Master - June 2008

2008 E4.40/SO20 Solutions

Key to letters on mark scheme: B=Bookwork, C=New computed example, A=New analysis

1. (a) If $|\mathcal{X}| = n$ then $q_i = n^{-1}\ \forall i$. From the Information Inequality (given in question)

$$0 \le D(\mathbf{p}\,\|\,\mathbf{q}) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) = -H(\mathbf{p}) + \sum_i p_i \log n \quad \Rightarrow \quad H(\mathbf{p}) \le \log n = H(\mathbf{q})$$ [3B]

(b) (i) From the question,

$$H(x) + H(y\,|\,x) + H(z\,|\,x,y) = H(x) + H(z\,|\,x) + H(y\,|\,x,z).$$

However because $z = x + y$, $H(z\,|\,x,y) = H(y\,|\,x,z) = 0$. Also, because x and y are independent, $H(y\,|\,x) = H(y)$. Hence

$$H(x) + H(y) = H(x) + H(z\,|\,x) \le H(x) + H(z)$$ [5A]

where the final inequality follows because conditioning reduces entropy. Subtracting $H(x)$ from both sides gives the required result.

(ii) If x is uniformly distributed and $y = 1 - x$, then $H(x) = H(y) = 1$ but $z \equiv 1$ so $H(z) = 0$. [3C]

(c) (i) $H(z) = \log 720 = 9.49$ bits. [1C]

(ii) The six wires R, …, W will form three pairs but you have no way of telling which pair is which. Thus there are 6 possible arrangements of the three pairs and two possible arrangements of the wires within each pair giving a total of $6 \times 2^3 = 48$ equally likely possibilities. Thus the entropy of z is $\log 48 = 5.58$ bits. [3C]

(iii) You now have three interconnection groups of sizes 1, 2 and 3. Because the sizes are all different, you can identify which is which and so your only uncertainty is the arrangement of wires within each group. Thus the number of possibilities is now $1! \times 2! \times 3! = 12$ which gives an entropy of $\log 12 = 3.58$ bits. [3C]

(iv) This final measurement uniquely identifies the wires since the members of any single group in the previous part are now in different groups. Thus the entropy is now 0. [2C]

2. (a) $\mathbf{p}_x = [0.65 \quad 0.35]$ [1C]

$$H(\mathcal{X}) = 0.65 \times H(0.05/0.65) + 0.35 \times H(0.05/0.35)$$
$$= 0.65 \times H(0.077) + 0.35 \times H(0.143)$$
$$= 0.65 \times 0.391 + 0.35 \times 0.592 = 0.461 \text{ bits}$$ [3C]

(b) Huffman Code for inputs [00, 11, 01, 10] with probs [0.6, 0.3, 0.05, 0.05] are [0, 10, 110, 111] giving an expected code length of 1.5 bits. [4C]

(c) The joint probability of (x_{i-1}, y_i) is

$$0.8\begin{pmatrix} .6 & .05 \\ .05 & .3 \end{pmatrix} + 0.2\begin{pmatrix} .05 & .6 \\ .3 & .05 \end{pmatrix} = \begin{pmatrix} .49 & .16 \\ .1 & .25 \end{pmatrix}$$ [3C]

and the joint probability of (y_{i-1}, y_i) is

$$0.8\begin{pmatrix} .49 & .16 \\ .1 & .25 \end{pmatrix} + 0.2\begin{pmatrix} .1 & .25 \\ .49 & .16 \end{pmatrix} = \begin{pmatrix} .412 & .178 \\ .178 & .232 \end{pmatrix}$$ [3C]

(d) We have

$$H(y_i \mid x_{i-1}) = 0.65 \times H(0.16/0.65) + 0.35 \times H(0.1/0.35)$$
$$= 0.65 \times H(0.246) + 0.35 \times H(0.286)$$
$$= 0.65 \times 0.8051 + 0.35 \times 0.8631 = 0.8254$$ [2C]

also, noting that $\mathbf{p}_y = [0.59 \quad 0.41]$,

$$H(y_i \mid y_{i-1}) = 0.59 \times H(0.178/0.59) + 0.41 \times H(0.178/0.41)$$
$$= 0.59 \times H(0.302) + 0.41 \times H(0.434)$$
$$= 0.59 \times 0.8834 + 0.41 \times 0.9875 = 0.926$$ [2C]

