Functions Alternative to the Likelihood

Brandon Willard
Department of Physics,
Wayne State University, Detroit, Michigan, 48202

August 13, 2007

Abstract

In this paper we determine the need for improved bias sensitive methods in particle physics and apply a technique that uses higher order derivatives of the likelihood, L, to improve efficiency in a narrow resonance fit scenario. The technique used is that described by G. Bonvicini in [1].

1 Introduction

Since bias is present to some extent in every analysis, methods to detect and potentially correct for bias are necessary. The χ^2 (when available) is the sole parameter used to characterize the goodness of fit, but, as we show, there are practical cases in particle physics that could benefit substantially from improvements. We find in the 2006 Particle Physics Data Book, PDB, that summary table quotes of mass and width quantities for sets of mesons display significant bias in their reduced χ^2 values, χ^2/dof, where dof is the number of experiments minus the number of free parameters (in this case, one). We also show that this bias cannot be explained by a change in the confidence interval production, i.e., through the L versus Feldman & Cousins, because differences between the two methods are of order $1/N$, N being the number of events considered. Having given cases for improvement we proceed to simulate experiments and apply the technique of [1]. In this paper, the case of a narrow resonance amidst background is analyzed. Biased fits are produced through a two signal convolution, then derivatives of L are used to improve bias detection and efficiency.

2 An Example.

For the purpose of illustration, we select a subsample of results from the Particle Data Book that, while being reasonably sound experimentally, can be expected to suffer from substantial theoretical biases.

The sample consists of the masses and widths of unflavored light mesons ($S = B = C = 0$). Their dynamics are usually modeled with QCD models that are not exact solutions of the QCD Lagrangian. Their large widths (after the cuts mentioned below) induce large interference effects.

We consider only measurements of particles known to exist (they are listed in the Summary Tables), where at least 5 experimental measurements are reported, for the purpose of extracting a good χ^2 from the data. To enhance the possibility of having large interference effects, we consider only particles with a measured width exceeding 10 MeV. We arrive at a sample of 28 widths and 28 masses.

To compute the reduced χ^2 values, which is sometimes also quoted in the PDB, we use the following equation

$$\chi^2/dof = \frac{1}{N_e - 1} \sum_{i}^{N_e} \frac{(y_i - y_i^*)^2}{\sigma_i^2} \tag{1}$$

in which y_i are the experiments, y_i^* the 2006 PDB summary table values, and σ_i^2 the quoted variance of y_i^*.

The resulting histograms can be expected, if the data were truly unbiased, to have approximately equal populations above and below the reduced χ^2 expectation value of 1. Instead, Figs. 1 and 2, only one of the 56 quantities is below, all others being above.

The most obvious explanation is that the bias is of the order of the quoted experimental errors, and therefore significant. Of note is the fact that these data are usually obtained through multi-parameter fits. The data samples had populations N ranging from 10^3 to 10^6, so that the bias was of the order 0.03 to 0.001 depending on the sample.

2.1 Confidence Interval Comparison

There may be a question of whether these discrepancies could be due to a choice in the fitting method (the standard likelihood). The bulk of the data considered in Figs. 1 and 2 was obtained in the 1970s and 1980s, whereas nowadays the Feldman & Cousins is used.

For the purpose of disproving this notion, we compare the confidence intervals constructed with the two methods, in the case of varying Poissonian

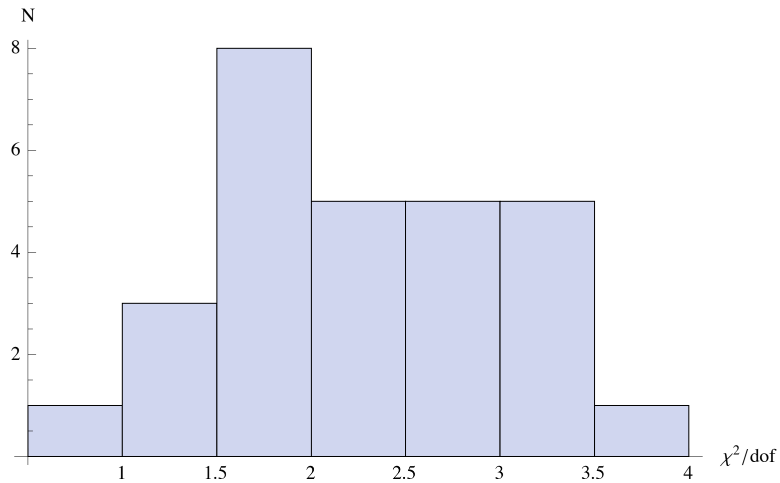

Figure 1: Reduced χ^2 plot for 2006 PDB summary table masses.

signal, N_p, with no background.

