Nuclear Instruments and Methods in Physics Research A 1082 (2026) 170857

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

Full Length Article

Asymmetric errors

Roger Barlow [a,*], Alessandra Rosalba Brazzale [b], Igor Volobouev [c]

[a] The University of Huddersfield, Huddersfield, UK
[b] University of Padova, Padova, Italy
[c] Texas Tech University, Lubbock, TX, USA

ARTICLE INFO

Keywords:
Systematic error
Statistical error
Asymmetric
Confidence region

ABSTRACT

We present a procedure for handling asymmetric errors. Many results in particle physics are presented as values with different positive and negative errors, and there is no consistent procedure for handling them. We consider the difference between errors quoted, using pdfs and using likelihoods, and the difference between the rms spread of a measurement and the 68% central confidence region. We provide a comprehensive analysis of the possibilities, and software tools to enable their use.

1. Introduction

1.1. Background and motivation

Results in particle physics are often given with errors which are asymmetric, of the form $R^{+\sigma^+}_{-\sigma^-}$For instance, taking some results from the latest (at the time of writing) EPS conference:

- ATLAS quote their measurement of the Higgs width as $\Gamma_H = 4.5^{+3.3}_{-2.5}$ MeV [1].
- CMS quote the same quantity as $\Gamma_H = 3.2^{+2.4}_{-1.7}$ MeV [2].
- NOvA have measured the neutrino CP violating parameter as $\delta_{CP} = 0.82^{+0.27}_{-0.87}\,\pi$ [3].
- Belle II quote the branching ratio for $B \to \rho^+\rho^0$ as $(23.2^{+2.2}_{-2.1}(stat.) \pm 2.7(sys.)) \times 10^{-6}$ [4].
- LHCb give the difference in the decay width for the neutral B mesons as $\Gamma_s - \Gamma_d = -0.0056^{+0.0013}_{-0.0015}(stat.) \pm 0.0014(sys.)\,\text{ps}^{-1}$ [5].

Despite their widespread use, their interpretation and their handling is unclear and discussion in the literature is limited [6–11]. This paper explores the reasons why they appear, and gives methods for their consistent handling. It addresses the basic question of how to use results presented as $R^{+\sigma^+}_{-\sigma^-}$, in the absence of any further information.

In practice such errors spring from three causes:

1. when systematic errors are being studied, often using the OPAT ("One Parameter At a Time") method when profiling out the nuisance parameters in the full likelihood is not feasible. In evaluating a systematic error due to the uncertainty (which can be Bayesian or frequentist) of a nuisance parameter ν one can change it by $\pm\sigma$ to observe the shift this produces in a result, and, as illustrated in Fig. 1, the upward and downward shifts are different. Alternatively, when it is technically possible, many values can be generated according to a Gaussian distribution, to find and quantify the distribution of the results, and this may not be symmetric.
2. from maximum likelihood estimation with errors given by the $\Delta \ln L = -\frac{1}{2}$ method, and the log likelihood is not a symmetric parabola. This is also shown in Fig. 1. (The $\Delta \ln L = -\frac{1}{2}$ method is an approximation, but its use is widespread, essentially universal, and in what follows this is what is meant by 'likelihood errors'.)
3. when a random variable is distributed according to a non-Gaussian distribution, such as a Poisson distribution with small mean. This includes as an important special case the replacement of a Gaussian random variable by another variable according to a non-trivial transformation, as happens on a plot with a logarithmic scale.

The third case does not hold conceptual problems as one has, in principle, full knowledge of the underlying distribution function. The challenges come with the first and second cases, where one has to handle the quoted value and errors with no further information about how they arose. This prompts us to consider more closely some fundamental facts about the nature of 'errors' which are normally hidden from view by the convenient properties of the Gaussian function.

* Corresponding author.
E-mail address: Roger.Barlow@cern.ch (R. Barlow).

https://doi.org/10.1016/j.nima.2025.170857
Received 11 December 2024; Received in revised form 28 June 2025; Accepted 14 July 2025
Available online 7 August 2025
0168-9002/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

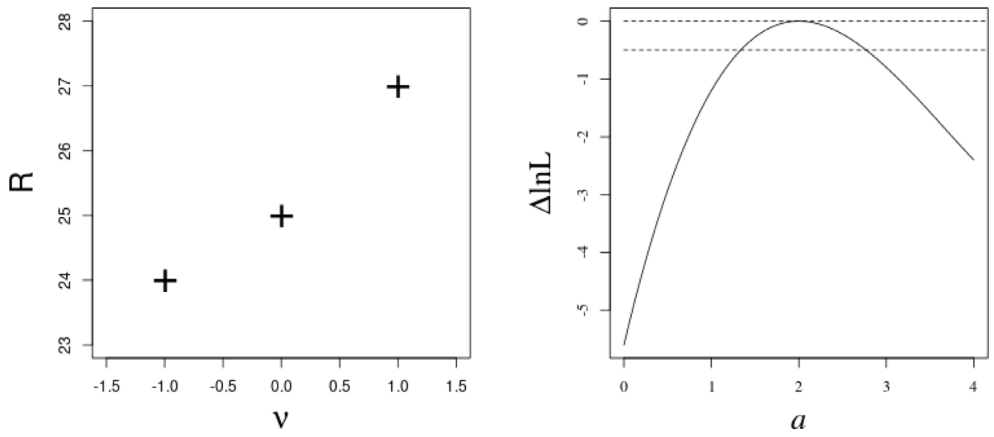

Fig. 1. The two classes of asymmetric error, from OPAT systematic studies of changes to a result R in response to a nuisance parameter ν (left) and maximum likelihood estimation for a parameter a (right).

For the avoidance of confusion we point out that in this paper σ^- is positive, like σ^+. And if R falls with ν (rather than rising as shown in Fig. 1), then σ^+ and σ^- are still the magnitudes of the positive and negative deviations in R, not in ν. If (as occasionally happens) both deviations are in the same direction then special care is needed, as described in the section on 'flipped' distributions, Section 2.1.1.

1.2. Beyond the Gaussian

The Gaussian ("Normal") distribution is described by the density function

$$\phi(x;a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-a)^2}{\sigma^2}} \tag{1}$$

for a measurement x of some value a. This is what is encoded in any reported result of the form $\hat{a} = x \pm \sigma$. When a result is reported as $\hat{a} = x^{+\sigma^+}_{-\sigma^-}$ then this is making a clear analogy with the Gaussian form. But this similarity conceals two important questions which cannot be evaded.

The first is whether we are talking about a function which is asymmetric in x or in a. Eq. (1) is, of course, symmetric in both. Sometimes we consider it as a function of x for a given a, in which case it must be normalized to 1 and we call it a pdf (probability density function). Sometimes we consider it as a function of a for fixed data x, in which case we call it a likelihood and normalization is irrelevant. So the first question to consider in any discussion of 'asymmetric errors' is: what is asymmetric, the pdf or the likelihood?

The second question is the meaning of the error. Specifically, are we talking about the rms spread of the distribution, or the 68% central confidence interval? These happen to be the same for a Gaussian distribution, but not in general. In Eq. (1) the parameter σ describes both the rms spread of the distribution about the mean, $\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$, and the 68% central confidence region for a: $[x - \sigma, x + \sigma]$, and we do not need to be concerned with the difference. But for non-Gaussian distributions we have to know which definition of σ we are using.

The difference between rms spread 'pdf errors' and confidence interval 'likelihood errors' requires clarification. They both refer to uncertainties, the former from random probability and the latter from statistical inference. The Poisson process provides a simple illustration (though for a discrete probability distribution rather than a continuous pdf): the former describes the uncertainty of predicting results from a particular true mean, the latter the uncertainty of inferring the true value of the mean from a particular result.

Pdf errors arise especially through measurements and their errors in experimental work. At high school and university the physicist is firmly taught never to quote an experimental result as a single number x, but *always* to give its error σ. That σ is the standard deviation of the Gaussian distribution, as given in (1), from which this value is a sample. If, for better accuracy, the mean of N measurements is quoted then it is accompanied by the standard error $\sigma/\sqrt{N}$ which is the standard deviation of the pdf of $\overline{x}$. The well-known formula for the combination of errors, given below as Eq. (2), known to statisticians as the delta method, builds on this. It describes the combination of measurement errors: if I is sampled from a Gaussian with standard deviation σ_I, and V with σ_V, then the resistance measurement $R = V/I$ is sampled from a Gaussian for which (assuming the function can legitimately be linearized) $\sigma_R^2 = \sigma_V^2/I^2 + V^2\sigma_I^2/I^4$. The quoted result, $V/I \pm \sigma_R$, is saying the value has been drawn from a Gaussian distribution of unknown mean and standard deviation σ_R.

Such measurements can be converted to confidence intervals using the Neyman confidence belt, as described in Section 40 of the Particle Data Group handbook [12]. For Gaussians this is trivial. In frequentist usage, to say "I have measured a to be $x \pm 3$". means "I have a measurement x from a device which returns values distributed about the true value a according to a Gaussian distribution with a standard deviation of 3: I therefore assert that x falls within the interval $[a-3, a+3]$ with probability 68%, which in the Gaussian case can be turned into saying that, given x, the interval $[x-3, x+3]$ covers the true value a with 68% confidence. (It does not mean "The true value of a lies within $x-3$ and $x+3$ with 68% probability"., that would be a Bayesian credible interval, though this subtle distinction may not always be appreciated.)

If the pdf is not Gaussian – and for asymmetric errors this must be the case – then this trivial equivalence breaks. 'Pdf errors' describe the standard deviation and 'likelihood errors' the 68% central confidence interval. To take an artificially exaggerated but hopefully illuminating example, suppose a measuring device has a 50% chance of recording a measurement below the true value and does so according to a half-Gaussian with $\sigma = 1$, and a 50% chance of recording a measurement above the true value, according to a half-Gaussian with $\sigma = 5$. (This is the dimidiated Gaussian, described later.) If it measures $x = 11.7$ this would be written as $11.7^{+5.0}_{-1.0}$, to describe the pdf. However the Neyman construction for the confidence belt would be $11.7^{+1.0}_{-5.0}$: if 11.7 is a downward fluctuation it cannot have come very far from the true value, whereas an upward fluctuation may have been produced by a true value one distance away. This is shown in Fig. 2.

It is tempting to suggest that any result presented with an asymmetric error should denote a confidence-level, and there is a case for doing so for final results. However in an analysis there will be intermediate values used to construct a final result using addition-in-quadrature (with the aid of the combination-of-errors formula), and this is done in the framework of pdfs and their moments. Variances add under convolution, but 68% central regions do not, requiring more complicated meta-analysis procedures. In a typical particle physics

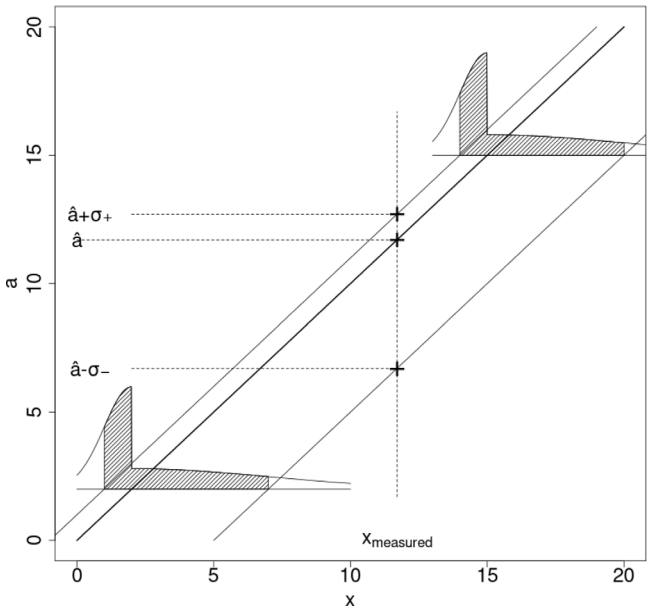

Fig. 2. The Neyman confidence belt referred to in the text, showing how a positive pdf asymmetry $11.7^{+5.0}_{-1.0}$ matches a negative likelihood asymmetry $11.7^{+1.0}_{-5.0}$. The shaded areas display 68% confidence regions of the pdf. This is an artificial example to make the point clear: real life is more complicated.

analysis, there are many intermediate steps between the raw data and the final quoted figure, and typically for these the errors quoted are properties of the pdf.

If we are considering a pdf $p(x;a)$, then we can integrate it to obtain the spread (as given by the variance) of the measured x for a given value of a and the 'errors' are expression of that spread. It tells us nothing about a, as that is given. If we are dealing with a likelihood function $L(a;x)$ we can get the 'best' estimate $\hat{a}$ and take the 68% CL central region from the $\Delta \ln L = -\frac{1}{2}$ points,[1] but we can say nothing more about x. It is true that a pdf can give 68% central limits, but in the language of the Neyman confidence belt they are the horizontal lines constructed on the plot shown in Fig. 2, not the vertical ones from which the confidence interval for a result is read off. So an 'asymmetric error' may refer to the variance and skewness of a pdf, or to the confidence region for a likelihood. Both cases arise in practice. We will therefore consider asymmetric pdfs and the variances (and higher moments) of x, and also, separately, asymmetric likelihoods and the confidence regions for a.

Usage varies in the literature, but within this paper we use the variable names x and sometimes y to refer to results of measurements, and a to refer to an ideal 'true' value. Thus pdfs are considered as functions of x and likelihoods as functions of a. Alternative names may be used in particular examples, as when R and ν (for 'Result' and 'nuisance') are used in the OPAT plot of Fig. 1 rather than $\hat{a}$ and x.

One then has to consider how the 'errors' are going to be used. Again there are two possibilities: Combination of Errors and Combination of Results.

1.2.1. Combination of errors

The simple formula for combination of errors is how most physicists first meet statistics, in the context of experimental errors and uncertainties. For some function $u = u(x, y)$ where the errors on$x$ and y are statistically independent and small (so that a linear approximation is valid),

$$\sigma_u^2 = \left(\frac{\partial u}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial u}{\partial y}\right)^2 \sigma_y^2 \,. \tag{2}$$

We need to extend it to cope with $\sigma_x^+, \sigma_x^-, \sigma_y^+, \sigma_y^-$ and thus σ_u^+, σ_u^-.

