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

Ling Li, ling@cs.caltech.edu

November 21, 2000

7.1 *Preprocessing the output.*

(a) Since $X \to Y \to g(Y)$, the data processing inequality or homework 4.2 says $I(X; g(Y)) \leq I(X; Y)$ with equality iff $X \to g(Y) \to Y$. Thus

$$\tilde{C} = \max_{p(x)} I(X; \tilde{Y}) = \max_{p(x)} I(X; g(Y)) \leq \max_{p(x)} I(X; Y) = C. \tag{1}$$

That is, using $\tilde{Y} = g(Y)$ will not strictly improve the capacity. □

(b) (1) says that for some probability mass function $p(x)$, $\tilde{C} = I(X; g(Y)) \leq I(X; Y) \leq C$. Thus $\tilde{C} = C$ iff for that $p(x)$, $X \to g(Y) \to Y$ and $I(X; Y) = \max_{p(x)} I(X; Y)$. When $\tilde{C} \neq C$, $\tilde{C}$ is strictly less than $C$. So only when for some distribution $p(x)$ that maximizes $I(X; g(Y))$

$$\boxed{X \to g(Y) \to Y \text{ and } I(X; Y) = \max_{p(x)} I(X; Y)},$$

using $\tilde{Y} = g(Y)$ doesn't strictly decrease the capacity. A simplified (sufficient but not necessary) condition is that $X \to g(Y) \to Y$ for any $p(x)$.

7.2 *Maximum likelihood decoding.* Let $w = w(y)$ be the weight of $y$, i.e., the number of 1's in $y$.

(a) $a_1$ is encoded as 000 and $a_2$ is encoded as 111. The probability that $a_1$ came out of the source given received sequence $y$ is

$$\begin{aligned}
\Pr\{X = a_1 | Y = y\} &= \frac{\Pr\{X = a_1, Y = y\}}{\Pr\{Y = y\}} \\
&= \frac{\Pr\{X = a_1, Y = y\}}{\Pr\{X = a_1, Y = y\} + \Pr\{X = a_2, Y = y\}} \\
&= \frac{\epsilon^w (1-\epsilon)^{3-w}}{\epsilon^w (1-\epsilon)^{3-w} + \epsilon^{3-w} (1-\epsilon)^w} = \frac{1}{1 + \left(\frac{\epsilon}{1-\epsilon}\right)^{3-2w}}.
\end{aligned}$$

Since $\epsilon < \frac{1}{2}$, that probability decreases when $w$ increases. Here're values of $\Pr\{X = a_1 | Y = y\}$ for each possible $y$:

| $y$ | 000 | 001, 010, 100 | 011, 101, 110 | 111 |
|---|---|---|---|---|
| $\Pr\{X = a_1 \| Y = y\}$ | $\frac{\epsilon^3}{\epsilon^3 + (1-\epsilon)^3}$ | $1-\epsilon$ | $\epsilon$ | $\frac{(1-\epsilon)^3}{\epsilon^3 + (1-\epsilon)^3}$ |

(b) Let $g(Y)$ be the decoding function. The probability of error is (slightly different definition of $P_e$ with the textbook)

$$P_e = 1 - \Pr\{X = g(Y)\} = 1 - \sum_y \Pr\{X = g(Y)|Y = y\} \Pr\{Y = y\}.$$

To minimize $P_e$, we can maximize $\Pr\{X = g(Y)|Y = y\}$ for each $y$. From the table in (a), $\Pr\{X = a_1|Y = y\} > 0.5$ when $w(y) \leq 1$; $\Pr\{X = a_1|Y = y\} < 0.5$ when $w(y) \geq 2$. Thus when $w(y) \leq 1$, we should let $g(Y) = a_1$, otherwise let $g(Y) = a_2$. This is just the decoding rule in the problem. So the decoding rule in the problem minimizes the probability of an incorrect decision. □

(c) When the coding in problem is used, we can calculate $P_e$ as

$$\begin{aligned}
P_e &= \Pr\{X = a_1, w(Y) \geq 2\} + \Pr\{X = a_2, w(Y) \leq 1\} \\
&= \frac{1}{2}\left[3\epsilon^2(1-\epsilon) + \epsilon^3\right] + \frac{1}{2}\left[\epsilon^3 + 3\epsilon^2(1-\epsilon)\right] \\
&= \boxed{3\epsilon^2 - 2\epsilon^3}.
\end{aligned}$$

(d) Similarly,

$$\Pr\{X = a_1|Y = y\} = \frac{1}{1 + \left(\frac{\epsilon}{1-\epsilon}\right)^{2(n-w)+1}} \left\{ \begin{array}{ll} > 0.5, & w \leq n; \\ < 0.5, & w > n. \end{array} \right.$$

