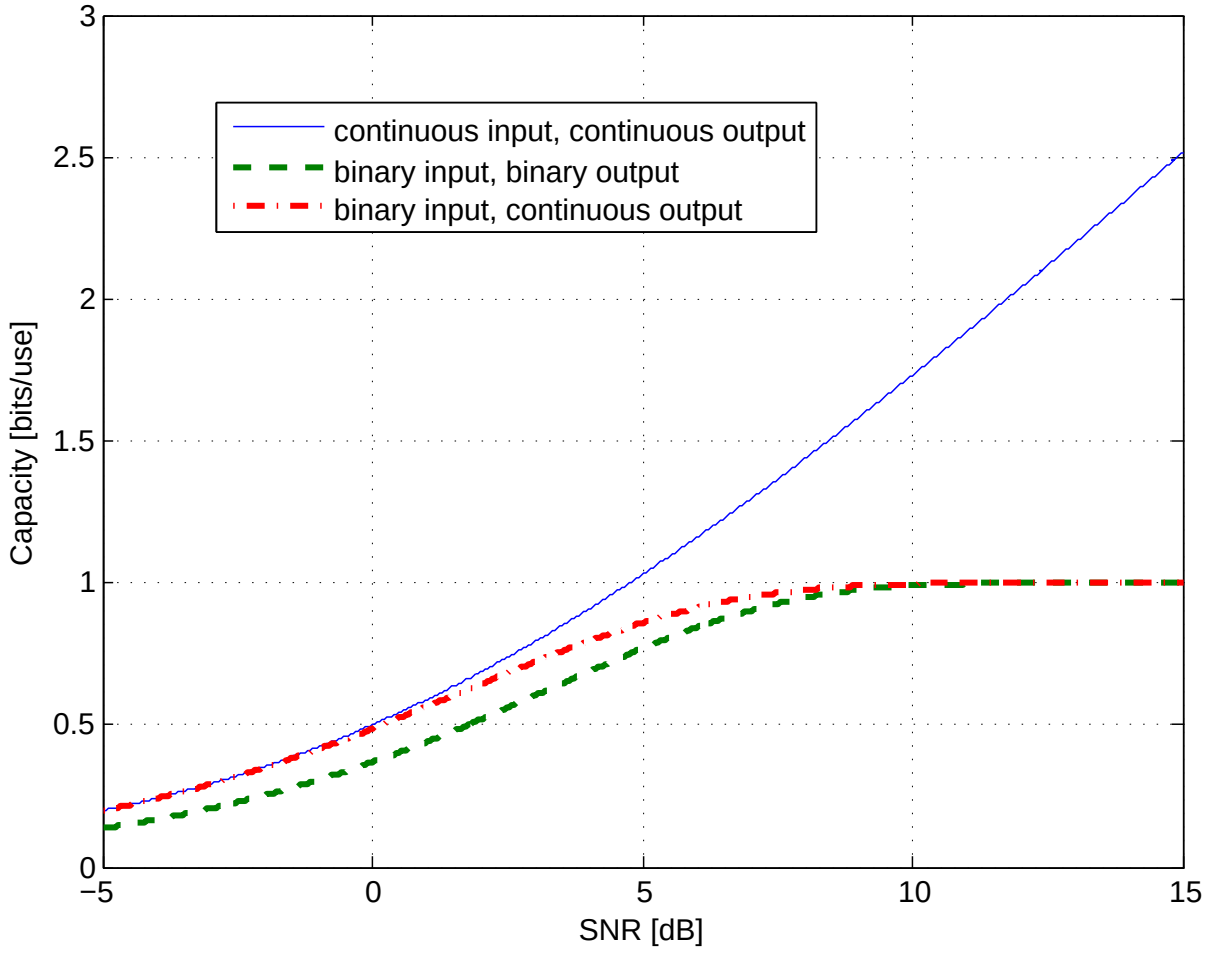
since the noise is uncorrelated with the signal, so $\sigma_Y^2 = \sigma_X^2 + \sigma_N^2$.
For the given channel, $\sigma_X / \sigma_N = 2/0.5 = 4$. Hence

$$I(Y;X) = \tfrac{1}{2} \log_2(1 + 4^2) = 2.0437 \text{ bit / sym}$$

and so the theoretical capacity of the Gaussian channel is

$$C_G = 100\text{k} \,.\, 2.0437 = 204.37 \text{ kbit/s} \quad \text{(over twice that of the binary channel)}$$

(c) The three capacities are plotted on the next page. There is no loss at low SNR for using binary signalling as long as the output remains continuous. As the SNR increases, all binary signalling methods hit a 1 bit/use ceiling whereas the capacity for continuous signalling continues to grow unbounded.


continuous input, continuous output
binary input, binary output
binary input, continuous output
Capacity [bits/use]
SNR [dB]
0
0.5
1
1.5
2
2.5
3
−5
0
5
10
15