Scale Correlations in a χ^2 Fit

Frank Porter
Caltech
October 11, 2001
Expanded March 21, 2002
Current version March 29, 2002

Introduction

We are sometimes faced with the problem of extracting several parameters from a cross section scan as a function of energy, for several simultaneous channels. The primary data are in the form:

$$N_{ij}, \quad i = 1, 2, \ldots, n_{ch}; \quad j = 1, 2, \ldots, n_{pt}, \tag{1}$$

where n_{ch} is the number of channels, and n_{pt} is the number of scan points. We assume here that the numbers of events are large enough that the normal approximation is valid, and that a χ^2 fit may be entertained. We assume that the N_{ij} are uncorrelated, although correlations can be handled if the correlation matrix is known. The χ^2 to be minimized, with respect to parameters θ is:

$$\chi^2 = \sum_{j=1}^{n_{pt}} \sum_{i=1}^{n_{ch}} \left[\frac{N_{ij} - f_{ij}(\theta)}{\sigma_{ij}} \right]^2. \tag{2}$$

A problem that may arise in such a fit is that there are additional quantities involved in relating the parameters of interest to the measurements, and these additional quantities may have uncertainties. Common examples are uncertainties in the acceptance and in the luminosity. Some of these uncertainties (*e. g.*, luminosity) will be common to all channels, and some (*e. g.*, acceptance) will be common to all scan points. We thus have the situation where we should write:

$$f_{ij} = \alpha_i \beta_j f_{ij}, \tag{3}$$

where we normalize α and β to have the measured values 1, but with some uncertainty. The question is, how should we include these correlated uncertainties in the fit, and the error estimate? The reason the answer is not completely straight forward is that the σ_{ij} also depend on the additional quantities.

We will assume for discussion a model in which the uncertainties σ_{ij} scale with α and β:

$$\sigma_{ij} = \alpha_i \beta_j \sigma_{ij}. \tag{4}$$

Note that this is probably not the right model when we are dealing with numbers of events as above, where we should use $\sigma_{ij} = \sqrt{\alpha_i \beta_j}\sigma_{ij}$. However, we are only illustrating the idea here. The last section deals with a more realistic situation. Let $\delta_{\alpha i}$ be the uncertainty in α_i, and $\delta_{\beta j}$ be the uncertainty in β_j.

Likelihood Function Analysis

We assume that the likelihood function may be written, up to unessential constants:

$$\begin{aligned}\mathcal{L}(\theta; \{N_{ij}, \alpha_i, \beta_j\}) = &\left\{ \prod_{i=1}^{n_{pt}} \exp\left[-\frac{1}{2} \left(\frac{1-\alpha_i}{\delta_{\alpha i}} \right)^2 \right] \right\} \left\{ \prod_{j=1}^{n_{ch}} \exp\left[-\frac{1}{2} \left(\frac{1-\beta_j}{\delta_{\beta j}} \right)^2 \right] \right\} \\ &\prod_{j=1}^{n_{pt}} \prod_{i=1}^{n_{ch}} \frac{1}{\alpha_i \beta_j} \exp\left\{ -\frac{1}{2} \left[\frac{N_{ij} - \alpha_i \beta_j f_{ij}(\theta)}{\alpha_i \beta_j \sigma_{ij}} \right]^2 \right\}.\end{aligned} \tag{5}$$

Taking the logarithm gives:

$$-2\ln\mathcal{L} = \sum_{j=1}^{n_{pt}}\sum_{i=1}^{n_{ch}}\left[\frac{N_{ij}-\alpha_i\beta_j f_{ij}(\theta)}{\alpha_i\beta_j\sigma_{ij}}\right]^2 + \sum_{i=1}^{n_{ch}}\left[\left(\frac{1-\alpha_i}{\delta_{\alpha i}}\right)^2 + 2n_{pt}\ln\alpha_i\right] + \sum_{j=1}^{n_{pt}}\left[\left(\frac{1-\beta_j}{\delta_{\beta j}}\right)^2 + 2n_{ch}\ln\beta_j\right]. \tag{6}$$

This function could be minimized as is, with respect to θ, α_i, and β_j. In the example discussed by D'Agostini,[1] the logarithmic terms are not present, but the formulas otherwise correspond.

Adding Further Complications

The discussion so far carries the assumption that the mix of event "types" does not depend on scan point, so that the uncertainty on the efficiencies can be taken to apply at all points. However, there may be contributions to each scan point from continuum and from resonance processes, and the characteristics of the events may differ depending on the source. As the mix of continuum and resonance may vary over the scan poiints, so will the overall efficiencies. Thus, we'll develop the formalism for this situation. At the same time, we should incorporate the treatment of backgrounds.

