EE/Ma 126a Information Theory - Homework Set #8

Ling Li, ling@cs.caltech.edu

November 26, 2000

8.1 *Channels with memory have higher capacity.* Here we use $X^n \oplus Z^n$ as a short for $X_1 \oplus Z_1, X_2 \oplus Z_2, \ldots, X_n \oplus Z_n$. Since $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent, by c.r.e.,

$$H(Z^n) = \sum_{i=1}^{n} H(Z_i|Z_1^{i-1}) \leq \sum_{i=1}^{n} H(Z_i) = nH(p, 1-p).$$

When $X_1, X_2, \ldots, X_n$ are i.i.d. and evenly distributed, we have $\Pr\{X^n = x^n\} = 2^{-n}$, and

$$\begin{aligned}
\Pr\{Y^n = y^n\} &= \sum_{z^n} \Pr\{X^n = y^n \oplus z^n, Z^n = z^n\} \\
&= \sum_{z^n} \Pr\{X^n = y^n \oplus z^n\} \Pr\{Z^n = z^n\} \text{ by } Z^n \text{ independent of } X^n \\
&= 2^{-n} \sum_{z^n} \Pr\{Z^n = z^n\} = 2^{-n}.
\end{aligned}$$

Thus $H(Y^n) = n$,

$$\begin{aligned}
I(X^n; Y^n) &= I(X^n; X^n \oplus Z^n) \\
&= H(X^n \oplus Z^n) - H(X^n \oplus Z^n | X^n) \\
&= H(Y^n) - H(Z^n) \text{ by } Z^n \text{ independent of } X^n \\
&\geq n - nH(p, 1-p) = nC.
\end{aligned}$$

So

$$\boxed{\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC}.$$

8.2 *Using two channels at once.* In the new channel, $p(y_1, y_2|x_1, x_2) = p(y_1|x_1) \times p(y_2|x_2)$. Thus

$$\begin{aligned}
H(Y_1, Y_2|X_1, X_2) &= \sum_{x_1, x_2, y_1, y_2} p(x_1, x_2, y_1, y_2) \log \frac{1}{p(y_1, y_2|x_1, x_2)} \\
&= \sum_{x_1, x_2, y_1, y_2} p(x_1, x_2, y_1, y_2) \left[\log \frac{1}{p(y_1|x_1)} + \log \frac{1}{p(y_2|x_2)}\right] \\
&= \sum_{x_1, y_1} p(x_1, y_1) \log \frac{1}{p(y_1|x_1)} + \sum_{x_2, y_2} p(x_2, y_2) \log \frac{1}{p(y_2|x_2)} \\
&= H(Y_1|X_1) + H(Y_2|X_2).
\end{aligned}$$

Together with $H(Y_1,Y_2) \le H(Y_1)+H(Y_2)$, we have

$$\begin{aligned} I(X_1,X_2;Y_1,Y_2) &= H(Y_1,Y_2) - H(Y_1,Y_2|X_1,X_2) \\ &\le H(Y_1) - H(Y_1|X_1) + H(Y_2) - H(Y_2|X_2) \\ &= I(X_1;Y_1) + I(X_2;Y_2) \end{aligned} \tag{1}$$

So the capacity C of the new channel satisfies

$$C = \max_{p(x_1,x_2)} I(X_1,X_2;Y_1,Y_2) \le \max_{p(x_1)} I(X_1;Y_1) + \max_{p(x_2)} I(X_2;Y_2) = C_1 + C_2. \tag{2}$$

In fact, the equality of (2) holds since we can make X_2 independent of X_1. When X_1 and X_2 are independent, $p(x_1,x_2) = p(x_1)p(x_2)$, we have

$$\begin{aligned} p(y_1,y_2) &= \sum_{x_1,x_2} p(y_1,y_2|x_1,x_2)p(x_1,x_2) \\ &= \sum_{x_1,x_2} p(y_1|x_1)p(y_2|x_2)p(x_1)p(x_2) \\ &= \sum_{x_1} p(y_1|x_1)p(x_1) \sum_{x_2} p(y_2|x_2)p(x_2) = p(y_1)p(y_2). \end{aligned}$$

