Prof. Dr. Rudolf Mathar, Dr.-Ing. Gholamreza Alirezaei, Emilio Balda, Vimal Radhakrishnan

Tutorial 3
- Proposed Solution -

Monday, November 12, 2018

Solution of Problem 1

a) The Entropy of distribution $\mathbf{p}$ can be calculated as

$$H(\mathbf{p}) = -\sum_i p_i \log p_i = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = 1.5 \text{ bits.}$$

Similarly, the Entropy of distribution $\mathbf{q}$,

$$H(\mathbf{q}) = -\sum_i q_i \log p_i = -\frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} = 1.5849 \text{ bits.}$$

For Kullbach-Leibler Divergence, we get

$$D(\mathbf{p}||\mathbf{q}) = \sum_i p_i \log\frac{p_i}{q_i} = \frac{1}{2}\log\frac{3}{2} + \frac{1}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{3}{4} = 0.0849$$

and

$$D(\mathbf{q}||\mathbf{p}) = \sum_i q_i \log\frac{q_i}{p_i} = \frac{1}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{4}{3} + \frac{1}{3}\log\frac{4}{3} = 0.08170$$

b) Consider two distributions $\mathbf{p}$ and $\mathbf{q}$ on a binary alphabet with probability mass function $(p, 1-p)$ and $(q, 1-q)$ respectively.

The relative entropies can be written as

$$D(\mathbf{p}||\mathbf{q}) = \sum_i p_i \log\frac{p_i}{q_i} = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \tag{1}$$

and

$$D(\mathbf{q}||\mathbf{p}) = \sum_i q_i \log\frac{q_i}{p_i} = q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p} \tag{2}$$

Equating (1) and (2), we write

$$\begin{aligned} D(\mathbf{p}||\mathbf{q}) &= D(\mathbf{q}||\mathbf{p}) \\ q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p} &= p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \\ (p+q)\log\frac{q}{p} &= (1-p+1-q)\log\frac{1-p}{1-q} \end{aligned}$$

We can clearly see that, the equality holds when $p = 1 - q$.

Solution of Problem 2

a) The minimum probability of error predictor when there is no information is $\hat{X} = 1$, the most probable value of X. In this case, the probability of error $P_e = 1 - p_1$. Hence if we fix P_e, we fix p_1.

In order to obtain an upper bound on the entropy for a given P_e, we maximize the entropy of X for a given P_e. The entropy can be written as

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

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

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

The above inequality is the unconditional form of Fano's inequality. Thus, an explicit lower bound for P_e can be written as

$$P_e \geq \frac{H(X) - \log 2}{\log(m-1)}.$$

b) From the above exercise it is clear that the maximum of entropy $H(X)$ or $H(\mathbf{p})$ is attained when $p_1 = 1 - P_e$ and $p_2, p_3, ...p_m$ corresponds to a uniform distribution. That is $p_2 = p_3.... = p_m = \frac{P_e}{m-1}$. Hence the probability vector $\mathbf{p}$ for which Fano's inequality is sharp can be written as

$$\mathbf{p} = (1 - P_e, \frac{P_e}{m-1}, ...\frac{P_e}{m-1}).$$

This can be easily verified by calculating entropy with probability vector $\mathbf{p} = (1 - P_e, \frac{P_e}{m-1}, ...\frac{P_e}{m-1})$, we get

$$\begin{aligned} H(\mathbf{p}) &= -p_1 \log p_1 - \sum_{i=2}^{m} \frac{P_e}{m-1} \log \frac{P_e}{m-1} \\ &= -(1 - P_e) \log(1 - P_e) - P_e \log \frac{P_e}{m-1} \\ &= -(1 - P_e) \log(1 - P_e) - P_e \log P_e + P_e \log(m-1) \\ &= H(P_e) + P_e \log(m-1), \end{aligned}$$

the equality holds.

Solution of Problem 3

a) From data processing inequality, we can write

$$\begin{aligned} I(X_1;X_3) &\leq I(X_1;X_2) \\ &= H(X_2) - H(X_2|X_1) \\ &\leq H(X_2) \quad \text{(Since } H(X_2|X_1) \geq 0) \\ &\leq \log k \quad \text{(maximum entropy of an uniform distribution)} \end{aligned} \tag{3}$$

Thus,the dependence between X_1 and X_3 is limited by the size of the bottleneck. That is $I(X_1;X_3) \leq \log k$.

b) For $k = 1$, $I(X_1;X_3) \leq \log 1 = 0$ and since $I(X_1;X_3) \geq 0$, we get $I(X_1;X_3) = 0$. So, for $k = 1$, X_1 and X_3 are independent.

Solution of Problem 4

Define for $t > 0$

$$f(t) = \ln t - t + 1. \tag{4}$$

Taking the first derivative and equating to zero, we get

$$f'(t) = \frac{1}{t} - 1 = 0 \implies t = 1. \tag{5}$$

we get a maximal point, since second derivative $f''(t) = -\frac{1}{t^2} < 0$ for $\forall t > 0$. Also

$$\lim_{t\to 0^+} f(t) = -\infty = \lim_{t\to\infty} f(t) \tag{6}$$

implies the above attained point is a global maximal point. Thus,

$$\forall t > 0,\ f(t) \leq f(1) = \ln 1 - 1 + 1 = 0. \tag{7}$$

For $t = 0$,

$$\ln 0 < 0 - 1 \to -\infty < -1.$$

Consequently, we can write

$$\ln t \leq t - 1, \quad \forall t \geq 0. \text{ (Hence proved)} \tag{8}$$