Problem set - I

Information Theory for Wireless Communications (Part-I, Spring 2012)

1) For given constants $\delta \geq \sqrt{MM'}$ and $K_0 \geq 2$, let

$$\mathcal{B}^{\delta,K_0} \triangleq \Big\{(x^n, y^n) \;\;|\;\; y^n \text{ is } \delta\text{-generated by } x^n \text{ and } x^n \text{ is } K_0\text{-input typical } \Big\} \tag{1}$$

a) For the system in Fig. 1, show that

$$\Pr\Big((\boldsymbol{X}^n, \boldsymbol{Y}^n) \in \mathcal{B}^{\delta,K_0}\Big) > (1 - \frac{MM'}{\delta^2})(1 - \frac{1}{K_0}) \tag{2}$$

b) Show that for any $(x^n, y^n) \in \mathcal{B}^{\delta,K_0}$

$$2^{-n\left(H(\boldsymbol{X},\boldsymbol{Y}) + \frac{K_1}{\sqrt{n}}\right)} \;<\; \Pr\Big(\boldsymbol{X}^n = x^n \;,\; \boldsymbol{Y}^n = y^n\Big) \;<\; 2^{-n\left(H(\boldsymbol{X},\boldsymbol{Y}) - \frac{K_1}{\sqrt{n}}\right)} \tag{3}$$

where $K_1 > 0$ is a constant which depends only on δ, M, M', K_0, and is independent of $p(i), p(j \mid i)$. $H(\boldsymbol{X}, \boldsymbol{Y}) \triangleq - \sum_{i=1}^{M} \sum_{j=1}^{M'} p(i,j) \log_2(p(i,j))$ is the joint entropy of the r.v's $\boldsymbol{X}$ and $\boldsymbol{Y}$.

c) Using the results in a) and b), show that

$$2^{n\left(H(\boldsymbol{X},\boldsymbol{Y}) - \frac{K_2}{\sqrt{n}}\right)} \;<\; \Big|\mathcal{B}^{\delta,K_0}\Big| \;<\; 2^{n\left(H(\boldsymbol{X},\boldsymbol{Y}) + \frac{K_2}{\sqrt{n}}\right)} \tag{4}$$

where $K_2 > 0$ is a constant which depends only on δ, M, M', K_0, and is independent of $p(i), p(j \mid i)$.

2) For a given K_0-input typical sequence x^n, consider the set

$$\mathcal{B}_2^{\delta,K_0}(x^n) \triangleq \Big\{y^n \;\;|\;\; y^n \text{ is } \delta\text{-generated by } x^n\Big\}. \tag{5}$$

For the random variable $\boldsymbol{Y}^n$ in Fig. 1, show that

$$2^{-n\left(I(\boldsymbol{X};\boldsymbol{Y}) + \frac{K_{10}}{\sqrt{n}}\right)} \;<\; \Pr\Big(\boldsymbol{Y}^n \;\in\; \mathcal{B}_2^{\delta,K_0}(x^n)\Big) \;<\; 2^{-n\left(I(\boldsymbol{X};\boldsymbol{Y}) - \frac{K_{10}}{\sqrt{n}}\right)} \tag{6}$$

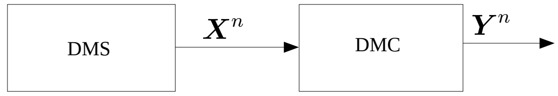

Fig. 1.

where $K_{10} > 0$ is a constant which depends only on δ, M, M', K_0, and is independent of $p(i), p(j \mid i)$. $I(\boldsymbol{X};\boldsymbol{Y}) \triangleq \sum_{i=1}^{M}\sum_{j=1}^{M'} p(i,j)\log_2\left(\frac{p(i,j)}{p(i)p(j)}\right)$ is the mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

3) Entropy of a finite memory stationary source.