So $H(Y_1,Y_2) = H(Y_1)+H(Y_2)$, the equality of (1) holds. Thus the equality of (2) also holds,

$$\boxed{C = C_1 + C_2}.$$

8.3 *Cascade of Binary Symmetric Channels.* A raw error ($X_n \neq X_0$) happens iff there are raw errors with an odd number of channels. Thus the probability of $X_n \neq X_0$ is

$$\begin{aligned} p^{(n)} &= \sum_{i=1}^{\lceil n/2 \rceil} \binom{n}{2i-1} p^{2i-1}(1-p)^{n-(2i-1)} \\ &= \frac{1}{2}\left[\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i=0}^{n} \binom{n}{i} (-p)^i (1-p)^{n-i}\right] \\ &= \frac{1}{2}\left[(p+1-p)^n - (-p+1-p)^n\right] \\ &= \frac{1}{2}\left[1-(1-2p)^n\right]. \end{aligned}$$

Thus the cascade of n identical BSC's is equivalent to a single BSC with raw error probability

$$\boxed{p^{(n)} = \frac{1}{2}\left[1-(1-2p)^n\right]}.$$

By the properties of BSC, the capacity of the cascade channel is

$$C^{(n)} = \max_{p(x_0)} I(X_0;X_n) = 1 - H(p^{(n)}).$$

If $p \neq 0,1$, $|1-2p| < 1$, thus $\lim\limits_{n\to\infty} p^{(n)} = \frac{1}{2}$, and

$$\lim_{n\to\infty} C^{(n)} = 1 - H(\frac{1}{2}) = 0.$$

So $\boxed{\text{the capacity of the cascade channel tends to zero}}$. $0 \leq I(X_0; X_n) \leq C^{(n)}$ also gives

$$\boxed{\lim_{n\to\infty} I(X_0; X_n) = 0}.$$

8.4 *The Z Channel.* Let $p = \Pr\{X = 1\}$. From the given $p(y|x)$, we have

$$\begin{aligned} H(Y|X) &= \Pr\{X=0\}H(Y|X=0) + \Pr\{X=1\}H(Y|X=1) \\ &= (1-p)H(1,0) + pH(\frac{1}{2}, \frac{1}{2}) = p, \\ \Pr\{Y=1\} &= \frac{1}{2}p. \end{aligned}$$

Thus

$$I(X;Y) = H(Y) - H(Y|X) = H(\frac{p}{2}) - p.$$

Since $H(p)$ is a strictly concave function of p and p is a convex function of p, $I(X;Y)$ is a strictly concave function of p. The maximum of $I(X;Y)$ is achieved when $\frac{dI(X;Y)}{dp} = 0$. From $\frac{dH(p)}{dp} = -\log\frac{p}{1-p}$, we have

$$\frac{dI(X;Y)}{dp} = -\frac{1}{2}\log\frac{p}{2-p} - 1.$$

Thus $\frac{dI(X;Y)}{dp} = 0$ gives $p = \frac{2}{5}$. So the maximizing input probability distribution is

$$\boxed{p^*(0) = \frac{3}{5}, p^*(1) = \frac{2}{5},}$$

and

$$\boxed{C = \max_{p(x)} I(X;Y) = H(\frac{1}{5}) - \frac{2}{5} = \log 5 - 2}.$$

8.5 *Zero-error capacity.* This is a symmetric channel since in the table of $p(y|x)$ below,

$X \backslash Y$	0	1	2	3	4
0	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$
1	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
2	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0
3	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
4	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0

every row is a permutation of every other row and every column is also a permutation of every other column.

(a) Thus the capacity of this channel is $\boxed{C = \log|\mathcal{Y}| - H(\frac{1}{2}, \frac{1}{2}) = \log 5 - 1 \approx 1.3219}$ bits.

(b) Consider a code of length 2, with alphabet $\mathcal{X}^{(2)} = \{00, 12, 24, 31, 43\}$. The possible output Y^2 of every input $X^2 \in \mathcal{X}^{(2)}$ for the channel is

X^2	possible Y^2			
00	11	14	41	44
12	01	03	21	23
24	10	13	30	33
31	20	22	40	42
43	02	04	32	34

We can see that there's no Y^2 that can simultaneously be the output of different X^2. Thus, given Y^2, we can determine a unique X^2. So this is a zero-error code, and we know that the zero-error capacity C_{ZE} is no less than the number of bits per channel use that can be transmitted by this code, i.e.,

$$\boxed{C_{\text{ZE}} \geq \frac{1}{2} \log \left|\mathcal{X}^{(2)}\right| = \frac{\log 5}{2} \approx 1.1180 > 1}.$$

(c) From [1], the Shannon capacity of this pentagon is $\sqrt{5}$, which means the exact value of the zero-error capacity is

$$\boxed{C_{\text{ZE}} = \log \sqrt{5} = \frac{\log 5}{2}}.$$

References

[1] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, IT-25:1-7, 1979.