We still assume that at scan point j and channel i we observe N_{ij} events. That is, we assume that we do not attempt to make any continuum/resonance separation at each scan point. The relative amount of continuum and resonance will be varied in the fit. The cross section at scan point j for channel i is:

$$\sigma_{ij} = \sigma^r_{ij} + \sigma^c_{ij}, \tag{7}$$

where we use r to denote the resonance contribution, and c to denote the continuum contribution.*

We define a number of quantities and assumptions:

1. Let N_{ij} be the total number of events observed at scan point j and channel i. This is assumed to be sampled from a probability distribution which may be characterized by the mean expected ($\langle N_{ij}\rangle \equiv N^0_{ij}$), and the standard deviation of the distribution ($\delta_{N_{ij}}$). These distribution parameters are related to the parameters of interest in the measurement, as well as to other parameters describing the backgrounds, efficiencies, and luminosities.
2. Let N^b_{ij} be the estimated number of background events in the dataset at scan point j and channel i. This is assumed to be sampled from a probability distribution which may be characterized by the mean expected ($\langle N^b_{ij}\rangle \equiv N^{b0}_{ij}$), and the standard deviation of the distribution ($\delta_{N^b_{ij}}$). Note that the estimated number of signal events is thus: $N^s_{ij} = N_{ij} - N^b_{ij}$. We assume that $\delta_{N^b_{ij}}$ is known.
3. Let $\mathcal{L}_j$ be the estimated integrated luminosity at scan point j. This is assumed to be sampled from a probability distribution which may be characterized by the mean expected ($\langle\mathcal{L}_j\rangle \equiv \mathcal{L}^0_j$), and the standard deviation of the distribution ($\delta_{\mathcal{L}_j}$). We assume that $\delta_{\mathcal{L}_j}$ is known.
4. Let ϵ^r_i be the estimated efficiency for resonance events in channel i. This is assumed to be sampled from a probability distribution which may be characterized by the mean expected ($\langle\epsilon^r_i\rangle \equiv \epsilon^{r0}_i$), and the standard deviation of the distribution ($\delta_{\epsilon^r_i}$). We assume that $\delta_{\epsilon^r_i}$ is known.
5. Let ϵ^c_i be the estimated efficiency for continuum events in channel i. This is assumed to be sampled from a probability distribution which may be characterized by the mean expected ($\langle\epsilon^c_i\rangle \equiv \epsilon^{c0}_i$), and the standard deviation of the distribution ($\delta_{\epsilon^c_i}$). We assume that $\delta_{\epsilon^c_i}$ is known.
6. Let θ stand for the parameters of interest to us, *i. e.*, the resonance partial widths, *etc.*

We assume that the overall sampling distribution for the "experiment" may be factored according to:

[1] G. D'Agostini, "On the use of the covariance matrix to fit correlated data", *Nucl. Inst. Meth. in Phys. Res. A*, **346** (1994) 306-311.

* We could, of course, extend this discussion to more than two categories of event types.

$$
\begin{aligned}
&P\left(\{N\},\{N^b\},\{\mathcal{L}\},\{\epsilon^r\},\{\epsilon^c\};\{\theta\},\{N^{b0}\},\{\mathcal{L}^0\},\{\epsilon^{r0}\},\{\epsilon^{c0}\}\right) = \\
&\qquad = \left\{\prod_{i=1}^{n_{ch}}\prod_{j=1}^{n_{pt}} P_N(N_{ij};N_{ij}^0,\delta_{N_{ij}})P_{N^b}(N_{ij}^b;N_{ij}^{bo},\delta_{N_{ij}^b})\right\} \\
&\qquad \times \left\{\prod_{i=1}^{n_{ch}} P_{\epsilon^r}(\epsilon_i^r;\epsilon_i^{r0},\delta_{\epsilon_i^r})P_{\epsilon^c}(\epsilon_i^c;\epsilon_i^{c0},\delta_{\epsilon_i^c})\right\} \\
&\qquad \times \left\{\prod_{j=1}^{n_{pt}} P_{\mathcal{L}}(\mathcal{L}_j;\mathcal{L}_j^0,\delta_{\mathcal{L}_j})\right\}. \qquad (8)
\end{aligned}
$$

We are implicitly assuming that there are no further correlations among the sampled quantities, though such additional correlations could also be incorporated along similar lines or in terms of full covariance matrices, depending on the situation. For example, the uncertainties in ϵ_i^r and ϵ_i^c might be correlated. This correlation may be included with 2×2 covariance matrices in the Gaussian treatment below.

