# Homework 03

**Assigned:** Fri 03 Oct 2008
**Due:** Fri 10 Oct 2008

**Instructions:**

- Feel free to work with others on this assignment. However, you must acknowledge with whom you worked, and you must write up your own solutions.

## Problem 1 [15 pts]: Cover and Thomas, 2nd Edition, Problems 2.2, 2.4, and 2.5.

*Clarification:* For each part of Problem 2.2, state that $H(Y) \sim H(X)$ where $\sim$ is replaced with one of $\{<, \leq, =, \geq, >\}$ and justify your answer.

## Problem 2 [30 pts]: Jensen-Shannon Divergence

Let $p_1$ and $p_2$ be probability distrubutions over a discrete space $\mathcal{X}$ and define the *average* of these distributions, $\overline{p}_{12}$, as follows:

$$\overline{p}_{12}(x) = \frac{p_1(x) + p_2(x)}{2} \ \forall x \in \mathcal{X}.$$

When the underlying distrubutions are clear, we shall drop the subscripts; hence, $\overline{p} = \overline{p}_{12}$.

The Jensen-Shannon divergence between two distributions $p_1$ and $p_2$ is defined as follows:

$$\mathrm{JS}(p_1, p_2) = \frac{D(p_1 \| \overline{p}) + D(p_2 \| \overline{p})}{2}.$$

In other words, the Jensen-Shannon divergence is the *average* of the KL-distances to the *average* distribution.

**i.** Show that

$$\mathrm{JS}(p_1, p_2) = H(\overline{p}) - \frac{H(p_1) + H(p_2)}{2}.$$

In other words, the Jensen-Shannon divergence is the entropy of the *average* minus the *average* of the entropies.

**ii.** Show that

- $\mathrm{JS}(p_1, p_2) \geq 0$,
- $\mathrm{JS}(p_1, p_2) = \mathrm{JS}(p_2, p_1)$, and
- $\mathrm{JS}(p_1, p_2) = 0$ if and only if $p_1 = p_2$.

These are three of the four properties necessary for a metric, the fourth property being triangle inequality. Additionally, show that

- $\mathrm{JS}(p_1, p_2) \leq 1$.

*Hint:* To prove this last property, argue that $D(p_i \| \overline{p}) \leq 1$ for both $i = 1$ and 2, then appeal to the definition of Jensen-Shannon.

**iii.** Let $p_1$, $p_2$, and $p_3$ be distributions over a discrete space $\mathcal{X}$. For Jensen-Shannon to be a metric, it must satisfy the triangle inequality property

$$\mathrm{JS}(p_1, p_2) + \mathrm{JS}(p_2, p_3) \geq \mathrm{JS}(p_1, p_3)$$

in addition to the first three properties described in part (ii) above.

- Show that Jensen-Shannon satisfies the triangle inequality if and only if

$$H(\overline{p}_{12}) + H(\overline{p}_{23}) \geq H(\overline{p}_{13}) + H(p_2).$$

  *Hint:* Use the result from part (i) above.

- Use the above result to show that Jensen-Shannon is *not* a metric by constructing three simple distributions $p_1$, $p_2$, and $p_3$ for which the above inequality does not hold. (Distributions over a discrete space of size two suffice.)

*Aside:* While the Jensen-Shannon divergence is not a metric, it can be shown that the square root of the Jensen-Shannon divergence is a metric. The Jensen-Shannon divergence can also be generalized to allow for weighted averages among distributions and to generate a divergence for an arbitrary number of distributions. The most general form of Jensen-Shannon is as follows: Let $p_1, p_2, \ldots, p_n$ be $n$ distributions over a discrete space $\mathcal{X}$, and let $\vec{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be a distribution over $\{1, 2, \ldots, n\}$. Define the weighted average distribution $\overline{p}(x) = \sum_i \lambda_i p_i(x)\ \forall x \in \mathcal{X}$. Then the generalized Jensen-Shannon divergence among $p_1, p_2, \ldots, p_n$ with respect to $\vec{\lambda}$ is

$$\begin{aligned} \mathrm{JS}(p_1, p_2, \ldots, p_n) &= \sum_i \lambda_i D(p_i \| \overline{p}) \\ &= H(\overline{p}) - \sum_i \lambda_i H(p_i). \end{aligned}$$

**Problem 3** [25 pts]**:** Cover and Thomas, 2nd Edition, Problem 2.9.

*Hint:* Construct the Venn diagram for intuition.

**Problem 4** [30 pts]**:** Cover and Thomas, 2nd Edition, Problem 2.30.

*Hint:* This is a constrained optimization problem that you should solve using Lagrange multipliers. You will have two constraints that need to be simultaneously satisfied: one which ensures that the probabilities form a distribution and a second which ensures that the distribution has a specific mean.

*Hint:* You should obtain $H(X) = (1 + A)\lg(1 + A) - A\lg A$.