For a hidden markov process, we have $H(y_i \mid y_{i-1}, x_j) \le H(\mathcal{Y}) \le H(y_i \mid y_{i-1})$ for any $j \le i-1$ and, in particular for $j = i-1$. But since y_{i-1} depends only on x_{i-1}, we have $H(y_i \mid y_{i-1}, x_{i-1}) = H(y_i \mid x_{i-1})$. [2C]

3. (a) (i) This is the definition of mutual information. [1B]

(ii) We can decompose conditional entropy as a weighted sum of entropy conditional on specific values. [1B]

(iii) For both $y=0$ and $y=1$, z is a Bernoulli variable with probability vector $[1-g \quad g]$ and so its entropy is $H(g)$. [1B]

$H(z)=H(p_z)$ and, as given in the question preamble, $p_z = g+(1-2g)p_y$.

The value of $I(y;z)$ is maximized by making $g+(1-2g)p_y = 0.5$ which occurs when $p_y = ½$. The channel capacity is therefore $1-H(g)=1-0.469=0.531$ bits. [2B]

(b) In a similar way, we have

$$\begin{aligned} I(x;y) &= H(y)-H(y\,|\,x) \\ &= H(y)-H(y\,|\,x=0)(1-p_x)-H(y\,|\,x=1)p_x \\ &= H(y)-0-H(f)p_x \\ &= H\big(p_x(1-f)\big)-p_xH(f) \end{aligned}$$ [3A]

Setting the derivative with respect to p_x to zero gives

$$\begin{aligned} 0 = {dI}/{dp} &= (1-f)\log(p^{-1}(1-f)^{-1}-1)-H(f) \\ &= 0.875\log(1.143p^{-1}-1)-0.5436 \end{aligned}$$

$$\Rightarrow \quad 1.143p^{-1}-1 = 2^{0.5436/0.875} = 2^{0.6212} = 1.5382$$

$$\Rightarrow \quad p = 1.143/2.5382 = 0.4503$$ [2C]

$$\begin{aligned} \Rightarrow \quad I &= H(0.4503\times0.875)-0.4503\times0.5436 \\ &= H(0.394)-0.2447 = 0.9673-0.2447 = 0.7226 \text{ bits} \end{aligned}$$ [2C]

(c) The transition probabilities of the combined channel are

	$z=0$	$z=1$
$x=0$	$1-g=0.9$	$g=0.1$
$x=1$	$f+g-2fg=0.2$	1-f-g+2fg=0.8

Following the previous derivation, and noting that $p_z = 0.1 + 0.7 p_x$, the channel capacity is [2C]

$C = H(p_z) - p_x H(0.2) - (1-p_x)H(0.1)$ [1A]

Setting the derivative w.r.t p_x to zero gives

$$0 = {dC}/{dp} = 0.7\log\left(p_z^{-1} - 1\right) - 0.7219 + 0.469$$

$$\Rightarrow \quad p_z^{-1} - 1 = 2^{0.3613} = 1.2846 \quad \Rightarrow \quad p_z = 0.4337$$ [2C]

$$\Rightarrow \quad p = (0.4377 - 0.1)/0.7 = 0.4824$$

$$\Rightarrow \quad C = H(0.4377) - 0.4824 \times 0.7219 - 0.5176 \times 0.469$$

$$= 0.9888 - 0.3483 - 0.2427$$ [2C]

$$= 0.3978 \text{ bits}$$

(d) It would be possible to increase to the minimum capacity of the sub channels, namely 0.531 bits. [1A]

4. (a) (i) Definition of mutual information.

(ii) Definition of y.

(iii) Trnaslation invariance by a conditional constant.

(iv) z is independent of x.

