**Information for students**

*Notation:*

(a) Random variables are shown in Tahoma font. $x$, $\mathbf{x}$, $\mathbf{X}$ denote a random scalar, vector and matrix respectively.

(b) The size of a set $A$ is denoted by $|A|$.

(c) By default, the logarithm is to the base 2.

(d) $\oplus$ denotes the exclusive-or operation, or modulo-2 addition.

(e) "i.i.d." means "independent identically distributed".

(f) $H(p)$ is the entropy function.

(g) $C(x) = \frac{1}{2}\log_2(1+x)$ is the capacity function for the Gaussian channel in bits/channel use.

# The Questions

1. Basics of information theory.

   a) $x$ and $y$ are correlated binary random variables with $p(x=y=0)=0$ and all other joint probabilities equal to 1/3. Calculate $H(x)$, $H(y)$, $H(x \mid y)$, $H(y \mid x)$, $H(x,y)$, $I(x\,;y)$.

   [6]

   b) Suppose $x_1$ and $x_2$ are i.i.d. Bernoulli random variables taking values of 0 and 1 with equal probabilities ($p$ = 0.5). Let $y_1 = x_2$, $y_2 = x_1$, and $y_3 = x_1 \oplus x_2$. Compute the following mutual information:

   i) $I(x_1\,;\,y_1)$

   ii) $I(x_2\,;\,y_2)$

   iii) $I(x_{1:2}\,;\,y_{1:2})$

   iv) $I(x_1\,;\,x_2 \mid y_3)$

   [8]

   c) Consider a Markov process with two states, 0 and 1, and transition matrix

   $$T = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

   i) Determine the stationary distribution.

   ii) Calculate the entropy rate, $H(\mathrm{X})$.

   iii) Find the values of $p$ and $q$ that maximize $H(\mathrm{X})$.

   [11]

2. Source coding.

a) Fano's inequality. Consider the Markov chain shown in Fig. 2.1, where $x$ and $y$ are discrete random variables, and $\hat{x}$ is the estimate of $x$.

$$x \longrightarrow y \longrightarrow \hat{x}$$

Fig. 2.1. Markov chain arising in Fano's inequality.

i) Define a random variable $e = (\hat{x} \neq x) \in \{0,1\}$. Justify each step of the following derivations.

$$H(e,x\,|\,y) \overset{(1)}{=} H(x\,|\,y) + H(e\,|\,x,y) \overset{(2)}{=} H(e\,|\,y) + H(x\,|\,e,y)$$
$$\overset{(3)}{\Rightarrow} H(x\,|\,y) + 0 \leq H(e) + H(x\,|\,e,y)$$
$$\overset{(4)}{=} H(e) + H(x\,|\,y, e=0)(1-p_e) + H(x\,|\,y, e=1)\,p_e$$
$$\overset{(5)}{\leq} H(p_e) + 0 \times (1-p_e) + \log(|\mathrm{X}|-1)\,p_e$$
$$\overset{(6)}{\Rightarrow} p_e \geq \frac{\left(H(x\,|\,y) - H(p_e)\right)}{\log\left(|\mathrm{X}|-1\right)} \overset{(7)}{\geq} \frac{\left(H(x\,|\,y) - 1\right)}{\log\left(|\mathrm{X}|-1\right)}$$

[8]

ii) Given the following joint distribution

| $x$ \ $y$ | a | b | c |
|---|---|---|---|
| 1 | **1/6** | **1/12** | **1/12** |
| 2 | **1/12** | **1/6** | **1/12** |
| 3 | **1/12** | **1/12** | **1/6** |

Find the minimum error probability corresponding to the optimum estimator and compare with Fano's inequality for this problem.

[7]

b) Upper bound on the rate-distortion function. For the case of a continuous random variable $X$ with mean zero and variance $\sigma^2$ and squared-error distortion, show that the Gaussian distribution has the largest rate-distortion function, i.e., the rate-distortion function for $X$ is bounded as follows:

$$R(D) \leq \frac{1}{2}\log\frac{\sigma^2}{D}.$$

Hint: use the following joint distribution of $X$ and $\hat{X}$ in Fig. 2.2.

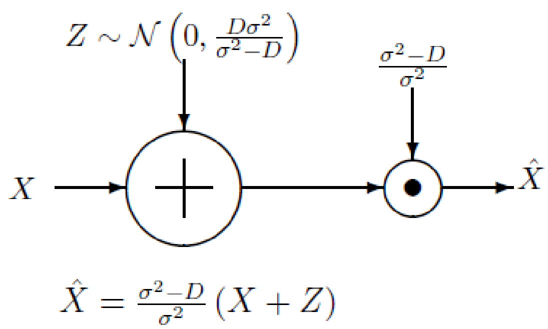


Fig. 2.2. Joint distribution of $X$ and $\hat{X}$. $X$ and $Z$ are independent.

