ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 19 | Information Theory and Coding
Solution 8 | November 23, 2010, SG1 – 15:15-17:00

Problem 1 (Binary Erasure Channel)**.** (a) Let the input distribution be $\Pr\{X = 1\} = \alpha$ and $\Pr\{X = 0\} = 1 - \alpha$.

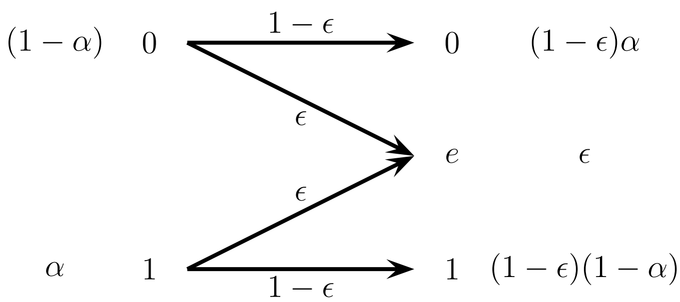

Figure 1: Binary Erasure Channel

(i)

$$
\begin{aligned}
H(Y|X) &= \Pr\{X=0\}\underbrace{H(Y|X=0)}_{=H_b(\epsilon)} + \Pr\{X=1\}\underbrace{H(Y|X=1)}_{=H_b(\epsilon)} \\
&= (1-\alpha)H_b(\epsilon) + \alpha H_b(\epsilon) \\
&= H_b(\epsilon).
\end{aligned}
$$

(ii) From Figure 1, it can be seen that $\Pr\{Y = 0\} = (1 - \epsilon)(1 - \alpha)$, $\Pr\{Y = e\} = \epsilon$, and $\Pr\{Y = 1\} = (1 - \epsilon)\alpha$. Therefore,

$$
\begin{aligned}
H(Y) &= -(1-\epsilon)(1-\alpha)\log((1-\epsilon)(1-\alpha)) - \epsilon\log\epsilon - (1-\epsilon)\alpha\log((1-\epsilon)\alpha) \\
&= H_b(\epsilon) + (1-\epsilon)H_b(\alpha).
\end{aligned}
$$

(iii) $I_\alpha(X;Y) = H(Y) - H(Y|X) = (1-\epsilon)H_b(\alpha)$.

(b) $C_{BEC} = \max_{0\le\alpha\le1} I(X;Y) = \max_{0\le\alpha\le1}(1-\epsilon)H_b(\alpha) = (1-\epsilon)$. The capacity achieving distribution corresponds to $\alpha^* = 1/2$, which is the uniform distribution.

Problem 2 (Z Channel)**.** Let the input distribution be $\Pr\{X = 1\} = \alpha$ and $\Pr\{X = 0\} = 1 - \alpha$.

(a) (i)

$$
\begin{aligned}
H(Y|X) &= \Pr\{X=0\}\underbrace{H(Y|X=0)}_{=0} + \Pr\{X=1\}\underbrace{H(Y|X=1)}_{=H_b(\epsilon)} \\
&= \alpha H_b(\epsilon).
\end{aligned}
$$

(ii) From Figure 2, it can be seen that $\Pr\{Y = 0\} = 1 - \alpha(1 - \epsilon)$ and $\Pr\{Y = 1\} = \alpha(1 - \epsilon)$. Therefore, $H(Y) = H_b(\alpha(1 - \epsilon))$.

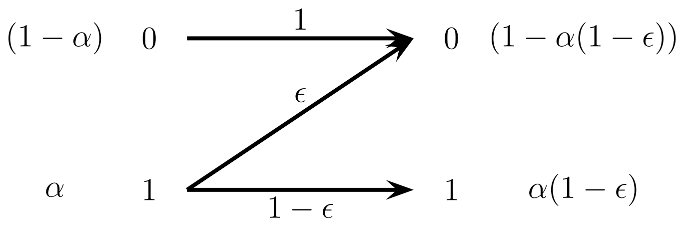

Figure 2: Z Channel

(b) $C_Z := \max_{0\leq\alpha\leq 1} I_\alpha(X;Y) = \max_{0\leq\alpha\leq 1} H_b(\alpha(1-\epsilon)) - \alpha H_b(\epsilon)$.