Fig. 3 shows that the difference between the two methods tends to a constant as N_p increases. Thus, the difference in the obtained upper and lower limits goes like $1/N$. The same result, not shown, was obtained for the 90% CI.

3 Simulation

To give an example of how the technique may work, we simulate an experiment, with statistics similar to current CLEO experiments. A narrow resonance with $N_s = 7000$ events is considered. The part of the spectrum considered for fitting extends 7 standard deviations above and below the resonance, with 50 bins in the histogram. The resonance is situated on top of a flat background of 3000 events per bin. When fitted in an unbiased way, the correct signal is recovered with an error of about 2.4%.

The bias is inserted by hypothesizing that the "true" resonance has two Gaussian resolution contributions, a main one contributing 90% of the statistics and a smaller one contributing 10% of the statistics, with width twice as large as the 90% contribution. Many a narrow peak are fitted with a single Gaussian when in fact the experimental resolution is the convolution of many Gaussians, due to the dependence of the spectrometer resolution on multiple scattering and track angle amongst other things. However, the fitting routine is written so that only one gaussian width is given as a free parameter to fit the data.

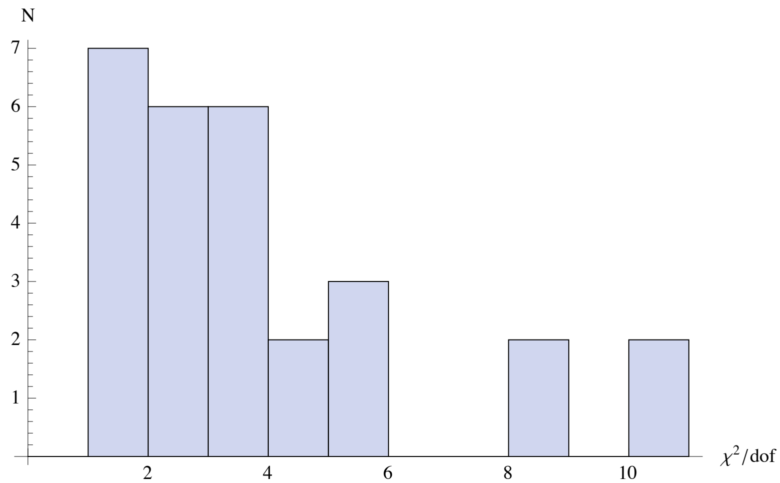

Figure 2: Reduced χ^2 plot for 2006 PDB summary table widths.

Therefore we have a signal parameter (the outcome of our experiment), and two nuisance parameters (the background, and the width of the peak). The bias enters only through the mis-representation of the functional form of the peak. Errors are considered to be Gaussian.

From Fig. 4, one can see that for the signal the outcome of many such biased fits is consistently biased down by about 3% (1.25 times the quoted signal error). From Fig. 4, one can also see that the χ^2 test does not allow good separation between biased (dashed) and unbiased (solid) fits. However, most quantities in the covariance matrix also show some separation (Fig. 5) between biased and unbiased fits which can be used to reject the biased fits. In Fig. 6, we also show the expectation probability for the correlation coefficients, based on the equation

$$\begin{aligned} \mu(\alpha, \beta) &= E\left[\frac{\partial \log L}{\partial \alpha}\frac{\partial \log L}{\partial \beta}\right] \\ &= \left(\frac{1}{\sigma^2}\right)^2 \left(\sigma^2 \sum_{i=1}^{N} \frac{\partial f_i}{\partial \alpha}\frac{\partial f_i}{\partial \beta} + \sum_{i \neq j}^{N} f_j f_i \frac{\partial f_i}{\partial \alpha}\frac{\partial f_j}{\partial \beta}\right) \end{aligned} \quad (2)$$

in which α, β are fit parameters for the function $f_i = f(x_i; \alpha, \beta)$, uniform event errors σ, and N number of events.