[10]

3. Channel coding.

a) Consider a channel with input $x_{1:n}$ and output $y_{1:n}$.

i) Justify each step of the following proof. Firstly, the independence bound for the conditional entropy

$$H(x_{1:n} \mid y_{1:n}) \overset{(1)}{=} \sum_{i=1}^{n} H(x_i \mid x_{1:i-1}, y_{1:n}) \overset{(2)}{\le} \sum_{i=1}^{n} H(x_i \mid y_i)$$

If all $x_i$'s are independent, then we have the following bound

$$\begin{aligned} I(x_{1:n}; y_{1:n}) &\overset{(3)}{=} H(x_{1:n}) - H(x_{1:n} \mid y_{1:n}) \overset{(4)}{=} \sum_{i=1}^{n} H(x_i) - H(x_{1:n} \mid y_{1:n}) \\ &\overset{(5)}{\ge} \sum_{i=1}^{n} H(x_i) - \sum_{i=1}^{n} H(x_i \mid y_i) \overset{(6)}{=} \sum_{i=1}^{n} I(x_i; y_i). \end{aligned}$$

On the other hand, if the channel is memoryless, then

$$\begin{aligned} I(x_{1:n}; y_{1:n}) &\overset{(7)}{=} H(y_{1:n}) - H(y_{1:n} \mid x_{1:n}) \overset{(8)}{=} H(y_{1:n}) - \sum_{i=1}^{n} H(y_i \mid x_i) \\ &\overset{(9)}{=} \sum_{i=1}^{n} H(y_i \mid y_{1:i-1}) - \sum_{i=1}^{n} H(y_i \mid x_i) \overset{(10)}{\le} \sum_{i=1}^{n} H(y_i) - \sum_{i=1}^{n} H(y_i \mid x_i) \\ &= \sum_{i=1}^{n} I(x_i; y_i) \end{aligned}$$

[10]

ii) Then, show that a channel with memory has higher capacity than the corresponding memoryless channel.

[6]

b) Calculate the capacity of the following channels with probability transition matrix

i) $Q = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix} \quad x \in \{0,1\} \quad y \in \{0,1,2\}$

ii) $Q = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \quad x, y \in \{0,1,2\}$

iii) $Q = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix} \quad x, y \in \{0,1,2,3\}$

[9]

4. Network information theory.

a) Consider the inference channel in Fig. 4.1. There are two senders with equal power $P$, two receivers, with crosstalk coefficient $a$. The noise is Gaussian with zero mean and variance $N$. Show that the capacity under very strong interference (i.e., $a^2 \geq 1 + P/N$) is equal to the capacity under no interference at all.

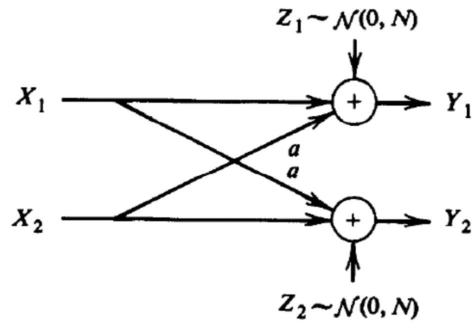


Fig. 4.1. Interference channel.

[10]

b) Slepian-Wolf coding. Two senders know random variables $U_1$ and $U_2$ respectively. Let the random variables $(U_1, U_2)$ have the following joint distribution:

| $U_1 \backslash U_2$ | $0$ | $1$ | $2$ | $\cdots$ | $m-1$ |
|---|---|---|---|---|---|
| $0$ | $\alpha$ | $\frac{\beta}{m-1}$ | $\frac{\beta}{m-1}$ | $\cdots$ | $\frac{\beta}{m-1}$ |
| $1$ | $\frac{\gamma}{m-1}$ | $0$ | $0$ | $\cdots$ | $0$ |
| $2$ | $\frac{\gamma}{m-1}$ | $0$ | $0$ | $\cdots$ | $0$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| $m-1$ | $\frac{\gamma}{m-1}$ | $0$ | $0$ | $\cdots$ | $0$ |

where $\alpha+\beta+\gamma = 1$. Find the region of rates $(R_1, R_2)$ that allow a common receiver to decode both random variables reliably.

[15]