CDF/MEMO/STATISTICS/PUBLIC/8661
April 28, 1999

Equivalence of the best-fit and covariance-matrix methods for comparing binned data with a model in the presence of correlated systematic uncertainties

Luc Demortier

Laboratory of Experimental High-Energy Physics
The Rockefeller University

Abstract

Two methods are known for comparing binned data to predictions in the presence of correlated systematic uncertainties. The first method uses a chisquared expression with a covariance matrix that incorporates statistical and systematic uncertainties and their correlations. For the second method, known as best-fit, one writes down a chisquared with diagonal covariance matrix and one fit parameter for each systematic uncertainty. This chisquared is then minimized with respect to the free parameters. In this note we show that the two methods are equivalent.

1 Introduction

In the best-fit formulation one minimizes the following χ^2:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(d_i - t_i - \sum_{j=1}^{K} \alpha_j s_{ji}\right)^2}{\sigma_i^2} + \sum_{j=1}^{K} \alpha_j^2, \tag{1}$$

where:

- d_i = content of data bin i;
- t_i = model prediction for bin i;
- s_{ji} = systematic uncertainty from source j on the contents of bin i;
- σ_i = statistical uncertainty on the contents of bin i;
- α_j = fit parameter;
- N = number of bins;
- K = number of sources of systematic uncertainties.

The purpose of this note is to calculate the value $\chi^2_{\rm min}$ of the above χ^2 at the minimum with respect to all the α_k, and to show that it can be written as:

$$\chi^2_{\rm min} \;=\; \Delta^T C^{-1} \Delta, \tag{2}$$

where Δ is a column matrix with elements $d_i - t_i$ and C is the covariance matrix of the measurements d_i, taking into account statistical and systematic uncertainties as well as their correlations. The only assumption we will need to make is that the χ^2 of equation (1) has a unique analytical minimum.

2 Matrix Algebra

Calculations can be made more transparent by working with matrices instead of scalars. Define the following quantities:

- a $K \times 1$ matrix α: $\alpha_k \equiv k^{th}$ fit parameter,
- a $K \times N$ matrix R: $R_{ji} \equiv \frac{s_{ji}}{\sigma_i}$,
- an $N \times 1$ matrix u: $u_i \equiv \frac{d_i - t_i}{\sigma_i}$.

With this notation the χ^2 can be rewritten as:

$$\begin{aligned}
\chi^2 &= \left(u - R^T\alpha\right)^T \left(u - R^T\alpha\right) + \alpha^T\alpha && (3)\\
&= \left(u^T - \alpha^T R\right)\left(u - R^T\alpha\right) + \alpha^T\alpha && (4)\\
&= u^T u - u^T R^T \alpha - \alpha^T R u + \alpha^T R R^T \alpha + \alpha^T \alpha && (5)\\
&= u^T u - 2\,\alpha^T R u + \alpha^T R R^T \alpha + \alpha^T \alpha. && (6)
\end{aligned}$$

The minimum of the χ^2 satisfies $\partial\chi^2/\partial\alpha_k = 0$, or:

$$-Ru + RR^T\alpha_{\rm min} + \alpha_{\rm min} \;=\; 0. \tag{7}$$

Left-multiplying this equation by $\alpha_{\rm min}^T$ yields:

$$-\alpha_{\rm min}^T Ru + \alpha_{\rm min}^T RR^T\alpha_{\rm min} + \alpha_{\rm min}^T\alpha_{\rm min} \;=\; 0, \tag{8}$$

which shows that the last three terms of equation (5) cancel at the minimum:

$$\chi^2_{\rm min} \;=\; u^T u - u^T R^T \alpha_{\rm min}. \tag{9}$$

Next, we solve equation (7) for $\alpha_{\rm min}$:

$$\alpha_{\rm min} \;=\; \left(1_K + RR^T\right)^{-1} Ru, \tag{10}$$

where 1_K is the unit matrix in $K \times K$ dimensions. The existence of an inverse for the matrix $(1_K + RR^T)$ follows from the assumption that the χ^2 of equation (1) has a unique analytical minimum. Before going any further, we need to show that if $(1_K + RR^T)$ is invertible, then so is $(1_N + R^T R)$:

$$\left(1_N + R^T R\right)^{-1} \;=\; 1_N - R^T\left(1_K + RR^T\right)^{-1} R \tag{11}$$

