EE4-40
EE9-SO20

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2017

MSc and EEE/EIE PART IV: MEng and ACGI

Corrected copy

INFORMATION THEORY

Friday, 19 May 10:00 am

Time allowed: 3:00 hours

There are FOUR questions on this paper.

Answer ALL questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : C. Ling

Second Marker(s) : D. Gunduz

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) $H(\cdot)$ 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) Suppose X_1 and X_2 are i.i.d. Bernoulli random variables taking values of 0 and 1 with equal probabilities, i.e., $P(X_i = 0) = 0.5$. Let $Y = \max(X_1, X_2)$. Compute the following entropy or mutual information terms:

      i) $H(Y)$

      ii) $I(X_1 ; Y)$

      iii) $I(X_1, X_2; Y)$

   [9]

   b) For $\mathbf{p} = [\frac{1}{2}, \frac{1}{4}, \frac{1}{4}]$ and $\mathbf{q} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$, compute the relative entropy terms $D(\mathbf{p}||\mathbf{q})$ and $D(\mathbf{q}||\mathbf{p})$.

   [6]

   c) Given two random variables X and Y, assume that X is uniformly distributed over the set $\mathcal{X} = \{1, ..., M\}$. Prove the following inequality that relates the mutual information $I(X ; Y)$ to the probability $P(X = Y)$:

   $$I(X ; Y) \geq P(X = Y) \log M - H(P(X = Y))$$

   Hint: use Fano's inequality $H(X|Y) \leq P(X \neq Y) \log M + H(P(X \neq Y))$.

   [10]

2. Typical sequences.

$\mathbf{x}$ and $\mathbf{y}$ are discrete-valued random variables of length n where each pair (X_i, Y_i) is drawn independently from the joint probability distribution function $p_{XY}(x,y)$. The jointly typical set $J_\varepsilon^{(n)}$ is the set of vector pairs satisfying the following conditions:

$$J_\varepsilon^{(n)} = \left\{\mathbf{x},\mathbf{y} : \left|-n^{-1}\log p_{\mathbf{x}}(\mathbf{x}) - H(X)\right| < \varepsilon, \right.$$
$$\left|-n^{-1}\log p_{\mathbf{y}}(\mathbf{y}) - H(Y)\right| < \varepsilon,$$
$$\left.\left|-n^{-1}\log p_{\mathbf{xy}}(\mathbf{x},\mathbf{y}) - H(X,Y)\right| < \varepsilon\right\}$$

where $p_X(x)$ and $p_Y(x)$ are the marginal probability distribution functions of X_i and Y_i, respectively. The probability $p_{\mathbf{x}}(\mathbf{x}) = \prod_{i=1}^{n} p_X(x_i)$, and similarly for $p_{\mathbf{y}}(\mathbf{x})$ and $p_{\mathbf{xy}}(\mathbf{x},\mathbf{y})$.

a) Justify each step in the following derivation of the size of the jointly typical set:

$$(1-\varepsilon)2^{n(H(X,Y)-\varepsilon)} \overset{n>N_\varepsilon}{<} \left|J_\varepsilon^{(n)}\right| \le 2^{n(H(X,Y)+\varepsilon)}.$$

$$1-\varepsilon \overset{(1)}{<} \sum_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(2)}{\le} \left|J_\varepsilon^{(n)}\right| \max_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(3)}{\le} \left|J_\varepsilon^{(n)}\right| 2^{-n(H(X,Y)-\varepsilon)} \overset{(4)}{\Rightarrow} \left|J_\varepsilon^{(n)}\right| \ge (1-\varepsilon)2^{n(H(X,Y)-\varepsilon)}$$

$$1 \overset{(5)}{\ge} \sum_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(6)}{\ge} \left|J_\varepsilon^{(n)}\right| \min_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(7)}{\ge} \left|J_\varepsilon^{(n)}\right| 2^{-n(H(X,Y)+\varepsilon)} \overset{(8)}{\Rightarrow} \left|J_\varepsilon^{(n)}\right| \le 2^{n(H(X,Y)+\varepsilon)}$$

