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) Let $\mathbf{p} = (p_1, p_2, p_3)$ be a probability distribution on three elements. Define a new distribution $\mathbf{q}$ on two elements as $q_1 = p_1$, $q_2 = p_2 + p_3$. Show that

$$H(\mathbf{p}) = H(\mathbf{q}) + q_2 H\left(\frac{p_2}{q_2}, \frac{p_3}{q_2}\right)$$

[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 $Y = \min(X_1, X_2)$. Compute the following entropy or mutual information:

i) $H(Y)$

ii) $I(X_1 ; Y)$

iii) $I(X_{1:2} ; Y)$

[9]

c) A fair coin is flipped until the first head occurs. Let X denote the number of flips required. Find the entropy $H(X)$ in bits. The following equalities may be useful.

$$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r} \qquad \sum_{n=1}^{\infty} n r^n = \frac{r}{(1-r)^2} \qquad |r| < 1 .$$

[10]

2. Source coding.

a) Typical set.

i) Given the joint probability distribution function $p(x, y)$ defined as below

$x \backslash y$	0	1
0	1/8	1/8
1	1/8	5/8

Let $\varepsilon = 0.2$. Are the sequences $\mathbf{x} = 11100111$ and $\mathbf{y} = 01111110$ individually typical with respect to ε? Are they jointly typical with respect to ε? Your answers need to be justified.

[10]

ii) Justify each step in the following proof of the fact that the typical set $T_\varepsilon^{(n)}$ cannot be smaller. $\overline{T_\varepsilon^{(n)}}$ denotes the complement of $T_\varepsilon^{(n)}$.

For any $0 < \varepsilon < 1$, choose N_ε such that typicality holds, and choose $N_0 = -\varepsilon^{-1}\log\varepsilon$. Then for any $n > \max(N_0, N_\varepsilon)$ and any subset $S^{(n)}$ satisfying $\left|S^{(n)}\right| < 2^{n(H(x)-2\varepsilon)}$, we have

$$
\begin{aligned}
p\left(\mathbf{x} \in S^{(n)}\right) &\overset{(1)}{=} p\left(\mathbf{x} \in S^{(n)} \cap T_\varepsilon^{(n)}\right) + p\left(\mathbf{x} \in S^{(n)} \cap \overline{T_\varepsilon^{(n)}}\right) \\
&\overset{(2)}{<} \left|S^{(n)}\right| \max_{\mathbf{x} \in T_\varepsilon^{(n)}} p(\mathbf{x}) + p\left(\mathbf{x} \in \overline{T_\varepsilon^{(n)}}\right) \\
&\overset{(3)}{<} 2^{n(H(x)-2\varepsilon)} 2^{-n(H(x)-\varepsilon)} + \varepsilon \qquad \text{for } n > N_\varepsilon \\
&\overset{(4)}{=} 2^{-n\varepsilon} + \varepsilon \overset{(5)}{<} 2\varepsilon \qquad \text{for } n > N_0
\end{aligned}
$$

[6]

b) Parallel Gaussian sources and reverse waterfilling.

Consider three Gaussian random variables X_1, X_2, X_3 with variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$, respectively. Assume that $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > 0$. The average distortion is given by $D = (D_1 + D_2 + D_3)/3$. At what average distortion does the lossy source encoder behave like an encoder for

i) a single source with noise variance σ_1^2?

ii) a pair of sources with noise variances σ_1^2 and σ_2^2?

iii) three sources with noise variances σ_1^2, σ_2^2 and σ_3^2?

iv) Find the rates for cases i), ii), and iii).

[9]

3. Channel coding.

a) Justify each step in the following proof of the coding theorem for discrete memoryless channels.