Indeed:

$$\begin{aligned}
&\left(1_N + R^T R\right)\left[1_N - R^T\left(1_K + RR^T\right)^{-1} R\right] \\
&= 1_N + R^T R - \left(1_N + R^T R\right) R^T \left(1_K + RR^T\right)^{-1} R \\
&= 1_N + R^T R - \left(R^T + R^T RR^T\right)\left(1_K + RR^T\right)^{-1} R \\
&= 1_N + R^T R - R^T\left(1_K + RR^T\right)\left(1_K + RR^T\right)^{-1} R \\
&= 1_N + R^T R - R^T R \\
&= 1_N
\end{aligned}$$

The equality of the left-inverse and the right-inverse follows from the symmetry of the matrix $(1_N + R^T R)$. Having proved the existence of $(1_N + R^T R)^{-1}$, we can rewrite the expression for $\alpha_{\rm min}$ as follows:

$$\begin{aligned}
\alpha_{\rm min} &= \left(1_K + RR^T\right)^{-1} R\left(1_N + R^T R\right)\left(1_N + R^T R\right)^{-1} u, &(12)\\
&= \left(1_K + RR^T\right)^{-1}\left(R + RR^T R\right)\left(1_N + R^T R\right)^{-1} u, &(13)\\
&= \left(1_K + RR^T\right)^{-1}\left(1_K + RR^T\right) R\left(1_N + R^T R\right)^{-1} u, &(14)\\
&= R\left(1_N + R^T R\right)^{-1} u. &(15)
\end{aligned}$$

Substituting this expression in (9) yields:

$$\begin{aligned}
\chi^2_{\rm min} &= u^T u - u^T R^T R \left(1_N + R^T R\right)^{-1} u, && (16)\\
&= u^T \left[1_N - R^T R \left(1_N + R^T R\right)^{-1}\right] u, && (17)\\
&= u^T \left[\left(1_N + R^T R\right) - R^T R\right] \left(1_N + R^T R\right)^{-1} u, && (18)\\
&= u^T \left(1_N + R^T R\right)^{-1} u. && (19)
\end{aligned}$$

Finally, define:

- an $N \times 1$ matrix Δ: $\Delta_i \equiv d_i - t_i$,
- an $N \times N$ matrix S: $S_{ij} \equiv \sigma_i \delta_{ij}$,

so that:

$$u = S^{-1}\Delta \tag{20}$$

and

$$\begin{aligned}
\chi^2_{\rm min} &= \Delta^T S^{-1T} \left[1_N + R^T R\right]^{-1} S^{-1}\Delta, && (21)\\
&= \Delta^T \left[S \left(1_N + R^T R\right) S^T\right]^{-1} \Delta, && (22)\\
&= \Delta^T C^{-1} \Delta, && (23)
\end{aligned}$$

with:

$$C \equiv S \left(1_N + R^T R\right) S^T. \tag{24}$$

The elements of the matrix C are:

$$\begin{aligned}
C_{ij} &= \sum_{k=1}^{N} \sum_{l=1}^{N} \sigma_i \delta_{ik} \left(\delta_{kl} + \sum_{m=1}^{K} \frac{s_{mk}}{\sigma_k} \frac{s_{ml}}{\sigma l}\right) \sigma_l \delta_{lj}, && (25)\\
&= \sigma_i^2 \delta_{ij} + \sum_{m=1}^{K} s_{mi} s_{mj}. && (26)
\end{aligned}$$

Hence C is simply the covariance matrix of the measurements.

3 Conclusion

The above derivation shows the equivalence of the best-fit and covariance-matrix methods for the case where the χ^2 fit is *unconstrained*. This is not necessarily a trivial condition. For example, one may want to require that

$$t_i + \sum_{j=1}^{K} \alpha_j s_{ji} \geq 0 \quad \text{for all } i \tag{27}$$

in order to obtain a physically meaningful solution. In such a case the solution is no longer given by equation (10), and $\chi^2_{\rm min}$ no longer equals the covariance-matrix expression.