Engineering Tripos Part IIA THIRD YEAR

3F1: Signals and Systems

INFORMATION THEORY

Examples Paper CRIB 1/1

Straightforward questions are marked †.
Tripos standard (but not necessarily Tripos length) questions are marked ∗.

1. **The output of a discrete memoryless source consists of the possible letters $X_1, X_2, \cdots, X_n$, which occur with probabilities $P_1, P_2, \cdots, P_n$, respectively. Prove that the entropy $H(X)$ of the source is at most $\log_2(n)$.**

   Want to show that $H(S) \le \log_2(N)$ (i.e. $H(S) - \log_2(N) \le 0$)

$$\begin{aligned} H(S) - \log_2(N) &= \sum_{i=1}^{N} p_i \left(-\log_2(p_i) - \log_2(N)\right) \\ &= \sum_{i=1}^{N} p_i \log_2(\frac{1}{Np_i}) \\ &= \sum_{i=1}^{N} p_i \log_2(e)\ln(\frac{1}{Np_i}) \\ &\le \log_2(e) \sum_{i=1}^{N} p_i \left(\frac{1}{Np_i} - 1\right) \qquad \text{since } \ln(x) \le (x-1) \\ &= \log_2(e) \left(\sum_{i=1}^{N} \frac{1}{N} - \sum_{i=1}^{N} p_i\right) \\ &= \log_2(e)(1-1) \\ &= 0 \end{aligned}$$

2. **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.

```
                                                  (6)      (7)
0.25 -------------------------------------------- 0.58 --- 1.0
                        (5)                     /        /
0.20 ----------------- 0.42 ------------------)-------/
                     /                (4)     /
0.15 ---------------)--------------- 0.33 --/
           (3)     /               /
0.12 ----- 0.22 --/               /
        /                        /
0.10 --/                        /
                    (2)        /
0.08 ---------------- 0.18 --/
           (1)     /
0.05 ----- 0.10 --/
        /
0.05 --/
```

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

Supervisors should 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 (number, R, of binary digits per source letter)**

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 compare it with R.**

$H(S) = -\frac{1}{\ln(e)}(0.25 * \ln(0.25) + 0.2 * \ln(0.2) + 0.15 * \ln(0.15) + 0.12 * \ln(0.12) + 0.1 * \ln(0.1) + 0.05 * \ln(0.05) + 0.05 * \ln(0.05)) = 2.798$ bits.

The average codeword length is greater than the entropy (as expected). Efficiency = 98.9%.

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

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

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

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

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

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

by induction this gives

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

4. **The optimum four level non-uniform quantizer for a Gaussian distributed signal amplitude results in the four levels a_1, a_2, a_3, a_4 with corresponding probabilities of occurrence $p_1 = p_2 = 0.3365$ and $p_3 = p_4 = 0.1635$.**

   (a) **Design a Huffman code that encodes a single level at a time and determine the average codeword length.**

```
p1 0.3365 --------------------------------- 1.0
                                          /
p2 0.3365 ------------------- 0.6635 --/
                              /
p3 0.1635 ------ 0.327 --/
             /
p4 0.1635 --/
```

giving code:

Symbol	Code
p_1	0
p_2	10
p_3	110
p_4	111

average bit rate = 1 * 0.3365 + 2 * 0.3365 + 3 * 0.327 = 1.9905

   (b) **Design a Huffman code that encodes two output levels at a time and determine the average codeword length.**

```
3 p1p1 0.1132 -----------------------------\
                                            \
3 p1p2 0.1132 ---- 0.2264 --------------------)-------\
                /                              \       \
3 p2p1 0.1132 -/                                \       \
                                                 \       \
3 p2p2 0.1132 --------------- 0.2232 -\           \       \
                           /           \           \       \
4 p1p3 0.055 ------ 0.11 -/             \           \       \
                /                        \           \       \
4 p1p4 0.055 --/                          \           \       \
                                           \           \       \
4 p2p3 0.055 ------ 0.11------- 0.22 -------- 0.4432 ---)-------)--------\
                /            /                          \       \         \
4 p2p4 0.055 --/            /                            \       \         \
                           /                              \       \         \
4 p3p1 0.055 ------ 0.11 -/                                \       \         \
                /                                           \       \         \
4 p3p2 0.055 --/                                             \       \         \
                                                              \       \         \
5 p4p1 0.055 ------ 0.11 ----------------------- 0.2168 ------- 0.33 -- 0.5564 -- 0.9998 (=1.0)
                /                              /
5 p4p2 0.055 --/                              /
                                             /
6 p3p3 0.0267 ----- 0.0534 -------- 0.1068 -/
                 /              /
6 p3p4 0.0267 --/              /
                              /
6 p4p3 0.0267 ----- 0.0534 --/
                 /
6 p4p4 0.0267 --/
```

Codeword lengths are shown on the left. The average codeword length = 0.1132 * 12 + 0.055 * 34 + 0.0267 * 24 = 3.8692 which is less than twice the average codeword length for coding one output at a time (2 * 1.9905 = 3.981).

