EE/Ma 126b Information Theory - Homework Set #7

Ling Li, ling@cs.caltech.edu

March 7, 2001

7.1 *Broadcast capacity depends only on the conditional marginals.* $(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)$ iff $\hat{W}_1 \neq W_1$ or $\hat{W}_2 \neq W_2$. That is, the event $(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)$ is the union of the two events $\hat{W}_1 \neq W_1$ and $\hat{W}_2 \neq W_2$. Thus

$$\max\left\{P_1^{(n)}, P_2^{(n)}\right\} \leq P^{(n)} \leq P_1^{(n)} + P_2^{(n)}.$$

Hence $P^{(n)} \to 0$ when $n \to \infty$ iff $P_1^{(n)} \to 0$ and $P_2^{(n)} \to 0$ when $n \to \infty$.

The probability of error $P^{(n)}$ does depend on the conditional joint distribution $p(y_1, y_2|x)$. However, $P_1^{(n)}$ and $P_2^{(n)}$ only depends on the conditional marginal distributions $p(y_1|x)$ and $p(y_2|x)$ respectively. Hence if for a particular broadcast channel, we have a sequence of codes with $P^{(n)} \to 0$, (so $P_1^{(n)} \to 0$ and $P_2^{(n)} \to 0$,) then using the same codes for any broadcast channel with the same conditional marginals will also ensure $P^{(n)} \to 0$ for that channel. Thus the capacity region for a broadcast channel depends only on the conditional marginals.

7.2 *Converse for the degraded broadcast channel.*

$$\begin{aligned}
nR_2 &\leq_{\text{Fano}} I(W_2; Z^n) \\
&\overset{(a)}{=} \sum_{i=1}^{n} I(W_2; Z_i|Z^{i-1}) \\
&\overset{(b)}{=} \sum_i \left(H(Z_i|Z^{i-1}) - H(Z_i|W_2, Z^{i-1})\right) \\
&\overset{(c)}{\leq} \sum_i \left(H(Z_i) - H(Z_i|W_2, Z^{i-1}, Y^{i-1})\right) \\
&\overset{(d)}{=} \sum_i \left(H(Z_i) - H(Z_i|W_2, Y^{i-1})\right) \\
&\overset{(e)}{=} \sum_{i=1}^{n} I(U_i; Z_i)
\end{aligned}$$

(a) from the chain rule for mutual information,
(b) from the definition of conditional mutual information,
(c) from conditioning reduces entropy, (removing conditioning increases entropy as well,)
(d) from the fact that Z_i is conditional independent of Z^{i-1} given Y^{i-1},
(e) from the definitions of U_i and mutual information.

Continuation of converse.

$$\begin{aligned} nR_1 &\leq_{\text{Fano}} I(W_1;Y^n) \\ &\overset{(f)}{\leq} I(W_1;Y^n,W_2) \\ &\overset{(g)}{=} I(W_1;Y^n|W_2) \\ &\overset{(h)}{=} \sum_{i=1}^{n} I(W_1;Y_i|Y^{i-1},W_2) \\ &\overset{(i)}{\leq} \sum_{i=1}^{n} I(X_i;Y_i|U_i). \end{aligned}$$

(f) from $I(W_1;Y^n,W_2) - I(W_1;Y^n) = I(W_1;W_2|Y^n) \geq 0$,
(g) from the chain rule for mutual information and the fact that W_1 and W_2 are independent,
(h) from the chain rule for mutual information,
(i) from $W_1 \to X_i \to Y_i$ is a Markov chain given $U_i = (Y^{i-1}, W_2)$ and the data processing inequality.

7.3 *Degraded broadcast channel.* The cardinality of the auxiliary random variable U is binary since X and Y_1 are binary and Y_2 are ternary. By symmetry, we connect U to X by another BSC with parameter β.

Let $q = \Pr\{X = 1\}$. The distribution of Y_2 is

$$\Pr\{Y_2\} = \{(1-\alpha)(1-q*p), \alpha, (1-\alpha)(q*p)\},$$

where $q*p = q(1-p) + (1-q)p$. Hence

$$\begin{aligned} H(Y_2) &= -(1-\alpha)(1-q*p)[\log(1-\alpha) + \log(1-q*p)] - \alpha\log\alpha \\ &\quad - (1-\alpha)(q*p)[\log(1-\alpha) + \log(q*p)] \\ &= -\alpha\log\alpha - (1-\alpha)\log(1-\alpha) \\ &\quad - (1-\alpha)[(1-q*p)\log(1-q*p) + (q*p)\log(q*p)] \\ &= H(\alpha) + (1-\alpha)H(q*p), \end{aligned} \tag{1}$$

It is clear that $q = \frac{1}{2}$ gives the maximal $H(Y_2)$ since now $q*p = \frac{1}{2}$, and when $\alpha < 1$ and $p \neq \frac{1}{2}$, maximizing $H(Y_2)$ requires $q = \frac{1}{2}$. And $q = \frac{1}{2}$ iff U is uniformly distributed on $\{0,1\}$. Also by (1), we get

$$\begin{aligned} H(Y_2|U) &= \Pr\{U=0\}H(Y_2)_{q=\beta} + \Pr\{U=1\}H(Y_2)_{q=1-\beta} \\ &= H(\alpha) + (1-\alpha)H(\beta*p), \end{aligned}$$

since $(1-\beta)*p = 1-\beta*p$ thus $H((1-\beta)*p) = H(\beta*p)$. Thus

$$I(U;Y_2) = H(Y_2) - H(Y_2|U) = (1-\alpha)(H(q*p) - H(\beta*p)).$$

And the uniform distribution of U gives $I(U;Y_2) = (1-\alpha)(1-H(\beta*p))$.