A stationary source (random process) $\boldsymbol{X}$ with alphabet χ is said to have order $m \geq 0$ memory, if the distribution of the n-th source output X_n depends only on the previous m source outputs $X_{n-1}, \cdots, X_{n-m}$. Hence, for $n \geq m$

$$p(X_1,\cdots,X_n) = p(X_m,\cdots,X_1)\prod_{i=m+1}^{n} p(X_i \mid X_{i-1},\cdots,X_{i-m}). \tag{7}$$

a) Show that $-\frac{\log_2 p(X_1,X_2,\cdots,X_n)}{n}$ converges to $\mathcal{H}(\boldsymbol{X}) \triangleq H(X_i \mid X_{i-1},\cdots,X_{i-m})$ in probability, as $n \to \infty$.

b) For a given constant $\epsilon > 0$, define the ϵ-entropy typical set $\mathcal{A}_\epsilon^n$ for the source output $\boldsymbol{X}^n$ as

$$\mathcal{A}_\epsilon^n = \left\{x^n \in \chi^n \;\middle|\; \left|-\frac{\log_2 p(X_1 = x_1, X_2 = x_2,\cdots,X_n = x_n)}{n} - \mathcal{H}(\boldsymbol{X})\right| < \epsilon\right\} \tag{8}$$

and show that for sufficiently large n, $\boldsymbol{X}^n$ belongs to this set with probability greater than $1-\epsilon$.

c) Show that $|\mathcal{A}_\epsilon^n| < 2^{n\left(\mathcal{H}(\boldsymbol{X})+\epsilon\right)}$ and $|\mathcal{A}_\epsilon^n| > 2^{n\left(\mathcal{H}(\boldsymbol{X})-\epsilon+\frac{\log_2(1-\epsilon)}{n}\right)}$ for sufficiently large n. $\mathcal{H}(\boldsymbol{X})$ is also referred to as the entropy-rate of the source.

4) Entropy maximization for a discrete memoryless source (DMS) with mean-square constraint.

For a M-ary DMS $\boldsymbol{X}$, we know that the entropy $H(\boldsymbol{X})$ is maximized when the input distribution is uniform i.e., $p(i) = 1/M, i = 1,2,\ldots,M$. In this problem we are interested in maximizing the source entropy subject to a mean-square constraint, i.e.,

$$(p_1^\star, p_2^\star,\cdots,p_M^\star) = \arg\max_{(p_1,\cdots,p_M)\ \mid\ \sum_i p_i = 1,\, \sum_i p_i a_i^2 \leq P} -\sum_{i=1}^{M} p_i \log_2 p_i \tag{9}$$

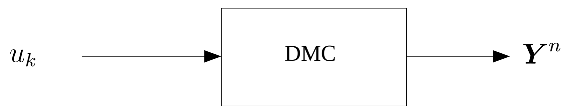

Fig. 2.

where $P > 0$, and $\chi = \{a_1, a_2, \cdots, a_M\}$, $a_i \in \mathbb{R}$.

a) Frame the problem in (9) as a convex optimization problem and list the K.K.T. conditions.

b) Based on the K.K.T. conditions show that the uniform probability distribution is the entropy maximizer if $P \geq \sum_i a_i^2/M$ (in which case the entropy is $\log_2 M$).

c) How will the optimal probability distribution change when $P < \sum_i a_i^2/M$? Will the optimal distribution still be uniform, or more probability will be assigned to those a_i which are smaller in magnitude ?

d) To understand the optimal probability distribution when $\sum_i a_i^2/M > P$, let us consider the special case of a PAM-type alphabet, where $M = 2KN + 1$ ($K, N \in \mathbb{Z}$) and $a_i = -K + ((i-1)/N)$, $i = 1, 2, \ldots, 2KN + 1$.

Fix $K = 5$ and $P = 5$. Numerically find the optimal input distribution for $N = 1, 2$ and comment on its shape. Also compare the shape of the optimal input distribution with that of a Gaussian random variable with mean 0 and variance $P = 5$.

