ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 17 — Information Theory and Coding
Solutions to homework 7 — November 26, 2007

PROBLEM 1.

(a) Observe that with P_3 defined as in the problem, whatever distribution we choose for X, the random variables X, Y, Z form a Markov chain, i.e., given Y, the random variables X and Z are independent. The data processing theorem then yields:

$$I(X;Z) \leq I(X;Y) \leq C_1$$
$$I(X;Z) \leq I(Y;Z) \leq C_2$$

and thus $I(X;Z) \leq \min\{C_1, C_2\}$ for any distribution on X. We then conclude that $C_3 = \max_{p_X} I(X;Z) \leq \min\{C_1, C_2\}$.

(b) The statistician calculates $\tilde{Y} = g(Y)$.

(b1) Since $X \to Y \to \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every distribution on X,

$$I(X;Y) \geq I(X;\tilde{Y}).$$

Let $\tilde{p}(x)$ be the distribution on x that maximizes $I(X;\tilde{Y})$. Then

$$C = \max_{p(x)} I(X;Y) \geq I(X;Y)_{p(x)=\tilde{p}(x)} \geq I(X;\tilde{Y})_{p(x)=\tilde{p}(x)} = \max_{p(x)} I(X;\tilde{Y}) = \tilde{C}.$$

Thus, the statistician is wrong and processing the output does not increase capacity.

(b2) We have equality (no decrease in capacity) in the above sequence of inequalities only if we have equality in data processing inequality, i.e., for the distribution that maximizes $I(X;\tilde{Y})$, we have $X \to \tilde{Y} \to Y$ forming a Markov chain, in other words if given $\tilde{Y}$, X and Y are independent.

PROBLEM 2.

$$Y = X + Z \qquad X \in \{0,1\}, \quad Z \in \{0,a\}$$

We have to distinguish various cases depending on the values of a.

$a = 0$ In this case, $Y = X$, and $\max I(X;Y) = \max H(X) = 1$. Hence the capacity is 1 bit per transmission.

$a \neq 0, \pm 1$ In this case, Y has four possible values $0, 1, a$ and $1+a$. Knowing Y, we know the X which was sent, and hence $H(X|Y) = 0$. Hence $\max I(X;Y) = \max H(X) = 1$, achieved for an uniform distribution on the input X.

$a = \pm 1$ In the case $a = 1$, Y has three possible output values, $0, 1$ and 2, and the channel is identical to the binary erasure channel discussed in class, with $\epsilon = 1/2$. As derived in class, the capacity of this channel is $1 - \epsilon = 1/2$ bit per transmission. The case of $a = -1$ is essentially the same and the capacity here is also $1/2$ bit per transmission.

PROBLEM 3. Since given X, one can determine Y from Z and vice versa, $H(Y|X) = H(Z|X) = H(Z) = \log 3$, regardless of the distribution of X. Hence the capacity of the channel is

$$\begin{aligned}
C &= \max_{p_X} I(X;Y) \\
&= \max_{p_X} H(Y) - H(Y|X) \\
&= \log 11 - \log 3
\end{aligned}$$

which is attained when X has uniform distribution. The same result can also be seen by observing that this channel is symmetric.

PROBLEM 4.

(a) By Bayes rule, for any events A and B,

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)}.$$

In this case, we wish to calculate the conditional probability of a_1 given the channel output. Thus we take the event A to the event that the source produced a_1, and B to be the event corresponding to one of the 8 possible output sequences. Thus $\Pr(A) = 1/2$, and $\Pr(B|A) = \epsilon^i(1-\epsilon)^{3-i}$, where i is the number of ones in the received sequence. $\Pr(B)$ can then be calculated as $\Pr(B) = \Pr(a_1)\Pr(B|a_1) + \Pr(a_2)\Pr(B|a_2)$. Thus we can calculate

$$\begin{aligned}
\Pr(a_1|000) &= \frac{\frac{1}{2}(1-\epsilon)^3}{\frac{1}{2}(1-\epsilon)^3 + \frac{1}{2}\epsilon^3} \\
\Pr(a_1|100) = \Pr(a_1|010) = \Pr(a_1|001) &= \frac{\frac{1}{2}(1-\epsilon)^2\epsilon}{\frac{1}{2}(1-\epsilon)^2\epsilon + \frac{1}{2}\epsilon^2(1-\epsilon)} \\
\Pr(a_1|110) = \Pr(a_1|011) = \Pr(a_1|101) &= \frac{\frac{1}{2}(1-\epsilon)\epsilon^2}{\frac{1}{2}(1-\epsilon)\epsilon^2 + \frac{1}{2}\epsilon(1-\epsilon)^2} \\
\Pr(a_1|111) &= \frac{\frac{1}{2}\epsilon^3}{\frac{1}{2}\epsilon^3 + \frac{1}{2}(1-\epsilon)^3}
\end{aligned}$$