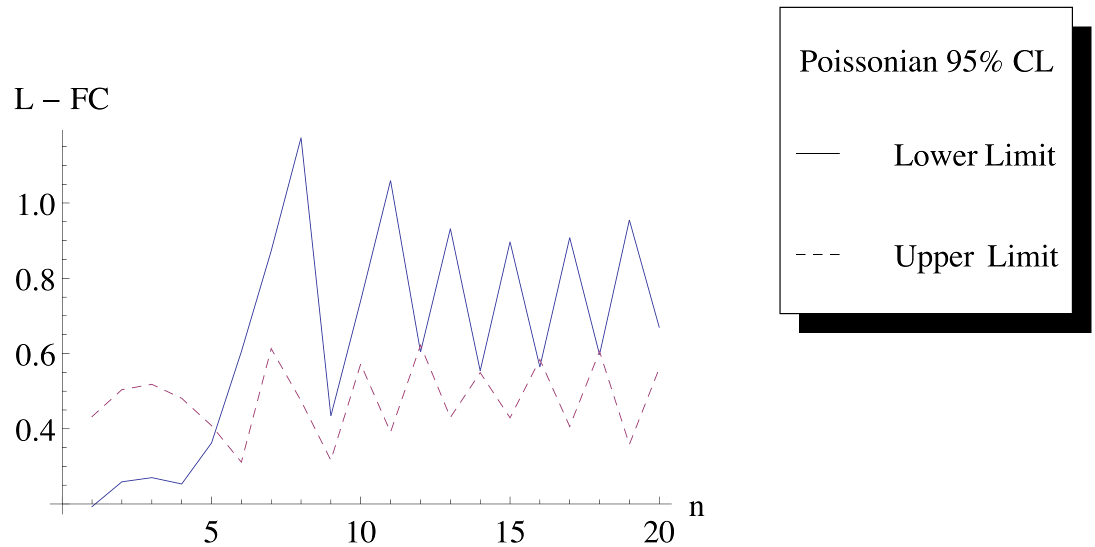

Figure 3: Difference between L and Feldman & Cousins (FC) methods.

3.1 Improvements

Here we address how this information is used to improve efficiency. First, in Table I we show the efficiency for unbiased fits when different cuts are applied on the χ^2 (first line) as well as the six members of the covariance matrix. For the χ^2, $E[\chi^2] = 47$ with $\sqrt{D[\chi^2]} = 9.75$, so the 2.0 cut corresponds to a $47 + 2 * 9.75 = 66.5$ cut. For all other variables, the distribution is so close to Gaussian that the cut is applied symmetrically. If a fit is more than 2 standard deviations above or below $E[\rho]$ (as given by Eq. 2), for example, it will fail the 2.0 cut.

In Table II, we give the same table for the biased fits. One can see that, for example, a χ^2 cut that eliminates 3.4% of the unbiased fits (column 2, Table I) will eliminate 12.9% of the biased fits. A mixed cut, using the sixth column for the χ^2 and the fifth column for the cov_{ij} (the selection of columns to obtain approximately equal ineficiency for all cuts) gives 3.2% inefficiency for unbiased fits but 27.0% rejection for biased fits. Already the rejection power has more than doubled.

In fact, one may argue that, because the result of the experiment is only the signal, the other two parameters being nuisance, one should construct the goodness of fit for that parameter alone. It does not matter what the biasing information is about the nuisance parameter, since they are there only to steer a line through the data. If we do that (using only the first, second, and

