An Algebraic Comment* on Using Large $\Delta\chi^2$

Bob Cousins, UCLA

PDF4LHC, 22 Feb 2008

*Already well known to those who know it well.

HEP Standard for Discrepant Data: The PDG Scale factor

http://pdg.lbl.gov/2007/reviews/textrpp.pdf, page 10:

Finally, if $\chi^2/(N-1)$ is greater than 1, but not greatly so, we still average the data, but then also do the following:

(a) We increase our quoted error, $\delta\overline{x}$ in Eq. (1), by a scale factor S defined as

$$S = \left[\chi^2/(N-1)\right]^{1/2} . \tag{2}$$

Our reasoning is as follows. The large value of the χ^2 is likely to be due to underestimation of errors in at least one of the experiments. Not knowing which of the errors are underestimated, we assume they are all underestimated by the same factor S. If we scale up all the input errors by this factor, the χ^2 becomes $N-1$, and of course the output error $\delta\overline{x}$ scales up by the same factor. See Ref. 3.

When combining data with widely varying errors, we modify this procedure slightly.

This is a very blunt procedure, but a long-standing and reasonable one given the difficulty of the problem. Still, one is well-advised to use extreme caution in interpreting results with large scale factors. The PDG always lists the scale factors in its summaries, wallet card, etc.. –BC.

From Robert Thorne's talk at PhyStat-LHC

http://phystat-lhc.web.cern.ch/phystat-lhc/program.html

$$\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} (D_i - T_i(a)) C_{ij}^{-1} (D_j - T_j(a))$$

Slide 5

$\Delta\chi^2 = 1$ is not a sensible option.

Slide 56

CTEQ use $\Delta\chi^2 \sim 100$ and MRST/MSTW use $\Delta\chi^2 \sim 50$.

Slide 12

[N.B. This is for ~90% C.L., for which the "book value" is $\Delta\chi^2 = 2.7$. –BC]

My comment, which I have been making since the 2002 Durham workshop but which still surprises people (and so I was invited to say it here) is:

For $\Delta\chi^2$ of the above form, using $\Delta\chi^2$ which is a factor of S^2 larger than the book value (whatever the motivation) is mathematically the same a using the book value after multiplying the covariance matrix C by S^2, i.e., *the same as using a PDG scale factor S* …with all that implies!

Thus, using $\Delta\chi^2$ = 50 or 100 rather than $\Delta\chi^2$ = 2.7 corresponds to a PDG scale factor *S* = sqrt(50/2.7) = 4.3 or *S* = sqrt(100/2.7) = 6, i.e., to scaling the errors on *all* input measurements by this factor.

…and as the PDG notes, of course all output errors scale up by the same factor.

I would assume that all the experts at this meeting know this*, but my experience is that consumers of PDF uncertainties are **completely** ***unaware of this, and are a bit shocked when told.***

And my question is: For the fancier modifications of χ^2 that CTEQ and MRST use (and extra weight factors for some experiments, etc.), how is this simple statement modified?

My recommendation is that the effective PDG scale factor be better advertised by the purveyors of PDFs.

****S* is proportional to the tolerance parameter *T* of J. Pumplin, et al., Phys Rev. D65, 014012 (2002), arXiv:hep-ph/0101032**

BACKUP

Postcript: PDG (cont.)

5.4. ***Discussion*** : The problem of averaging data containing discrepant values is nicely discussed by Taylor in Ref. 4. He considers a number of algorithms that attempt to incorporate inconsistent data into a meaningful average. However, it is difficult to develop a procedure that handles simultaneously in a reasonable way two basic types of situations: (a) data that lie apart from the main body of the data are incorrect (contain unreported errors); and (b) the opposite—it is the main body of data that is incorrect. Unfortunately, as Taylor shows, case (b) is not infrequent. He concludes that the choice of procedure is less significant than the initial choice of data to include or exclude.

3. A.H. Rosenfeld, Ann. Rev. Nucl. Sci. **25**, 555 (1975).
4. B.N. Taylor, "Numerical Comparisons of Several Algorithms for Treating Inconsistent Data in a Least-Squares Adjustment of the Fundamental Constants," U.S. National Bureau of Standards NBSIR 81-2426 (1982).