Mean of a Dataset with Distinct Variances

Bill Menke, February 15, 2013

Assumption. Each datum d_i is drawn from a different Normal p.d.f., $p(d_i)$. These p.d.f.'s are uncorrelated, have distinct (and known) variances, s_i^2, but the same (and unknown) mean, m.

Estimate of the mean and its variance. The model equation is based on the statement that each datum equals the mean, $d_i=m$, with each row of weighted by its certainty, s_i^{-1}:

$$\mathbf{F}\mathrm{m} = \mathbf{f} \quad \text{or} \quad \begin{bmatrix} s_1^{-1} \\ \dots \\ s_N^{-1} \end{bmatrix} m = \begin{bmatrix} s_1^{-1} d_1 \\ \dots \\ s_N^{-1} d_N \end{bmatrix}$$

Note that this equation is normalized, in the sense that the covariance $\mathbf{C_f} = \mathbf{I}$. Both the generalized least-squares method and the maximum likelihood method lead to the same equation for m^{est}, namely:

$$\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathrm{m}^{\mathrm{est}} = \mathbf{F}^{\mathrm{T}}\mathbf{f}$$

$$\begin{bmatrix} s_1^{-1} & \dots & s_N^{-1} \end{bmatrix} \begin{bmatrix} s_1^{-1} \\ \dots \\ s_N^{-1} \end{bmatrix} m^{est} = \begin{bmatrix} s_1^{-1} & \dots & s_N^{-1} \end{bmatrix} \begin{bmatrix} s_1^{-1} d_1 \\ \dots \\ s_N^{-1} d_N \end{bmatrix}$$

This equation has solution

$$m^{est} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}\mathbf{f} \ \text{ or } m^{est} = \left(\sum_{i=1}^{N} s_i^{-2} \right)^{-1} \sum_{i=1}^{N} s_i^{-2}\, d_i$$

Note that the estimated mean is a linear function of the data, with the form $m^{est} = \mathbf{Mf}$, with $\mathbf{M} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}$. By the standard rule of error propagation, variance of m^{est} is:

$$\mathrm{var}(m^{est}) = \mathbf{M}\mathbf{C_f}\mathbf{M}^{\mathbf{T}} =$$

$$\{[\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}\}\mathbf{C_f}\{[\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}\}^{\mathbf{T}} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-\mathbf{1}} = \left(\sum_{i=1}^{N} s_i^{-2}\right)^{-1}$$

(since $\mathbf{C_f} = \mathbf{I}$).

If all the variances are equal, $s_i = s$, and these equations reduces to:

$$m^{est} = \left(\sum_{i=1}^{N} s^{-2}\right)^{-1} \sum_{i=1}^{N} s^{-2}\, d_i \approx N^{-1} \sum_{i=1}^{N} d_i$$

$$\mathrm{var}(m^{est}) = \left(\sum_{i=1}^{N} s^{-2}\right)^{-1} \approx \frac{s^2}{N}$$

which are the usual formulas for the estimated mean and its variance.

If one datum, say d_k, has a variance that is much smaller than all the others, then:

$$m^{est} = \left(s_k^{-2} + \sum_{i \neq k}^{N} s_i^{-2}\right)^{-1} \left(s_k^{-2} d_k + \sum_{i \neq k}^{N} s_i^{-2}\, d_i\right) \approx d_k$$

$$\mathrm{var}(m^{est}) = \left(s_k^{-2} + \sum_{i \neq k}^{N} s_i^{-2}\right)^{-1} \approx s_k^2$$

That is, only the most certain datum counts.