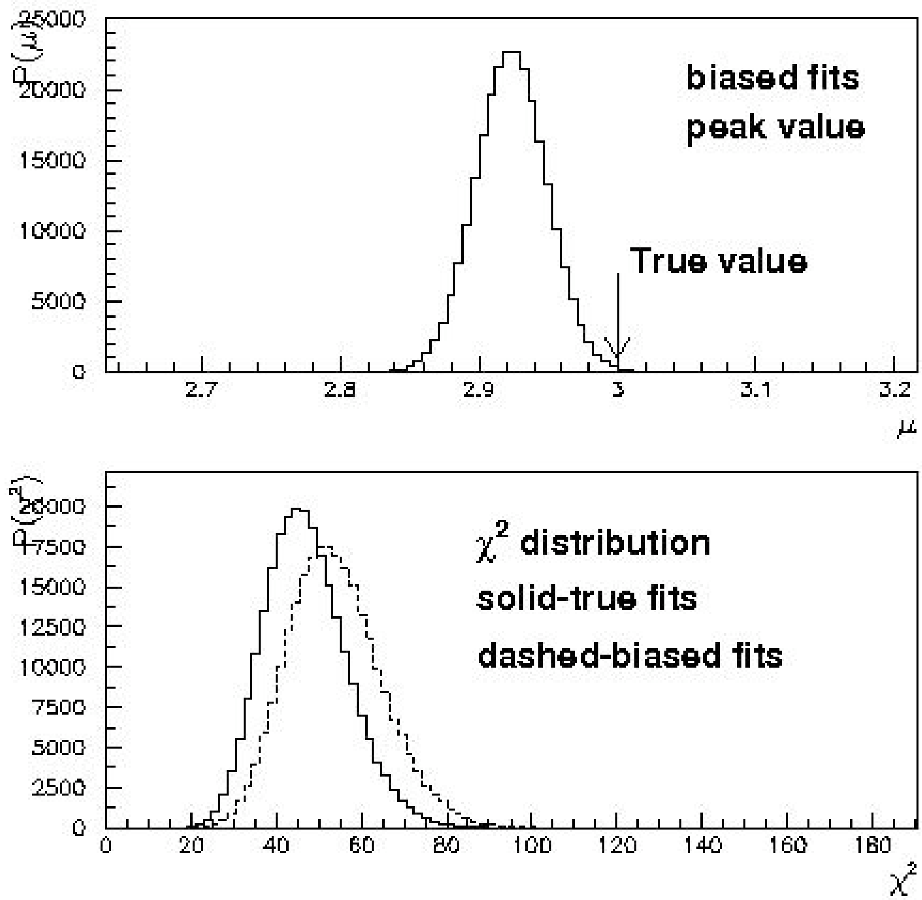

Figure 4: Above: Effect of bias on the signal fit (μ is the integral of the signal). Below: Effect of bias on χ^2.

fifth row, fourth column) we obtain a 3.9% inefficiency for unbiased data and a 35.5% rejection for biased data. The rejection power has almost tripled compared to the χ^2 cut alone, for about the same inefficiency.

4 Outlook and Ongoing Work

As we can see from the narrow signal case, the derivatives contain useful information. An alternative method that uses this information can be understood as one that maximizes not L, but the residual population. Through this method one can gain more information about the residual population from higher order derivatives, and, as was demonstrated, the first order derivatives

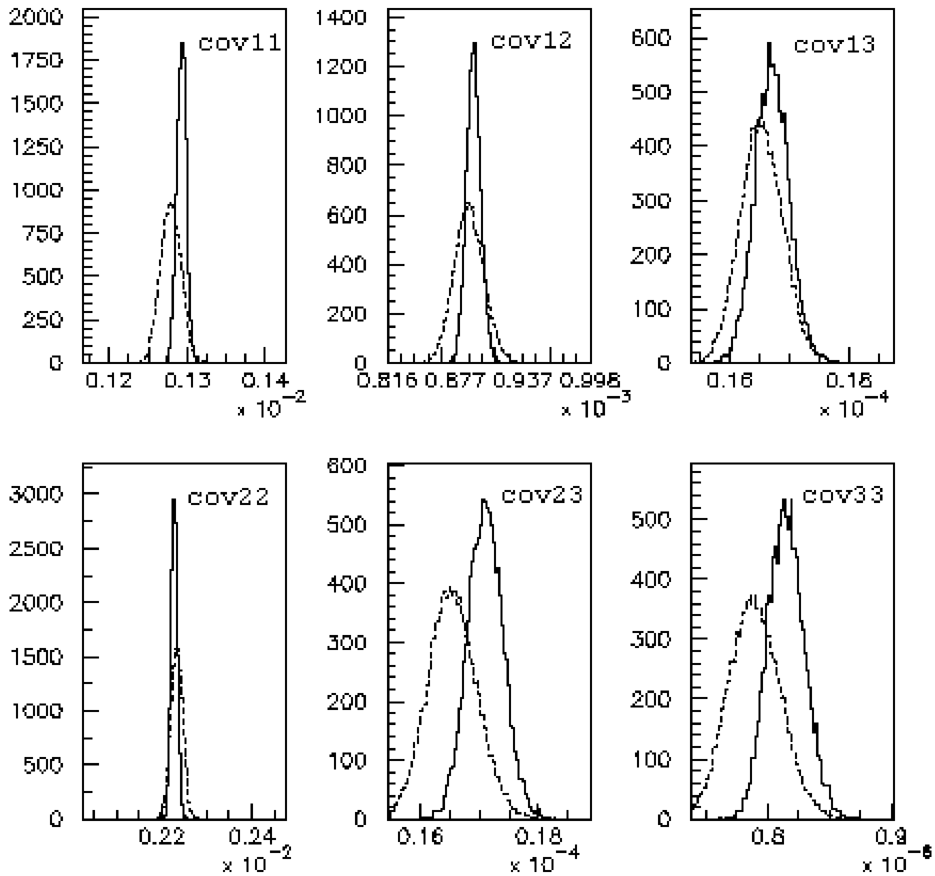

Figure 5: Covariance elements for biased (solid line) and unbiased (dashed line) fits [1-signal, 2-background, 3-width of peak]

cut	1.80	2.00	2.20	2.40	2.60	2.80	3.00
χ^2	0.95246	0.96602	0.97592	0.98364	0.98926	0.99310	0.99544
cov_{11}	0.92946	0.95626	0.97328	0.98450	0.99106	0.99516	0.99746
cov_{22}	0.92698	0.95338	0.97218	0.98282	0.99030	0.99434	0.99694
cov_{33}	0.92878	0.95466	0.97214	0.98354	0.99054	0.99464	0.99700
cov_{12}	0.93234	0.95806	0.97448	0.98506	0.99164	0.99546	0.99782
cov_{13}	0.92774	0.95522	0.97250	0.98386	0.99080	0.99496	0.99728
cov_{23}	0.93098	0.95584	0.97316	0.98428	0.99098	0.99512	0.99754

