JENSEN INEQUALITY BASICS

This is a revision of material that appeared earlier in SEP 37 and reprinted in PVI.

Let f be a function with a positive second derivative. Such a function is called "convex" and satisfies the **inequality**

$$\frac{f(a) \ + \ f(b)}{2} \ - \ f\left(\frac{a+b}{2}\right) \quad \geq \quad 0 \tag{1}$$

Equation (1) relates a function of an average to an average of the function. The average can be weighted, for example,

$$\frac{1}{3} f(a) \ + \ \frac{2}{3} f(b) \ - \ f\left(\frac{1}{3}a + \frac{2}{3}b\right) \quad \geq \quad 0 \tag{2}$$

Figure 1 is a graphical interpretation of equation (2) for the function $f = x^2$. There

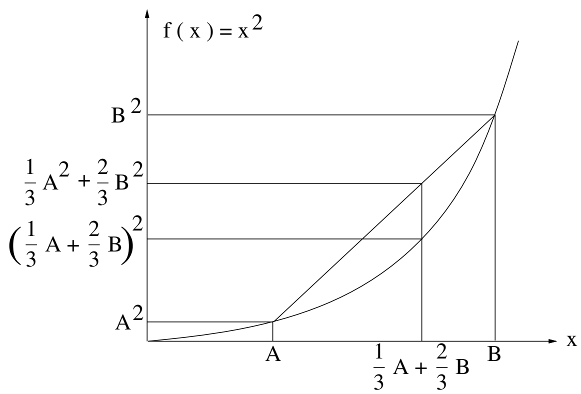

Figure 1: Sketch of $y = x^2$ for interpreting equation (2).

is nothing special about $f = x^2$, except that it is convex. Given three numbers a, b, and c, the inequality (2) can first be applied to a and b, then c with the average of a and b. Thus, recursively, an inequality like (2) can be built for a weighted average of three or more numbers. Define weights $w_j \geq 0$ that are normalized ($\sum_j w_j = 1$). The general result is

$$F(p_j) \quad = \quad \sum_{j=1}^{N} w_j f(p_j) \ - \ f\left(\sum_{j=1}^{N} w_j p_j\right) \quad \geq \quad 0 \tag{3}$$

If all the p_j are the same, say $\bar{p}$, then the two terms in (3) both become $f(\bar{p})$ so the inequality becomes an equality. Thus, minimizing F is like urging all the p_j to be identical. Equilibrium is when F is reduced to the smallest possible value which satisfies any constraints that may be applicable. An experimentalist naturally wonders which $f()$ is best for any particular application. Let's look at some.

Examples of Jensen inequalities

The most familiar example of a Jensen inequality occurs when the weights are all equal to $1/N$ and the convex function is $f(x) = x^2$. In this case the Jensen inequality gives the familiar result that the mean of the squares exceeds the square of the mean:

$$Q \quad = \quad \frac{1}{N}\sum_{i=1}^{N} x_i^2 \; - \; \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2 \quad \geq \quad 0 \tag{4}$$

In many applications the population consists of positive members only, so the function $f(p)$ need have a positive second derivative only for positive values of p. The function $f(p) = 1/p$ yields a Jensen inequality for the **harmonic mean**:

$$H \quad = \quad \sum \frac{w_i}{p_i} \; - \; \frac{1}{\sum w_i p_i} \quad \geq \quad 0 \tag{5}$$

A more important case is the **geometric inequality**. Here $f(p) = -\ln(p)$, and

$$G \quad = \quad -\sum w_i \ln p_i \; + \; \ln \sum w_i p_i \quad \geq \quad 0 \tag{6}$$

The more familiar form of the geometric inequality results from exponentiation and a choice of weights equal to 1/N:

$$\frac{1}{N}\sum_{i=1}^{N} p_i \quad \geq \quad \prod_{i=1}^{N} p_i^{1/N} \tag{7}$$

In other words, the product of square roots of two values is smaller than half the sum of the values.

The function $f(p) = p\ln(p)$ is also convex. That's not obvious, so let us check. First, $f' = 1 + \ln(p)$. Then $f'' = 1/p$, so yes it is convex for $|p| > 0$. The average of the function minus the function of the average is $< p \ln p > - < p > \ln(< p >)$, more explicitly:

$$S_{\text{extrinsic}} \quad = \quad \overline{p \ln p} \; - \; \overline{p} \; \ln \overline{p} \quad \geq \quad 0 \tag{8}$$

$$S_{\text{extrinsic}} \quad = \quad \sum w_i p_i \ln p_i \; - \; \left(\sum w_i p_i\right) \; \ln \sum w_i p_i \quad \geq \quad 0 \tag{9}$$

$$S_{\text{intrinsic}} \quad = \quad \frac{\sum w_i p_i \ln p_i}{\sum w_i p_i} \; - \; \ln \sum w_i p_i \quad \geq \quad 0 \tag{10}$$

This inequality is similar to what we may find in Physics and Information Theory. It might be exactly that, but they tend to use integrals instead of sums, so it is not easy to find it expressed in the "programmer ready form" there. No worries at $p = 0$. The logarithm diverges, but p is stronger so the product $p\ln(p)$ is zero.

DISCUSSION

I first derived the entropy expression in SEP 37 and also published it in PVI. But I derived it the hard way starting from the convex function $f = |r|^{1+\epsilon}$ as $\epsilon \to 0$ while the method I use above was suggested to me by a book "Inequalities," by Hardy, Littlewood, and Polya, Cambridge University Press, 1934 who also have the result in the "programmer ready form" like mine.

Most likely this result is also implicit in the work of Jensen, of Gibbs, or of Shannon, but I haven't seen it there or elsewhere in the self contained "programmer ready form." It belongs in Wikipedia.

Stew Levin agreed to check my algebra, but better than that, he dug up the Hardy book which provides us the much simpler derivation above. Actually, the Hardy gang expressed it for weights that are not necessarily normalized, $\sum_i w_i \neq 1$, which offers a slight programming advantage. For example, initially I omitted weights choosing them identical and inverse to their number. But when time came to knock out the bums, I had to renormalize. Here is the Hardy result in my notation:

$$S \quad = \quad \frac{\sum w_i p_i \ln p_i}{\sum w_i p_i} \; - \; \ln \frac{\sum w_i p_i}{\sum w_i} \quad \geq \quad 0 \tag{11}$$