In introductory texts this formula is used in problems like determining the speed from the distance traveled in a given time using $v = x/t$, or the acceleration due to gravity from the length and period of a simple pendulum using $g = 4\pi^2 L/T^2$. In more advanced experimental work it is used in instances like calculating a branching ratio, $Br = \frac{N-b}{\eta N_T}$, from some number N of observed decays into the channel of interest out of a total of N_T decays, with a background b and an efficiency η, or in fitting a straight line (or some more sophisticated function) to a set of values $\{x_i, y_i \pm \sigma_i\}$.

As a very basic pedagogical example we consider the measurement of the length ℓ of a rod by measuring the positions x_1 and x_2 of the two ends, with $\ell = x_2 - x_1$. Working with pdfs, we have a two dimensional pdf $p(x_1, x_2)$. The pdf for ℓ is given by the convolution of the two individual pdfs

$$p(\ell) = \iint p_1(x_1)p_2(x_2)\delta(\ell - x_2 + x_1)\, dx_1 dx_2 = \int p_1(x_1)p_2(\ell + x_1)\, dx_1, \tag{3}$$

where the two random variables X_1 and X_2 are assumed independent. If the pdfs are Gaussian, as shown in the left hand plot of Fig. 3, where the dotted lines are the lines of constant $x_2 - x_1$, then the integral gives a Gaussian for ℓ. This Gaussian has standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$, as variances add. This is what is expressed by Eq. (2).

In a likelihood framework, we suppose that x_1 and x_2 are measurements of (unknown) true values a_1 and a_2 and consider the joint likelihood, as shown in the right hand plot of Fig. 3. For the (symmetric) Gaussian this is the same as the left hand plot, apart from the labels on the axes, but in general it will be different. This two-dimensional plot is reduced to one dimension by profiling: writing $\ell = a_2 - a_1$ and $u = a_2 + a_1$, for each value of ℓ we find the maximum value of $L_1(a_1)L_2(a_2)$, which lies along the dashed line,

$$\hat{u} = \frac{\ell(\sigma_2^2 - \sigma_1^2) + 2(\hat{x}_2\sigma_2^2 + \hat{x}_1\sigma_2^2)}{\sigma_1^2 + \sigma_2^2},$$

and the resulting profile likelihood for ℓ is just

$$-\frac{1}{2}\frac{\left(\ell - (\hat{x}_2 - \hat{x}_1)\right)^2}{\sigma_1^2 + \sigma_2^2}.$$

From this we read off the peak at $\ell = x_2 - x_1$, and the $\Delta \ln L = -\frac{1}{2}$ errors at $\pm\sqrt{\sigma_1^2 + \sigma_2^2}$. This is exactly the same result, for the best value and the 'error', as obtained from the combination of errors formula. So when dealing with Gaussians, the same combination of errors formula is found using pdfs (and convolution) or using likelihoods (and profiling), even though the meaning is subtly different in the two cases, because in the first the resulting error is the root of a variance which happens to span a 68% probability interval, and in the second it is the half-width of a 68% confidence interval. For non-Gaussians we have to consider them separately.

Combination of errors is usually a matter for pdfs, because variances add, even for non-Gaussian pdfs. This is dealt with in Section 2.2. A typical analysis will consider many sources of error, usually labeled as systematic, and the combination of these is a major concern of this paper. In such cases the Central Limit Theorem can be useful as when a large number of uncertainties is combined the overall distribution may be adequately described by a Gaussian, even though several of the contributions are asymmetric. However, there are instances where combination of errors needs to be done using likelihoods, and this is considered in Section 3.4.

[1] This is only an approximation to the Neyman construction, but this point is generally not considered. For a discussion see Ref. [13].

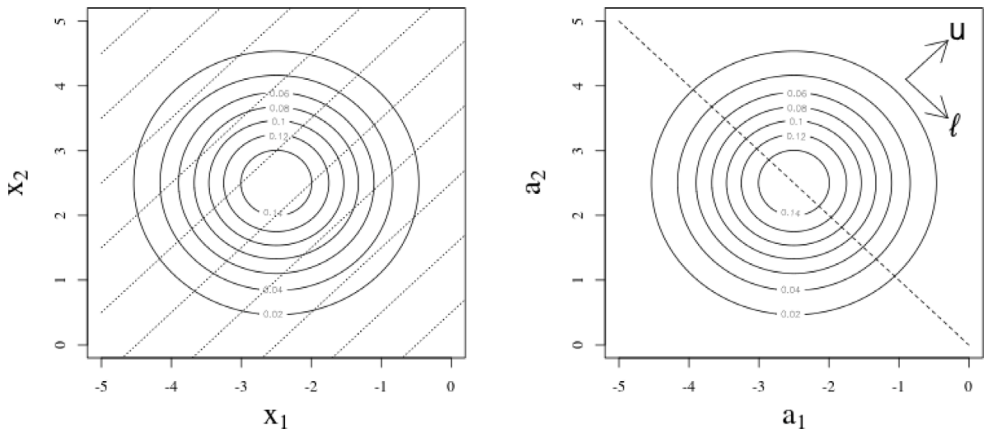

Fig. 3. Combination of errors: $\ell = (2.5 \pm 1.0) - (-2.5 \pm 1.0)$ using pdfs (left) and likelihoods (right).

1.2.2. Combination of results

The second use we need to consider is the combination of results, also known as meta-analysis. Given a set of measurements of the same quantity one wishes to find the appropriate combined best value and its error — and usually one would also want some goodness of fit statistic to describe whether the different results are compatible. This can readily arise when the results are presented in the form of likelihoods (or the maximum likelihood estimator and the 68% central interval obtained from the likelihood function). It is conceptually simple to form the complete likelihood as the product of the individual ones, and find the location of the maximum and the 68% confidence band; there are technical challenges, and Section 3 deals with those. And this is the usual form of the problem. If these are all simple Gaussian measurements this requires the maximization of $-\frac{1}{2}\sum_i \left(\frac{a_i - \hat{a}}{\sigma_i}\right)^2$, which occurs at $\hat{a} = \frac{\sum a_i/\sigma_i^2}{\sum 1/\sigma_i^2}$. (The index i runs over the results being combined; each result may involve many data points but that is not relevant here.)

But it is also possible that one has a set of results presented as pdfs: the $\hat{a}_i$ are estimators of the true a, they are functions of the data x_i and are therefore described by pdfs. If they are unbiased then any weighted sum $\sum_i w_i \hat{a}_i$ is an unbiased estimate provided $\sum_i w_i = 1$and one has the freedom to choose the w_i to minimize the variance on this, using the combination of errors formula, Eq. (2). For Gaussian errors this gives the well known prescription $w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$: the weight is proportional to the inverse square of the error. Hence for Gaussian errors the combination results using pdfs and using likelihoods gives the same answer, even though the meaning of the answer is subtly different. Again, for non-Gaussian errors this is not the case and we must consider both separately.

1.3. Summary on combination-of-errors and combination-of-results

The distinction between combination of errors and combination of results turns out to require clarification, perhaps because both are considered as 'the error analysis'. The fundamental difference is that, in the first case, we want to quantify the precision of a combination, which involves different quantities, while in the second case, we compute the precision of an updated measurement of the same quantity, which we obtain by combining different measurements of the same quantity. In the first case, precision goes down, because uncertainty adds. In the second case, precision goes up because we measure something with more data. (In statistics, the same name is kept for the precision, but different names are used for the outcome: measurement in the first case, estimate in the second.)

Combination-of-errors fits naturally into the framework of pdf errors, though combination of results is perfectly possible where the result is a weighted sum of measured values. Conversely combination-of-results emerges very naturally from likelihoods, although combination of errors is perfectly possible through profiling.

For clarification, the distinctions are shown in Table 1.

2. Pdf errors

2.1. Modeling pdf errors

The question as to how a model function, whether considered as a pdf or as a likelihood, should best be written in a non-Gaussian way is a very open one. In this section we consider pdfs: likelihoods will be dealt with in Section 3. If we restrict ourselves to distributions which look similar to a Gaussian, we can expect them to require an additional third parameter, expressing the asymmetry in some way: we do not consider extensions to more than 3 parameters, but the formalism is open should they be needed.

The three parameters may be specified in various ways, we may use:

1. the moments, namely the mean $\mu = \langle x \rangle$, the variance $V = \langle x^2 \rangle - \langle x \rangle^2$, and the unnormalized skewness, $\gamma = \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3$.
2. the quantiles. The median M is the middle quantile $x_{0.50}$, and the 68% (one sigma) central confidence probability region $[x_{0.16}, x_{0.84}]$ can be written as $[M - \sigma^-, M + \sigma^+]$. This can be expressed in the form $M^{+\sigma^+}_{-\sigma^-}$ and will be referred to as "the quantile parameters" in what follows. (The region is actually bounded by the values of the cumulative Gaussian $\Phi(-1)$ and $\Phi(+1)$: though here and elsewhere only 2 significant figures are shown, the exact 1 sigma region is always meant.)
3. parameters appropriate to the particular model.

It might be thought that the mean could be used as an alternative to the median in combination with the bounds of the 68% central confidence region, writing it as $[x_{0.16}, x_{0.84}] = [\mu - \sigma^-, \mu + \sigma^+]$. However, this parameterization has an unhelpful behavior in some models, as will be noted in Appendices A.8 and A.9, and it will not be pursued further.

Any implementation must provide functions to convert any of the 3 parameterizations into any of the others. The algebra for doing so is given in Appendices A.1 to A.11, and the software in Appendices C and D.

Two such models are readily obtained by considering the OPAT plot of Fig. 1 and supposing that the dependence of R on ν be described by two straight lines, or by a quadratic. Under the first assumption, the Gaussian in ν becomes a dimidiated Gaussian: two equal-area half-Gaussians, under the second a somewhat distorted Gaussian. These models have the advantage of being clearly motivated and giving simple algebraic relations between parameters (details are in Appendices A.1 and A.2).

Typical examples are shown in Fig. 4 for two cases, one with a small positive skewness and one with a slightly larger negative skewness. The

Table 1
Combination of errors and results using pdf errors and likelihood errors.

Combining \ Through	PDF (Section 2)	LIKELIHOOD (Section 3)	Notes
ERROR (Sections 2.2 and 3.4) Want to quantify variability of a measurement obtained by combining measurements of **different** quantities.	$X \sim p_x(x; a_x, \sigma_x^+, \sigma_x^-)$ $Y \sim p_y(y; a_y, \sigma_y^+, \sigma_y^-)$ Parameter of interest: $\psi = a_x/a_y$, fitted by $\hat{\psi} = x/y$ (for example). Uncertainty of $\hat{\psi}$ is quantified through $U = X/Y \sim p_u(u; \psi, \sigma_u^+, \sigma_u^-)$. Bayesian view: $A_x \sim p_1(a_x \mid x, \sigma_x^+, \sigma_x^-)$ $A_y \sim p_2(a_y \mid y, \sigma_y^+, \sigma_y^-)$ combined to yield $U = A_x/A_y \sim p_u(u \mid x, y, \sigma_u^+, \sigma_u^-)$	$L(a_1; x)$ and $L(a_2; y)$ $\Downarrow$ $L(a_1, a_2; x, y) = L(\psi, \lambda; x, y)$, with $\psi = a_1/a_2$ parameter of interest (for example) and λ any function of (a_1, a_2) such that (ψ, λ) is a reparametrization of (a_1, a_2)	• The parameter of interest ψ is a combination of other model parameters (say, a_1 and a_2). • n represents the number of random variables (here, X and Y) which provide information (here on a_1 and a_2). These are then combined to provide information on ψ. • Uncertainty goes up! • The procedure is usually based on PDFs (where σ represents an rms error). • If based on LIKELIHOODS (σ then typically represents the spread of an 68% CI), this requires profiling out the nuisance parameters λ which complete the reparametrization from (a_1, a_2) to (ψ, λ).
RESULT (Sections 2.4 and 3.2) Want to quantify confidence of measurement obtained by combining measurements of **the same** quantity.	Weighted average (meta-analysis) $\hat{a} = \sum_i w_i \hat{a}_i$, with $\sum w_i = 1$	Combine $\hat{a}(x_1)^{+\sigma^+(x_1)}_{-\sigma^-(x_1)}$ from $L(a; x_1)$ $\hat{a}(x_2)^{+\sigma^+(x_2)}_{-\sigma^-(x_2)}$ from $L(a; x_2)$ to yield $\hat{a}(x)^{+\sigma^+(x)}_{-\sigma^-(x)}$ from $L(a; x)$, with $x = (x_1, x_2)$.	• There is an unique parameter of interest a. • n is the number of observations available on a. • Uncertainty goes down! • The procedure is usually based on LIKELIHOODS (where σ represents the spread of an 68% CI). • It can be based on PDFs (where σ represents an rms error) using meta-analysis. • Check for goodness-of-fit!
Notes	• σ measures **variability** (e.g. precision of measurement device). • Approximation needs be normalized (to yield a PDF).	• σ measures the **confidence** region (e.g. as expressed by Neyman confidence belt). • Approximation needs be "nearly" parabolic (to work well).	

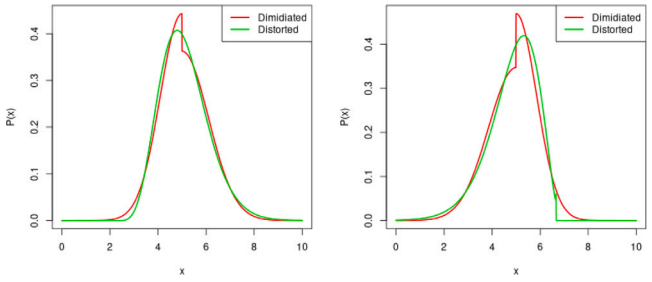

Fig. 4. The dimidiated and distorted Gaussians for $5.00^{+1.0}_{-0.9}$ and $5.00^{+0.85}_{-1.15}$.

two shapes are broadly similar but differ in detail, as one would expect. The dimidiated form has a discontinuity at the central value, but is well behaved elsewhere. The distorted form appears better behaved at the center, but the turnover of the parabola can give a cutoff, as appears on the positive side of the second plot.

These forms are not 'correct': there can be no such guarantee. They give reasonable answers in most practical cases, and we have introduced them here to aid the discussion of ideas. Other models will be described in Section 2.1.2.

2.1.1. *Flipped distributions*

It can happen that an OPAT analysis results in both deviations having the same sign, as illustrated in Fig. 5.

This is implausible but not impossible. If the error concerned is important then this presents a major problem: clearly something significant is going on, and the standard use of linear approximations in evaluating errors is inappropriate. Fortunately such cases are rare. A more typical situation where this arises, and it does arise, is in the evaluation of the contribution of many systematic uncertainties to the total error. Among the many small contributions there may be one or two where both deviations go the same way, though this is often due to statistical fluctuations on a small and unimportant quantity whose treatment will make no difference to the final quoted result. Nevertheless one needs a way of dealing gracefully with such cases. The best advice to the user is to repeat the analysis with more points and more data — but this extra effort may not be justifiable.

