Asymmetric Errors

Roger Barlow
Huddersfield University
Preliminary version not intended for presentation

BIRS workshop on systematic errors

26^{th} April 2023

Banff International Research Station
for Mathematical Innovation and Discovery

Many particle physics results have asymmetric errors.

ATLAS EXPERIMENT

$\sigma(t\bar{t}t\bar{t}) = 24^{+7}_{-6}$ fb

JHEP 11 (2021) 118

From Shabalina's ATLAS Moriond talk

	μ
$WH\ p_T^V < 150$ GeV	$1.5^{+1.0}_{-0.9}$
$WH\ p_T^V > 150$ GeV	$3.6^{+1.8}_{-1.6}$
$ZH\ p_T^V < 150$ GeV	$3.4^{+1.1}_{-1.0}$
$ZH\ p_T^V > 150$ GeV	$0.8^{+1.2}_{-0.9}$

From Calandri's CMS Moriond talk

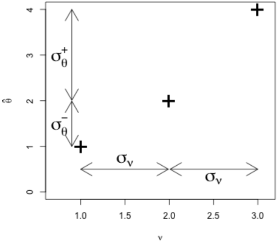
How should these be handled? The experts don't know.

Some ground rules for the talk+discussion

1. The question requires an answer within the frequentist framework. Once we have that, a Bayesian analysis will be interesting, but until then it will just be confusing.
2. Functions which are known to be asymmetric (Poisson, logNormal...) are not part of the problem, as for them we have full information. (They are useful for checking).
3. We are working in the fairly-large N region. Not every distribution is normal, but they are recognisable distortions.
4. Adding + and - sigma separately in quadrature is obviously wrong

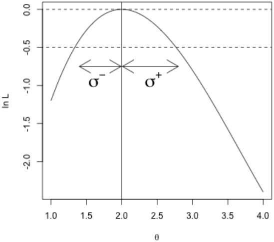
Two Reasons for Asymmetric Errors

“Systematic” OPAT systematics evaluation



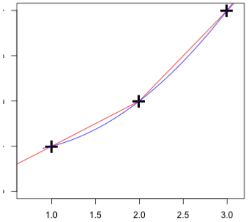
ν effects the likelihood $L(\theta, \nu|x)$
(typically an MC tuning parameter)
It is known with some well-behaved Gaussian uncertainty $\nu = \nu_0 \pm \sigma_\nu$
$\hat\theta$ from maximising $\ln L(\theta, \nu_0|x)$
Errors from maximising $\ln L(\theta, \nu_0 \pm \sigma_\nu|x)$
If not equally spaced about $\hat\theta$, report asymmetric errors

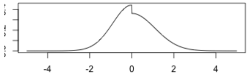
“Statistical” From ML estimation

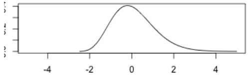


Likelihood as function of θ
Read off $\hat\theta$ from the position of the peak, and the errors from the $\Delta \ln L = \pm\frac{1}{2}$ points
If curve is a parabola, these are equidistant.
If not equidistant, report asymmetric errors

"Systematic" AsymmetrIc Errors







Can parametrise dependence of $\hat{\theta}$ on ν as

Model 1) Two straight lines

Model 2) A quadratic: $y = y_0 + \frac{\sigma^+ + \sigma^-}{2\sigma_\nu}(x - x_0) + \frac{\sigma^+ - \sigma^-}{2\sigma_\nu^2}(x - x_0)^2$

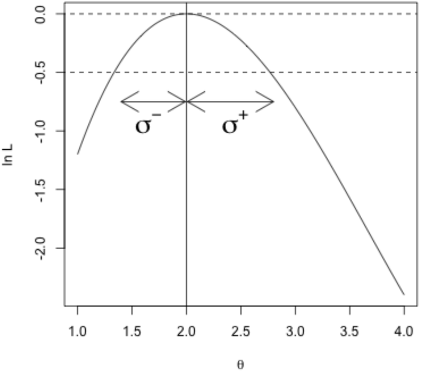
Neither is very satisfactory but you can't do much with 3 points. Typically evaluation of $\hat{\theta}$ with a different ν is computationally intensive (involving generation of a large MC sample) so more points are not an option.