Table 1: Unbiased efficiencies

alone yield improvements over χ^2. In general, the technique presented here

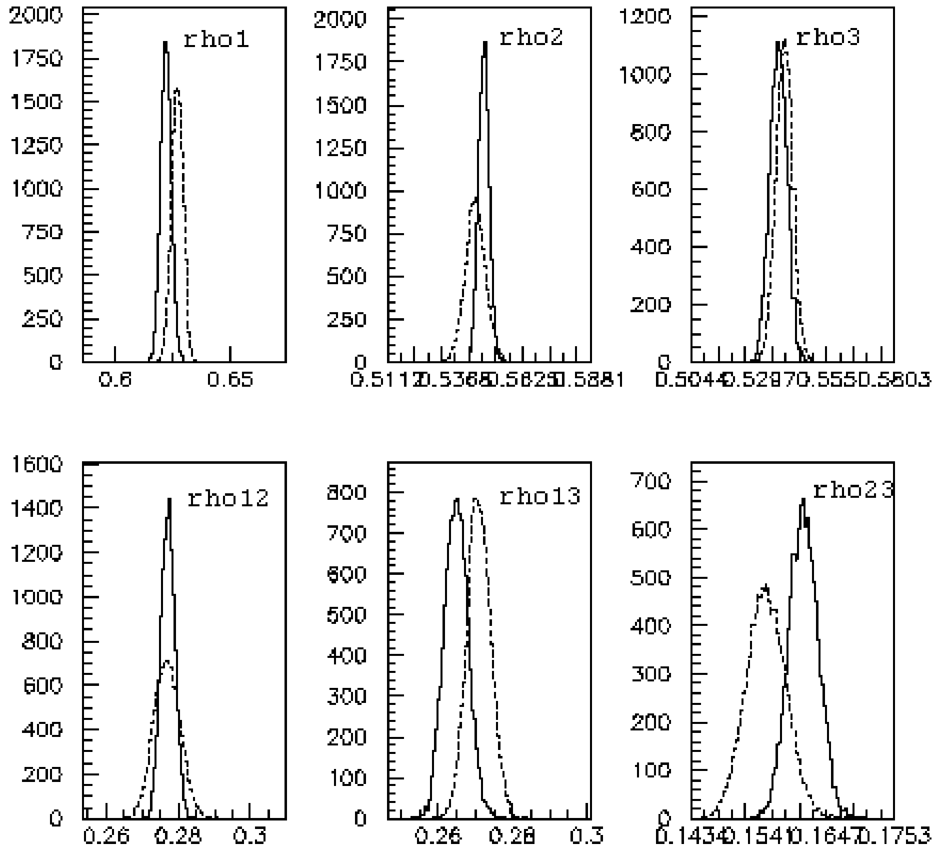

Figure 6: Correlation elements for biased (solid line) and unbiased (dashed line) [1-signal, 2-background, 3-width of peak]

cut	1.80	2.00	2.20	2.40	2.60	2.80	3.00
χ^2	0.83410	0.87087	0.90017	0.92425	0.94370	0.95887	0.97061
cov_{11}	0.94627	0.97772	0.99177	0.99709	0.99940	0.99983	1.00000
cov_{22}	0.99957	0.99991	1.00000	1.00000	1.00000	1.00000	1.00000
cov_{33}	0.79374	0.86358	0.91585	0.95150	0.97258	0.98586	0.99332
cov_{12}	0.44953	0.54422	0.63299	0.71320	0.78886	0.85476	0.90197
cov_{13}	0.96590	0.98646	0.99503	0.99846	0.99966	0.99991	1.00000
cov_{23}	0.85244	0.89477	0.92682	0.95321	0.97078	0.98235	0.99023

Table 2: Biased efficiencies

entails a new goodness of fit, as described in [1].

Acknowledgments

I would like to thank Giovanni Bonvicini for all the support and guidance, Rich Galik, and the NSF (grant PHY-0353994).

References

[1] G.Bonvicini, Nuclear Instruments and Methods in Physics Research A 552 (2005) 522558.