The distorted model can handle such a situation. The minimum or maximum is now inside the 68% central confidence region so the need to consider both branches of the parabola forthe pdf is very significant and the Jacobian spike is considerable, but all the algebra holds provided care is taken of the signs of σ^- and σ^+.

For the dimidiated model the distribution in R is the sum of two half-Gaussians, with the same nominal mean, which is actually the extreme or cut-off value $\mathcal{M}$. If the 'errors' are σ_1 and σ_2 and are both

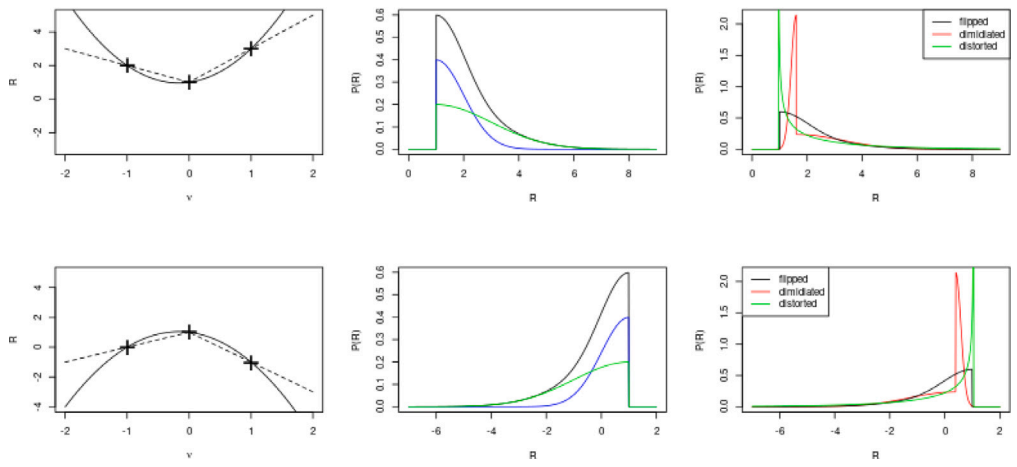

Fig. 5. The top row shows an OPAT analysis where the deviations are both positive, though of differing magnitude, and the dimidiated and distorted fits. The second plot shows the dimidiated distribution: the total (black) is the sum of the two half-Gaussians (green and blue). The third plot also shows this pdf, together with the pdf obtained from the same 3 points using the distorted model, in green, and that from the equivalent dimidiatedpdf in red. The bottom row shows the samefor two negative deviations.

positive, as in the upper row of Fig. 5,then, from the properties of the Gaussian, the combined distribution has moments

$$\mu = \mathcal{M} + \frac{\sigma_1 + \sigma_2}{\sqrt{2\pi}}, \quad V = \frac{\sigma_1^2 + \sigma_2^2}{2} - \frac{(\sigma_1 + \sigma_2)^2}{2\pi},$$
$$\gamma = 2\frac{\sigma_1^3 + \sigma_2^3}{\sqrt{2\pi}} - 3\frac{\sigma_1^2 + \sigma_2^2}{2}\frac{\sigma_1 + \sigma_2}{\sqrt{2\pi}} + 2\left(\frac{\sigma_2 + \sigma_2}{\sqrt{2\pi}}\right)^3. \tag{4}$$

It would be possible to introduce a specific flipped-dimidiated model using this algebra, but this seems an unnecessary complication to handle a situation which is either so serious that a simple 3-point analysis is inadequate, or so trivial that its impact will be small. We suggest instead that, when a flipped distribution is encountered, it be replaced by a conventional dimidiated distribution with the same moments as given by Eq. (4), and that the software (Appendix C–D) should provide a tool for doing this. This, the red curve in Fig. 5, has the same mean, variance and skewness as the black curve: they may appear different to the eye but this difference is not as large as it looks, and is similar to the differences between the dimidiated and distorted model (in green).

Another scenario in which this formalism might be applied can occur with a discrete nuisance parameter; for example one might need to consider both the (favored) normal and the (disfavored but possible) inverted neutrino mass hierarchy. Suppose a result R_1 is obtained from the basic model or assumption, whereas a less-favored alternative gives R_2. If, say $R_2 > R_1$, one might wish to quote this as $R_{1\,-0}^{\,+|R_2-R_1|}$. We are not recommending this, but if someone chooses to do so we can accommodate it as a flipped pdf where σ^- and σ^+ are equal (or opposite: the signs require careful handling in the software) and given by $|R_1 - R_2|$.

For some models this cannot be handled, and they should not be deployed. The relation between v and R is no longer monotonic, so the quantiles of v do not match the quantiles of R.

2.1.2. Comparison of the models

Many different 3-parameter forms can be used to model asymmetric pdfs, and we have considered several. These are briefly listed in Table 2 and fully described in Appendix A.

The Asymmetry A is here defined as $\frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}$. Many models can handle arbitrary values of σ^+ and σ^- and thus arbitrary A. Others, such as the Fechner and the skew normal, for which the extreme case is a half-Gaussian, cannot model the largest $|A|$. The dimidiated Gaussian can handle arbitrary A but not arbitrary $\gamma/V^{3/2}$.

The upper row of Fig. 6 shows several models, with parameters chosen to match quantile values (on the left) and moments (on the right). For these moderate asymmetries the behavior appears sensible and there is generally little difference between the shapes and the curves coincide,except for the dimidiated Gaussian near the center (where it probably does not matter very much).

Two examples of larger asymmetry are shown in the lower row, and differences appear. Azzalini's skew normal distribution cannot handle these values. The Edgeworth form goes negative and develops a second bump, which is surely not physical. The distorted Gaussian shows a Jacobian spike which, again, is probably an undesired artifact. The railway Gaussian resembles the distorted but avoids the spike.

Other examples can be tried, and show the same behavior: for moderate skewness the functions are well behaved and results are similar; for large skewness they need to be handled carefully.

As a further exploration, we can use some examples of the third original category of Section 1.1, where a Gaussian-distributed variable is transformed to make it non-Gaussian. We apply the various models to the moments of this distribution, and compare the model with the true original. Examples are shown in Fig. 7, taking the square of a Gaussian-distributed variable, the square root, the exponential and the logarithm. (Different means and standard deviations were used, for presentational reasons; square roots and logarithms of negative numbers are ignored.) For small to moderate asymmetries there is general agreement among the models (apart from the dimidiated model near the center, as before) and they match the original. The fourth example shown, panel D, considers the logarithm of a Gaussian variable, and is more interesting in that the models generally disagree with each other and the original.

It is strongly recommended that any calculations with asymmetric errors use (at least) two models, as a check for robustness. The dimidiated and distorted models have the advantage of simplicity, even though some aspects are clearly unrealistic. The railway model is similar to the distorted and avoids the Jacobian spike. On the other hand the Edgeworth model utilizing only three leading cumulants has an inbuilt problem in that it always gives a negative (unphysical) probability for some arguments. Azzalini's skew normal is restricted to small asymmetries. It seems to us that the Edgeworth and skew normal models are not suitable for general-purpose use, which have to handle a wide range of possible asymmetries: they can be used if there is a specific reason for favoring them, and it is known *a priori* that the asymmetries will not be large.

2.2. Combination of pdf errors

The combination of two (or more) errors requires the convolution of the relevant pdfs: given $p_x(x)$ and $p_y(y)$ with $u = u(x, y)$ we require $p_u(u)$. One assumes that the errors are 'small' so the linear approximation

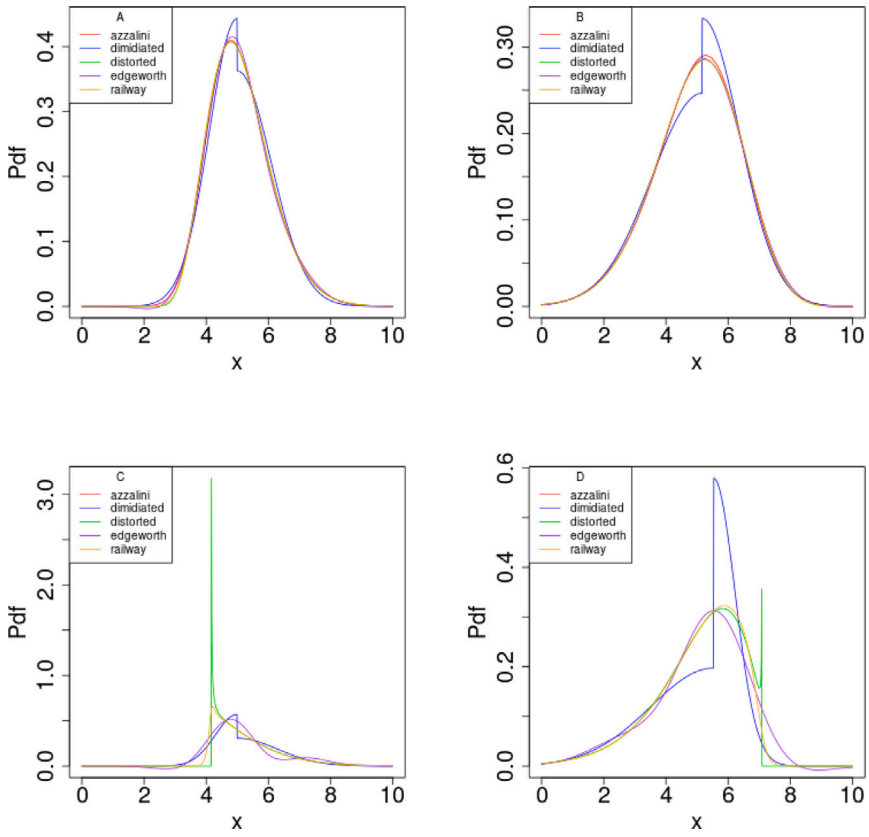

Fig. 6. Pdfs for a fewdifferent models, listed in the legends.Panel A is for $5.0^{+1.1}_{-0.9}$, B is for moments: $\mu = 5.0, V = 2.0, \gamma = -1.0$, C is for $5.0^{+1.3}_{-0.7}$ and panel D for $\mu = 5, V = 2.0, \gamma = -3.0$.

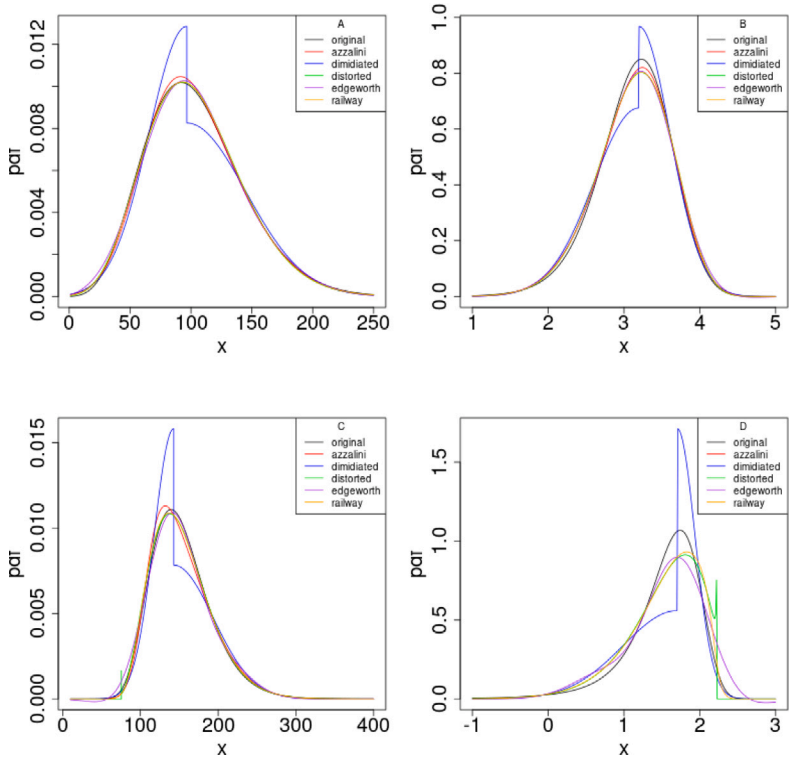

Fig. 7. Pdfs from transformations of Gaussian distributions, and the model fits to them. Panel A is the distribution for x^2, where x is distributed according to $N(10.0, 2.0)$. B is for $\sqrt{x}$ with x from $N(10.0, 3.0)$. C is e^x with x from $N(5.0, 0.25)$ and panel D is $\ln x$ with x from $N(5.0, 2.0)$.

Table 2
Functions that can model asymmetric pdfs.

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[a]	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

[a] Subsequent tables utilize the notation "Symmetric beta Gaussian(p, K)" for this model. This notation supplies shape parameters p and $h = 0.1K$ detailed in Appendix A.5.

Table 3
Combining errors from various pdfs using various models. Δ is the combined median.

