CIE6020 Assignment 4

Due: 23:55, 10 April 2019

1. *Fano's inequality.* Consider a random variable X over $\{1, 2, \dots, m\}$ with $\Pr(X = i) = p_i, i = 1, 2, ..., m$, where $p_1 \geq p_2 \geq \cdots \geq p_m$. The minimal probability of error predictor when there is no information about the instance of X is $\hat{X} = 1$, the most probable value of X, with resulting probability of error $P_e = 1 - p_1$. Maximize $H(X)$ subject to the constraint $1 - p_1 = P_e$ to find a bound on P_e in terms of $H(X)$. This is Fano's inequality in the absence of conditioning.

Solution: The minimal probability of error predictor of X is $\hat{X} = 1$. The probability of error is $P_e = 1 - p_1$. We maximize $H(X)$ for a given P_e. Write

$$
\begin{aligned}
H(X) &= -p_1 \log p_1 - \sum_{i=2}^{m} p_i \log p_i \\
&= -p_1 \log p_1 - \sum_{i=2}^{m} P_e \frac{p_i}{P_e} \log \frac{p_i}{P_e} - P_e \log P_e \\
&= H(P_e) + P_e H\left(\frac{p_2}{P_e}, \frac{p_3}{P_e}, \cdots, \frac{p_m}{P_e}\right) \\
&\leq H(P_e) + P_e \log(m-1),
\end{aligned}
$$

since the maximum of $H(\frac{p_2}{P_e}, \frac{p_3}{P_e}, \cdots, \frac{p_m}{P_e})$ is attained by an uniform distribution. Hence any X that can be predicted with a probability of error P_e must satisfy

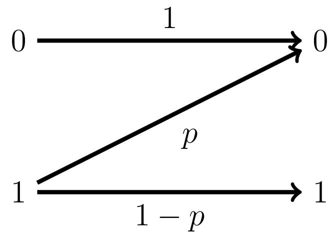
$$H(X) \leq H(P_e) + P_e \log(m-1),$$

which is the unconditional form of Fano's inequality. We can weaken this inequality to obtain an explicit lower bound for P_e,

$$P_e \geq \frac{H(X) - 1}{\log(m-1)}.$$

2. *Z-channel.* The Z-channel is a binary input and binary output channel with the transition probabilities $W(y|x)$ given by

$$W = \begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}.$$



See an illustration in the figure below.

Find the capacity of the Z-channel and the maximizing input probability distribution.

Solution: Let X be the input and Y be the output of the Z-channel, where $f_X(0) = q$ and $f_X(1) = 1 - q$. We can calcualte that $f_Y(0) = f_X(0) + pf_X(1) = q + p(1-q) = p + q - pq$ and $f_Y(1) = (1-p)f_X(1) = (1-p)(1-q)$.

Thus, the mutual information

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(q+p-pq) - f_X(0)H(Y|X=0) - f_X(1)H(Y|X=1) \\ &= h(q+p-pq) - (1-q)h(p), \end{aligned}$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function.

By $h'(p) = \ln\frac{1-p}{p}$, we have

$$\frac{dI(X;Y)}{dq} = (1-p)\log\frac{1-p-q+pq}{q+p-pq} + h(p).$$

By solving $\frac{dI(X;Y)}{dq} = 0$, we get

$$q = \frac{1}{1-p}\left(\frac{1}{1+2^{-\frac{h(p)}{1-p}}} - p\right).$$

Substituting the above value of q into $I(X;Y)$, we have

$$C = h\left(\frac{1}{1+2^{-\frac{h(p)}{1-p}}}\right) - h(p) + \frac{h(p)}{1-p}\left(\frac{1}{1+2^{-\frac{h(p)}{1-p}}} - p\right).$$

3. Consider the discrete memoryless channel $Y = X + Z \pmod{11}$, where the alphabet of X is $\{0, 1, \ldots, 10\}$ and Z is uniformly distributed on $\{1, 2, 3\}$. Assume Z is independent of X. Find the capacity of this channel and the maximizing input probability distribution.

Solution: The mutual information between X and Y is

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(X+Z|X) = H(Y) - H(Z|X)$$

since Z is independent of X, then $H(Z|X) = H(Z) = log3$
$H(Y) = -\sum_{i=1}^{11} P(Y = i-1) log P(Y = i-1)$
Simply, we have $H(Y)$ achieve maximum when Y is uniformly distributed, which

$$C = \max_{p(x)} I(X;Y) = -\sum_{i=1}^{11} \frac{1}{11} log \frac{1}{11} - log3 = log\frac{11}{3}$$

this can be achieved when Y is uniformly distributed, where by symmetric X is uniformly distributed.

4. Consider a channel with the input and output alphabet $\{0,1\}$. The ith input X_i and the ith output Y_i, $i = 1, 2, \ldots$ are related by

$$Y_i = X_i + U_i$$

where the addition is modulo 2 and U_i has distribution $\Pr\{U_i = 1\} = 1 - \Pr\{U_i = 0\} = q$. Here U_j and $(X_i, i = 1, \ldots)$ are independent.