5. **Given two random variables X and Y, $I(X;Y)$ is defined as:**

$$I(X;Y) = \sum_{x \in X, y \in Y} P(x,y) \log_2 \left(\frac{P(x|y)}{P(x)} \right)$$

Show that I(X;Y) = I(Y;X)

$$\begin{aligned}
\frac{P(x|y)}{P(x)} &= \frac{P(x|y)P(y)}{P(x)P(y)} \\
&= \frac{P(x,y)}{P(x)P(y)} \\
&= \frac{P(y|x)P(x)}{P(x)P(y)} \\
&= \frac{P(y|x)}{P(y)}
\end{aligned}$$

Hence

$$\begin{aligned}
\sum_{x \in X, y \in Y} P(x,y) \log_2 \left(\frac{P(x|y)}{P(x)} \right) &= \sum_{x \in X, y \in Y} P(x,y) \log_2 \left(\frac{P(y|x)}{P(y)} \right) \\
&= I(Y;X)
\end{aligned}$$

Students should also appreciate that this gives the same quantity as defined in lectures:

$$\begin{aligned}
I(X;Y) &= \sum_{x \in X, y \in Y} P(x,y) \log_2 \left(\frac{P(x|y)}{P(x)} \right) \\
&= \sum_{x \in X, y \in Y} P(x,y) \left(\log_2(P(x|y)) - \log_2(P(x)) \right) \\
&= -\sum_{x \in X} P(x) \log_2(P(x) + \sum_{x \in X, y \in Y} P(x,y) \log_2(P(x|y)) \\
&= H(X) - H(X|Y)
\end{aligned}$$

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

(a) $P(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{a} & 0 < x < a \\ 0 & x > a \end{cases}$

$$
\begin{aligned}
H(X) &= -\int_0^a \tfrac{1}{a}\log_2(\tfrac{1}{a})dx \\
&= -log_2(\tfrac{1}{a}) \\
&= log_2(a)
\end{aligned}
$$

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

$$
\begin{aligned}
H(X) &= -\frac{1}{\ln(2)}\int_{-\infty}^{\infty} \frac{\lambda}{2}e^{-\lambda|x|}\ln\left(\frac{\lambda}{2}e^{-\lambda|x|}\right)dx \\
&= -\frac{2}{\ln(2)}\int_0^{\infty} \frac{\lambda}{2}e^{-\lambda x}\ln\left(\frac{\lambda}{2}e^{-\lambda x}\right)dx \\
&= -\frac{2}{\ln(2)}\int_0^{\infty} \frac{\lambda}{2}e^{-\lambda x}\left(\ln\left(\frac{\lambda}{2}\right) - \lambda x\right)dx \\
&= -\frac{\lambda\ln(\lambda/2)}{\ln(2)}\int_0^{\infty} e^{-\lambda x}dx + \frac{\lambda^2}{\ln(2)}\int_0^{\infty} xe^{-\lambda x}dx \\
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{\lambda}{\ln(2)}\int_0^{\infty} \lambda xe^{-\lambda x}dx
\end{aligned}
$$

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

$$
\begin{aligned}
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{\lambda}{\ln(2)}\left(\left[-xe^{-\lambda x}\right]_0^{\infty} - \int_0^{\infty} -e^{-\lambda x}dx\right) \\
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{\lambda}{\ln(2)}\left(0 + \frac{1}{\lambda}\right) \\
&= -\frac{\ln(\lambda/2)}{\ln(2)} + \frac{1}{\ln(2)} \\
&= \log_2(2e/\lambda)
\end{aligned}
$$

(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 $v = -e^{-x^2/2\sigma^2}$:

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

7. $*$ **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

$$H(Z) = \log_2(\sqrt{4\pi}) + \frac{\log_2(e)}{2}$$

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

$$\begin{aligned} I(Z;X) &= \log_2(\sqrt{4\pi}) - \log_2(\sqrt{2\pi}) \\ &= \log_2(\frac{\sqrt{4\pi}}{\sqrt{2\pi}}) \\ &= \log_2(\sqrt{2}) \\ &= 0.5 \end{aligned}$$