	Errors											
	$\sigma_x^- = 1.0,\ \sigma_x^+ = 1.0$ $\sigma_y^- = 0.8,\ \sigma_y^+ = 1.2$			$\sigma_x^- = 0.8,\ \sigma_x^+ = 1.2$ $\sigma_y^- = 0.8,\ \sigma_y^+ = 1.2$			$\sigma_x^- = 0.5,\ \sigma_x^+ = 1.5$ $\sigma_y^- = 0.8,\ \sigma_y^+ = 1.2$			$\sigma_x^- = 0.5,\ \sigma_x^+ = 1.5$ $\sigma_y^- = 0.5,\ \sigma_y^+ = 1.5$		
Models	σ^-	σ^+	Δ	σ^-	σ^+	Δ	σ^-	σ^+	Δ	σ^-	σ^+	Δ
Dimidiated Gaussian	1.32	1.52	0.080	1.22	1.62	0.160	1.09	1.78	0.284	0.97	1.93	0.413
Distorted Gaussian	1.33	1.54	0.098	1.25	1.64	0.203	1.17	1.88	0.349	1.12	2.07	0.534
Railway Gaussian	1.34	1.53	0.098	1.25	1.64	0.199	1.18	1.86	0.325	1.13	2.04	0.484
Double cubic Gaussian	1.32	1.52	0.080	1.23	1.62	0.160	1.11	1.79	0.285	1.00	1.95	0.419
Symmetric beta Gaussian(1,10)	1.32	1.52	0.080	1.23	1.62	0.162	1.12	1.80	0.288	1.01	1.95	0.423
Symmetric beta Gaussian(1,30)	1.33	1.53	0.096	1.24	1.64	0.195	1.17	1.86	0.343	1.11	2.03	0.519
Symmetric beta Gaussian(4,10)	1.32	1.52	0.078	1.22	1.62	0.157	1.11	1.79	0.280	0.99	1.94	0.409
Symmetric beta Gaussian(4,30)	1.33	1.53	0.090	1.23	1.63	0.182	1.15	1.83	0.323	1.07	1.99	0.482
QVW Gaussian	1.32	1.52	0.083	1.23	1.62	0.167	1.12	1.80	0.298	1.03	1.96	0.440
Fechner distribution	1.30	1.52	0.092	1.19	1.62	0.186	–			–		
Edgeworth expansion	1.31	1.52	0.095	1.21	1.62	0.195	–			–		
Skew normal	1.33	1.49	0.101	1.22	1.60	0.190	–			–		
Johnson system	1.24	1.68	−0.006	1.28	1.83	0.156	–			–		
Log normal	1.33	1.59	0.075	1.26	1.70	0.188	0.73	1.93	0.115	0.91	2.42	0.351

can be used. Near the nominal point (x_0, y_0) we write $u(x, y) \approx u_0 + \left(\frac{\partial u}{\partial x}\right)(x - x_0) + \left(\frac{\partial u}{\partial y}\right)(y - y_0)$. Scaling to absorb the differentials we write $\tilde{u} = u - u_0, \tilde{x} = \left(\frac{\partial u}{\partial x}\right)(x - x_0), \tilde{y} = \left(\frac{\partial u}{\partial y}\right)(y - y_0)$, which gives $\tilde{u} = \tilde{x} + \tilde{y}$. So within the limits of the approximation we can express any general $u = u(x, y)$ as a simple sum and for simplicity in the following algebra for error manipulations we will work with $u = x + y$, knowing that the general form can be reclaimed when and if desired. Using this, $p_u(u)$ is given by Eq. (3).In some cases this can be evaluated algebraically, in others it needs to be done numerically.

This presents a problem, in that, although the convolution of two Gaussians gives another Gaussian, this is not in general true for other functions. The convolution of two pdfs, both belonging to one of the parameterizations in Table 2, does not give a function described by that parameterization. See Appendix A.1.1 for a detailed discussion using the dimidiated Gaussian.

To proceed, we note that the moments μ, V, γ add under convolution. To provide a consistent procedure for 'adding errors' we can evaluate the individual moments using Eqs. (17), (25), or their equivalents for other models, sum them to obtain the totals, and then use Eqs. (19), (26), or their equivalents, to find the model parameters and, if desired, the quantile parameters that give a pdf that has these moments.

Table 3 illustrates the results of combining various typical errors, using different models. The behavior is much as one would expect. The combined errors increase in all cases, but the asymmetry falls. This is brought out in the rightmost entry, where two very asymmetric errors ($\sigma^+ = 3\sigma^-$) combine to give an error with only $\sigma^+ \approx 2\sigma^-$. For moderate asymmetries, as in the first two columns, there is reasonable agreement between the models, to two significant figures, but not three, which gives an estimate of how far the precision of such calculations can be trusted. For large asymmetries, as in the third and fourth columns, the agreement is lost, showing that in such cases the accuracy should not be considered as being definite. (Several models fail to accommodate the examples with very high asymmetries, so the entries are blank). After all, any method will break down somewhere.

An important point to note is that in all cases the nominal value shifts: this is the number shown as Δ. If two asymmetric pdfs are convolved, then the median of the result is not the sum of the individual medians. A procedure for combining errors using asymmetric distributions must, to be consistent, include a shift in the central value. If practitioners are not prepared to make such an adjustment, then they should perhaps reconsider the use of asymmetric errors and revert to the usual symmetric form, averaging their σ^+ and σ^- values. The values of σ^+ and σ^- in the table, and returned by the software, give, within the framework of the model concerned, the 68% central region around the nominal point: when the Δ shift is applied the region boundaries should be adjusted accordingly.

2.3. Why the 'usual procedure' is wrong

It is common practice (though we have not been able to find a reference) to combine asymmetric errors by combining all the positive errors in quadrature, and likewise all the negative errors, and then using the dimidiated ("bifurcated") Gaussian. This is obviously wrong. Suppose one has N sources of error, and that they all have the same positive error σ^+ and negative error σ^-. The combined error given by this procedure will have positive error $\sqrt{N}\sigma^+$ and negative error $\sqrt{N}\sigma^-$. The pdf of the combination has a width that increases like $\sqrt{N}$ but *does not change its shape*. But this contradicts the Central Limit

Theorem, which declares that at large N the distribution of a sum tends to a Gaussian (symmetric) shape, whatever the distributions of the components.

Examples of how this can give wrong results are in Section 4.1.2.

2.4. Combination of pdf results

As discussed in Section 1.1, the use of likelihoods to combine results is a very standard procedure and their use to produce a combined error is less mainstream, whereas for pdfs the case is reversed. In fact, for pdfs the combination of results can be considered as an instance of the combination of errors. The measurements are all of the same quantity and combined by taking a weighted mean, rather than distinct quantities combined with some general function.

If several independent estimates $\hat{r}_i$ with pdfs of variances V_i are obtained by different experiments for the same result r, then an obvious way to form a combined result is to take a weighted sum:

$$\hat{r} = \sum_i w_i \hat{r}_i \qquad \text{with} \qquad \sum w_i = 1, \tag{5}$$

where the requirement for the sum of the weights ensures that if the individual estimates are unbiased, so is the combination. For an efficient estimator one wants to minimize the variance of $\hat{r}$. This is given by

$$V_{\hat{r}} = \sum w_i^2 V_i, \tag{6}$$

which is minimized by taking

$$w_i \propto \frac{1}{V_i}, \tag{7}$$

the familiar result that results should be weighted by the inverse of their variance: the point here is that it holds even for non-Gaussian distributions.

For $\hat{r}_i$ we must take the mean of the distribution (as V is the variance about the mean). If the median is given, it must be converted, according to the model being used, before averaging. The skewness of the result is just $\sum_i w_i^3 \gamma_i$ and the resulting moments can be used to give the parameter set of whatever model is being used.

As an example, suppose that a variable x is distributed according to a Gaussian distribution with mean $\mu = 5$ and standard deviation $\sigma = 1/\sqrt{2}$. However the variable of interest is $r = x^2$, which accordingly has quantiles $M = 25$, $\sigma^+ = 7.571$ and $\sigma^- = 6.571$. Suppose that x is sampled twice and, as it happens, the two values are at $\mu - \sigma$ and $\mu + \sigma$. They will be added with equal weight, and the parameters of the pdf of the result extracted using the combination of errors procedure, assuming one of the models for asymmetric pdfs. The results of this are shown in Table 4, together with the results if the two samples happen to be at $\mu \pm 2\sigma$. (Results are shown to more significant figures than the data would normally warrant, to enable comparison of small differences between models.) This example illustrates the point that despite imposing $\sum w_i = 1$, bias in the estimate is essentially inevitable. Unbiased measurements of x will not give unbiased estimates of μ. Any non-trivial transformation of the variable of interest will entail a bias.

Continuing with this example, we use a toy Monte Carlo to generate pairs of 'samples' in x according to a random Gaussian process as described above,combine the two, and histogram the results. This is shown in Fig. 8. The mean, variance and skewness can be found from each histogram and compared with the values expected from the quoted results, shown in Table 5. It can be seen that there is good agreement for the variances in all cases and fair agreement for the skewness, except that the dimidiated Gaussian underestimates the skewness of the pdf of the combination, because the dimidiated model is less influenced by extreme values than the others.

This framework may be unrealistic, in that it assumes that the (asymmetric) errors are independent of the measurement. The quoted errors on the 'input' column in Table 4 are all the same, because the true mean of x is known. This will be complicated by additional information from the experiments which does not appear in the quoted values and errors. This is considered in detail by Schmelling [6] using a second order approximation (the "distorted Gaussian" in this work). Rather than the ideal case where all true quantities are known, he considers the more realistic case where the errors are taken from measurements (as happens when $\sqrt{N_{observed}}$ errors are used for Poisson statistics), and shows that as well as the bias from the difference between median and mean there is a bias from the use of an estimator $\hat{V}$ for the variance, and a further significant bias from the correlation between V and $\hat{r}$. If the parameter of interest is, for example, the square of a Gaussian quantity, then upward fluctuations will be ascribed larger errors, and thus smaller weights.

A combination of results generally brings with it the question of whether such a combination is valid: is it plausible that these are really measurements of the same quantity? This is generally expressed as a χ^2 value, or some other goodness of fit measure. Again, this is readily done when using likelihoods but is rather contrived when using pdfs.

For a given measurement $x_i{}^{+\sigma_i^+}_{-\sigma_i^-}$ we can express its compatibility with some quoted x_c using the p-value or its equivalent expression

Table 4
Examples of combination of errors with pdfs using different models.

Input		Combined result using model	
r_1	r_2	Dimidiated Gaussian	Distorted Gaussian
$32.571^{+7.571}_{-6.571}$	$18.429^{+7.571}_{-6.571}$	$25.700^{+5.252}_{-4.752}$	$25.750^{+5.262}_{-4.763}$
		Railway Gaussian	Double Cubic Gaussian
		$25.749^{+5.261}_{-4.765}$	$25.699^{+5.254}_{-4.754}$
		QVW Gaussian	Fechner Distribution
		$25.707^{+5.255}_{-4.755}$	$25.712^{+5.252}_{-4.749}$
		Edgeworth expansion	Skew normal
		$25.749^{+5.251}_{-4.749}$	$25.741^{+5.255}_{-4.777}$
		Johnson System	Log normal
		$25.742^{+5.333}_{-4.809}$	$25.748^{+5.282}_{-4.776}$
		Sym. beta Gaussian(1,10)	Sym. beta Gaussian(1,30)
		$25.700^{+5.254}_{-4.755}$	$25.741^{+5.260}_{-4.761}$
		Sym. beta Gaussian(4,10)	Sym. beta Gaussian(4,30)
		$25.695^{+5.253}_{-4.754}$	$25.725^{+5.257}_{-4.758}$
r_1	r_2	Dimidiated Gaussian	Distorted Gaussian
$41.142^{+7.571}_{-6.571}$	$12.878^{+7.571}_{-6.571}$	$27.200^{+5.252}_{-4.752}$	$27.250^{+5.262}_{-4.763}$
		Railway Gaussian	Double Cubic Gaussian
		$27.249^{+5.261}_{-4.765}$	$27.199^{+5.254}_{-4.754}$
		QVW Gaussian	Fechner Distribution
		$27.207^{+5.255}_{-4.755}$	$27.212^{+5.252}_{-4.749}$
		Edgeworth expansion	Skew normal
		$27.249^{+5.251}_{-4.74}$	$27.241^{+5.255}_{-4.777}$
		Johnson System	Log normal
		$27.242^{+5.333}_{-4.809}$	$27.248^{+5.282}_{-4.776}$
		Sym. beta Gaussian(1,10)	Sym. beta Gaussian(1,30)
		$27.200^{+5.254}_{-4.755}$	$27.241^{+5.260}_{-4.761}$
		Sym. beta Gaussian(4,10)	Sym. beta Gaussian(4,30)
		$27.195^{+5.253}_{-4.754}$	$27.225^{+5.257}_{-4.758}$

Table 5
Comparison of quoted quantities and toy Monte Carlo results based upon 10^6 toys. The toy results are not identical for all models but agree to five significant digits, $V_{MC} = 25.288$ and $\gamma_{MC} = 38.388$.

Model	V_{pred}	$\frac{V_{pred}-V_{MC}}{V_{MC}}$	γ_{pred}	$\frac{\gamma_{pred}-\gamma_{MC}}{\gamma_{MC}}$
Dimidiated Gaussian	25.045	−0.0096	14.967	−0.610
Distorted Gaussian	25.250	−0.0015	37.750	−0.015
Railway Gaussian	25.249	−0.0015	36.867	−0.038
Double cubic Gaussian	25.083	−0.0081	20.765	−0.458
Symmetric beta Gaussian(1,10)	25.090	−0.0078	21.804	−0.431
Symmetric beta Gaussian(1,30)	25.206	−0.0032	34.141	−0.109
Symmetric beta Gaussian(4,10)	25.071	−0.0086	19.007	−0.504
Symmetric beta Gaussian(4,30)	25.150	−0.0055	28.741	−0.250
QVW Gaussian	25.103	−0.0073	23.277	−0.393
Fechner distribution	25.016	−0.0107	25.157	−0.344
Edgeworth expansion	25.000	−0.0114	37.751	−0.015
Skew normal	25.583	0.0117	33.617	−0.123
Johnson system	26.555	0.0501	44.119	0.151
Log normal	25.593	0.0121	39.367	0.027

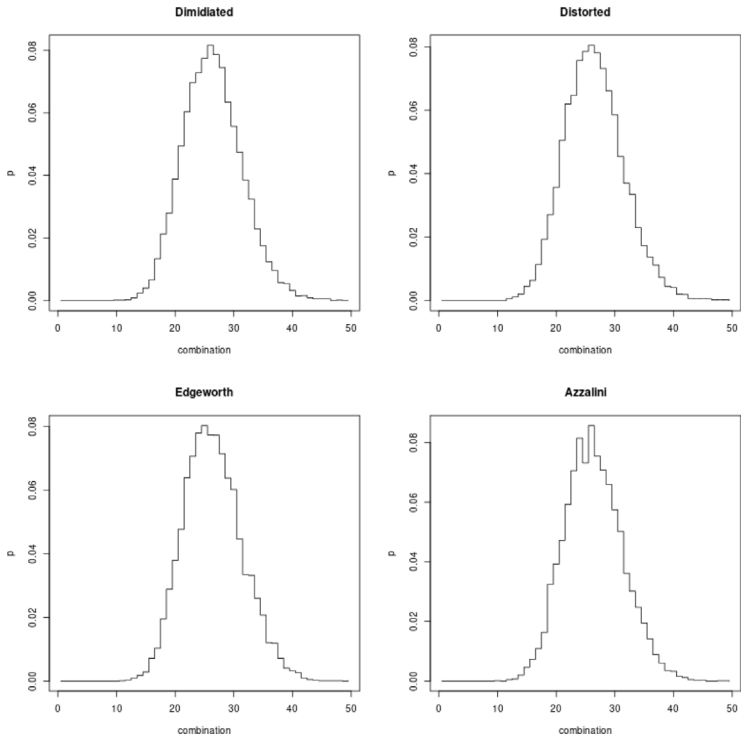

Fig. 8. Example, using Monte Carlo simulations with 10,000 samples, of combination of results using pdfs, taking two random Gaussian numbers and squaring them, both being ascribed errors $^{+7.57}_{-6.57}$, and then combined using different pdf models. From such distributions the quantities V_{MC} and γ_{MC} in Table 5 are calculated.

in terms of sigmas. With the familiar Gaussian, we would say that a measurement of 12.7 ± 0.1 was 5 sigma away from some nominal value of 12.2, and accordingly treat it with strong reservation. A similar approach for asymmetric errors will use σ_i^+ and σ_i^-, but care must be taken over the direction of the deviation. Suppose the proposed true value is 12.2 and the measurement is $12.7^{+0.1}_{-0.2}$. Using the dimidiated model, this deviation is a 5 sigma, not 2.5 sigma, as the quoted measurement implies that a process which may go up a little or down a lot has given a result of 12.7. For other models the arithmetic is not so simple, but the p-value can readily be determined from the pdf.

