Small Subset Correlations

Frank Porter
Caltech
August 15, 2002

We are interested in the case where we do a scan over several energy points, measuring two cross sections:

1. A "total" cross section.
2. A small (*e. g.*, exclusive channel) subset of the total cross section.

Let N_{Tj} be the total number of events observed at scan point j, and N_{Ej} be the number of events observed in the subset reactions at scan point j. We investigate the sampling distribution, and show that if $N_{Tj} \gg N_{Ej}$, it may be all right to neglect the correlations between the two samples.

The sampling probability distribution, at scan point j, is

$$P(N_{Tj}, N_{Ej}; \theta_{Tj}, \theta_{Ej}) = \frac{(\theta_{Tj} - \theta_{Ej})^{N_{Tj}-N_{Ej}}}{(N_{Tj} - N_{Ej})!} \exp(-\theta_{Tj} + \theta_{Ej}) \frac{\theta_{Ej}^{N_{Ej}}}{N_{Ej}!} \exp(-\theta_{Ej}).$$

If the normal approximation is acceptable, then the sampling distribution is

$$P(N_{Tj}, N_{Ej}; \theta_{Tj}, \theta_{Ej}) = \frac{1}{\sqrt{2\pi(\theta_{Tj} - \theta_{Ej})}} \exp\left[-\frac{(N_{Tj} - N_{Ej} - \theta_{Tj} + \theta_{Ej})^2}{2(\theta_{Tj} - \theta_{Ej})}\right] \frac{1}{\sqrt{2\pi\theta_{Ej}}} \exp\left[-\frac{(N_{Ej} - \theta_{Ej})^2}{2\theta_{Ej}}\right].$$

We may rewrite this in the form:

$$P(N_{Tj}, N_{Ej}; \theta_{Tj}, \theta_{Ej}) = A \exp\left[-\frac{1}{2}\left(N_{Tj} - \theta_{Tj}, N_{Ej} - \theta_{Ej}\right) M^{-1} \begin{pmatrix} N_{Tj} - \theta_{Tj} \\ N_{Ej} - \theta_{Ej} \end{pmatrix}\right],$$

where the inverse moment matrix is:

$$M^{-1} = \begin{pmatrix} \frac{1}{\theta_{Tj}-\theta_{Ej}} & -\frac{1}{\theta_{Tj}-\theta_{Ej}} \\ -\frac{1}{\theta_{Tj}-\theta_{Ej}} & \frac{1}{\theta_{Tj}-\theta_{Ej}} + \frac{1}{\theta_{Ej}} \end{pmatrix}.$$

If $\theta_{Tj} \gg \theta_{Ej}$ then we have the approximate matrix:

$$M^{-1} \approx \frac{1}{\theta_{Tj}} \begin{pmatrix} 1 & -1 \\ -1 & \frac{\theta_{Tj}}{\theta_{Ej}} \end{pmatrix}.$$

In this approximation, the sampling distribution is:

$$P(N_{Tj}, N_{Ej}; \theta_{Tj}, \theta_{Ej}) = A \exp\left\{-\frac{1}{2}\left[\frac{(N_{Tj} - \theta_{Tj})^2}{\theta_{Tj}} + \frac{(N_{Ej} - \theta_{Ej})^2}{\theta_{Ej}} - 2\frac{(N_{Tj} - \theta_{Tj})(N_{Ej} - \theta_{Ej})}{\theta_{Tj}}\right]\right\}.$$

A chi-square fit would minimize

$$\chi^2 = \sum_j \left[(N_{Tj} - \theta_{Tj})^2/\theta_{Tj} + (N_{Ej} - \theta_{Ej})^2/\theta_{Ej} - 2(N_{Tj} - \theta_{Tj})(N_{Ej} - \theta_{Ej})/\theta_{Tj}\right].$$

Let us estimate the relative importance of the correlation term. The sampling fluctuations will be such that:

$$N_{Tj} - \theta_{Tj} \sim \mathcal{O}(\sqrt{\theta_{Tj}})$$
$$N_{Ej} - \theta_{Ej} \sim \mathcal{O}(\sqrt{\theta_{Ej}}).$$

Thus, the first two terms in the χ^2 expression will typically be of order 1, for each j. The correlation term is of order:

$$2(N_{Tj} - \theta_{Tj})(N_{Tj} - \theta_{Tj})/\theta_{Tj} \sim \mathcal{O}\left(2\sqrt{\theta_{Ej}/\theta_{Tj}}\right).$$

This is typically small compared with the first two terms, in the limit $\theta_{Tj} \gg \theta_{Ej}$.