Problem Set 4G
Due: March 2, 2005

G0. Omit Question 1 in Problem Set 4.

G1. **Entropy and uncertainty.** Consider a random variable X that can take n values, $x_1, \ldots, x_n$ with corresponding probabilities $p_1, \ldots, p_n$. The **entropy** of X is defined to be

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i,$$

and is a measure of the uncertainty about the experimental value of X. To get a sense of this, note that $H(X) \geq 0$ and that $H(X)$ is very close to 0 when X is "nearly deterministic," i.e., takes one of its possible values with probability very close to 1 (since we have $p \log p \approx 0$ if either $p \approx 0$ or $p \approx 1$). The notion of entropy is fundamental in information theory, which originated with C. Shannon's famous work and is described in many specialized textbooks. For example, it can be shown that $H(X)$ is a lower bound to the average number of yes-no questions (such as "is $x = x_1$?" or "is $x < x_5$?") that must be asked in order to determine the experimental value of X that has occurred. Furthermore, with a suitable strategy for asking questions and assuming a very long string of experimental values of X, the average number of questions required per value can be made as close to $H(X)$ as desired.

(a) Show that if $q_1, \ldots, q_n$ are nonnegative numbers such that $\sum_{i=1}^{n} q_i = 1$, then

$$H(X) \leq -\sum_{i=1}^{n} p_i \log q_i,$$

with equality if and only if $p_i = q_i$ for all i. As a special case, show that

$$H(X) \leq \log n,$$

with equality if and only if $p_i = 1/n$ for all i. *Hint*: Use the inequality $\ln a \leq a - 1$, for $a > 0$, which holds with equality if and only if $a = 1$. (To see this, write $\ln a = \int_1^a b^{-1} db < \int_1^a db = a - 1$ for $a > 1$, and write $\ln a = -\int_a^1 b^{-1} db < -\int_a^1 db = a - 1$ for $0 < a < 1$.)

(b) Let X and Y be random variables taking a finite number of values, and having joint PMF $p_{X,Y}(x, y)$. Define

$$I(X;Y) = \sum_x \sum_y p_{X,Y}(x, y) \log \left(\frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)} \right).$$

Show that $I(X;Y) \geq 0$, and that $I(X;Y) = 0$ if and only if X and Y are independent.

(c) Show that

$$I(X;Y) = H(X) + H(Y) - H(X, Y),$$

where

$$\begin{aligned} H(X,Y) &= -\sum_x \sum_y p_{X,Y}(x, y) \log p_{X,Y}(x, y) \\ H(X) &= -\sum_x p_X(x) \log p_X(x) \\ H(Y) &= -\sum_y p_Y(y) \log p_Y(y). \end{aligned}$$

(d) Show that

$$I(X;Y) = H(X) - H(X \mid Y),$$

where

$$H(X \mid Y) = -\sum_x \sum_y p_Y(y) p_{X|Y}(x \mid y) \log p_{X|Y}(x \mid y).$$

[Note that $H(X \mid Y)$ may be viewed as the conditional entropy of X given Y. If we interpret $I(X;Y)$ as the information about X conveyed by Y, we can use the formula $H(X \mid Y) = H(X) - I(X;Y)$ to view the conditional entropy $H(X \mid Y)$ as the uncertainty about X reduced by the information about X conveyed by Y.]