If we ask about the compatibility of $x_{i\,-\sigma_i^-}^{+\sigma_i^+}$ with some $x_{c\,-\sigma_c^-}^{+\sigma_c^+}$, as opposed to a single value, this can be obtained from the convolution of the two pdfs. This can be done numerically, within the framework of the model adopted.

To consider the agreement of a whole set of N measurements one can convert the individual p-values into their equivalent χ^2 numbers and sum them to get a total χ^2. If x_c were imposed externally, this would have N degrees of freedom; as x_c has been obtained from the measurements it is probably reasonable to use $N-1$ in obtaining the overall p-value.

3. Likelihood errors

3.1. Modeling non-parabolic likelihoods

For the Gaussian the log likelihood function $\ln L$ is a parabola with two parameters, the location and scale. (The parameter giving the value at the peak is irrelevant.) For an asymmetric form a reasonable guess at the full likelihood function will be some curve with three parameters, approximately parabolic, for which the maximum likelihood peak occurs at some $a = \hat{a}$ and $\Delta \ln L = -\frac{1}{2}$ at $a = \hat{a} + \sigma^+$ and $a = \hat{a} - \sigma^-$. Thus the curve must go through 3 points, having a maximum at the middle one. The 3 parameters will be obtainable from the values of $\hat{a}, \sigma^+$ and σ^-. It should also have a 'reasonable' behavior elsewhere; in particular it should go to $-\infty$ at large positive and negative a. Provision of such models is somewhat easier for log likelihoods than for pdfs, as there are only two sets of parameters deployed: these $\hat{a}, \sigma^+, \sigma^-$ and the 3 parameters of the specific model. The software must convert between the two, but there is no third set corresponding to the moments of the pdfs.

Two models which are simple to use and turn out to be generally applicable arise from the suggestion by Bartlett that a likelihood function is described by a Gaussian whose width varies as a function of the argument a [13–15]. With such an assumption one might suppose that the variation is linear, either for the standard deviation ('linear sigma model') or the variance ('linear variance model')

$$\ln L(a) = -\frac{1}{2}\left(\frac{a-\hat{a}}{\sigma + (a-\hat{a})\sigma'}\right)^2 \qquad \text{or} \qquad -\frac{1}{2}\frac{(a-\hat{a})^2}{V + (a-\hat{a})V'}. \tag{8}$$

The parameters, σ, σ' or V, V', which are readily given by the requirement that the curve go through the two $-\frac{1}{2}$ points, are

$$\sigma = \frac{2\sigma^+\sigma^-}{\sigma^+ + \sigma^-}, \qquad \sigma' = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}, \qquad V = \sigma^+\sigma^-, \qquad V' = \sigma^+ - \sigma^-. \tag{9}$$

More details are in Appendices B.14 and B.15.

The use of these two models is illustrated in Fig. 9 which shows the likelihood curves resulting from a small positive asymmetry, $5.0^{+1.1}_{-0.9}$ and a large negative one, $5.0^{+0.5}_{-1.5}$. All curves go through the 3 defining points, by construction. In the central region there is little difference, even in the very asymmetric instance. Differences do appear further from the peak. As a general rule, the choice of model makes little difference for small asymmetries, but can become significant at large excursions for large asymmetries.

If any information is provided, or can be deduced, about the form of the likelihood then that should be used to guide the choice of model — there is no general 'best model', but some models do better for particular classes of result. By the same token, if a result with an asymmetric error is quoted, any further available information should be provided. If all the information available to a meta-analyst is the 3 numbers of the reported measurement, then the wide range of available models can potentially give a wide range of outcomes.

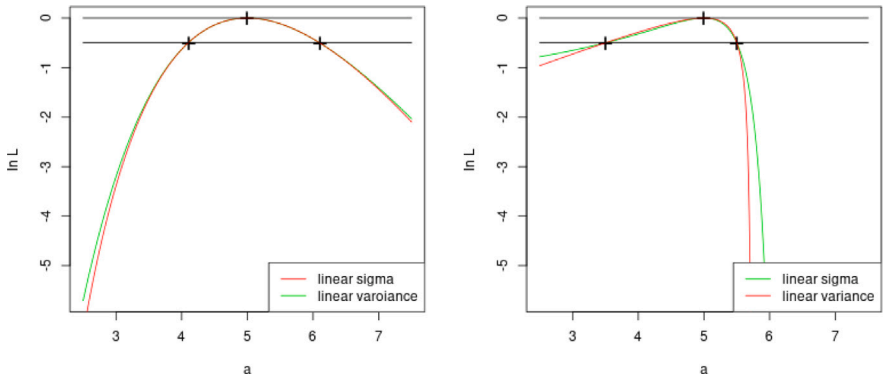

Fig. 9. Using the linear sigma and linear variance models of the log likelihood arising from $5.0^{+1.1}_{-0.9}$ (left) and $5.0^{+0.5}_{-1.5}$ (right).

Table 6
Functions for modeling asymmetric log likelihoods. For simplicity, formulæ assume $\hat{a} = 0$. The 'central region' is between $\hat{a} - \sigma^-$ and $\hat{a} + \sigma^+$. The symbol $\sigma_\pm$ is used to denote "σ^+ for positive values and σ^- for negative values".

Name	Description	Notes
Cubic	$-\frac{1}{2}(\alpha a^2 + \beta a^3)$	Badly behaved, $\ln L$ turns over.
Broken parabola	$-\frac{1}{2}\frac{a^2}{\sigma_\pm^2}$	Simple, but L'' discontinuity at peak
Constrained quartic	$-\frac{1}{2}(\alpha + \beta a)^2$	Quartic constrained at single peak
Molded quartic	$-\frac{1}{2}(\alpha a^4 + \beta a^3 + \gamma a^2)$	Quartic is best match to broken parabola in central region.
Matched quintic	Quintic centrally, $\sigma_\pm$ outside	Quintic coeffts from matching L and L'' at $\sigma_\pm$
Interpolated 7th degree	7th order centrally, $\sigma_\pm$ outside	Coeffts from matching L, L', L''
Simple double quartic	Central quartics, quadratics outside $\sigma_0 = \sqrt{\sigma^-\sigma^+}$	Coeffts from matching
Molded double quartic	Central quartics, quadratics outside $\sigma_0 = \sqrt{(\sigma^{-4} + \sigma^{+4})/(\sigma^{-2} + \sigma^{+2})}$	Coeffts from matching
Simple double quintic	Central quintics, quadratics outside	
Conservative spline	Cubic between spline points, quadratics outside	Spline points need to be chosen in central region.
Logistic beta	$-\alpha \ln(1 + e^{-x}) - \beta \ln(1 + e^{x}) - \ln B(\alpha, \beta)$	Log of Type IV generalized logistic
Logarithmic	$-\frac{1}{2}(\frac{\ln(1+\gamma a)}{\ln \beta})^2$	Parabola subject to scaled expansion. Limited domain
Generalized Poisson	$-\alpha a + \mathcal{N}(1 + \frac{\alpha a}{\mathcal{N}})$	Poisson, but $\mathcal{N}$ need not be integer.
Linear sigma	$\sigma(a) = \sigma + a\sigma'$	Good general-purpose use Finite valid domain
Linear variance	$V(a) = V + aV'$	Good general-purpose use. Finite domain
Linear log sigma	$\sigma = e^{\alpha Q(a)+\beta}$	$Q(a)$ is the cdf of the equivalent Fechner distribution.
Double cubic log sigma	$\sigma = e^{cubics}$ centrally, $\sigma_\pm$ outside	Different cubics for $a \le 0$ and $a \ge 0$
Quintic log sigma	$\sigma = e^{quintic}$ centrally, σ_{σ^+} outside	Quintic by matching L, L', L''
PDG method	$\sigma(a) = \sigma + a\sigma'$ or $\sigma_\pm$	Linear σ centrally, σ^+ and σ^- above and below. Discontinuous L'
Edgeworth expansion	Log of Eq. (41)	Numerically messy
Skew normal	Log of Eq. (42)	Numerically messy

3.1.1. Comparison of models and recommendations

As well as these two models, many others may be used as 3-parameter descriptions of not-quite parabolic log likelihoods. We list some in Table 6, their full descriptions are in Appendix B

As an illustration we take the result $a = 5.0000^{+2.5811}_{-1.9159}$, which would be reported from a Poisson measurement of 5 events. This is shown in Fig. 10. The true likelihood is $L(a) = e^{-a}a^5/5!$, and we compare this with the results of the likelihoods given by the models. All curves, by construction, peak at $a = 5$ and go through the $\Delta \ln L = \pm\frac{1}{2}$ points, and their interpolations in the central region are very similar. At larger values their behaviors are very different. The cubic turns over, which is unacceptable. The quartic provides a better match to the original Poisson likelihood, especially on the negative side. The logarithmic form does fairly well, the Poisson does perfectly, as it should in this particular example. The two Bartlett forms, labeled Linear Variance and Linear Sigma, both do well, the linear variance is somewhat better than the linear sigma. The PDG form does badly, as it assumes a Gaussian behavior outside the central interval. The Edgeworth form does badly at both small and large values, and is troubled by logarithms of negative numbers at large values and/or asymmetries. The skew normal is not as bad as Edgeworth's approximation, but it does not do very well. These comparisons assume the fundamental $a = 5.0000^{+2.5811}_{-1.9159}$ measurement comes from a Poisson measurement: if it does not, then conclusions might be different.

As another example, this time with a negative skewness, we take the logarithm of a parameter with a likelihood described by a Gaussian with $\mu = 8.0$ and $\sigma = 3.0$. This is shown in Fig. 11. The reported result is $a = 2.08^{+0.32}_{-0.47}$. Again the cubic does predictably badly, and the quartic surprisingly well. The logarithmic does quite well, and the Poisson is not good. The PDG form is again bad, but the Bartlett forms are good: this time the linear sigma does a bit better than the linear variance. The Edgeworth form does badly and the skew normal form does not do particularly well.

All models appear to agree closely with each other and with the originals in the central region, and this is borne out by closer examination, in Fig. 12. Outside the central region the models fall into two classes: those like the PDG model, the molded quartic and the matched quintic, which assume quadratic behavior outside, and those which do not. Considering the general form as a Gaussian with varying sigma, it is very counterintuitive to assume that sigma varies in the central region but is constant outside it. On the other hand, if it is varying, one has to guess correctly whether it is increasing or decreasing.

On the basis of these two examples it looks as if the two Bartlett forms generally behave well. There is not a lot to choose between them, but one will not go far wrong using the linear variance with the linear sigma as a sanity check, though the other way round would be quite acceptable — particularly if the measurement is basically a Poisson process. Other models may be useful in particular applications.

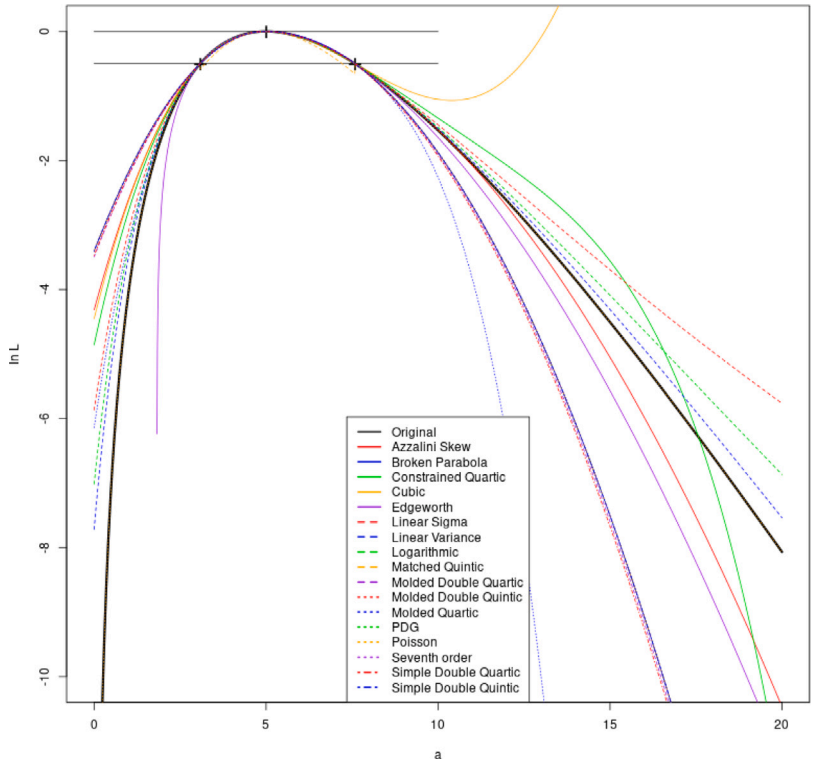

Fig. 10. Approximations to a Poisson likelihood.