(b) If $\epsilon < 1/2$, then the probability of a_1 given 000,001,010 or 100 is greater than $1/2$, and the probability of a_2 given 110,011,101 or 111 is greater than $1/2$. Therefore, the decoding rule above chooses the source symbol that has maximum probability given the observed output. This is the *maximum a posteriori* decoding rule, and is optimal in that it minimizes the probability of error. To see that this is true, let the input source symbol be X, let the output of the channel be denoted by Y and the decoded symbol be $\hat{X}(Y)$. Then

$$\begin{aligned}
\Pr(E) &= \Pr(X \neq \hat{X}) \\
&= \sum_y \Pr(Y=y) \Pr(X \neq \hat{X} | Y=y) \\
&= \sum_y \Pr(Y=y) \sum_{x \neq \hat{x}(y)} \Pr(x|Y=y) \\
&= \sum_y \Pr(Y=y)\left(1 - \Pr(\hat{x}(y)|Y=y)\right) \\
&= \sum_y \Pr(Y=y) - \sum_y Pr(Y=y) \Pr(\hat{x}(y)|Y=y) \\
&= 1 - \sum_y Pr(Y=y) \Pr(\hat{x}(y)|Y=y)
\end{aligned}$$

and thus to minimize the probability of error, we have to maximize the second term, which is maximized by choosing $\hat{x}(y)$ to the the symbol that maximizes the conditional probability of the source symbol given the output.

(c) The probability of error can also be expanded

$$\begin{aligned}
\Pr(E) &= \Pr(X \neq \hat{X}) \\
&= \sum_x \Pr(x) \Pr(\hat{X} \neq x) \\
&= \Pr(a_1) \Pr(Y = 011, 110, 101, \text{ or } 111) \\
&\quad + \Pr(a_2) \Pr(Y = 000, 001, 010 \text{ or } 100) \\
&= \frac{1}{2}\left(3\epsilon^2(1-\epsilon) + \epsilon^3\right) + \frac{1}{2}\left(3\epsilon^2(1-\epsilon) + \epsilon^3\right) \\
&= 3\epsilon^2(1-\epsilon) + \epsilon^3.
\end{aligned}$$

(d) By extending the same arguments, it is easy to see that the decoding rule that minimizes the probability of error is the maximum a posteriori decoding rule, which in this case is the same as the maximum likelihood decoding rule (since the two input symbols are equally likely). So we choose the source symbol that is most likely to have produced the given output. This corresponds to choosing a_1 if the number of 1's in the received sequence is n or less, and choosing a_2 otherwise. The probability of error is then equal to (by symmetry) the probability of error given that a_1 was sent, which is the probability that $n+1$ or more 0's have been changed to 1's by the channel. This probability is

$$\Pr(E) = \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} \epsilon^i (1-\epsilon)^{2n+1-i}$$

This probability goes to 0 as $n \to \infty$, since this is the probability that the number of 1's is $n+1$ or more, and since the expected proportion of 1's is $n\epsilon < n+1$, by the weak law of large numbers the above probability goes to 0 as $n \to \infty$.

Problem 5.

First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $x = \Pr(X = 1)$:

$$\begin{aligned}
H(Y|X) &= x\,\mathcal{H}(\varepsilon)\\
H(Y) &= \mathcal{H}(\Pr(Y = 1)) = \mathcal{H}((1-\varepsilon)x)\\
I(X;Y) &= H(Y) - H(Y|X) = \mathcal{H}((1-\varepsilon)x) - x\,\mathcal{H}(\varepsilon) \qquad (1)
\end{aligned}$$

We deduce that if $\varepsilon = 0$, the capacity equals 1 bit/symbol and is attained for $x = 1/2$. If $\varepsilon = 1$, then $I(X;Y) = 0$ for every $0 \leq x \leq 1$. Hence, the capacity is equal to zero and any value of x achieves it. From now on we assume $\varepsilon \neq 0, 1$.

Using elementary calculus, we have that

$$\frac{d}{dx}I(X;Y) = (1-\varepsilon)\log\left(\frac{1-(1-\varepsilon)x}{(1-\varepsilon)x}\right) - \mathcal{H}(\varepsilon)\ .$$

Imposing the condition $\frac{d}{dx}I(X;Y) = 0$ yields to the unique solution

$$x^*(\varepsilon) = \left((1-\varepsilon)(2^{\frac{\mathcal{H}(\varepsilon)}{1-\varepsilon}} + 1)\right)^{-1}\ .$$

From (1) we have $I(X;Y) = 0$ for $x = 0$ and $x = 1$, and therefore the maximum of the mutual information is achieved for $x = x^*(\varepsilon)$. The capacity $C(\varepsilon)$ is given by

$$C(\varepsilon) = \mathcal{H}((1-\varepsilon)x^*(\varepsilon)) - x^*(\varepsilon)\,\mathcal{H}(\varepsilon) \text{ bits/symbol.}$$