Similarly,

$$\begin{aligned} I(X;Y_1|U) &= H(Y_1|U) - H(Y_1|X,U) \\ &= H(Y_1|U) - H(Y_1|X) \text{ since } Y_1 \text{ and } U \text{ are independent given } X \\ &= H(\beta*p) - H(p). \end{aligned}$$

Thus the capacity region for this channel is the convex hull of all rate pairs (R_1, R_2) satisfying

$$\boxed{R_1 \le H(\beta * p) - H(p), \quad R_2 \le (1-\alpha)\left(1 - H(\beta * p)\right),}$$

for some β. Note that $\beta * p$ can be any value between p and $(1-p)$, so $H(\beta * p)$ can be any value between $H(p)$ and 1. So the capacity region is exactly a triangular (when $\alpha < 1$):

$$\boxed{R_1 + \frac{1}{1-\alpha} R_2 \le 1 - H(p).}$$

When $\alpha = 1$, no information can be sent ($R_2 = 0$) since Y_2 is a constant. And the capacity region now is just $\boxed{R_1 \le H(\beta * p) - H(p) \text{ and } R_2 = 0}$.

7.4 *Channels with unknown parameters.* Without loss of generality, assume $p_1 < p_2$. First, devise two codes X_1^n and X_2^n for p_1 and p_2 respectively, under the assumption that the receiver *knows* which one will be used. The alphabet size is $2^{nC(p_1)}$ for X_1^n and $2^{nC(p_2)}$ for X_2^n. Then insert a sequence of 1's before each codewords in X_1^n and X_2^n, as the prefix. The number of 1's inserted for each codeword is $K = \lceil \log n \rceil$. The receiver count the number of 1's in the first K bits of the received codewords (denote the number as X). If $X < \frac{p_1+p_2}{2} K$, then the receiver uses $p = p_1$; otherwise uses $p = p_2$. From the Chebyshev's inequality, when p_1 is used in the channel,

$$\Pr\left\{X \ge \frac{p_1+p_2}{2} K\right\} \le \Pr\left\{|X - p_1 K| \ge \frac{p_2 - p_1}{2} K\right\} \le \frac{p_1(1-p_1)}{(p_2-p_1)^2 K},$$

and while p_2 is used,

$$\Pr\left\{X < \frac{p_1+p_2}{2} K\right\} \le \Pr\left\{|X - p_2 K| \ge \frac{p_2 - p_1}{2} K\right\} \le \frac{p_2(1-p_2)}{(p_2-p_1)^2 K}.$$

Thus the probability that the receiver takes a wrong p is tends to 0 when $n \to \infty$. And the rate is

$$\frac{nC(p_i)}{n+K} = C(p_i) \cdot \frac{1}{1 + \frac{\log n}{n}} \to C(p_i) \text{ when } n \to \infty,$$

if $p = p_i$ is used $(i = 1, 2)$.

7.5 *Two-way channel.*

(a) For some product distribution $p(x_1)p(x_2)p(y_1, y_2|x_1, x_2)$, randomly design two independent codes for W_1 and W_2. Since receiver 1 (Y_1) could know X_1 exactly, the channel from the standpoint of Y_1 is a multiple access channel with inputs X_1, X_2 and output (X_1, Y_1), in which X_1 is transmitted through an error-free channel. Applying the analysis for achievability in a multiple access channel here, we get the following rate region is achievable:

$$\begin{aligned} R_1 &< I(X_1; X_1, Y_1|X_2), && (2) \\ R_2 &< I(X_2; X_1, Y_1|X_1), && (3) \\ R_1 + R_2 &< I(X_1, X_2; X_1, Y_1). && (4) \end{aligned}$$

(4) is superfluous since

$$\begin{aligned}
I(X_1, X_2; X_1, Y_1) &= I(X_1; X_1, Y_1) + I(X_2; X_1, Y_1|X_1) \\
&\geq I(X_1; X_1, Y_1|X_2) + I(X_2; X_1, Y_1|X_1).
\end{aligned}$$

Hence (4) is satisfied if (2) and (3) are satisfied. By $I(X_1; Y_1|X_1, X_2) = 0$ and X_1 and X_2 are independet,

$$I(X_1; X_1, Y_1|X_2) = I(X_1; X_1|X_2) + I(X_1; Y_1|X_1, X_2) = H(X_1).$$

So (2) becomes $R_1 < H(X_1)$, which must be satisfied if $\boxed{R_1 < I(X_1; Y_2|X_2)}$, since $I(X_1; Y_2|X_2) \leq H(X_1|X_2) = H(X_1)$. Similarly,

$$I(X_2; X_1, Y_1|X_1) = I(X_2; Y_1|X_1) + I(X_2; X_1|X_1, Y_1) = I(X_2; Y_1|X_1).$$

So (3) becomes $\boxed{R_2 < I(X_2; Y_1|X_1)}$.

The same analysis can be repeated from the standpoint of Y_2. Thus we conclude that the following rate region is achievable:

$$\begin{aligned}
R_1 &< I(X_1; Y_2|X_2), \\
R_2 &< I(X_2; Y_1|X_1).
\end{aligned}$$

(b) Using Theorem 14.10.1 in Cover's book, we have

$$\begin{aligned}
R_1 &\leq I(X_1; Y_2|X_2), \\
R_2 &\leq I(X_2; Y_1|X_1),
\end{aligned}$$

if the rate pair (R_1, R_2) is achievable.