Glen Cowan
16 January, 2007

Robust track fitting

This note describes first a method for estimating tracking errors in the presence of outliers by exploiting the mode of the χ^2 distribution. Then we describe a procedure for improving the estimates of the tracking parameters by identifying tracks with outlying coordinates, which can then be eliminated from the tracking fit.

1 Estimating the tracking error

Suppose we have a tracking chamber that measures coordinates $y_1, \ldots, y_n$. The predicted values are $f(x_i; \boldsymbol{\theta})$ where x_i is ith value of a control variable, e.g., corresponding to the position of the tracking chamber, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ is a vector of tracking parameters.

In this note we will consider the special case where $n - m = 2$, i.e., we have two degrees of freedom. For example, we could have four measured coordinates and we are fitting a straight line. Let us also suppose that *a priori* we do not know the standard deviations of the measurements y_i, but we can assume it is the same value σ for all of the measurements.

If we knew the standard deviation σ, we would be able to estimate the track parameters by minimizing the quantity

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i; \boldsymbol{\theta}))^2}{\sigma^2} \ . \tag{1}$$

If the assumptions of our model are correct, the minimized value of (1) should follow a chi-square distribution with a number of degrees of freedom $n_{\rm d}$ equal to the number of measurements n minus the number of fitted parameters m. If we write the argument of the chi-square distribution as z, then for $n_{\rm d}$ degrees of freedom this is [1]

$$f(z; n_{\rm d}) = \frac{1}{2\Gamma(n_{\rm d}/2)} \left(\frac{z}{2}\right)^{\frac{n_{\rm d}}{2}-1} e^{-z/2} \ . \tag{2}$$

This has a mean equal to $n_{\rm d}$ and a mode of $n_{\rm d} - 2$. For $n_{\rm d} = 2$ this is an exponential pdf with a mean of 2.

Unfortunately we do not know σ, but since we are assuming it is a constant, we will obtain the same estimators for the track parameters by minimizing the sum of squares

$$S^2 = \sum_{i=1}^{n} (y_i - f(x_i; \boldsymbol{\theta}))^2 \ . \tag{3}$$

If the assumptions of our model are correct, the minimized value of S^2 should follow something shaped like a chi-square distribution, but the horizontal scale will differ by a factor of σ^2. If we were confident in our model assumptions, we could estimate σ using

$$\hat{\sigma} = \sqrt{\frac{S^2}{n_{\rm d}}} \ , \tag{4}$$

where $\overline{S^2}$ denotes the arithmetic average of the minimized S^2 values from our data sample. Equation (4) gives the value of σ such that one would obtain an average value for χ^2 equal to n_{d}.

Now if we assume that our track sample includes cases where one of the coordinates is very poorly measured because, say, of a malfunction or sparking, then the distribution of S^2 values will show a tail toward high values. Because of these high values, we can no longer use equation (4) to estimate the tracking error σ.

The poorly measured tracks should not have a large influence, however, on the mode of the distribution of minimized χ^2 or S^2 values. We could also try to estimate σ by using the value that would force the mode of the χ^2 distribution equal to $n_{\text{d}} - 2$,

$$\hat{\sigma} = \sqrt{\frac{\text{mode}[S^2]}{n_{\text{d}} - 2}} \; . \tag{5}$$

This estimate of σ should be more robust than equation (4), since the mode of a pdf is relatively insensitive to outliers.

This method should work for $n_{\text{d}} \geq 3$, but it cannot be applied for only two degrees of freedom. In that case, however, we can exploit the fact that the parameters that minimize the χ^2 will be the same that minimize its square root,

$$\chi = \sqrt{\sum_{i=1}^{n} \frac{(y_i - f(x_i; \boldsymbol{\theta}))^2}{\sigma^2}} \; . \tag{6}$$

The chi-distribution is given by

$$f(y; n_{\text{d}}) = \frac{1}{\Gamma(n_{\text{d}}/2)} y^{n_{\text{d}}-1} \left(\frac{1}{2}\right)^{\frac{n_{\text{d}}}{2}-1} e^{-y^2/2} \; , \tag{7}$$

which has a mode equal to $\sqrt{n_{\text{d}} - 1}$ (see [2] Section 8.14).

We can first estimate the track parameters by minimizing the square root of S^2,

$$S = \sqrt{\sum_{i=1}^{n} (y_i - f(x_i; \boldsymbol{\theta}))^2} \; . \tag{8}$$

Then we estimate σ using

$$\hat{\sigma} = \frac{\text{mode}[S]}{\sqrt{n_{\text{d}} - 1}} \; , \tag{9}$$

where mode$[S]$ is the mode of the distribution of S values from our data sample.

2 Eliminating outliers from tracking fits

Once the tracking error σ has been estimated, the track fitting procedure can be repeated in a way that identifies and eliminates those tracks with very poorly measured coordinates. In the case where there is only one outlier, we expect a large reduction in the χ^2 value when this coordinate is eliminated from the fit. Furthermore, the χ^2 values for the remaining combinations should be roughly equal to the number of degrees of freedom.

Let us define

$$\chi_m^2 = \min \left[\sum_{i=1}^{n} \frac{(y_i - f(x_i; \boldsymbol{\theta}))^2}{\sigma^2} \right] \tag{10}$$

as the minimum χ^2 value obtained when only fitting with a subset of m out of the n available coordinates. If we then look at the distribution of

$$\delta^2 = \chi_n^2 - \chi_{n-1}^2 \,, \tag{11}$$

then this should have a tail toward high values for tracks containing one very poorly measured coordinate. An improved prescription for estimating the track parameters would be to make a cut on the quantity δ^2 such that if it exceeds a certain threshold, one eliminates the rogue coordinate and takes the parameters from the minimum of χ_{n-1}^2. If δ^2 is below the threshold, the minimum of χ_n^2 is used.

References

[1] G. Cowan, *Statistical Data Analysis*, Oxford University Press, 1998.

[2] Christian Walck, *Handbook on Statistical Distributions for Experimentalists*, University of Stockholm, Particle Physics Group, Internal Report SUF-PFY/96-01.