5) In this problem, we consider a scenario where the transmitter has imperfect knowledge of the channel probability function ($p(j \mid i)$). We will denote the transmitter's knowledge of the c.p.f. by $p'(j \mid i)$.

a) Suppose that the transmitter uses a (n, N, λ) code (designed using the code construction that we discussed in Lecture 3 to prove the channel coding theorem). For any codeword $u_k, k = 1, 2, \ldots, N$ we know that the corresponding decoding set $\mathcal{A}_i \subset \mathcal{B}_2^{\delta, K_0}(u_k)$. Note that, due to imperfect knowledge of the c.p.f. at the transmitter, the δ-generated set $\mathcal{B}_2^{\delta, K_0}(u_k)$ is given by

$$\mathcal{B}_2^{\delta, K_0}(u_k) \triangleq \Big\{ y^n \;\mid\; y^n \text{ is } \delta\text{-generated by } u_k \Big\}. \tag{10}$$

where an output sequence y^n is said to be δ-generated by an input sequence x^n if

$$\left| \frac{f_{i,j}(x^n, y^n) - f_i(x^n)p'(j \mid i)}{\sqrt{f_i(x^n)p'(j \mid i)(1-p'(j \mid i))}} \right| < \delta \ , \ i = 1, \ldots, M \ , \ j = 1, \ldots, M'. \tag{11}$$

From Lemma 2.6 (third lecture), we also know that

$$2^{n(H'(\boldsymbol{Y} \mid \boldsymbol{X}) - \frac{K_6}{\sqrt{n}})} < \left| \mathcal{B}_2^{\delta, K_0}(u_k) \right| < 2^{n(H'(\boldsymbol{Y} \mid \boldsymbol{X}) + \frac{K_6}{\sqrt{n}})} \tag{12}$$

where

$$H'(\boldsymbol{Y} \mid \boldsymbol{X}) \triangleq - \sum_{i=1}^{M} \sum_{j=1}^{M'} p(i)p'(j \mid i) \log_2 p'(j \mid i). \tag{13}$$

a)

For the system in Fig. 2 show that for any $y^n \in \mathcal{B}_2^{\delta, K_0}(u_k)$

$$2^{-n(-\sum_{i=1}^{M}\sum_{j=1}^{M'} p(i)p'(j|i)\log_2 p(j|i) + \frac{K_{11}}{\sqrt{n}})} < \Pr(\boldsymbol{Y}^n = y^n | \boldsymbol{X}^n = u_k) < 2^{-n(-\sum_{i=1}^{M}\sum_{j=1}^{M'} p(i)p'(j|i)\log_2 p(j|i) - \frac{K_{11}}{\sqrt{n}})} \tag{14}$$

where $K_{11} > 0$ is a constant which depends only on M, M', δ, K_0 (assume that $p(j \mid i) \neq 0$ for all i, j).

b)

Using (12) and (14) show that for the system in Fig. 2

$$2^{-n(\sum_{i=1}^{M} p_i D_i(p' \| p) + K_{12}/\sqrt{n})} < \Pr(\boldsymbol{Y}^n \in \mathcal{B}_2^{\delta, K_0}(u_k) \mid u_k \text{ was tx'mt}) < 2^{-n(\sum_{i=1}^{M} p_i D_i(p' \| p) - K_{12}/\sqrt{n})} \tag{15}$$

where $K_{12} > 0$ is a constant which depends only on M, M', δ, K_0 (assume that $p(j \mid i) \neq 0$ for all i, j). Further

$$D_i(p' \| p) \triangleq \sum_{j=1}^{M'} p'(j \mid i) \log_2 \frac{p'(j \mid i)}{p(j \mid i)}. \tag{16}$$

Note that for a given i, $D_i(p' \| p)$ is the Kullback-Leibler (KL) divergence between the distributions $p'(j|i)$ and $p(j|i)$.

Show that $D_i(p' \| p) \geq 0$ and that $D_i(p' \| p) = 0$ if and only if $p'(j \mid i) = p(j \mid i)$ for all $j = 1, 2, \ldots, M'$.

Therefore, conclude that as $n \to \infty$, the probability of error goes to 1. Hence, a mismatched channel code will result in high error probability. The speed with which the error probability converges to 1 as $n \to \infty$ will depend upon the KL divergence between the true c.p.f. and the c.p.f. known to the transmitter.

c) We will now discuss a code design for reliable communication in the scenario where the transmitter has imperfect c.p.f. knowledge.