ν is gaussian so $\hat{\theta}$ is distributed with a dimidated (or bifurcated, or...) gaussian (Model 1) or a distorted gaussian (Model 2)

This enables us to handle the errors. Not perfectly, but adequately. Details in R.B. *Asymmetric Systematic Errors*.arXiv:physics/0306138v1 (2003).

"Statistical" AsymmetrIc Errors

Possible distortions of a parabola
Try cubic (but turns over)
Try restricted quartic, also
generalised Poisson and log-normal
Best results from scaled parabola
$f = -\frac{1}{2}\frac{(x-x_0)^2}{V+V'(x-x_0)}$
or $f = -\frac{1}{2}\frac{(x-x_0)^2}{(\sigma+\sigma'(x-x_0))^2}$



Using $\sigma = \frac{2\sigma^+\sigma^-}{\sigma^+ + \sigma^-}, \sigma' = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}$ or $V = \sigma^+\sigma^-, V' = \sigma^+ - \sigma^-$
This enables us to handle the errors. Not perfectly, but adequately. Details in R.B. *Asymmetric Statistical Errors* arXiv;physics/0406120v1 (2004)

Why this has never been sent to a journal?

Fear

Are all asymmetric errors really one of these two types? (Plus the known-asymmetric-function cases mentioned earlier.) Or are there more out there that I havn't considered?

Hope

Why are there two different types? Why are they different? Are they linked by some duality?
Can we bring them together in some unified scheme?

Other questions

More choices...

1. What do we mean by 'the error'? The 68% central CL or the variance of the estimator?
2. Are we talking about asymmetries in the pdf (fixed θ) or the likelihood (fixed $\hat{\theta}$)?
3. What do we mean by 'handle the errors'? Combination-of-errors or combination-of-results? Is either a special case of the other?

What is an error? Think carefully before answering!

Statistician's Definition (Wikipedia)

The difference between an observation and the true value: $\hat{\theta} - \theta$

Physicist's definition(1)

The rms expectation value of the statistician's definition $\sqrt{\left\langle (\hat{\theta} - \theta)^2 \right\rangle}$

Physicist's definition(2)

The 68% central confidence region: θ lies between $\hat{\theta} - \sigma$ and $\hat{\theta} + \sigma$

Equivalent for Gaussians but which is right for a non-Gaussian case?
Definition (2) preferred. We want our result $\theta = 12.34 \pm 0.56$ to be statement about θ, not something about the mechanism that got us here.
But adding in quadrature only applies to definition (1).
Typical analysis evaluates many (systematic) errors and adds them in quadrature to get the final figure.

Asymmetries in pdfs and in likelihoods

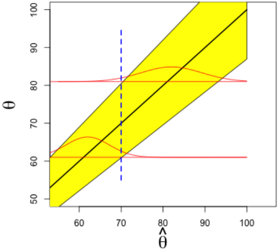
In a confidence-belt construction, pdfs run horizontally and likelihoods run vertically

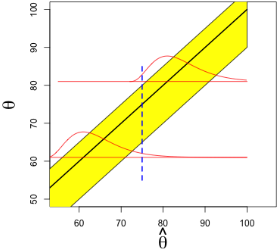
You can have a symmetric pdf but an asymmetric likelihood - e.g. proportional Gaussian

An asymmetric pdf leads to an asymmetric likelihood, but with the opposite skew

A $V(\hat{\theta})$ error relates to the pdf

A 68% CL error relates to the likelihood ($\Delta \ln L = -\frac{1}{2}$ handles the confidence belt construction. Somehow.)