Choose large enough block length n such that joint typicality holds; choose $\mathbf{p}_{\mathcal{X}}$ so that $I(\mathcal{X};\mathcal{Y})$ equals the capacity; from this distribution a random code of rate R is generated. The decoding error probability is given by

$$p(E) \overset{(1)}{=} \sum_C p(C) 2^{-nR} \sum_{w=1}^{2^{nR}} \lambda_w(C) \overset{(2)}{=} 2^{-nR} \sum_{w=1}^{2^{nR}} \sum_C p(C)\lambda_w(C)$$
$$\overset{(3)}{=} \sum_C p(C)\lambda_1(C) \overset{(4)}{=} p(E \mid w=1)$$

Let e_w denote the event that received vector $\boldsymbol{y}$ is jointly typical with codeword $\boldsymbol{x}(w)$. The decoder uses joint typicality decoding, so

$$p(E) = p(E \mid W=1) \overset{(5)}{=} p\left(\overline{e_1} \cup e_2 \cup e_3 \cup \cdots \cup e_{2^{nR}}\right) \overset{(6)}{\le} p\left(\overline{e_1}\right) + \sum_{w=2}^{2^{nR}} p(e_w)$$
$$\overset{(7)}{\le} \varepsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(\mathcal{X};\mathcal{Y})-3\varepsilon)} \overset{(8)}{<} \varepsilon + 2^{nR} 2^{-n(I(\mathcal{X};\mathcal{Y})-3\varepsilon)}$$
$$\overset{(9)}{\le} \varepsilon + 2^{-n(C-R-3\varepsilon)} \overset{(10)}{\le} 2\varepsilon \quad \text{for} \quad R < C - 3\varepsilon \text{ and } n > -\frac{\log \varepsilon}{C - R - 3\varepsilon}$$

Since average of $P(E)$ over all codes is $\le 2\varepsilon$, there must be at least one code for which

$$2^{-nR} \sum_w \lambda_w \overset{(11)}{\le} 2\varepsilon$$

Now throw away the worst half of the codewords; the remaining ones must all have

$$\lambda_w \overset{(12)}{\le} 4\varepsilon.$$

The resultant code has rate

$$\overset{(13)}{=} R - n^{-1} \cong R.$$

[13]

b) Consider the Gaussian channel shown in the following figure, where the transmitted signal x of power P is received by two antennas:

$$y_1 = x + z_1$$
$$y_2 = x + z_2$$

where z_1 and z_2 are independent Gaussian noises of power N_1 and N_2, respectively ($N_1 < N_2$). Moreover, the signals at the two antennas are combined as $y = \alpha y_1 + (1 - \alpha) y_2$ before decoding ($0 \leq \alpha \leq 1$).

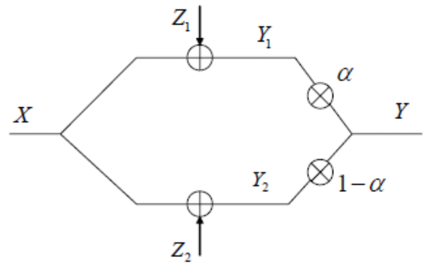

Fig. 3.1. Signal received at two antennas.

i) Find the capacity of the channel for a given α.

[6]

ii) Find the optimal α that maximizes the capacity and write down the corresponding maximum capacity.

[6]

4. Network information theory.

a) Slepian-Wolf coding.
Let X be i.i.d. Bernoulli(p), p = 0.5. Let Z be i.i.d. Bernoulli(r), r = 0.1, and let Z be independent of X. Finally, let $Y = X \oplus Z$ (mod 2 addition). Let X be encoded at rate R_1 and Y be encoded at rate R_2. What region of rates allows recovery of X and Y with probability of error tending to zero? Sketch this Slepian-Wolf rate region.

[9]

b) Consider the following degraded broadcast channel, where Y_1 and Y_2 are two receivers, and E denotes Erasure.

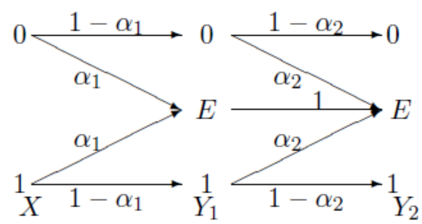

Fig. 4.1. Degraded broadcast channel, where E denotes Erasure.

i) What is the capacity of the channel from X to Y_1?

[2]

ii) What is the capacity of the channel from X to Y_2?

[4]

iii) What is the capacity region of all (R_1, R_2) achievable rate pairs for this broadcast channel? Sketch the capacity region.
Hint: the capacity region of a degraded broadcast channel is given by

$$R_1 = I(X; Y_1|U)$$
$$R_2 = I(U; Y_2)$$

For this problem, the auxiliary random variable U is binary and uniformly distributed on {0, 1}. It is connected to X by another binary symmetric channel of parameter β.

[10]