PHY340 Data Analysis Feedback  
Group P08 doing Problem P4

# Data Analysis

The main problem with the data analysis is of course that it is completely wrong. A mass of 145 GeV/*c*2 and a standard deviation of 125 GeV/*c*2, with a tail extending out to 1400 GeV/*c*2, is defi­nitely not consistent with expectation: your figure 1 shows that the CMA should produce a peak in the right place, with a standard deviation of a few tens of GeV. This does not seem to have occurred to you: there is no evidence in your report that you realise that there is anything wrong. In fact, it was not difficult to find the error in your code, which I think represents an er­ror in your understanding rather than a bug (the code is doing what you intended it to do, but what you intended it to do is wrong). Whoever wrote the code simply doesn’t understand four-vectors: the dot product of a four-vector is : it is ***not*** as you have written. Therefore, all your masses are simply wrong. (It should be obvious that, with this definition of , is not in fact negligible as you claim in your derivation, which should have alerted you to the fact that you have a problem.)

You do not really explain your fitting procedure: your code calculates the mean and variance of your data directly, without a fit, but you do not mention this in your text, where you refer to a Gaussian fit. It is therefore not clear whether your quoted mass is from the fit or from the direct calculation. It is also not at all clear where your “systematic error” comes from: you claim that your Gaussian fit “program also gives the systematic error”, but this certainly is not true—the whole *point* of systematic errors is that they are *not* identifiable in this way, because they bias the data in a coherent, systematic fashion (hence the name). I have no idea what it is that you have interpreted as a systematic error, except that if it’s something returned by the fit routine it definitely isn’t a systematic error. The GNUplot documentation for fit says that it “reports 'stdfit', the standard deviation of the fit, which is the rms of the residuals, and the variance of the residuals, also called 'reduced chisquare' when the data points are weighted.” Since the stan­dard deviation is simply the square root of the variance, I don’t know why it bothers to return both, but neither of them is a “systematic error”—they are measures of the goodness of fit. In ad­dition, it returns error estimates for the parameters, and a covariance matrix.

If I have correctly interpreted your GNUplot commands, you did not properly weight your data: as the number of entries in a histogram bin is a count, it is a Poisson variable, with a statistical error equal to the square root of the number of entries (provided that the number of entries is large enough to approximate the Poisson distribution to a Gaussian). This means that your fit is probably not in fact the best fit to your data, and makes the number of significant figures you have quoted look even siller than it did already (remember that you should be quoting at most two significant figures in an error estimate—not 4).

Your comparison with literature is meaningless because your data are nonsense, but it is worth noting that higher energy taus would *not* “skew the result towards a higher energy”: the invari­ant mass of a system is *an invariant*, and does not depend on the boost. In fact, as the collinear mass approximation improves for higher-energy taus (because the decay products of a higher energy tau are more nearly collinear), you would expect the inclusion of higher energy events to *improve* the precision of your mass peak. As you were explicitly told that the data file consisted of H → τ+τ− events, it is also pointless to consider the effects of non-existent Z → τ+τ− events (which would, in any case, cause an excess of *low* mass events, since the Z mass is lower than the Higgs mass—again, see your figure 1). As you do seem to be aware that your results are not really satisfactory, it is very surprising that it never seems to have occurred to you that your code might be wrong—always the most likely cause of a large discrepancy (and not just for un­der­graduates).

Although your results are completely invalidated by the error in your analysis, you have ap­proached the analysis in a generally sensible fashion. You just need to have a more self-critical attitude when examining your results: if a result looks very different from the literature values, by far the most likely explanation is that *you have made a mistake*. Only when you have ruled this out should you consider possible systematic errors.

Average mark for this section: 27.75/50

# Data Presentation

Your figures have appropriate numbers and captions, and in the case of figure 1 an appropriate reference to the source. The legend of figure 2 could be more helpful: “f(x)” is *not* an informa­tive label (you needed something like “Gaussian fit”). You should provide a better explanation of what you fit actually *does*: is it a least-squares fit or maximum likelihood; if the latter, is it weighted or unweighted; what minimisation algorithm is used? This information is *not* difficult to find: the GNUplot online documentation tells me that the fit command is “an implementation of the nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm.” This is what the reader needs to know: the fact that your particular Marquardt-Levenberg nonlinear least-squares fit is implemented in GNUplot rather than in Python (the Python curve\_fit routine uses exactly the same algorithm) is not important, assuming that you are using a reputable piece of code which is not likely to contain major bugs.

You are quoting far too many significant figures: your mass should be 146±13±8 (if quoting both statistical and systematic er­rors, one always puts the statistical error first), and the standard deviation 125±9. I have no idea why the error values are in brackets, and—as noted above—I don’t know what the thing you are calling “systematic error” actually *is*, but it certainly is *not* a systematic error since you say it is an output from the fit. If it’s stdfit, then the fact that it’s 8.7 indicates an exceptionally poor fit—which is not too surprising, as your fit in figure 2 does look poor (there is, of course, no reason to expect the output of your meaningless calcula­tion to be Gaussian, so this is not surprising).

Average mark for this section: 19/30

# Style

The overall structure of the report is not bad, although it is rather short and lacking in useful de­tail. However, it needed better proof-reading, both to remove typos such as “lepontic” for “lep­tonic” on page 1 and to get rid of gibberish like “Now equate these two equate equations” in the appendix. There are also some spelling errors (“it’s” means “it is” or “it has”—if you mean “of or belonging to it” then the word is “its”) and some bad grammar (sentences that aren’t sentences; use of “however” to join sentences together—what you want in such a case is simply “but”).

Average mark for this section: 12.65/20

Overall average mark: 59.4%.