More Examples

Symmetric Normal

"$x = 1.23 \pm 0.34$" means: "I have measured x as 1.23 using a method which returns a value distributed normally about the true x_0 with a σ of 0.34. On that basis I say with 68% confidence that x_0 lies within 0.34 of 1.23"

Proportional Gaussian

Suppose pdf is Gaussian with $\sigma = 0.1x_0$. ('measured to 10%..') From measured $x = 100.0$ I say with 68% confidence that x_0 lies between 91.1 and 111.1

Symmetric pdf but skew likelihood

Negative Skew pdf

Suppose pdf has 45% chance of returning x within x_0 and $x_0 + 1$, and 23% chance of returning x between $x_0 - 2$ and x_0. From measurement of 100 I say with 68% confidence that x_0 lies between 99 and 102

Positive Skew Likelihood

Poisson measurements

P has positive skew (cannot fluctuate below zero)

likelihood $e^{-\mu}\mu^r$ has positive skew

Positive skew in likelihood driven by increase of σ with r, NOT by skew in pdf.

An unhelpful example

Working with pdfs. 1/3: Combination of errors

The classic combination-of-errors formula for $f(x, y)$:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\rho \left(\frac{\partial f}{\partial x}\right) \left(\frac{\partial f}{\partial y}\right) \sigma_x \sigma_y$$

is a statement about pdfs. $\sigma_f^2 \equiv \left\langle f^2 \right\rangle - \left\langle f \right\rangle^2$

For non-Gaussian distributions, it is still true that variances add.

Care necessary as asymmetric pdf is biassed: $\theta(< \nu >) \neq < \theta >$. Central value is not the mean (but it is the median)
Even for non-Gaussian distributions, the biasses add, and so does the un-normalised skew: $\gamma = \left\langle x^3 \right\rangle - 3 \left\langle x \right\rangle \left\langle x^2 \right\rangle + 2 \left\langle x \right\rangle^3$
Suggested recipe: for each component, evaluate bias, variance and skew from σ^+ and σ^-
Add to get total bias, variance and skew.
Translate back into σ^+ and σ^- and bias.

Working with pdfs. 1/3: Combination of errors(cont)

Formulæ from integrating Gaussians:

	Two Straight lines	Quadratic
Bias	$\frac{\sigma^+-\sigma^-}{\sqrt{2\pi}}$	$\frac{\sigma^+-\sigma^-}{2}$
Variance	$\frac{{\sigma^+}^2+{\sigma^-}^2}{2} - \frac{(\sigma^+-\sigma^-)^2}{2\pi}$	$\frac{(\sigma^++\sigma^-)^2}{4} + \frac{(\sigma^+-\sigma^-)^2}{2}$
Skew	$\frac{1}{\sqrt{2\pi}}\left[2({\sigma^+}^3 - {\sigma^-}^3) - \frac{3}{2}(\sigma^+ - \sigma^-)({\sigma^+}^2 + {\sigma^-}^2) + \frac{1}{\pi}(\sigma^+ - \sigma^-)^3\right]$	$\frac{3}{4}(\sigma^+ + \sigma^-)^2(\sigma^+ - \sigma^-)$ $. +(\sigma^+ - \sigma^-)^3$

Given σ^+ and σ^-, readily find bias, variance and skew

Given variance and skew, can determine σ^+, σ^- numerically. Bias should be incorporated.

Working with pdfs. 2/3: χ^2

Given some $\mu^{+\sigma^+}_{-\sigma^-}$ and some x, for straight-line model,
$\chi^2 = \left(\frac{x-\mu}{\sigma^+}\right)^2$ for $x > \mu$ or $\left(\frac{x-\mu}{\sigma^-}\right)^2$ for $x < \mu$
For the parabolic model, after some algebra and approximations, one has
$\chi^2 = (x-\mu)^2 \left(\frac{{\sigma^+}^3 + {\sigma^-}^3}{{\sigma^+}^2 {\sigma^-}^2 (\sigma^+ + \sigma^-)}\right) \left(1 - (x-\mu)\frac{{\sigma^+}^2 - {\sigma^-}^2}{{\sigma^+}^3 + {\sigma^-}^3}\right)$

This can be used to answer the question "Is x compatible with μ, based on the pdf?"