$$\begin{aligned} I_\alpha(X;Y) &= H_b(\alpha(1-\epsilon)) - \alpha H_b(\epsilon) \\ &= -\alpha\bar{\epsilon}\ln(\alpha\bar{\epsilon}) - (1-\alpha\bar{\epsilon})\ln(1-\alpha\bar{\epsilon}) - \alpha H_b(\epsilon), \end{aligned} \tag{1}$$

where $\bar{\epsilon} := (1-\epsilon)$. Differentiating $I_\alpha(X;Y)$ with respect to α we get

$$\frac{d}{d\alpha} I_\alpha(X;Y) = -\bar{\epsilon}\ln(\alpha\bar{\epsilon}) - \bar{\epsilon} + \bar{\epsilon}\ln(1-\alpha\bar{\epsilon}) + \bar{\epsilon} - H_b(\epsilon). \tag{2}$$

Setting the above expression to zero, we get

$$\begin{aligned} \bar{\epsilon}\ln\left(\frac{1-\alpha^*\bar{\epsilon}}{\alpha^*\bar{\epsilon}}\right) &= H_b(\epsilon) = -\epsilon\ln\epsilon - \bar{\epsilon}\ln\bar{\epsilon} \\ \frac{1-\alpha^*\bar{\epsilon}}{\alpha^*\bar{\epsilon}} &= e^{-\frac{\epsilon}{\bar{\epsilon}}\ln\epsilon - \frac{\bar{\epsilon}}{\bar{\epsilon}}\ln\bar{\epsilon}} = \frac{1}{\bar{\epsilon}\epsilon^{\frac{\epsilon}{\bar{\epsilon}}}} \\ \alpha^* &= \frac{\epsilon^{\frac{\epsilon}{\bar{\epsilon}}}}{1+\bar{\epsilon}\epsilon^{\frac{\epsilon}{\bar{\epsilon}}}} \end{aligned}$$

The above α^* corresponds to the capacity achieving input distribution. From (1) and (2), we note that

$$\begin{aligned} C_Z &= I_{\alpha^*}(X;Y) = \alpha^* \left.\frac{d}{d\alpha} I_\alpha(X;Y)\right|_{\alpha=\alpha^*} - \ln(1-\alpha^*\bar{\epsilon}) \\ &= -\ln(1-\alpha^*\bar{\epsilon}) \\ &= -\ln\left(1 - \frac{\bar{\epsilon}\epsilon^{\frac{\epsilon}{\bar{\epsilon}}}}{1+\bar{\epsilon}\epsilon^{\frac{\epsilon}{\bar{\epsilon}}}}\right) \\ &= \log\left(1+\bar{\epsilon}\epsilon^{\frac{\epsilon}{\bar{\epsilon}}}\right). \end{aligned}$$

.

(c) Figure 3 shows the plots of capacity C_Z and the information rate $I_{1/2}(X;Y)$ versus ϵ. We lose approximately a maximum of 6% on the information rate by using the uniform input distribution instead of the capacity achieving distribution. In fact, the following general result holds: for any binary input discrete memoryless channel, the uniform input distribution achieves at least a fraction $\frac{1}{2}e\ln 2 \approx 0.942$ of the channel capacity of that channel. [Refer E. E. Majani and H. Rumsey, Jr., *Two results on binary-input discrete memoryless channels*, in Proc. of the IEEE Int. Symposium on Information Theory, June 1991, p. 104]

Problem 3 (Symmetric Channels)**.** Let $\mathbf{r}$ be a row of the transition matrix. Then

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(\mathbf{r}) \\ &\leq \log n - H(\mathbf{r}), \end{aligned}$$

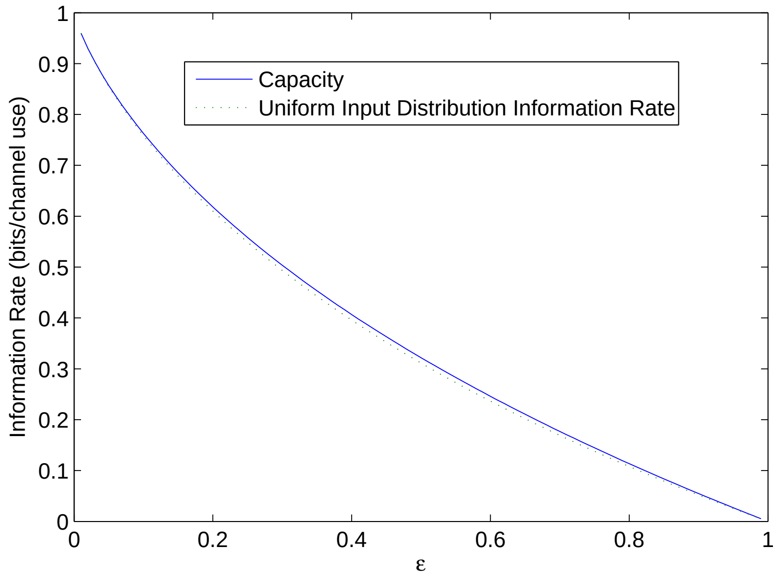

Figure 3: Comparison of C_Z and the information rate $I_{1/2}(X;Y)$ for the Z channel.