Assume that the transmitter knows that, for each i the KL divergence $D_i(p\|p') \leq \mu$. We now try to construct a code (just as in the proof of the coding theorem). We however define the δ-generated set corresponding to a K_0-input typical sequence u_k (k-th codeword in our code) to be

$$\mathcal{B}_2^{\delta,K_0}(u_k) \triangleq \bigcup_{\Pi(j|i)\ i,j=1,\cdots,M,M' \ |\ D_i(\Pi\|p')\leq\mu\ \forall i=1,\ldots,M} \mathcal{B}_2^{\delta,K_0,\Pi}(u_k) \tag{17}$$

where

$$\mathcal{B}_2^{\delta,K_0,\Pi}(u_k) \triangleq \left\{ y^n \ \middle|\ \left| \frac{f_{i,j}(u_k,y^n) - f_i(u_k)\Pi(j \mid i)}{\sqrt{f_i(u_k)\Pi(j \mid i)(1-\Pi(j \mid i))}} \right| \ < \ \delta \ , \ \ i=1,\ldots,M \ , \ j=1,\ldots,M' \right\}. \tag{18}$$

With $\mathcal{B}_2^{\delta,K_0}(u_k)$ defined as in (17), show that for the system in Fig. 2

$$\Pr(Y^n \in \mathcal{B}_2^{\delta,K_0}(u_k) \mid u_k \text{ was tx'mt}) > (1 - \frac{MM'}{\delta^2}) \tag{19}$$

Further show that, for any $y^n \in \mathcal{B}_2^{\delta,K_0}(u_k)$ and the system in Fig. 2

$$\Pr(Y^n = y^n \mid X^n = u_k) > 2^{-n(H_{p'}(\boldsymbol{Y}|\boldsymbol{X}) + \frac{K_{13}}{\sqrt{n}})} \tag{20}$$

where $K_{13} > 0$ is a constant which depends only on M, M', δ, K_0, and

$$H_{p'}(\boldsymbol{Y}|\boldsymbol{X}) \triangleq \max_{\Pi(j|i)\ i,j=1,\cdots,M,M' \ |\ D_i(\Pi\|p')\leq\mu\ \forall i=1,\ldots,M} \sum_{i=1}^{M}\sum_{j=1}^{M'} p(i)\Pi(j|i)\log_2\frac{1}{p(j|i)} \tag{21}$$

The subscript in $H_{p'}(\cdot)$ emphasizes the dependency on p'.

Following the steps as in the proof for the coding theorem, show that for any $0 < \lambda \leq 1$, for every $n > 1$ there exists a (n, N, λ) code with

$$N > 2^{n(I'(\boldsymbol{X};\boldsymbol{Y}) - \frac{K_{14}}{\sqrt{n}})} \tag{22}$$

where $K_{14} > 0$ is a constant which depends only on $\lambda, M, M', \delta, K_0$, and

$$I'(\boldsymbol{X};\boldsymbol{Y}) \triangleq \max(0, H(\boldsymbol{Y}) - H_{p'}(\boldsymbol{Y}|\boldsymbol{X})). \tag{23}$$

Therefore as $n \to \infty$, we observe that even with imperfect knowledge of c.p.f., reliable communication is still possible at a rate equal to $I'(\boldsymbol{X};\boldsymbol{Y})$.

d)

Show that, for any given input distribution $p(i)$,

$$I'(\boldsymbol{X};\boldsymbol{Y}) \leq I(\boldsymbol{X};\boldsymbol{Y}) \tag{24}$$

where $I(\boldsymbol{X};\boldsymbol{Y})$ is information rate achievable with perfect c.p.f. knowledge at the transmitter.

To assess the loss in information rate, it would be interesting to study the loss $I(\boldsymbol{X};\boldsymbol{Y}) - I'(\boldsymbol{X};\boldsymbol{Y}) = H_{p'}(\boldsymbol{Y}|\boldsymbol{X}) - H(\boldsymbol{Y}|\boldsymbol{X})$ as a function of μ. Note that the maximization in (21) is a convex optimization problem.

For the special case of a binary symmetric channel ($M = M' = 2$) with true cross-over probability $q < 1/2$, characterize the information loss due to imperfect c.p.f. knowledge as a function of the estimated crossover probability $0 < \epsilon \leq 1/2$. For a given ϵ, set $\mu = q \log_2(q/\epsilon) + (1-q)\log_2((1-q)/(1-\epsilon))$ and then solve (21). (Take the input probability distribution to be uniform).