Mean of dataset with distinct variances

for P. Pollisar 02-15-13

Assumption. Each datum d_i is drawn from a different p.d.f., $p(d_i)$. These p.d.f.'s are uncorrelated, have distinct variances, s_i^2, but the same 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}:

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

Note that this equation is normalized, in the sense that the covariance $\mathbf{C_f} = \mathbf{I}$. The generalized least-squares equation is:

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

Which has solution

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

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

$$\mathrm{var}(m^{est}) = \mathbf{M}\mathbf{C}_\mathrm{f}\mathbf{M}^\mathrm{T} =$$

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

(since $\mathbf{C}_\mathrm{f} = \mathbf{I}$).