where the last step follows from the fact that the entropy of a discrete random variable is always less than the logarithm of the alphabet size of the random variable. The condition for equality is when the distribution of Y is uniform on its alphabet. If there exists a input distribution that makes the distribution of Y uniform on its alphabet, it will be the capacity-achieving distribution and the RHS in the above inequality will be the capacity. Let us check if that is true for weakly symmetric channels with uniform distribution on the input alphabet:

$$\Pr\{Y=y\} = \sum_{x=1}^{m} \Pr\{Y=y|X=x\} \underbrace{\Pr\{X=x\}}_{=1/m} = \frac{1}{m} \underbrace{\sum_{x=1}^{m} \Pr\{Y=y|X=x\}}_{=:c} = \frac{c}{m},$$

where the second equality arises because the input is uniformly distributed, and the third equality is due to the property of weakly symmetric channels (sum of rows of the transition matrix are the same).

Problem 4 (Fano Inequality). (a) By inspection, we see that the estimator that minimizes P_e would be

$$\hat{X}(y) = \begin{cases} 1 & \text{if } y=a \\ 2 & \text{if } y=b \\ 3 & \text{if } y=c \end{cases}$$

The associated P_e is the sum of $\Pr\{X=x, Y=y\}$, $x \neq \hat{X}(y)$. Therefore, $P_e = 1/2$.

(b) One form of Fano's inequality is:

$$P_e \geq \frac{H(X|Y)-1}{\log_2 |\mathcal{X}|},$$

where $\mathcal{X}$ is the alphabet of X, which is our case is $\mathcal{X} = \{1, 2, 3\}$. We have

$$\begin{aligned}
H(X|Y) &= H(X|Y=a)\Pr\{Y=a\} + H(X|Y=b)\Pr\{Y=b\} + H(X|Y=c)\Pr\{Y=c\} \\
&= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)(\Pr\{Y=a\} + \Pr\{Y=b\} + \Pr\{Y=c\}) \\
&= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) = 1.5 \text{ bits.}
\end{aligned}$$

Therefore,

$$P_e \geq \frac{1.5 - 1}{\log_2 3} = 0.316.$$

The estimator $\hat{X}(Y)$ is not very close to the above form of Fano's bound (because $P_e = 1/2$ for the estimator).

The reason is as follows: In the derivation of Fano's inequality, we had to bound $H(X|E, \hat{X})$, where $E = 0$ if $\hat{X} = X$ and $E = 1$ if $\hat{X} \neq X$:

$$\begin{aligned}
H(X|E, \hat{X}) &= \Pr\{E=0\} \underbrace{H(X|\hat{X}, E=0)}_{=0} + \underbrace{\Pr\{E=1\}}_{=P_e} H(X|\hat{X}, E=1) \\
&= P_e H(X|\hat{X}, E=1),
\end{aligned}$$

where the first entropy is zero because, given that $E = 0$, X is completely determined by $\hat{X}$. If the alphabet of $\hat{X}$ is not the same as the alphabet of X, then in general, the entropy $H(X|\hat{X}, E = 1)$ can be bounded by $\log_2 |\mathcal{X}|$. However, if the alphabets of $\hat{X}$ and X are the same, then given that $E = 1$, we can rule out the possibility of $X = \hat{X} \in \mathcal{X}$. Therefore, the entropy $H(X|\hat{X}, E = 1)$ can be bounded by $\log_2(|\mathcal{X}|-1)$, and a tighter bound can be obtained:

$$P_e \geq \frac{H(X|Y) - 1}{\log_2(|\mathcal{X}| - 1)}.$$

By this tighter bound, we get

$$P_e \geq \frac{1.5 - 1}{\log_2 2} = \frac{1}{2}.$$

Therefore the estimator in (a) is actually quite good.