Asymmetric Uncertainties

Roger Barlow
The University of Huddersfield

Terascale Statistics School, DESY, Hamburg

27th February 2025

PHYSICS AT THE TERA SCALE
Helmholtz Alliance

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 they 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 might be interesting.
2. Functions which are known to be asymmetric (Poisson, logNormal...) are not part of this problem, as for them we have full information.
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

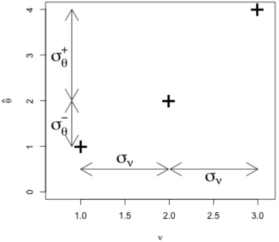
Why adding positive and negative sigma separately is manifestly wrong.

Let $x = x_1 + x_2 + ...x_N$, and let all the x_i have the same errors:
$\sigma^+ = 2.0, \sigma^- = 3.0$
Adding separately in quadrature gives $\sigma_x^+ = 2.0\sqrt{N}, \sigma_x^- = 3.0\sqrt{N}$.
So the distribution for x is the same as the original for 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 negative sigmas. But also one may go positive while the other goes negative (50% chance) which fills in the central region of the distribution, making it more Gaussian.

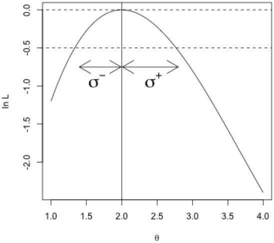
Two sources of Asymmetric Errors

"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

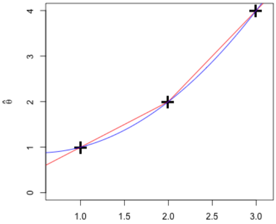
"Statistical"
From ML
estimation

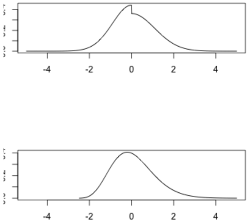


Likelihood as function of θ
Read off $\hat{\theta}$ from the position of the peak, and the errors from the $\Delta \ln L = \pm\frac{1}{2}$ points
If curve is a parabola, these are equidistant.
If not equidistant, report asymmetric errors

"Systematic" AsymmetrIc Errors

R.B. *Asymmetric Systematic Errors*.arXiv:physics/0306138v1 (2003).





Consider 2 models for 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.

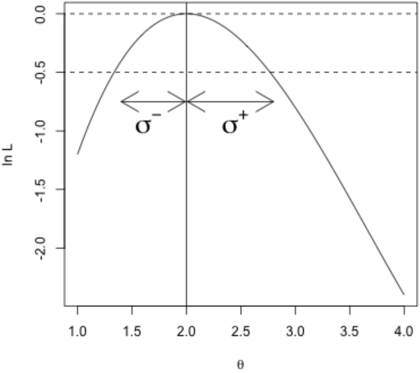
ν is gaussian so $\hat{\theta}$ is distributed with a dimidiated (or bifurcated, or...) gaussian (Model 1) or a distorted gaussian (Model 2)

This enables us to handle the errors. Not perfectly, but adequately.

“Statistical” AsymmetrIc Errors

R.B. *Asymmetric Statistical Errors* arXiv;physics/0406120v1 (2004)

Possible distortions of a parabola
Cubic, restricted quartic, generalised Poisson, log-normal, and PDG recipe (parabolas outside $[\sigma^-, \sigma^+]$, linear interpolation inside)
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.

Combining the two types

RB, Alessandra Brazzale, Igor Volobouyev, *Asymmetric Errors*, arXiv:2411.15499 (2024)

3 questions

What is an error?

Are you talking about σ^2 as variance or about $[\mu - \sigma, \mu + \sigma]$ as the 68% confidence region?

Pdf or likelihood?

The Gaussian $L(x;a) = \frac{1}{\sqrt{2\pi}\sigma} exp^{-\frac{1}{2}(x-a)^2/\sigma^2}$ is symmetric in x and in a. Is your asymmetry talking about a pdf ("systematic") or a likelihood ("statistical")?

What are you using them for

Combination-of-errors or combination-of-results (meta-analysis)

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)

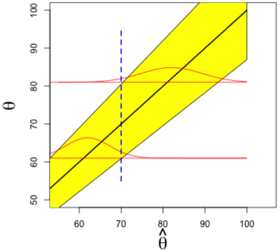
The 68% central confidence region: θ lies between $\hat{\theta} - \sigma$ and $\hat{\theta} + \sigma$

Equivalent for Gaussians but not for non-Gaussians?
Definition (2) is good as 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 in quadrature.
So we need both.