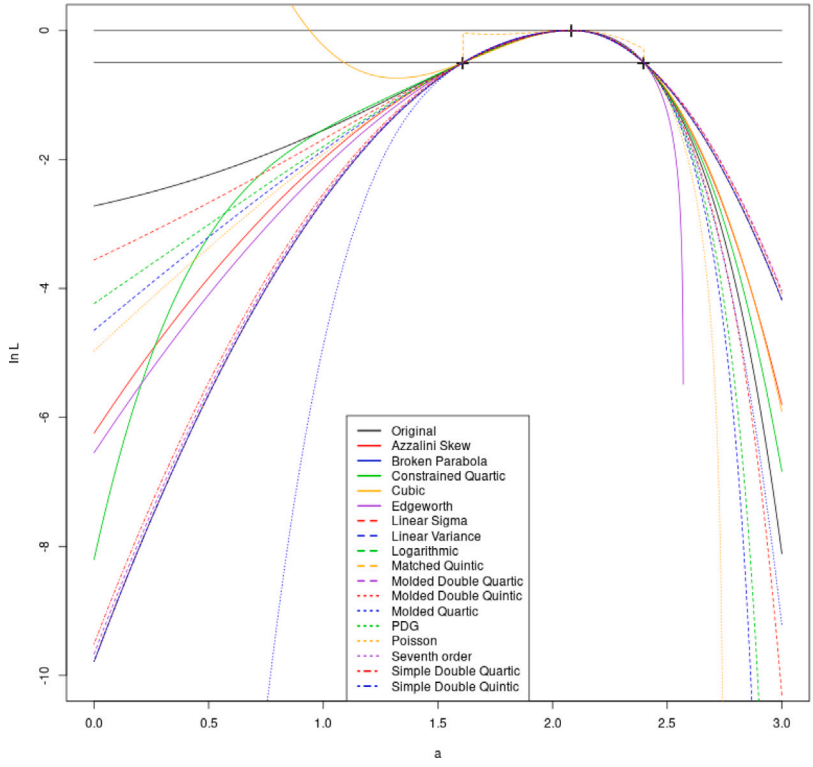

Fig. 11. Possible models of log likelihoods for the measurement $2.08^{+0.32}_{-.047}$, which can arise from $\ln(8 \pm 3)$.

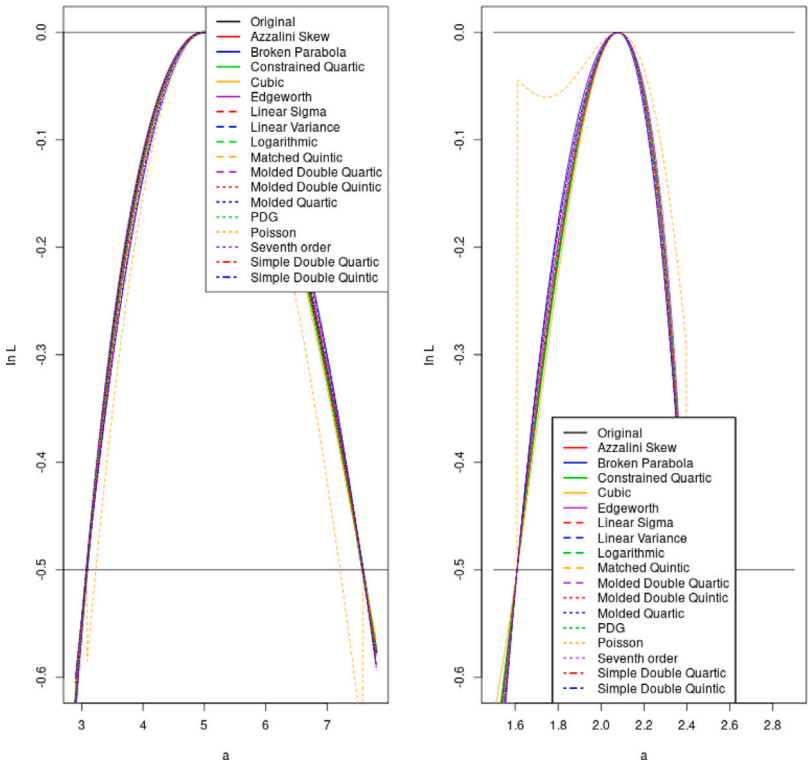

Fig. 12. The central regions for Figs. 10 and 11.

3.2. Combination of results using likelihood

Likelihood-based errors are greatly used in the combination of results. The combination of errors will be dealt with in Section 3.4.

If one has several results measuring the same quantity, then the total likelihood is the product of the individual likelihoods, and the log likelihood is just the sum of the individual log likelihoods. From that, one can read off the overall $\hat{a}$ and the $\Delta \ln L = -\frac{1}{2}$ errors. If we do not know the full forms for the likelihoods, but are only given the peak values and errors, then we can use the same approach, but modeling the likelihoods by one of the methods above. In general a numerical maximization can be used to give a solution. For the two Bartlett models this can be done efficiently: the likelihood is

$$\ln L(a) = -\frac{1}{2}\sum_i \left(\frac{a-\hat{a}_i}{\sigma_i(a)}\right)^2, \tag{10}$$

where $\hat{a}_i$ are the individual maximum likelihood estimates of a. For the linear sigma form, differentiating and setting to zero leads to

$$\hat{a}\sum_i w_i = \sum_i \hat{a}_i w_i \qquad \text{with} \qquad w_i = \frac{\sigma_i}{(\sigma_i + \sigma_i'(a-\hat{a}_i))^3}. \tag{11}$$

The equivalent for the linear variance model is

$$\hat{a}\sum_i w_i = \sum_i w_i(\hat{a}_i - \frac{V_i'}{2V_i}(\hat{a}-\hat{a}_i)^2) \qquad \text{with} \qquad w_i = \frac{V_i}{(V_i + V_i'(a-\hat{a}_i))^2}. \tag{12}$$

Solutions can be found by iteration. In all cases one can then find the errors on the final result by numerical root finding, to discover where the log likelihood falls by $\frac{1}{2}$ from its peak value.

An illustration is shown in Fig. 13. The 3 values $1.9^{+0.7}_{-0.5}$, $2.4^{+0.6}_{-0.8}$ and $3.1^{+0.5}_{-0.4}$ are combined (using the linear variance model) to give $2.754^{+0.286}_{-0.263}$. For the linear sigma model the result is $2.758^{+0.293}_{-0.272}$ and the plot is identical to the eye. Indeed other models give very similar results, as shown in Table 7. The likelihood plots all look very similar to Fig. 13, apart from that using the cubic model.

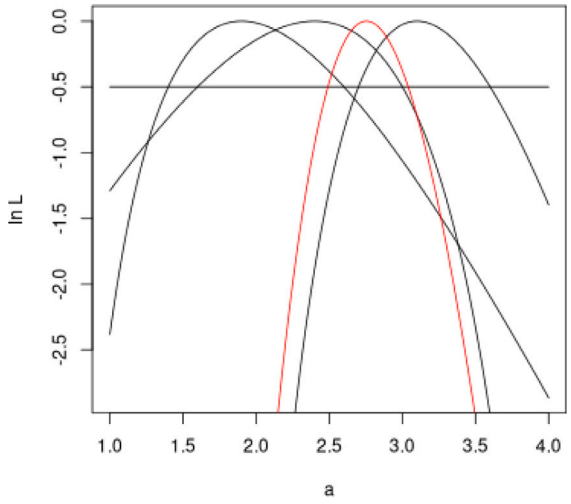

Fig. 13. Combining three asymmetric errors (black solid curves) to yield a result (red solid curve).

It is worth pointing out how the combined error depends on the values being combined, as well as their quoted errors. If one of the values in the combination is changed, this affects not only the combined central value but also the quoted error. In the symmetric case the quoted errors are independent of the values.

3.3. Goodness of fit with likelihood

According to Wilks' theorem, under the null hypothesis the improvement in likelihood due to the introduction of n extra parameters is such that $-2\Delta \ln L$ is distributed according to a χ^2 distribution. On that basis, the change in the log likelihood when using only 1 free parameter (the combined result) rather then n fitted values (the individual results)

Table 7
Results of combining the same 3 results, $1.9^{+0.7}_{-0.5}$, $2.4^{+0.6}_{-0.8}$ and $3.1^{+0.5}_{-0.4}$,according to different models.

Model	Result	Model	Result
Cubic	fails	Truncated cubic	$2.792^{+0.342}_{-0.306}$
Broken parabola	$2.703^{+0.301}_{-0.301}$	Symmetrized parabola	$2.666^{+0.321}_{-0.321}$
Constrained quartic	$2.765^{+0.303}_{-0.285}$	Molded quartic	$2.721^{+0.246}_{-0.240}$
Matched quintic	$2.728^{+0.290}_{-0.300}$	Interpolated 7th degree	$2.702^{+0.301}_{-0.296}$
Simple double quartic	$2.720^{+0.289}_{-0.299}$	Molded double quartic	$2.718^{+0.286}_{-0.298}$
Simple double quintic	$2.702^{+0.301}_{-0.298}$	Molded double quintic	$2.701^{+0.302}_{-0.297}$
Conservative spline(0.05)	$2.720^{+0.291}_{-0.292}$	Conservative spline(0.1)	$2.732^{+0.285}_{-0.287}$
Conservative spline(0.15)	$2.738^{+0.285}_{-0.281}$	Conservative spline(0.20)	$2.742^{+0.287}_{-0.281}$
Conservative spline(max)	$2.742^{+0.287}_{-0.282}$	Log logistic beta	$2.763^{+0.336}_{-0.366}$
Logarithmic	$2.755^{+0.288}_{-0.266}$	Generalized Poisson	$2.753^{+0.283}_{-0.258}$
Variable sigma	$2.758^{+0.293}_{-0.272}$	Variable variance	$2.754^{+0.286}_{-0.263}$
Variable log sigma	$2.743^{+0.289}_{-0.291}$	Molded cubic log sigma	$2.702^{+0.301}_{-0.297}$
Quintic log sigma	$2.702^{+0.301}_{-0.296}$	PDG	$2.726^{+0.273}_{-0.309}$
Skew normal	$2.749^{+0.293}_{-0.285}$		
Railway Gaussian	$2.753^{+0.277}_{-0.250}$	Double cubic Gaussian	$2.731^{+0.294}_{-0.316}$
Symmetric beta Gaussian(1,10)	$2.745^{+0.282}_{-0.314}$	Symmetric beta Gaussian(1,30)	$2.751^{+0.281}_{-0.260}$
Symmetric beta Gaussian(4,10)	$2.701^{+0.309}_{-0.309}$	Symmetric beta Gaussian(4,30)	$2.748^{+0.286}_{-0.278}$
QVW Gaussian	$2.741^{+0.290}_{-0.297}$	Johnson system	–
Log normal	$2.755^{+0.288}_{-0.266}$		

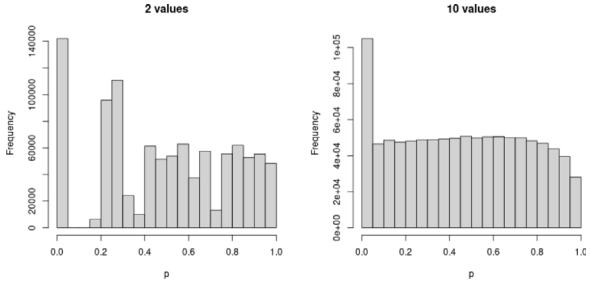

Fig. 14. Goodness of fit distributions for the simulations described in the text from combining 2 and 10 similar results.

gives a measure of goodness of fit such that $-2\Delta \ln L$ is χ^2 with $n-1$ degrees of freedom.

Wilks' theorem holds only in the asymptotic limit,and provided the first model is contained bythe second. The first point depends clearly on the number and precision of the results being combined. So this needs to be investigated empirically.

As an illustration, we have performed a simulation in which two values were drawn from a Poisson distribution of mean 5.0. These were encoded as two values with asymmetric errors (using the $\Delta \ln L = -\frac{1}{2}$ prescription), which were then combined with the linear variance model. The total likelihood at the combined best value was converted to a χ^2 using Wilks' theorem. This was repeated many times. If the assumptions are valid this quantity will be distributed according to a χ^2 distribution with one degree of freedom. We actually show, in the left hand plot of Fig. 14, the histogram of the corresponding p-value, as this is easier to interpret: for a true χ^2 distribution this histogram would be flat. It is not flat but it does its best, given that the space of possible results is discrete (the observations must be integer). The right hand plot shows the result of a similar simulation, combining 10 such values at a time, and the outcome has almost (but not quite) the ideal flat shape.

On examination, the fact that the shape is not quite flat is caused by a small overall positive shift in the $-2\Delta \ln L$ distribution, as compared to the ideal χ^2: this produces the spike at small values and the dip at large ones. This can be ascribed to the fact that a Poisson of mean 5 is not Gaussian. If the simulations are repeated for means of 10, 20 and 30 then the distribution improves, becoming apparently perfectly flat.

It therefore appears that $\Delta \ln L$ – which emerges as part of the combination procedure – can be used as a criterion for the acceptability of the combination, though it should not be treated as an exact quantity as the conditions for the validity of Wilks' theorem may not be met.

3.4. Combination of errors with likelihood

As discussed earlier, one expects that most manipulation of errors from likelihoods will be concerned with the combination of results. However there will be cases where the combination of errors is required. If we had full knowledge of the likelihood we would use the profiling technique to obtain the error on the combination. Instead we do this as best we can using one of the models for the shapes of the asymmetric likelihoods.

The concept is shown in an example in Fig. 15. Some total background B is the sum of B_1 and B_2 which were measured as 4 and 5 Poisson counts. The dotted lines show contours of constant $B = B_1 + B_2$, and for each value of B the maximum $\ln L$ is found, indicated by the crosses. These points give the combined likelihood curve, shown on the right, from which the $\Delta \ln L = -\frac{1}{2}$ errors may be read off.

If we have many variables, $\{a_1 \ldots a_n\}$ and are concerned with the form of the maximum $\ln L(u) = \sum_i \ln L_i(a_i)$ for a given $u = a_1 + a_2 \cdots + a_n$, this can be found using Lagrange's method of undetermined multipliers $\sum \frac{\partial \ln L_i}{\partial a_i} + \lambda \frac{\partial u}{\partial a_i} = 0$. In our simple case $\frac{\partial u}{\partial a_i}$ is just 1, and we can write $\frac{\partial \ln L_i}{\partial a_i} = -\frac{a_i}{w_i}$, where w_i is approximately constant, as the form is roughly parabolic. This gives $a_i = \lambda w_i$ and the constraint reads $u = \lambda \sum w_i$,

$$a_i = u\frac{w_i}{\sum_i w_i} \qquad w_i = -\left(\frac{1}{a_i}\frac{\partial \ln L_i}{\partial a_i}\right)^{-1}. \tag{13}$$

For the linear sigma model, $w_i = \frac{(\sigma_i + \sigma_i' a_i)^3}{\sigma_i}$, for the linear variance model $w_i = \frac{(V_i + a_i V_i')^2}{2V_i + V_i' a_i}$, and similarly for other models. If the cancellation of the a_i factorcannot be done algebraically and numerical evaluation gives 0/0 at the peak, one can use $w = -\left(\frac{\partial^2 L}{\partial a^2}\right)^{-1}$ for small a.