(v) Gaussian bound for $h(y)$ in terms of its variance. Known $h(z)$ since Gaussian.

(vi) Algebra [6B]

We have equality in step (v) only if y is Gaussian. Since we know that z is Gaussian, this implies that x is also Gaussian.

(b) (i) Rearranging the definition of Q gives $h(z) = \frac{1}{2}\log 2\pi e Q$ and the result follows if this is inserted into the previous part after step (iv). [2A]

(ii) From the previous part, $2^{2h(z)} = 2\pi e Q$ and, if we choose x to be Gaussian, $2^{2h(x)} = 2\pi e P$. Hence the power inequality becomes

$$2^{2h(y)} \geq 2\pi e(P+Q) \quad \Rightarrow \quad h(y) \geq \tfrac{1}{2}\log\left(2\pi e(P+Q)\right)$$ [2A]

Hence, from part (a),

$$\begin{aligned} I(x;y) &= h(y) - h(z) \\ &\geq \tfrac{1}{2}\log 2\pi e(P+Q) - \tfrac{1}{2}\log 2\pi e Q = \tfrac{1}{2}\log\frac{P+Q}{Q} \end{aligned}$$ [2A]

Thus this is a lower bound on the capacity with Gaussian x so the capacity with arbitrary x must be at least this large. [2A]

(c) (i) For a uniform distribution $h(z) = \log 6 = 2.585$. This gives $Q = 36(2\pi e)^{-1} = 2.1078$. The variance of z is $N = 36/12 = 3$. Hence

$$1.8153 = \tfrac{1}{2}\log\frac{P+Q}{Q} \leq C \leq \tfrac{1}{2}\log\frac{P+N}{Q} = 1.8396$$ [3A]

(ii) $H(x) = \log 3 = 1.585$. In this case, we can achieve error-free decoding since the inputs are separated by twice the maximum noise amplitude. Hence $H(x \mid y) = 0$ and $I(x;y) = 1.585$ bits. [3A]

5. (a) (i) Total probability cannot exceed 1.

(ii) A finite sum cannot exceed the minimum summand multiplied by the number of summands.

(iii) The minimum value of $P(\mathbf{x},\mathbf{y})$ within the jointly typical set is given in its definition.

(iv) Algebra [4B]

(b) (i) $\max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x},\mathbf{z}) = \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x})P(\mathbf{z}) \le \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x}) \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{z}) \le 2^{-nH(X)+n\varepsilon}2^{-nH(Y)+n\varepsilon}$ [2B]

(ii) We can write

$$\sum_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x},\mathbf{z}) \le \left|J_\varepsilon^{(n)}\right| \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x},\mathbf{z}) \le 2^{nH(X,Y)+n\varepsilon}2^{-nH(X)-nH(Y)+2n\varepsilon}$$ [4B]

$$= 2^{nH(X,Y)-nH(X)-nH(Y)+3n\varepsilon}$$

(c) (i) $H(X) = -(7/11)\log(7/11) - (4/11)\log(4/11)$. Therefore $\mathbf{x}\in T_{\mathbf{x}}$ if and only if $P(\mathbf{x}) = 2^{-11H(X)} = (7/11)^7(4/11)^4$. This can only happen if exactly 4 of the x_i are equal to 1. The number of ways of choosing 4 out of 11 x_i is C_{11}^4 and so the result follows. [2A]

(ii) If, in addition, $\mathbf{x},\mathbf{y}\in J_0^{(11)}$, then we require that $y_i = 1$ for 2 out of the 7 i for which $x_i = 0$ and for 3 out of the 4 i for which $x_i = 0$. This gives the required expression. [2A]

(iii) $P\left(\mathbf{x},\mathbf{y}\in J_0^{(11)}\right) = P\left(\mathbf{x},\mathbf{y}\in J_0^{(11)} \mid \mathbf{x}\in T_{\mathbf{x}}\right)P\left(\mathbf{x}\in T_{\mathbf{x}}\right) = 0.1345\times 0.2438 = 0.0328$. [2A]