Thus let $g(Y) = a_1$ when $w \leq n$ and $g(Y) = a_2$ when $w > n$. The probability of error is

$$\begin{aligned}
P_e^{(2n+1)} &= \Pr\{X = a_1, w(Y) > n\} + \Pr\{X = a_2, w(Y) \leq n\} \\
&= \sum_{w=n+1}^{2n+1} \binom{2n+1}{w} \epsilon^w (1-\epsilon)^{2n+1-w} \\
&= \sum_{w=n+1}^{2n+1} \binom{2n+1}{w} \epsilon^n (1-\epsilon)^n \left(\frac{\epsilon}{1-\epsilon}\right)^{w-n} (1-\epsilon) \\
&\leq \sum_{w=n+1}^{2n+1} \binom{2n+1}{w} \epsilon^n (1-\epsilon)^n \\
&= 2^{2n}\epsilon^n(1-\epsilon)^n = \left[1 - (1-2\epsilon)^2\right]^n.
\end{aligned}$$

Since $0 \leq \epsilon < \frac{1}{2}$, we have $0 \leq 1 - (1-2\epsilon)^2 < 1$. So $\boxed{P_e^{(2n+1)} \to 0 \text{ when } n \to \infty}$.

7.3 *An additive noise channel.* When $a \neq 0, \pm 1$, $\{a, a+1\} \cap \{0, 1\} = \emptyset$, and $\mathcal{Y} = \{0, 1, a, a+1\}$. $X$ can be fully determined by given $Y$ since $\{0, a\} \cap \{1, 1+a\} = \emptyset$, or we can write

$$X = \left\{ \begin{array}{ll} 0, & \text{when } Y = 0 \text{ or } a; \\ 1, & \text{when } Y = 1 \text{ or } 1+a. \end{array} \right.$$

Thus $H(X|Y) = 0$ and

$$\boxed{C = \max_{p(x)} I(X;Y) = \max_{p(x)} [H(X) - H(X|Y)] = \max_{p(x)} H(X) = 1}$$

with equality when $X$ is equally probable.

When $a = 1$, $\mathcal{Y} = \{0, 1, 2\}$ and the joint-probability distribution is

| $X \backslash Y$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | $\frac{1}{2}p$ | $\frac{1}{2}p$ | 0 |
| 1 | 0 | $\frac{1}{2}(1-p)$ | $\frac{1}{2}(1-p)$ |

where $p = \Pr\{X = 1\}$. Thus we have

$$\boxed{C = \max_{p(x)} [H(X) - H(X|Y)] = \max_{p(x)} \frac{1}{2} H(X) = \frac{1}{2}}$$

with equality when $X$ is equally probable. When $a = -1$, we can similarly get $C = \frac{1}{2}$. $a$ can not be 0 since we can not say $\Pr\{Z = 0\} = \Pr\{Z = a\} = \frac{1}{2}$ when $a = 0$.

7.4 *Channel capacity.*

(a) $H(Y|X) = H(X + Z|X) = H(Z|X) = H(Z) = \log 3$ since $Z$ is independent of $X$. Thus

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \log 3 \le \log 11 - \log 3, \tag{2}$$

with equality iff $Y$ is evenly distributed. From $Y = X + Z$ we get

$$\Pr\{Y = y\} = \sum_{z=1}^{3} \Pr\{Z = z\} \Pr\{X = y - z\} = \frac{1}{3} \sum_{z=1}^{3} \Pr\{X = y - z\}.$$

Thus the equality of (2) holds iff $X$ is evenly distributed. So

$$\boxed{C = \max_{p(x)} I(X;Y) = \log 11 - \log 3}.$$

(b) From (a), the $p(x)$ that maximizes the $I(X;Y)$ is $\boxed{p^*(x) = \frac{1}{|\mathcal{X}|} = \frac{1}{11}}$.

7.5 *Noisy typewriter.*

(a) Now $Y = X$, $I(X;Y) = H(X)$, so the capacity is $\boxed{C = \log |\mathcal{X}| = \log 26}$.

(b) Here, $Y$ is $X$ or the next letter to $X$ with equal probability. So $H(Y|X) = H(\frac{1}{2}, \frac{1}{2}) = 1$,

$$I(X;Y) = H(Y) - H(Y|X) \le \log |\mathcal{Y}| - 1 = \log 13,$$

with equality iff $Y$ is evenly distributed, i.e., $X$ is evenly distributed. Thus

$$\boxed{C = \max_{p(x)} I(X;Y) = \log 13}.$$

(c) The codes that consist of all the letters at odd (even) positions, i.e.,

$$\boxed{\{A, C, E, G, I, K, M, O, Q, S, U, W, Y\} \text{ or } \{B, D, F, H, J, L, N, P, R, T, V, X, Z\}},$$

are two highest rate codes with block length one that achieve zero probability of error. For example, for the first code, if the printed letter is $A$ or $B$, we know without doubt that the key pushed is $A$.