(a) When $U_i, i = 1, 2, \ldots$ and $(X_j, j = 1, \ldots)$ are independent, show the channel is a memoryless binary symmetric channel and give its capacity.
(Hint: show that for any integer $n > 0$,

$$\Pr\{Y_i = y_i, i = 1, \ldots, n | X_i = x_i, i = 1, \ldots, n\} = \prod_{i=1}^{n} \Pr\{Y_i = y_i | X_i = x_i\},$$

i.e., the channel is memoryless.)

Solution: First,

$$\begin{aligned}\Pr\{Y_i = y | X_i = x\} &= \Pr\{U_i = y - x | X_i = x\} \\ &= \Pr\{U_i = y - x\} \\ &\triangleq W(y|x).\end{aligned}$$

We know that $W(1|0) = W(0|1) = q$ and $W(0|0) = W(1|1) = 1 - q$.
Write

$$\begin{aligned}&\Pr\{Y_i = y_i, i = 1, \ldots, n | X_i = x_i, i = 1, \ldots, n\} \\ &= \frac{\Pr\{U_i = y_i - x_i, X_i, i = 1, \ldots, n\}}{\Pr\{X_i = x_i, i = 1, \ldots, n\}} \\ &= \Pr\{U_i = y_i - x_i, i = 1, \ldots, n\} \\ &= \prod_{i=1}^{n} \Pr\{U_i = y_i - x_i\} \\ &= \prod_{i=1}^{n} W(y_i|x_i).\end{aligned}$$

Therefore, the channel is memoryless and binary symmetry.
The capacity of the channel is $1 - H(q)$.

(b) When $U_i = U_{i+1}, i = 1, 3, 5, \ldots$, and $U_i, i = 1, 3, 5, \ldots$ and $(X_j, j = 1, \ldots)$ are independent, show that the channel is not memoryless.

(Hint: calculate $\Pr\{Y_1 = y_1, Y_2 = y_2 | X_1 = x_1, X_2 = x_2\}$ and show that it is not the same as $\Pr\{Y_1 = y_1 | X_1 = x_1\} \Pr\{Y_2 = y_2 | X_2 = x_2\}$.)

Solution: Let

$$\begin{aligned} W_2(y_1, y_2 | x_1, x_2) &\triangleq \Pr\{Y_1 = y_1, Y_2 = y_2 | X_1 = x_1, X_2 = x_2\} \\ &= \Pr\{U_1 = y_1 - x_1, U_2 = y_2 - x_2\}. \end{aligned}$$

As $U_1 = U_2$, we have for $x, y \in \{0, 1\}$,

$$\begin{aligned} W_2(x, y | x, y) &= 1 - q, \\ W_2(x + 1, y + 1 | x, y) &= q. \end{aligned}$$

As $W^2 \neq W_2$, the channel is not memoryless.

(c) Under the condition of (b), the channel can be equivalent to a DMC by combining two consecutive uses of the channel. Give the transition matrix of this DMC, and calculate its capacity.

Solution: W_2 as the transition matrix. Let X' and Y' be the input and output of this channel. We have

$$\begin{aligned} I(X'; Y') &= H(Y') - H(Y'|X') \\ &= H(Y') - H(q) \\ &\leq 2 - H(q). \end{aligned}$$

As the maximum can be achieved by the uniform distribution of X', the capacity of the channel is $2 - H(p)$.

(d) Assume you are given a set of capacity achieving codes for the memoryless binary symmetric channel under the condition of (a). Using these codes, construct a capacity achieving code for the channel under the condition of (b).

Solution: Consider an (n, M) code C for the binary symmetric DMC. We can modify this code to one for DMC $\{W_2\}$ of the same error probability and rate $1 + \log M/n$. For each codeword $(x_1, x_2, \ldots, x_n)$ of C, and n bits $(y_1, y_2, \ldots, y_n)$, we form a new codeword for W_2, where the i-th input is $(x_i, x_i + y_i)$.

This code has $M2^n$ codewords. Suppose the channel input is $(x_i, x_i + y_i), i = 1, \ldots, n$ and the corresponding output is $(u_i, v_i), i = 1, \ldots, n$. Then $y_i =$

$u_i + v_i$ and $(u_i, i = 1, \ldots n)$ can be used to decode $(x_i, i = 1, \ldots, n)$ using the decoding algorithm of C.

5. Consider a stochastic process $U_1, U_2, \ldots$ with $U_i \in \mathcal{U}$, a finite set, such that the entropy rate H exists and $-\frac{1}{n} \log p(U_1, U_2, \ldots, U_n) \to H$ in probability. For any integer $n > 0$ and real number $\epsilon > 0$, find a subset $A_\epsilon^{(n)} \subset \mathcal{U}^n$ such that $|A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$ and $\Pr\{(U_1, \ldots, U_n) \in A_\epsilon^{(n)}\} > 1 - \epsilon$ when n is sufficiently large.