(iv) We now have

$$P(\mathbf{x},\mathbf{z}\in J_0^{(11)} \mid \mathbf{x}\in T_{\mathbf{x}}) = C_7^2(5/11)^2(6/11)^5 C_4^3(5/11)^3(6/11) = 0.0429$$

Which gives

$$P\left(\mathbf{x},\mathbf{z}\in J_0^{(11)}\right) = P\left(\mathbf{x},\mathbf{y}\in J_0^{(11)} \mid \mathbf{x}\in T_{\mathbf{x}}\right)P\left(\mathbf{x}\in T_{\mathbf{x}}\right) = 0.0429\times 0.2438 = 0.0105$$ [4A]

6. (a) (i) Definition of mutual information

(ii) Translation invariance, since $\hat{X}$ is conditionally constant.

(iii) Conditioning reduces entropy. Equality if the quantization error is independent of $\hat{X}$.

(iv) Gaussian bound for entropy with a given variance. Equality if $(X-\hat{X})$ is Gaussian.

(v) $\mathrm{Var}(X-\hat{X}) \le E\left((X-\hat{X})^2\right) \le D$ and $\log()$ is monotonic. Equality if $(X-\hat{X})$ is zero-mean and distortion is maximum allowed. [6B]

(b) (i) Since X and Z are independent and zero mean, we have $E(XZ)=0$ and so can ignore the cross terms. We have,

$$E\left((X-\hat{X})^2\right) = E\left((X-kX-Z)^2\right) = (1-k)^2\sigma^2 + kD$$
$$= (D\sigma^{-2})^2\sigma^2 + (1-D\sigma^{-2})D = D^2\sigma^{-2} + D - D^2\sigma^{-2} = D$$ [2C]

(ii) We have

$$\begin{aligned}\mathrm{Var}(\hat{X}) &= k^2\,\mathrm{Var}(X) + \mathrm{Var}(Z) = k^2\sigma^2 + kD \\ &= k(k\sigma^2 + D) = k\left((1-D\sigma^{-2})\sigma^2 + D\right) \\ &= k(\sigma^2 - D + D) = k\sigma^2 = \sigma^2 - D\end{aligned}$$ [2C]

(iii) We have

$$\begin{aligned} I(X;\hat{X}) &= h(\hat{X}) - h(\hat{X}\mid X) \\ &\overset{\text{(i)}}{\le} \tfrac{1}{2}\log 2\pi e(\sigma^2 - D) - h(\hat{X}-X\mid X) \\ &\overset{\text{(ii)}}{=} \tfrac{1}{2}\log 2\pi e(\sigma^2 - D) - h(Z) \\ &= \tfrac{1}{2}\log 2\pi e(k\sigma^2) - \tfrac{1}{2}\log 2\pi e(kD) \\ &= \tfrac{1}{2}\log 2\pi e\,\sigma^2 D^{-1} \\ \Rightarrow\quad R(D) &\overset{\text{(iii)}}{\le} \tfrac{1}{2}\log 2\pi e\,\sigma^2 D^{-1} \end{aligned}$$ [5A]

(i) because of the Gaussian bound and translation invariance. (ii) because Z is independent of X. (iii) because $R(D)$ is the minimum and so cannot exceed any specific example.

(c) We have $\sigma^2 = 1/12$ and $h(X) = \log(1) = 0$. Also $(X-\hat{X}) \in \left[-\tfrac{1}{4}, +\tfrac{1}{4}\right]$ and is uniformly distributed so $D = E\left((X-\hat{X})^2\right) = 1/48$. [2C]

The lower bound from part (a) is $0 - \tfrac{1}{2}\log 2\pi e/48 = -\tfrac{1}{2}\log 0.3558 = 0.7454$ bits.

The upper bound from part (b) is $\tfrac{1}{2}\log(48/12) = 2$ bits. [2C]

The actual bit rate is necessarily above the lower bound. It also happens to be below the upper bound although this need not necessarily be true for a block length of only 1. [1C]