This is needed to read off the $\Delta \ln L = -\frac{1}{2}$ errors. This can be done by starting at the peak, $u = 0$, for which all a_i are zero, and increasing u in small steps, evaluating the w_i, and iterating Eq. (13) to

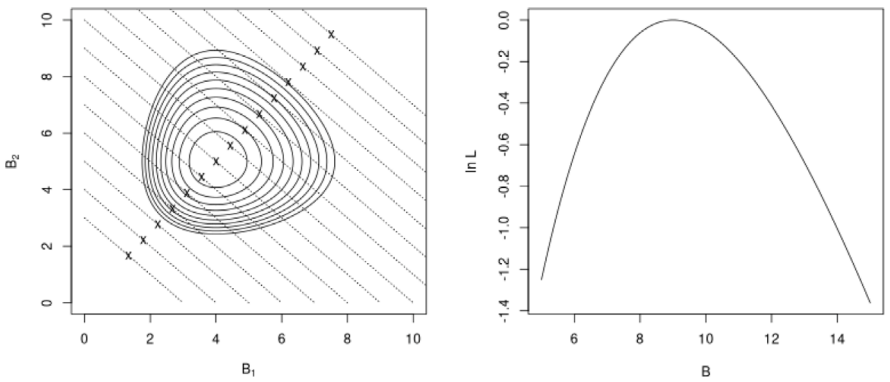

Fig. 15. Errors from profiled likelihoods, from 2D (left) to 1D (right).

convergence, and continuing until $\Delta \ln L = -\frac{1}{2}$ to find σ^+. The process is then repeated for σ^-.

4. Examples and case studies

4.1. Examples

Here we present some relevant examples where the true behavior behind the simple three numbers of the result(s) is known, and see how well the various models do when their outcome is compared with the exact behavior.

4.1.1. A signal with several backgrounds: combination of errors using likelihoods

Suppose an experiment counts some number of events. To extract the signal size, the number of background events is to be subtracted. There are several such background sources, and for simplicity we can suppose that each has been measured by an ancillary experiment, running the apparatus over the same time, so they do not need to be scaled.

If there are two such backgrounds and the ancillary experiments give 4 and 5 counts, then we could, using $\Delta \ln L = -\frac{1}{2}$ errors, report these as $4^{+2.346}_{-1.682}$ and $5^{+2.581}_{-1.916}$. In this case we happen to know that the two results can be combined to give a total background count of 9, with errors $^{+3.342}_{-2.676}$. But we might be unaware of this fact, and just been given the raw numbers. We could then combine the uncertainties using, for example, the linear variance method, and obtain errors of $^{+3.333}_{-2.668}$, in excellent agreement with the true value.

The same total background measurement of 9 might be obtained in various ways, such as 3+6 or 3+3+3 or even 9 separate measurements of 1 event. The same combination procedure also gives $^{+3.333}_{-2.668}$ for the 3+6 case, and changes only slightly to $^{+3.323}_{-2.659}$ for 3+3+3 and $^{+3.269}_{-2.610}$ for 1+1+1+1+1+1+1+1+1. Outcomes for other models and for other inputs are shown in Table 8, and show remarkable agreement with the true values for the errors.

4.1.2. A lifetime measurement: combination of results using likelihoods

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. (These numbers were generated from an exponential distribution with lifetime 1.0ensuring that this example is realistic.) Maximizing $\ln L$, which is $\ln \prod \exp(-t_i/\tau)/\tau$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 combined result $1.1325^{+0.6225}_{-0.3598}$. The likelihood used is the product of 6 separate exponential likelihoods, so we are using the full information.

Now suppose we take the two partial results separately (i.e. just the value and ± 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 full information 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 one.

These results are shown in Table 9 together with those from some other models. The Edgeworth model cannot cope with asymmetries this large, and the cubic does not find a positive error due to the turnover in the curve. It can be seen that the logarithmic and PDG models also do well, the quartic is less good, and the Poisson and skew normal models perform relatively poorly. The final row, labeled "wrong", takes the mean of the two results and combines the positive and negative errors separately in quadrature, and it can be seen that its error estimates are seriously discrepant.

4.1.3. Combining Poisson counts: combination of results using likelihoods

Suppose a counting experiment sees 5 events in an hour. The result is quoted, using $\Delta \ln L = -\frac{1}{2}$ errors even though this is a case where the full Neyman errors could be given, as $5^{+2.581}_{-1.916}$. This continues for another hour and, as it happens, 5 events are again seen. The total gives a result $10^{+3.504}_{-2.838}$ and with the knowledge we have of the way the experiment has been done, we can estimate the number of events per hour by dividing this by 2 to get $5^{+1.752}_{-1.419}$.

But it could be that this knowledge is suppressed, and we are just presented with two estimates (which happen to be the same) and we have to combine them as best we can. If we do this using the linear variance method, the result is $5.000^{+1.747}_{-1.415}$. This is an excellent match to the ideal value, with the errors differing only in the 4th significant figure. Using linear sigma we would get $5.000^{+1.737}_{-1.408}$ which is also very good.

Table 10 shows this and also the results obtained from other possible pairs of results with the same sum and thus the same ideal answer. It can be seen that the technique, especially for the linear variance model, works very well. It is worth pointing out that the slightly larger discrepancies in the final two rows arise from rather unlikely experimental circumstances — the probability of 10 events being split 9:1 or even 8:2 is small, and this shows up in a poor goodness of fit.

4.1.4. Strength of a known mass peak: models of a known likelihood

A common analysis method is to take measurements of a reconstructed particle mass, and fit them to a total of signal and background shapes where all parameters are known except for the normalizations. Fig. 16 shows such a situation, where the background is known to be flat and the signal is a Gaussian of mean 5.0 and standard deviation 1.0. One may then, as appropriate, work with the raw (unbinned) likelihood $\ln L = \sum_i \ln(N_B B(x_i) + N_S S(x_i))$, where the sum is over all events, or the binned likelihood $\ln L = \sum_j (F_j \ln n_j - F_j)$, where n_j is the contents of bin j and F_j is the prediction. The likelihood curves are shown in

Table 8
Examples of combination of errors using likelihoods. The 'correct' answer is $9^{+3.342}_{-2.676}$. Where models cannot accommodate the asymmetry (the $N = 1$ Poisson is strongly asymmetric and breaks several models, and the Johnson System cannot handle even moderate asymmetries) no result can be given.

Inputs	Broken parabola	Truncated cubic	Constrained quartic	Molded quartic	Matched quintic	Interpolated 7th degree
4 5	$9^{+3.488}_{-2.549}$	$9^{+3.218}_{-2.620}$	$9^{+3.272}_{-2.635}$	$9^{+3.439}_{-2.678}$	$9^{+3.283}_{-2.590}$	$9^{+3.425}_{-2.558}$
3 6	$9^{+3.483}_{-2.556}$	$9^{+3.224}_{-2.620}$	$9^{+3.274}_{-2.635}$	$9^{+3.433}_{-2.676}$	$9^{+3.275}_{-2.594}$	$9^{+3.402}_{-2.567}$
2 7	$9^{+3.471}_{-2.572}$	$9^{+3.238}_{-2.622}$	$9^{+3.278}_{-2.636}$	$9^{+3.417}_{-2.670}$	$9^{+3.264}_{-2.602}$	$9^{+3.345}_{-2.583}$
1 8	$9^{+3.450}_{-2.601}$	$9^{+3.264}_{-2.628}$	$9^{+3.290}_{-2.640}$	$9^{+3.368}_{-2.659}$	$9^{+3.269}_{-2.617}$	$9^{+3.279}_{-2.608}$
3 3 3	$9^{+3.603}_{-2.453}$	$9^{+3.105}_{-2.566}$	$9^{+3.206}_{-2.594}$	$9^{+3.500}_{-2.669}$	$9^{+3.161}_{-2.523}$	$9^{+3.356}_{-2.477}$
1 1 1 1 1 1 1 1 1	$9^{+4.073}_{-2.095}$	$9^{+2.605}_{-2.284}$	$9^{+2.894}_{-2.386}$	$9^{+3.389}_{-2.508}$	$9^{+2.552}_{-2.217}$	$9^{+2.576}_{-2.157}$
Inputs	Double quartic	Molded double quartic	Double quintic	Molded double quintic	Conservative spline (0.05)	Conservative spline (0.10)
4 5	$9^{+3.383}_{-2.600}$	$9^{+3.435}_{-2.617}$	$9^{+3.471}_{-2.558}$	$9^{+3.479}_{-2.561}$	$9^{+3.426}_{-2.592}$	$9^{+3.379}_{-2.618}$
3 6	$9^{+3.376}_{-2.606}$	$9^{+3.432}_{-2.623}$	$9^{+3.461}_{-2.568}$	$9^{+3.473}_{-2.573}$	$9^{+3.423}_{-2.596}$	$9^{+3.377}_{-2.620}$
2 7	$9^{+3.362}_{-2.617}$	$9^{+3.428}_{-2.635}$	$9^{+3.438}_{-2.588}$	$9^{+3.460}_{-2.595}$	$9^{+3.416}_{-2.606}$	$9^{+3.374}_{-2.626}$
1 8	$9^{+3.342}_{-2.636}$	$9^{+3.421}_{-2.651}$	$9^{+3.393}_{-2.619}$	$9^{+3.439}_{-2.627}$	$9^{+3.403}_{-2.624}$	$9^{+3.367}_{-2.636}$
3 3 3	$9^{+3.378}_{-2.552}$	$9^{+3.506}_{-2.589}$	$9^{+3.536}_{-2.481}$	$9^{+3.575}_{-2.492}$	$9^{+3.489}_{-2.526}$	$9^{+3.405}_{-2.569}$
1 1 1 1 1 1 1 1 1	$9^{+3.250}_{-2.380}$	$9^{+3.888}_{-2.509}$	$9^{+3.517}_{-2.245}$	$9^{+3.963}_{-2.306}$	$9^{+3.727}_{-2.263}$	$9^{+3.482}_{-2.355}$
Inputs	Conservative spline (0.15)	Conservative spline (0.20)	Conservative spline (max)	Log logistic-beta	Logarithmic	Generalized Poisson
4 5	$9^{+3.342}_{-2.628}$	$9^{+3.316}_{-2.627}$	$9^{+3.315}_{-2.627}$	$9^{+3.177}_{-2.545}$	$9^{+3.325}_{-2.663}$	$9^{+3.342}_{-2.676}$
3 6	$9^{+3.342}_{-2.629}$	$9^{+3.319}_{-2.629}$	$9^{+3.313}_{-2.628}$	$9^{+3.192}_{-2.555}$	$9^{+3.325}_{-2.663}$	$9^{+3.342}_{-2.676}$
2 7	$9^{+3.342}_{-2.632}$	$9^{+3.322}_{-2.632}$	$9^{+3.309}_{-2.631}$	$9^{+3.221}_{-2.576}$	$9^{+3.325}_{-2.663}$	$9^{+3.345}_{-2.583}$
1 8	$9^{+3.341}_{-2.640}$	$9^{+3.325}_{-2.640}$	$9^{+3.306}_{-2.638}$	$9^{+3.264}_{-2.609}$	$9^{+3.326}_{-2.663}$	$9^{+3.342}_{-2.676}$
3 3 3	$9^{+3.341}_{-2.585}$	$9^{+3.292}_{-2.584}$	$9^{+3.272}_{-2.582}$	$9^{+3.071}_{-2.459}$	$9^{+3.308}_{-2.649}$	$9^{+3.342}_{-2.676}$
1 1 1 1 1 1 1 1 1	$9^{+3.310}_{-2.383}$	$9^{+3.187}_{-2.386}$	$9^{+3.025}_{-2.370}$	$9^{+2.731}_{-2.178}$	$9^{+3.211}_{-2.572}$	$9^{+3.342}_{-2.676}$
Inputs	Linear sigma	Linear variance	Linear log sigma	Molded cubic log sigma	Quintic log sigma	PDG method
4 5	$9^{+3.310}_{-2.653}$	$9^{+3.333}_{-2.668}$	$9^{+3.322}_{-2.624}$	$9^{+3.475}_{-2.561}$	$9^{+3.443}_{-2.562}$	$9^{+3.310}_{-2.653}$
3 6	$9^{+3.310}_{-2.653}$	$9^{+3.333}_{-2.668}$	$9^{+3.318}_{-2.626}$	$9^{+3.467}_{-2.572}$	$9^{+3.427}_{-2.571}$	$9^{+3.311}_{-2.653}$
2 7	$9^{+3.311}_{-2.653}$	$9^{+3.333}_{-2.668}$	$9^{+3.311}_{-2.630}$	$9^{+3.449}_{-2.595}$	$9^{+3.390}_{-2.588}$	$9^{+3.311}_{-2.653}$
1 8	$9^{+3.313}_{-2.654}$	$9^{+3.333}_{-2.668}$	$9^{+3.304}_{-2.638}$	$9^{+3.416}_{-2.628}$	$9^{+3.315}_{-2.613}$	$9^{+3.313}_{-2.654}$
3 3 3	$9^{+3.278}_{-2.630}$	$9^{+3.323}_{-2.659}$	$9^{+3.278}_{-2.582}$	$9^{+3.555}_{-2.491}$	$9^{+3.437}_{-2.487}$	$9^{+3.278}_{-2.630}$
1 1 1 1 1 1 1 1 1	$9^{+3.098}_{-2.500}$	$9^{+3.269}_{-2.610}$	$9^{+2.971}_{-2.381}$	$9^{+3.735}_{-2.320}$	$9^{+2.910}_{-2.202}$	$9^{+3.098}_{-2.500}$
Inputs	Skew normal	Railway Gaussian	Double cubic Gaussian	Symmetric beta Gaussian(1,10)	Symmetric beta Gaussian(1,30)	Symmetric beta Gaussian(4,10)
4 5	$9^{+3.302}_{-2.628}$	$9^{+3.355}_{-2.705}$	$9^{+3.312}_{-2.579}$	$9^{+3.330}_{-2.626}$	$9^{+3.350}_{-2.678}$	$9^{+3.347}_{-2.518}$
3 6	$9^{+3.304}_{-2.630}$	$9^{+3.356}_{-2.707}$	$9^{+3.291}_{-2.584}$	$9^{+3.328}_{-2.627}$	$9^{+3.350}_{-2.679}$	$9^{+3.319}_{-2.528}$
2 7	$9^{+3.308}_{-2.635}$	$9^{+3.359}_{-2.711}$	$9^{+3.254}_{-2.597}$	$9^{+3.322}_{-2.630}$	$9^{+3.350}_{-2.680}$	$9^{+3.261}_{-2.548}$
1 8	$9^{+3.318}_{-2.645}$	$9^{+3.366}_{-2.720}$	–	$9^{+3.310}_{-2.634}$	$9^{+3.351}_{-2.683}$	–
3 3 3	$9^{+3.271}_{-2.591}$	$9^{+3.373}_{-2.740}$	$9^{+3.174}_{-2.509}$	$9^{+3.303}_{-2.587}$	$9^{+3.357}_{-2.683}$	$9^{+3.195}_{-2.399}$
1 1 1 1 1 1 1 1 1	$9^{+3.166}_{-2.444}$	$9^{+3.543}_{-3.032}$	–	$9^{+3.039}_{-2.350}$	$9^{+3.409}_{-2.740}$	–
Inputs	Symmetric beta Gaussian(4,30)	QVW Gaussian	Johnson system	Log normal		
4 5	$9^{+3.341}_{-2.652}$	$9^{+3.330}_{-2.620}$	–	$9^{+3.325}_{-2.663}$		
3 6	$9^{+3.339}_{-2.653}$	$9^{+3.326}_{-2.623}$	–	$9^{+3.325}_{-2.663}$		
2 7	$9^{+3.336}_{-2.655}$	$9^{+3.318}_{-2.628}$	–	$9^{+3.325}_{-2.663}$		
1 8	$9^{+3.332}_{-2.659}$	$9^{+3.308}_{-2.637}$	–	$9^{+3.326}_{-2.663}$		
3 3 3	$9^{+3.330}_{-2.634}$	$9^{+3.293}_{-2.578}$	–	$9^{+3.308}_{-2.649}$		
1 1 1 1 1 1 1 1 1	$9^{+3.238}_{-2.546}$	$9^{+3.013}_{-2.374}$	–	$9^{+3.211}_{-2.572}$		