Asymmetry in pdf or likelihood

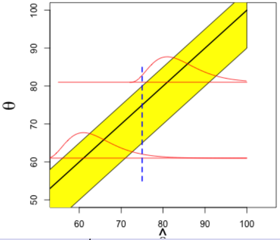
Strongly linked to previous question

In a Neyman confidence-belt, 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.

$V(\hat{\theta})$ has no meaning for the likelihood. You can define a 68% CL region for a pdf but that's not what's quoted.

Combination-of-errors and Combination-of-results

Weakly linked to previous question

Combination-of-errors emerges naturally for pdfs. As in the combination-of-errors formula and standard error analyses.

Combination-of-results from pdfs can be considered a special case of combination of errors. $\hat{\Theta} = \sum_i w_i \hat{\theta}_i$ with suitable weights w_i

Combination-of-results emerges naturally for likelihoods. The total $\ln L = \sum_i \ln L_i$ can be maximised and the $\Delta \ln L = -\frac{1}{2}$ points found,

Combination-of-errors can be considered for likelihoods as an instance of profiling. If $z = x - y$ where $L(x)$ and $L(y)$ are known then profile $x - y$ with $x + y$ as a nuisance parameter.

Working with pdfs. 1/3: Models

Many more models, some better than others.

Name	Description	Range of A	Handles flipping?	Notes
Dimidiated Gaussian	OPAT with 2 straight lines	± 1	Special case	Discontinuity which can give problems when fitting
Distorted Gaussian	OPAT with a parabola	± 0.57382	Yes	Limited range
Railway Gaussian	OPAT with parabola morphing to straight lines	± 0.60467	Yes	Arbitrary smoothing
Double cubic Gaussian	OPAT with two cubics morphing to straight lines	± 0.74538	Yes	Arbitrary smoothing
Symmetric beta Gaussian	OPAT with polynomial morphing to straight lines	± 1	Yes	Two arbitrary tuning parameters
QVW Gaussian	Gaussian with σ linear function of cumulant	± 0.68269	No	Messy numerically
Fechner distribution	Two half Gaussians of same height	± 0.21564	No	Little motivation
Edgeworth expansion	Edgeworth expansion	Very limited	No	Goes negative
Skew Normal	Azzalini's form	± 0.21564	No	
Maximum entropy Johnson	Johnson S_U	± 0.34719	No	Kurtosis from maximum entropy
Log normal	Takes exp of Gaussian distributed variable	± 1	No	Limited range

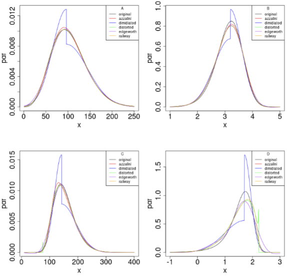
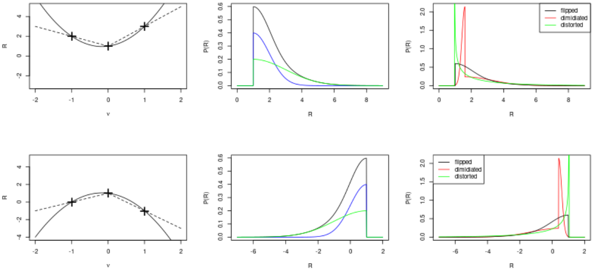


Figure 6: Pdfs from transformations of Gaussian distributions, and the model fits to them. Panel A is the distribution for x^2, where x is distributed according to $\phi(10.0, 2.0)$. B is for $\sqrt{x}$ with x from $\phi(10.0, 3.0)$. C is e^x with x from $\phi(5.0, 0.25)$ and panel D is $\ln x$ with x from $\phi(5.0, 2.0)$.

Check by taking Gaussian in x and pdfs for $x^2, \sqrt{x}, e^x, \ln(x)$, and attempting to reconstruct them from just the $\sigma_\pm$ points.
Generally OK agreement except for dimidiated near centre.

