Jensen inequality: modeling envelopes and spectra

Jon F. Claerbout

Envelopes

Time variable scales, such as $e^{\alpha t}$ or t^{γ} are often applied to seismograms. The choice of the constant α or γ may be made from the data itself. After scaling, the data population should be more uniform sized. Given a uniformity criterion, many values of α or γ can be tested and compared. This paper reviews a family of uniformity criteria based on Jensen inequalities.

The applications extend beyond gain control. Deconvolution amounts to spectral scaling. Going from one dimension to two, seismic data can be scaled in many more ways, not only by offset but also by radial parameter, $r = x/t$, or dip $p = dt/dx$, or in various other spaces. More generally, we are seeking to model all the time and space variable envelopes as well as envelopes in the various spectral domains. The transform to uniformity plays a role in both data display and in gathering statistics.

The Jensen inequality

Let f be a function with a positive second derivative. Such a function is called convex. By graphical interpretation we see that

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

Define weights $w_j \geq 0$ that are normalized $(\sum_j w_j = 1)$. Successive applications of (1) lead to a more general relation with weighted sums.

$$S(p_j) = \sum_{j=1}^{N} w_j f(p_j) - f\left(\sum_{j=1}^{N} w_j p_j\right) \geq 0 \tag{2}$$

If all the p_j are the same, then both of the two terms in S are the same, and S vanishes. So minimizing S is like urging all the p_j to be identical. Equilibrium is attained when S is reduced to the smallest possible value which satisfies any constraints which may be applicable. The function S defined by (2) is like the *entropy* defined in chemical thermodynamics.

Examples of Jensen inequalities

The most familiar example of a Jensen inequality arises 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 that the mean square exceeds the square of the mean.

$$Q \quad = \quad \frac{1}{N} \sum_{i=1}^{N} x_i^2 \; - \; \left[\frac{1}{N} \sum_{i=1}^{N} x_i \right]^2 \quad \geq \quad 0$$

In the rest of the applications that we will consider, the population will consist of positive members, 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.

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

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

$$G \quad = \quad -\sum w_i \ln p_i \; + \; \ln \sum w_i \, p_i \quad \geq \quad 0$$

The more familiar form of the geometric inequality arises from exponentiation and choice of weights equal to $1/N$.

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

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

A Jensen inequality with an adjustable parameter is suggested by $f(p) = p^\gamma$.

$$\Gamma_\gamma \quad = \quad \sum_{i=1}^{N} w_i \, p_i^\gamma \; - \; \left[\sum_{i=1}^{N} w_i \, p_i \right]^\gamma$$

Whether Γ is always positive or always negative depends upon the numerical value of γ. In practice you may see the dimensionless form where the ratio instead of the difference of the two terms is used.

A most important inequality in information theory and thermodynamics is the one based on $f(p) = p^{1+\varepsilon}$ where ε is a small positive number tending to zero. I call this the *weak* inequality. With some calculation we will quickly arrive at the limit.

$$\sum w_i p_i^{1+\varepsilon} \quad \geq \quad \left[\sum w_i p_i\right]^{1+\varepsilon}$$

Take logarithms

$$\ln \sum w_i p_i^{1+\varepsilon} \quad \geq \quad (1+\varepsilon) \ln \sum w_i p_i$$

Expand both sides in a Taylor series in powers of ε using

$$\frac{d}{d\varepsilon} a^u \quad = \quad \frac{du}{d\varepsilon} a^u \ln a$$

The leading term is identical on both sides and may be canceled. Divide both sides by ε and go to the limit $\varepsilon = 0$ getting

$$\frac{\sum w_i p_i \ln p_i}{\sum w_i p_i} \quad \geq \quad \ln \sum w_i p_i$$

We can now define a positive variable S' with or without a positive scaling factor $\sum w\, p$:

$$S'_{intensive} \quad = \quad \frac{\sum w_i p_i \ln p_i}{\sum w_i p_i} \quad - \quad \ln \sum w_i p_i \quad \geq \quad 0$$

$$S'_{extensive} \quad = \quad \sum w_i p_i \ln p_i \quad - \quad \left[\sum w_i p_i\right] \ln \left[\sum w_i p_i\right] \quad \geq \quad 0$$

Seismograms often contain zeros and gaps. Notice that a single zero p_i can upset the harmonic H or geometric G inequality, but a single zero has no major effect on S or Γ.

Priors and posteriors

You may have a prior distribution and a posterior distribution. Denote the prior by b (for before) and posterior by a (for after). Defining $p_i = a_i / b_i$ we try to uniformize the population of p_i thereby asking posteriors to tend to priors to the extent that the adjustable parameters will permit.

Definition of Jensen average

In minimizing some definition of entropy, the small values will tend to get larger while the large values tend to get smaller. For each population of values there is an average value, i.e. a value that neither tends to get larger nor smaller. The average depends not only on the population, but also on the definition of entropy.