Given an experiment, the probability function above may be turned into a likelihood function by evaluating at the sampled values of the random variables and considering as a function of the parameters. This (or its logarithm) may then be maximized over the parameter space to obtain a "best" fit, *etc.*. To make sure it is clear, the sampled quantities are $\{N\},\{N^b\},\{\mathcal{L}\},\{\epsilon^r\},\{\epsilon^c\}$, and the parameters to be estimated in the fit are $\{\theta\},\{N^{b0}\},\{\mathcal{L}^0\},\{\epsilon^{r0}\},\{\epsilon^{c0}\}$. If n_θ is the number of parameters $\{\theta\}$, then the number of degrees of freedom is $n_{ch}n_{pt}-n_\theta$.

It should probably be stressed that the "errors" (δ) correspond to the "statistical" uncertainties on the relevant quantities. Systematic effects, such as model dependencies, should be treated as a separate study, not as part of the fit procedure. Likewise, overall scale uncertainties, *e. g.*, in the luminosity measurement are best treated as separate systematic uncertainties in the results.

To be more explicit, let us suppose that it is a good approximation to treat all of the probability distributions as Gaussians. Then the logarithm (times -2) of the likelihood may be written, up to unessential constants:

$$
\begin{aligned}
-2\ln\mathcal{L} = &\sum_{i=1}^{n_{ch}}\sum_{j=1}^{n_{pt}}\left[\left(\frac{N_{ij}-N_{ij}^0}{\delta_{N_{ij}}}\right)^2 + 2\ln\delta_{N_{ij}} + \left(\frac{N_{ij}^b-N_{ij}^{b0}}{\delta_{N_{ij}^b}}\right)^2 + 2\ln\delta_{N_{ij}^b}\right] \\
&+\sum_{i=1}^{n_{ch}}\left[\left(\frac{\epsilon_i^r-\epsilon_i^{r0}}{\delta_{\epsilon_i^r}}\right)^2 + 2\ln\delta_{\epsilon_i^r} + \left(\frac{\epsilon_i^c-\epsilon_i^{c0}}{\delta_{\epsilon_i^c}}\right)^2 + 2\ln\delta_{\epsilon_i^c}\right] \\
&+\sum_{j=1}^{n_{pt}}\left[\left(\frac{\mathcal{L}_j-\mathcal{L}_j^0}{\delta_{\mathcal{L}_j}}\right)^2 + 2\ln\delta_{\mathcal{L}_j}\right]. \qquad (9)
\end{aligned}
$$

Since we are assuming that $\delta_{N_{ij}^b}$, $\delta_{\epsilon_i^r}$, $\delta_{\epsilon_i^c}$, and $\delta_{\mathcal{L}_j}$ are known (*i. e.*, do not depend on any of our fit parameters), the above expression may be simplified to:

$$
\begin{aligned}
-2\ln\mathcal{L} = &\sum_{i=1}^{n_{ch}}\sum_{j=1}^{n_{pt}}\left[\left(\frac{N_{ij}-N_{ij}^0}{\delta_{N_{ij}}}\right)^2 + 2\ln\delta_{N_{ij}} + \left(\frac{N_{ij}^b-N_{ij}^{b0}}{\delta_{N_{ij}^b}}\right)^2\right] \\
&+\sum_{i=1}^{n_{ch}}\left[\left(\frac{\epsilon_i^r-\epsilon_i^{r0}}{\delta_{\epsilon_i^r}}\right)^2 + \left(\frac{\epsilon_i^c-\epsilon_i^{c0}}{\delta_{\epsilon_i^c}}\right)^2\right] \\
&+\sum_{j=1}^{n_{pt}}\left(\frac{\mathcal{L}_j-\mathcal{L}_j^0}{\delta_{\mathcal{L}_j}}\right)^2. \qquad (10)
\end{aligned}
$$

We rewrite the expected numbers of events according to:

$$N^0_{ij} = \mathcal{L}^0_j \left[\sigma^r_{ij}(\theta)\epsilon^{r0}_i + \sigma^c_{ij}(\theta)\epsilon^{c0}_i\right] + N^{b0}_{ij}. \tag{11}$$

If we assume that the N_{ij} sampling distribution is the Gaussian approximation to a Poisson distribution, then we also have:

$$\delta^2_{N_{ij}} = \mathcal{L}^0_j \left[\sigma^r_{ij}(\theta)\epsilon^{r0}_i + \sigma^c_{ij}(\theta)\epsilon^{c0}_i\right] + N^{b0}_{ij}. \tag{12}$$

This completes the outline of the procedure which may be employed, which takes into account the correlated uncertainties, under the stated assumptions.