Working with pdfs. 1/3 (continued): Flipped cases

Sometimes the OPAT changes may have the same sign



The models can deal with this, sometimes automatically, sometimes as a special case. E.g. replace flipped dimidiated form by equivalent with same moments,

Not very satisfactory - but functional and hopefully good enough.

Working with pdfs. 2/3: 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, variances still add. So do biases and so does the un-normalised skew: $\gamma = \left\langle x^3 \right\rangle - 3 \left\langle x \right\rangle \left\langle x^2 \right\rangle + 2 \left\langle x \right\rangle^3$

Care necessary as asymmetric pdf is biassed: $\theta(< \nu >) \neq < \theta >$. Central value is not the mean (but it is the median)

Problem: Convolution of two model near-gaussians does not give curve from same family

Suggested recipe. Choose a model and then:
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. 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

$\hat{\hat{x}} = \sum w_i(x_i - b_i)$ with b_i from the model

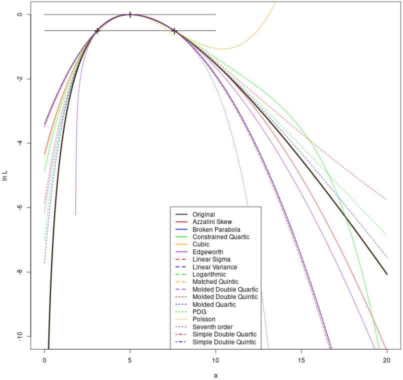
Minimisation leads to. $w_i = \frac{1/V_i}{\sum_j 1/V_j}$ with $V_i =$from the model

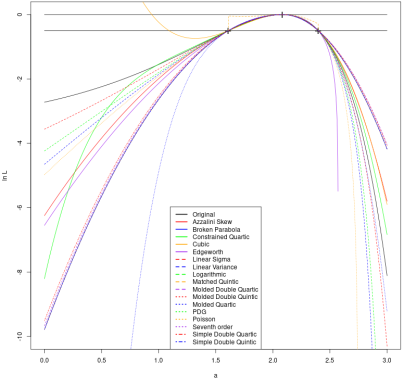
Suggested strategy

Work with one model for b_i, V_i, using another model as sanity check.

Working with likelihoods. 1/4: Models

We have many more models to choose from





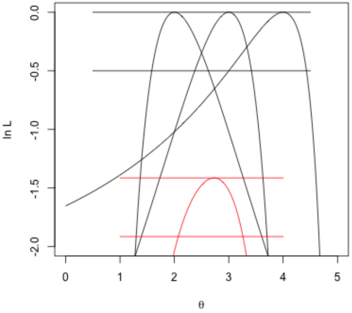
Modelling a Poisson likelihood from 5 events, quoted as $5.00^{+2.58}_{-1.92}$, and from $\log(8 \pm 3)$ coded as $2.08^{+0.32}_{-0.47}$

Working with likelihoods. 2/4: Combination of results

Likelihoods combine naturally

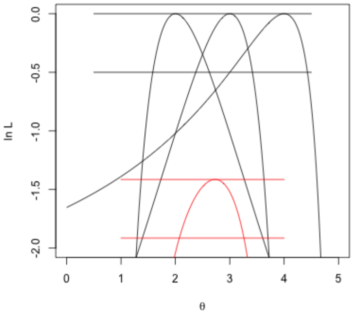
$$\ln L(\theta|\hat{\theta}_1, \hat{\theta}_2) = \ln L(\theta|\hat{\theta}_1) + \ln L(\theta|\hat{\theta}_2)$$

We do the same, using the models for the likelihoods as we don't have the originals
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.



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. 4/4: 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 $x + y$)

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

Example: lifetime measurements

Suppose an experiment measures a lifetime τ from 3 decays. The values happen to be 1.241, 0.592, and 0.988, in some time units. Maximising $\ln L$, which is $\ln \exp(-t/\tau)/\tau$, and using $\Delta \ln L = -\frac{1}{2}$ gives a result $\tau = 0.940^{+0.841}_{-0.385}$.

The experiment then measures 3 more lifetimes, which happen to be 0.834, 2.964, and 0.176, which combined on their own give $1.325^{+1.184}_{-0.542}$. If we combine all 6 values we get the best result $1.1325^{+0.6225}_{-0.3598}$. These are all 'correct' values in the sense that they use the fact that the likelihood is exponential.