Commonly, the p_j are positive and $\sum w_j p_j$ is a sort of an energy. Typically the total energy will be fixed. This can be included as a constraint, or we can find some other function to minimize. For example, divide both both terms in (2) by the second term and get an expression which is scale invariant, i.e. scaling p leaves (3) unchanged.

$$\frac{\sum_{j=1}^{N} w_j\, f(p_j)}{f\left[\sum_{j=1}^{N} w_j\, p_j\right]} \quad \geq \quad 1 \tag{3}$$

Because the expression exceeds unity we are tempted to take a logarithm and make a new function for minimization.

$$J \;=\; \log_e\left[\sum_{j=1}^{N} w_j\, f(p_j)\right] \;-\; \log_e\left[f\left[\sum_{j=1}^{N} w_j\, p_j\right]\right] \;\geq\; 0 \tag{4}$$

Given a population p_j of positive variates, and an inequality like (4), I am now prepared to define what I'll call the Jensen average $\bar{p}$. Suppose there is one element, say p_J, of the population p_j that may be given a first order perturbation, and only a second order perturbation in J will result. Such an element is in equilibrium and will be called the Jensen average $\bar{p}$.

$$0 \;=\; \left.\frac{\partial J}{\partial p_J}\right)_{p_J = \bar{p}} \tag{5}$$

Let f_p denote the derivative of f with respect to its argument. Inserting (4) into (5)

$$0 \;=\; \frac{\partial J}{\partial p_J} \;=\; \frac{w_J\, f_p(p_J)}{\sum_{j=1}^{N} w_j\, f(p_j)} \;-\; \frac{f_p(\sum_{j=1}^{N} w_j\, p_j)\, w_J}{f\left[\sum_{j=1}^{N} w_j\, p_j\right]} \tag{6}$$

Solving

$$\bar{p} \;=\; p_J \;=\; f_p^{-1}\left[e^J\, f_p(\sum_{j=1}^{N} w_j\, p_j)\right] \tag{7}$$

But where do we get the function f and what do we say about the equilibrium value? Maybe we can somehow derive f from the population. If we can't work out a general theory, perhaps we can at least find the constant γ assuming the functional form to be $f = p^{\gamma}$.

Additivity of envelope entropy to spectral entropy

In some of my efforts to fill in missing data with entropy criteria, I often based the entropy on the spectrum and then found that the envelope would misbehave. I came to believe that the definition of entropy should involve both the spectrum and the envelope. To get started, let us assume that the power of a seismic signal is a product of an envelope function times a spectral function. Say

$$u(\omega,t) \quad = \quad p(\omega)\, e(t) \tag{8}$$

Notice that this separability assumption resembles the stationarity concept. I am not defending the assumption (8), only suggesting that it is an improvement over each term separately. Let us examine some of the algebraic consequences. First evaluate the intensive entropy.

$$S'_{intensive} \quad = \quad = \quad \frac{\sum_t \sum_\omega u \ln u}{\sum_t \sum_\omega u} \; - \; \ln \frac{1}{N^2} \sum_t \sum_\omega u \quad \geq \quad 0$$

$$= \quad \frac{\sum \sum p\, e\, (\ln p + \ln e)}{(\sum p)(\sum e)} \; - \; \ln \left[\frac{1}{N} \sum_\omega p \;\; \frac{1}{N} \sum_t e \right]$$

$$= \quad \frac{\sum e \sum p \ln p \; + \; \sum p \sum e \ln e}{(\sum p)(\sum e)} \; - \; \ln \frac{1}{N} \sum p \; - \; \ln \frac{1}{N} \sum e$$

$$= \quad \left[\frac{\sum p \ln p}{\sum p} \; - \; \ln \frac{1}{N} \sum p \right] \; + \; \left[\frac{\sum e \ln e}{\sum e} \; - \; \ln \frac{1}{N} \sum e \right]$$

$$= \quad S(p) \; + \; S(e)$$

We have not even used the fact that $\sum p = \sum e$.

Now we will tackle the same calculation with the geometric inequality.

$$G \quad = \quad \ln \frac{1}{N^2} \sum \sum u \; - \; \frac{1}{N^2} \sum \sum \ln u$$

$$= \quad \ln \left[\left[\frac{1}{N} \sum_t e \right] \left[\frac{1}{N} \sum_\omega p \right] \right] \; - \; \frac{1}{N^2} \sum_t \sum_\omega (\ln p_\omega + \ln e_t)$$

$$= \ln \bar{e} + \ln \bar{p} - \frac{1}{N^2} \sum_t 1_t \sum_\omega \ln p_\omega - \frac{1}{N^2} \sum_\omega 1_\omega \sum_t \ln e_t$$

$$= \ln \bar{e} + \ln \bar{p} - \frac{1}{N} \sum_\omega \ln p - \frac{1}{N} \sum_t \ln e$$

$$= G(t) + G(\omega)$$

This result seems surprising and I wonder if it applies for any other Jensen inequalities.

Missing data studies by means of least squares always seem to make data envelopes that are too small. Perhaps these combined spectral/envelope criteria will be an improvement.

REFERENCES

Wiggins, R.A. 1977, Minimum entropy deconvolution: *Proceedings of the International Symposium on Computer-Aided Seismic Analysis and Discrimination*, June 9-10, Falmouth, Massachusetts, IEEE Computer Society: p. 7-14.

Grey, W.C, 1978, Norm exponential gain estimation: SEP-15, p. 183.

Claerbout, J.F., 1980, Convex inequalities and statistical mechanics: essay #7 from *Seven Essays on Minimum Entropy:* SEP-24, p. 157.

Claerbout, J.F., 1982, Envelope sensing decon: SEP-30, p. 131.