It does not apply to μ and $x^{+\sigma^+}_{-\sigma^-}$

It cannot be considered (Wilks' theorem) as a likelihood function for μ, unless you can show $\sigma^\pm$ are independent of μ

Working with pdfs. 3/3: Combination of results

Given $\{x_1{}^{+\sigma^+{}_1}_{-\sigma^-{}_1}, x_2{}^{+\sigma^+{}_2}_{-\sigma^-{}_2}, ...x_N{}^{+\sigma^+{}_N}_{-\sigma^-{}_N}\}$, combine them to get the 'best' value $\hat{\hat{x}}$

Compatibility check need not apply!

Could be finding the best value for the average height of students in a class

Can frame question as:

Choose w_i such that $\sum w_i x_i$ is unbiassed and has minimum variance

$$\hat{\hat{x}} = \sum w_i(x_i - b_i) \qquad \text{with } b_i = \frac{\sigma^+{}_i - \sigma^-{}_i}{\sqrt{2\pi}} \text{ or } \frac{\sigma^+{}_i - \sigma^-{}_i}{2}$$

Minimisation leads to.

$$w_i = \frac{1/V_i}{\sum_j 1/V_j} \text{ with } V_i = \frac{{\sigma_i^+}^2 + {\sigma_i^-}^2}{2} - \frac{(\sigma_i^+ - \sigma_i^-)^2}{2\pi} \text{ or } \frac{(\sigma_i^+ + \sigma_i^-)^2}{4} + \frac{(\sigma_i^+ - \sigma_i^-)^2}{2}$$

Suggested strategy

Work with quadratic model for b_i, V_i, use straight-line model as sanity check. Or vice versa.

Working with likelihoods. 1/3: Combination of results

Likelihoods combine naturally

$$\ln L(\theta|\hat{\theta}_1, \hat{\theta}_2) = \ln L(\theta|\hat{\theta}_1) + \ln L(\theta|\hat{\theta}_2)$$

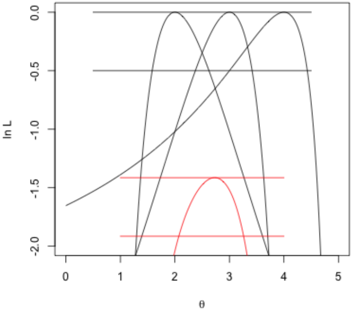
Minimise $\sum_i \left(\frac{\hat{\theta} - \hat{\theta}_i}{\sigma_i + \sigma'_i(\hat{\theta} - \hat{\theta}_i)} \right)^2$

or $\sum_i \frac{(\hat{\theta} - \hat{\theta}_i)^2}{V_i + V'_i(\hat{\theta} - \hat{\theta}_i)}$

σ_i, σ'_i or V_i, V'_i from $\sigma^+{}_i, \sigma^-{}_i$

Solution for $\hat{\theta}$ has to be found numerically, but is well behaved.

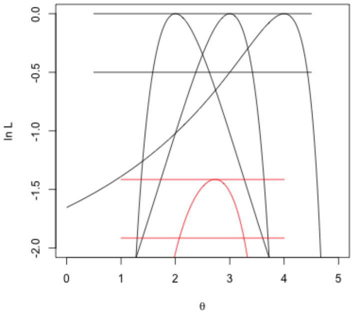
$\Delta \ln L = -\frac{1}{2}$ errors found similarly



Suggested strategy

Work with σ, σ' model, use V, V' model as sanity check. Or vice versa.

Working with likelihoods. 2/3: Goodness of fit



In such combinations, compatibility is essential - these are taken to be different measurements of the same thing.
Given by $\ln L(\hat{\theta})$ and Wilks' theorem ($N-1$ degrees of freedom)

Working with likelihoods. 3/3: Combination of Errors

Taking $f = x + y$ rather than $f(x, y)$ for simplicity:

You know $L(x|data)$ and $L(y|data)$, what is $L(x + y|Data)$?

Answer by taking $\nu \equiv x - y$ as a nuisance parameter and profiling
(or $\nu \equiv y$, or .... anything except $x + y$)
Can use Lagrange multiplier to make it symmetric.

Read off likelihood curve and find $\Delta \ln L = -\frac{1}{2}$ points

Why use different functions

Surely an approximate parabola gives an approximate Gaussian...?

