Systematic uncertainties from several models: a need for Stat WG guidance

Roger Barlow

LHCb Statistics Group Meeting

14^{th} December 2020

Reporting back from PHYSTAT2020

Survey of handling systematic errors in recent LHC results from ATLAS, CMS and LHCb

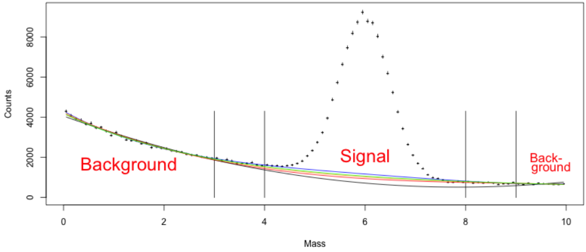
Lots of good stuff going on

But one persistent case of poor practice

Discrete uncertainties

When you have several alternative models

The set up: many models



Analysis

Count number of events in signal region.
Subtract background by fitting quadratic in background region (red curve)

Systematics

Try using cubic, quartic, exponential, exponential-quadratic ... functions to fit background

Points to note

- The background function has no theoretical grounding. It just fits the data.
- This is an uncertainty, not a check. You expect different functions to give different answers. This is a known unknown.
- Sometimes signal and background regions can be in different channels/experiments, but same logic applies.
- For the sake of the example, we suppose all functions give acceptable fits to data in the background region.
- If you know the likelihood function (including the peak) you can use the profile+envelope method [1], but not if you're just counting numbers

[1] P. D. Dauncey, M. Kenzie, N. Wardle and G. J. Davies, Journal of Instrumentation 10 p04015 (2015), arXiv:1408.6865v5

Possible choices

For your central analysis value, you can stick with the quadratic or you can use the average of all functions, depending on what you think of their status.

For the uncertainty you can take the RMS deviation of the values about the chosen central value. Or the mean absolute deviation.

If you take it as the maximum deviation then this is the wrong choice.

Why this is wrong

- It is over-conservative and inflates your errors
- It doesn't really represent a set of values. Mean, mode, median - but max!!
- It penalises diligence: if you consider many functions you are bound to make your errors larger

Plenty of recent examples

between 270 and 302 keV [2]. The maximal deviations in the ratios $\mathcal{R}^{\mathrm{X}}_{\mathrm{Y}}$ with respect to the baseline fit model are taken as systematic uncertainties for each of the systematic signal model sources. For the systematic uncertainty related to the modelling of the smooth

---

the first to the second order, separately for each fit component and each channel. In each case the ratio $\mathcal{R}^{\mathrm{X}}_{\mathrm{Y}}$ is computed and the maximal difference with respect to the baseline fit model is taken as a corresponding systematic uncertainty. For each choice of the fit

---

is found to be stable. It varies within 0.2% with respect to the efficiency computed for the phase-space model when the unknown phase Φ varies in the range $-\pi \leq \Phi < \pi$.

---

to ±20% change in the measured efficiency. The resulting variations in the efficiency ratios do not exceed 1%, which is taken as a corresponding systematic uncertainty. The last

When is it correct to take the maximum deviation?

Very seldom

- If you're an engineer. Then you quote tolerances. If you have a 99 ± 1 mm peg and a 101 ± 1 hole you want it to fit *every* time.
- If you define 'error' as '68% central confidence region' as opposed to 'rms spread'[2] and take an ultrafrequentist approach to '68% confidence means θ lies in the region at least 68% of the time', i.e. for absolutely all values of ν, rather than profiling or even marginalising. Note that if you take this route you can no longer add in quadrature

We must not be afraid of quoting a result that may be more than 1 sigma from the true value.

[2] For Gaussian distributions it's the same

Why do people do it?

"It's what we always do"

That is not a valid reason for doing anything

"If the result turns out to be outside the quoted error will be bad for my/our reputation"

32 % of our results should be outside their quoted error

"It's conservative"

For errors, *conservative* is another word for *wrong*

Summary

- This is a widespread example of poor statistical practice

Summary

- This is a widespread example of poor statistical practice, IMHO

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error.

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error. And this part of the analysis is generally done at the end and is wildly unexciting.

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error. And this part of the analysis is generally done at the end and is wildly unexciting.
- An estimate like this can never be rigorously justified, and numbers should not be taken too seriously

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error. And this part of the analysis is generally done at the end and is wildly unexciting.
- An estimate like this can never be rigorously justified, and numbers should not be taken too seriously
- There may be cases where using the maximum deviation can be justified.

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error. And this part of the analysis is generally done at the end and is wildly unexciting.
- An estimate like this can never be rigorously justified, and numbers should not be taken too seriously
- There may be cases where using the maximum deviation can be justified. We could discuss possible examples.

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error. And this part of the analysis is generally done at the end and is wildly unexciting.
- An estimate like this can never be rigorously justified, and numbers should not be taken too seriously
- There may be cases where using the maximum deviation can be justified. We could discuss possible examples.
- But should certainly not be regarded as standard operating procedure.

Summary

- This is a widespread example of poor statistical practice, IMHO
- Not a big deal, as usually a small part of total systematic error. And this part of the analysis is generally done at the end and is wildly unexciting.
- An estimate like this can never be rigorously justified, and numbers should not be taken too seriously
- There may be cases where using the maximum deviation can be justified. We could discuss possible examples.
- But should certainly not be regarded as standard operating procedure.

Suggestion

Add and publicise LHCb SWG guideline recommending that for uncertainties due to choice between several models, to use the RMS spread as measure of uncertainty, and explicitly not the 'conservative' maximum deviation