[8]

b) Now suppose $n = 7$, $\varepsilon = 0$, and $p_{XY}(x,y)$ given by

$p_{XY}(x,y)$	$y = 0$	$y = 1$
$x = 0$	3/7	1/7
$x = 1$	1/7	2/7

Define the typical set $T_{\mathbf{x}} = \{\mathbf{x} : -n^{-1}\log p_{\mathbf{x}}(\mathbf{x}) = H(X)\}$ for $\varepsilon = 0$.

i) Calculate the probability $P(\mathbf{x} \in T_{\mathbf{x}})$.

ii) Calculate $P(\mathbf{x},\mathbf{y} \in J_0^{(7)} \mid \mathbf{x} \in T_{\mathbf{x}})$.

iii) Determine the value of $P(\mathbf{x},\mathbf{y} \in J_0^{(7)})$.

iv) If $\mathbf{z}$ is a random vector, independent of $\mathbf{x}$, whose elements are independent Bernoulli variables with $P(Z_i = 0) = 4/7$, calculate $P(\mathbf{x},\mathbf{z} \in J_0^{(7)})$.

[17]

3. Source and channel coding.

a) Justify each step in the following proof that feedback does not increase the capacity of a discrete memoryless channel shown in Fig. 3.1.

Fig. 3.1. Discrete memoryless channel with feedback.

$$
\begin{aligned}
I(W;\mathbf{y}) &\overset{(1)}{=} H(\mathbf{y}) - H(\mathbf{y}\,|\,W) \\
&\overset{(2)}{=} H(\mathbf{y}) - \sum_{i=1}^{n} H(y_i\,|\,y_{1:i-1}, W) \\
&\overset{(3)}{=} H(\mathbf{y}) - \sum_{i=1}^{n} H(y_i\,|\,y_{1:i-1}, W, x_i) \\
&\overset{(4)}{=} H(\mathbf{y}) - \sum_{i=1}^{n} H(y_i\,|\,x_i) \\
&\overset{(5)}{\le} \sum_{i=1}^{n} H(y_i) - \sum_{i=1}^{n} H(y_i\,|\,x_i) \overset{(6)}{=} \sum_{i=1}^{n} I(x_i;y_i) \overset{(7)}{\le} nC
\end{aligned}
$$

Hence

$$
nR = H(W) = H(W\,|\,\mathbf{y}) + I(W;\mathbf{y}) \overset{(8)}{\le} 1 + nRP_e^{(n)} + nC
$$

$$
\overset{(9)}{\Rightarrow} P_e^{(n)} \ge \frac{R - C - n^{-1}}{R} \overset{(10)}{\Rightarrow} \text{Any rate } > C \text{ is unachievable}
$$

[10]

b) Reverse water filling. Consider lossy source coding of X_1, X_2, X_3, which are independent zero-mean Gaussian information sources with different variances $\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_3^2 = 4$. Find the values of the rate-distortion function $R(D)$ for the following cases (D is the average distortion):

i) $D = 0.5$.

ii) $D = 1$.

iii) $D = 2$.

[15]

4. Network information theory.

a) Broadcast channel. The Gaussian broadcast channel illustrated in Fig. 4.1 is a degraded broadcast channel, where Z_1, Z_2' are independent Gaussian noise components with zero means.

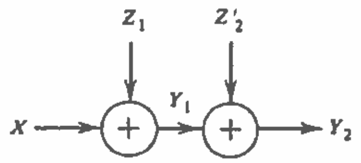

Fig. 4.1. The Gaussian broadcast channel.

i) Show that $X \to Y_1 \to Y_2$ forms a Markov chain.

ii) Describe an encoding and decoding strategy achieving the capacity region ($0<\alpha<1$)

$$R_1 \le C\left(\frac{\alpha P}{N_1}\right)$$

$$R_2 \le C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right)$$

[15]

b) Multiple-access channel. Consider a binary erasure multiple access channel. This multi-access channel has binary inputs $X_1, X_2 \in \{0,1\}$, and a ternary output $Y = X_1 + X_2 \in \{0,1,2\}$. There is no ambiguity in (X_1, X_2) if $Y = 0$ or $Y = 2$ is received; but $Y = 1$ can result from either (0, 1) or (1, 0). Find and sketch its capacity region.

[10]