E2 201 (Aug–Dec 2014)

Homework Assignment 2

Discussion: (to be decided) Quiz: Friday, Sept. 5

This assignment consists of two pages.

1. Let X and Y be discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively.
   (a) Show that $H(Y|X) = 0$ if and only if Y is a function of X, i.e., for each $x \in \mathcal{X}$ such that $p(x) > 0$, there exists a unique $y \in \mathcal{Y}$ such that $p(x, y) > 0$.
   (b) For any function f with domain $\mathcal{X}$, show that $H(f(X)) \le H(X)$.
   [*Hint*: There are many ways of showing this. One way is to consider the two different ways of expanding $H\big(X, f(X)\big)$ using the chain rule.]

2. Problem 2.8, Cover & Thomas, 2nd ed.

3. Problem 2.10, Cover & Thomas, 2nd ed.

4. Let X be a random variable taking values in $\{1, 2, 3, \ldots\}$, with mean $\mathbb{E}[X] = \mu$. Let Z be a geometric random variable also with mean μ, i.e., $\mathbb{E}[Z] = \mathbb{E}[X]$.
   (a) Show that $H(X) \le H(Z)$, with equality iff X has the same distribution as Z.
   [*Hint*: Let $\mathbf{p} = (p_n)$ be the pmf of X, and let $\boldsymbol{\gamma} = (\gamma_n)$ be the pmf of Z. Consider $D(\mathbf{p} \parallel \boldsymbol{\gamma})$.]

   Thus, the maximum entropy among all positive integer-valued random variables with a fixed mean μ is attained by a geometric random variable.

   (b) Prove that $H(X) \le \mathbb{E}[X]$. When does equality hold?
   [*Hint*: Evaluate $H(Z)$.]

5. Problem 2.25, Cover & Thomas, 2nd ed.

6. Find an example of pmfs $p(x)$, $q(x)$ and $r(x)$ defined on the same alphabet $\mathcal{X}$, such that $D(p \parallel q)$, $D(q \parallel r)$ and $D(p \parallel r)$ are all finite and satisfy

$$D(p \parallel q) + D(q \parallel r) < D(p \parallel r).$$

7. A non-negative matrix $W = \big[w_{i,j}\big]$ is called *doubly stochastic* if all its row sums and all its column sums are equal to 1: $\sum_j w_{i,j} = 1$ for all i, and $\sum_j w_{i,j} = 1$ for all j.

   Let $\mathbf{p} = (p_1, \ldots, p_m)$ be a probability mass function, and let W be an $m \times m$ doubly stochastic matrix.

   (a) Show that $\hat{\mathbf{p}} = \mathbf{p}W$ is also a probability mass function.
   (b) Let $\mathbf{q} = (q_1, \ldots, q_m)$ be another pmf. Show that

$$D(\mathbf{p}W \parallel \mathbf{q}W) \le D(\mathbf{p} \parallel \mathbf{q})$$

(c) From (b), deduce that $H(\mathbf{p}W) \geq H(\mathbf{p})$. Here, of course, $H(\mathbf{p}) = -\sum_i p_i \log p_i$, and $H(\mathbf{p}W)$ is the entropy corresponding to $\mathbf{p}W$.

[*Remark*: the result of part (b) is sometimes called the data processing inequality for relative entropy. In fact, your proof of (a) and (b) should only use the fact that the row sums of W are all equal to 1. Doubly stochastic will be needed for part (c).]

8. Let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be discrete random variables.

(a) Show that if $X_1, \ldots, X_n$ are independent rvs, then

$$I(X_1, \ldots, X_n\,;\, Y_1, \ldots, Y_n) \geq \sum_{i=1}^{n} I(X_i; Y_i).$$

Give a necessary-and-sufficient condition for equality to hold above.

(b) Show that if $p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} p(y_i \mid x_i)$, then

$$I(X_1, \ldots, X_n\,;\, Y_1, \ldots, Y_n) \leq \sum_{i=1}^{n} I(X_i; Y_i).$$

Give a necessary-and-sufficient condition for equality to hold above.