Now suppose we take the two partial results separately (i.e. just the value and $\pm$ errors, as quoted above) and combine them using the linear-sigma model. That gives $1.1323^{+0.6213}_{-0.3604}$. The linear-variance model gives $1.1318^{+0.6249}_{-0.3577}$. Both these agree well with the 'correct' value, both in the central value and the quoted errors. Both models (which know nothing about the fact that this was a lifetime measurement with an exponential likelihood) give a result very close to the full-information 'correct' one.

Example: Poisson counts

Experiment sees 5 events in an hour, quoted as $5^{+2.581}_{-1.916}$.
This continues for another hour and 5 events are again seen.
The total gives a result $10^{+3.504}_{-2.838}$ so number of events per hour is $5^{+1.752}_{-1.419}$.

But it could be that knowledge is suppressed, and we just have two estimates to be combined. Table shows results from 4 models

$\hat{a}_1$	$\hat{a}_2$	Linear σ	Linear V	Skew normal	Quartic
$5^{+2.581}_{-1.916}$	$5^{+2.581}_{-1.916}$	$5.000^{+1.737}_{-1.408}$	$5.000^{+1.748}_{-1.415}$	$5.000^{+1.732}_{-1.395}$	$5.000^{+1.719}_{-1.399}$
$6^{+2.794}_{-2.128}$	$4^{+2.346}_{-1.682}$	$4.998^{+1.778}_{-1.432}$	$5.000^{+1.759}_{-1.425}$	$5.009^{+1.766}_{-1.460}$	$4.992^{+1.815}_{-1.440}$
$7^{+2.989}_{-2.323}$	$3^{+2.080}_{-1.416}$	$5.038^{+1.937}_{-1.530}$	$5.009^{+1.794}_{-1.456}$	$5.002^{+1.744}_{-1.594}$	$5.107^{+2.147}_{-1.651}$
$8^{+3.171}_{-2.505}$	$2^{+1.765}_{-1.102}$	$5.401^{+2.368}_{-1.826}$	$5.054^{+1.856}_{-1.516}$	$4.689^{+1.635}_{-1.589}$	$5.282^{+1.277}_{-1.707}$
$9^{+3.342}_{-2.676}$	$1^{+1.358}_{-0.6983}$	$7.348^{+3.149}_{-2.549}$	$5.201^{+1.942}_{-1.605}$	$3.633^{+1.426}_{-1.414}$	$3.209^{+0.5942}_{-0.8277}$

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 sorts of asymmetric error

From PDFs

Error is variance of result

You are probably Combining Errors, in quadrature + skew

Goodness of fit is irrelevant

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

"Systematic" Asymmetric Error formulæ

From Likelihoods

Error is 68% central CL

You are probably Combining Results

Compatibility vital & straightforward

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

"Statistical" Asymmetric Error formulæ

Software

R code on
`https://barlow.web.cern.ch/programs/AsymmetricErrors.tar.gz`
Can download, or do a direct `install` from inside R
Instructions and examples in the paper.

Python code (front-end to C++) from
`https://github.com/igvgit/AsymmetricErrors` and
`https://github.com/igvgit/AsymmetricErrorsPy`
Instructions and examples with the code.

Conclusions

Avoid Asymmetric Errors when possible

If not possible, read the paper, download the software, and use it.

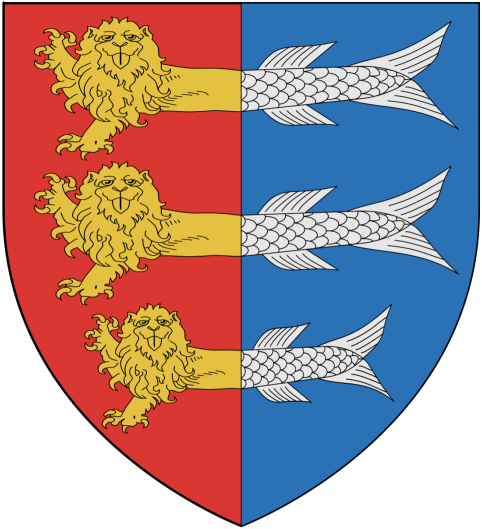
Need to try all these models in different scenarios to find the most useful ones.

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

Backup slides

Dimidation

The arms of Great Yarmouth



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?

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 μ, NOT by skew in pdf.

An unhelpful example