Using "Systematic" Gaussian approximations for "Statistical"

- Dimidated Gaussian has a discontinuity at the peak (from the $\frac{1}{\sigma^{\pm}\sqrt{2\pi}}$ factor) which will mess up maximum likelihood. (Could try 2-armed parabola but suspect it wouldn't do well.)
- Parabolic fit needs solution of quadratic (and both solutions). Messy

Using "Statistical" parabola approximations for "Systematic"

- Using linear σ or V makes integrals needed for $< x >$, $< x^2 >$, $< x^3 >$ impossible analytically

Bringing it all together

Allowed combinations

Responses to the questions 'What do you mean by an error?' and 'Is that a pdf or a likelihood?' are linked.

The likelihood $L(\theta|\hat{\theta})$ for fixed $\hat{\theta}$ can tell you nothing about $V(\hat{\theta})$

The pdf $P(\hat{\theta}|\theta)$ for fixed θ can tell you nothing about the 68% CL region for θ.

The difference between symmetric OPAT and asymmetric OPAT

Both say that $\hat{\theta}$ will lie within the $\pm\sigma$ limits for θ 68% of the time
To make the 68% CL statement about θ we have to assume that the lines on the confidence band plot are parallel
This is true for Gaussians, and CLT encourages us to treat everything as Gaussian until proved otherwise
Asymmetric OPAT clearly breaks this

Bringing it all together

Two sids of the coin

PDFs

Error is variance of result

You are probably combining Errors

Goodness of fit is irrelevant

You are probably not combining results (but you can if you work at it)

"Systematic" Asymmetric Error formulæ

Likelihoods

Error is 68% central CL

You are probabaly combining Results

Compatibility vital & straightforward

You are probably not combining errors (but you can if you work at it)

"Statistical" Asymmetric Error formulæ

Conclusions

This is where my thoughts have got to, and I now think I've got my head round the topic

I would really benefit from exploring these ideas with other practitioners and experts

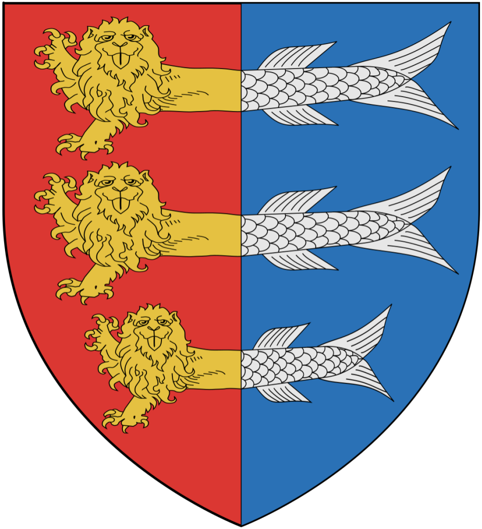
Discussion, helpful criticism, examples, further ideas, filling in details, and collaboration, all very welcome

Big and definitive paper on 'Asymmetric Errors' should be ready to go in a few months

Backup slides

Dimidation

The arms of Great Yarmouth



Why adding positive and negative sigma separately is manifestly wrong.

Let $x = x_1 + x_2 + ...x_N$, and let all the x_i have the same errors:
$\sigma^+ = 1.0, \sigma^- = 2.0$
Adding separately in quadrature gives $\sigma_x^+ = \sqrt{N}, \sigma^- = 2\sqrt{N}$.
The distribution for x has the same distribution as the original x_i, apart from a change in scale.
This breaks the central limit theorem. No matter how large N is, it will never become Gaussian.

Considering x_1 and x_2. They may both fluctuate positively, and this is described by the positive sigmas. Or they may both fluctuate negatively, according to the two positive sigmas. But also one may go positive while the other goes negatvie. (50% chance). which fills in the central region of the distribution, making it more Gaussian.

Open questions

1. Is $\Delta \ln L = -\frac{1}{2}$ appropriate?
2. What about other Gaussian-like functions (Johnson's SU functions, Azzolini's skew-normal...)?
3. Should we worry about second derivatives in combination-of-errors?

The PDG method