the second and third panels of Fig. 16. In this example the unbinned likelihood gives a result $N_S = 39.41^{+10.20}_{-10.13}$, and the binned likelihood gives $N_S = 40.43^{+10.40}_{-10.35}$.

From the quoted asymmetries one can model the log likelihoods and compare with the known true form. With these small asymmetries the modeling is good and the curves lie close to each other; it is more revealing to plot the difference between the modeled curve and the true one, as shown in Fig. 17.

Within the central region the modeling is excellent for all types except the Edgeworth form. Above and below the central region the agreement is not so good. Generally the modeled curves are too high, showing that they do not fall off fast enough, all with a very similar pattern. The exception (apart from the Edgeworth) is the PDG form, which is much too high below the peak, and too low above it. The PDG form is Gaussian outside the central region: the signal form clearly falls off increasingly fast below the peak and more gently above.

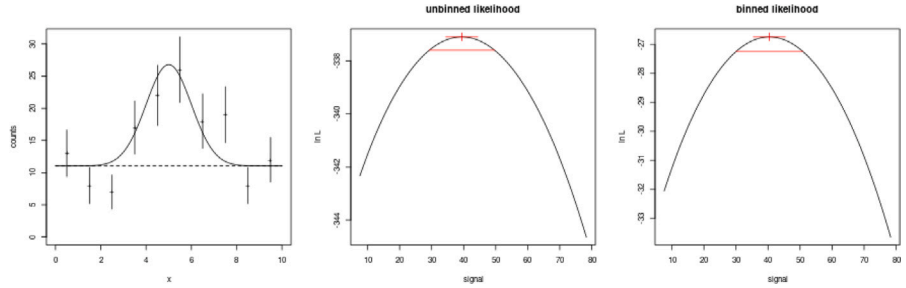

Fig. 16. Fitting a peak with known shapes: the data (first panel) and the likelihoods (second and third panels), unbinned and binned, as a function of supposed signal strength, showing the peak and the $\Delta \ln L = -\frac{1}{2}$ errors.

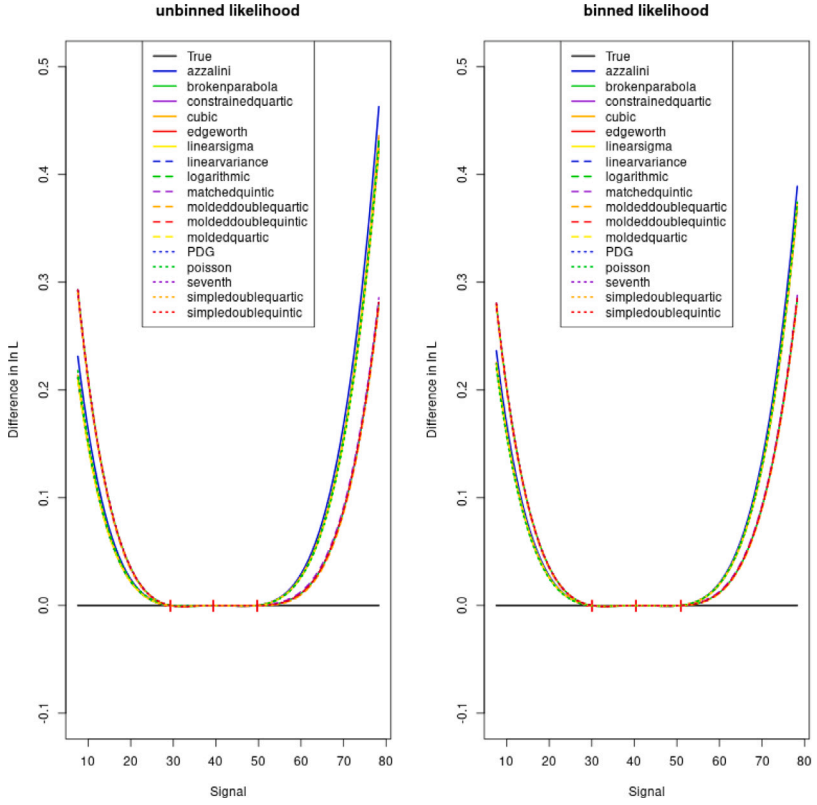

Fig. 17. Modeling the likelihoods in the case shown in Fig. 16, plotting the differences from the true curve. The different models give very similar functional forms, especially in the central region, making the point that results will be robust under choice of model, at least in this example. If the actual forms are shown, rather than the differences, the distinctions are even less apparent.

Table 9
Results from combining two lifetime experiments.

Model	Value	σ^+	σ^-
Full information	1.1325	0.6225	0.3598
Poisson	1.1285	0.6282	0.3533
Cubic	1.1074	NA	0.3433
Quartic	1.1335	0.6243	0.3637
Logarithmic	1.1319	0.6237	0.3586
Linear sigma	1.1323	0.6213	0.3604
Linear variance	1.1318	0.6249	0.3577
PDG method	1.1318	0.6249	0.3613
Skew normal	1.1494	0.6146	0.3764
Wrong	1.1325	0.7261	0.3326

Table 10
Combining results from two samples from the same Poisson distribution. The exact result is $5.000^{+1.752}_{-1.419}$.

$\hat{a}_1$	$\hat{a}_2$	Linear σ	Linear V	Skew normal	Quartic
$5^{+2.581}_{-1.916}$	$5^{+2.581}_{-1.916}$	$5.000^{+1.737}_{-1.408}$	$5.000^{+1.748}_{-1.415}$	$5.000^{+1.732}_{-1.395}$	$5.000^{+1.719}_{-1.399}$
$6^{+2.794}_{-2.128}$	$4^{+2.346}_{-1.682}$	$4.998^{+1.778}_{-1.432}$	$5.000^{+1.759}_{-1.425}$	$5.009^{+1.766}_{-1.460}$	$4.992^{+1.815}_{-1.440}$
$7^{+2.989}_{-2.323}$	$3^{+2.080}_{-1.416}$	$5.038^{+1.937}_{-1.530}$	$5.009^{+1.794}_{-1.456}$	$5.002^{+1.744}_{-1.594}$	$5.107^{+2.147}_{-1.651}$
$8^{+3.171}_{-2.505}$	$2^{+1.765}_{-1.102}$	$5.401^{+2.368}_{-1.826}$	$5.054^{+1.856}_{-1.516}$	$4.689^{+1.635}_{-1.589}$	$5.282^{+1.277}_{-1.707}$
$9^{+3.342}_{-2.676}$	$1^{+1.358}_{-0.698}$	$7.348^{+3.149}_{-2.549}$	$5.201^{+1.942}_{-1.605}$	$3.633^{+1.426}_{-1.414}$	$3.209^{+0.594}_{-0.828}$

4.1.5. A product with asymmetric uncertainties

The expected number of particle decays in some channel can be written as $N = \mathcal{L}\sigma F$, where $\mathcal{L}$ is the integrated luminosity (or equivalent), σ is the cross section,[2] and F is the branching fraction. N is the ideal number and the actual number observed will be Poisson-distributed with mean N. In a typical instance $\mathcal{L}$ will be accurately known, but σ and F may have asymmetric errors. What should we quote as the error for the expected N given, say, $\mathcal{L} = 1000\text{pb}^{-1}, \sigma = 12.3^{+0.4}_{-0.5}$pb, and $F = 0.12^{+0.01}_{-0.02}$?

The central value is $1000 \times 12.3 \times 0.12 = 1476$ events. Writing $\sigma = \sigma_0 + \Delta_\sigma, F = F_0 + \Delta_F$ then, to leading order in a Taylor expansion, $N = N_0 + (\mathcal{L}F_0)\Delta_\sigma + (\mathcal{L}\sigma_0)\Delta_F$, and the appropriate function (pdf or $\ln L$) is found by combining the errors, scaled by $\mathcal{L}F_0$ and $\mathcal{L}\sigma_0$ respectively. For the linear sigma model this gives 1476^{+136}_{-250}. Using the linear variance model it is 1476^{+137}_{-251}.

If, instead, one were calculating the cross section from some observed number of events using $\sigma = \frac{N}{\mathcal{L}F}$ one would scale the errors on N by $\frac{1}{\mathcal{L}_0 F_0}$ and on F by $\frac{1}{\mathcal{L}_0 F_0^2}$ and combine them.

5. Likelihood or pdf? variances or confidence?

5.1. Framework

An error quoted on a measurement may denote the rms of a pdf or a 68% central confidence region obtained from a likelihood. Previous sections have discussed the appropriate handling of these two separate cases. Now we consider the consequences if they are, inadvertently or deliberately, confused.

As mentioned in Section 1.2, a pdf can give 68% central limits, but they relate to the horizontal lines on the Neyman confidence belt, not the vertical ones, and the relationship is subtle. In the example discussed in Fig. 2 a pdf with positive skewness leads to a likelihood with negative asymmetry. Essentially the same problem occurs in the evaluation of confidence regions using the bootstrap. A pdf is approximated by taking repeated samples from the data, and these can be used to give a confidence region for some quantity. However this is not the confidence region for the true quantity, and to use it as such "amounts to looking up the wrong tables backwards", as Peter Hall puts it [16, p36] and [17, p928]. The proportional Gaussian has a symmetric pdf, $p(x) = \frac{1}{\sqrt{2\pi}\alpha a} e^{-\frac{1}{2}\left(\frac{x-a}{\alpha a}\right)^2}$, but a likelihood with positive asymmetry. The Poisson has a positive asymmetry for both probability and likelihood, as its variance increases with the mean. So there is no easy equivalence between the 68% confidence region obtained from the pdf and that from the likelihood.

5.2. What happens if you use the wrong formalism

There may be occasions where one has to work with asymmetric errors quoted with no indication of whether they are obtained from pdfs or from likelihoods.

As an example we consider the combination of errors from two results, both with error $\sigma^+ = 2, \sigma^- = 1$. These are shown in Table 11 for several models. Not surprisingly – this is a large asymmetry – the different models give a spread of different values for the combined σ^+, and likewise for σ^-. But the pdf models are clustered separately from the $\ln L$ models. The values of σ^+ and σ^- for the combination under the assumption that these come from pdfs are both larger than those under the assumption that these come from likelihoods. This discrepancy is not enormous, but it is larger than the differences between individual models. (The difference between any of these and the 'wrong' method of summing separately in quadrature is somewhat larger.) The conclusion to be tentatively drawn in this instance is that if a pdf error is (wrongly) treated as a likelihood error, or vice versa, this makes a meaningful, though not enormous, difference. But this should not be taken as being true in general, and individual cases will need to be justified.

Table 11
Different outcomes from combining errors $^{+2}_{-1}$ and $^{+2}_{-1}$.

Pdf models	σ^+	σ^-
Dimidiated	2.636	1.651
Distorted	2.727	1.760
Edgeworth	2.656	1.586
Railway	2.715	1.775
$\ln L$ models		
Linear sigma	2.467	1.526
Linear variance	2.562	1.562
Cubic	2.000	1.464
Logarithmic	2.530	1.550
Constrained quartic	2.363	1.492
Wrong	2.828	1.414

Table 12
Different outcomes from combining results $1.0^{+2.0}_{-1.0}$ and $2.0^{+2.0}_{-1.0}$.

	Result	σ^+	σ^-
$\ln L$ models			
Azzalini	1.703	1.232	0.796
Broken parabola	1.800	1.166	0.892
Constrained quartic	1.676	1.239	0.784
Cubic	1.634	1.193	0.771
Linear sigma	1.673	1.244	0.760
Linear variance	1.668	1.255	0.738
Logarithmic	1.670	1.251	0.745
Matched quintic	1.655	1.232	0.809
Molded double quartic	1.729	1.217	0.808
Molded double quintic	1.767	1.188	0.856
Molded quartic	1.729	1.234	0.809
PDG	1.673	1.244	0.791
Poisson	1.661	1.262	0.720
Seventh	1.730	1.226	0.881
Simple double quartic	1.715	1.222	0.822
Simple double quintic	1.775	1.186	0.878
Pdf models			
Dimidiated	1.703	1.318	0.825
Distorted	1.758	1.363	0.880
Edgeworth	1.732	1.328	0.793
Railway	1.744	1.357	0.888

Similarly Table 12 shows the outcomes from combining two results: $1.0^{+2.0}_{-1.0}$ and $2.0^{+2.0}_{-1.0}$. Again, the result to be quoted depends on the choice of model, which is arbitrary (though the practitioner may choose to reject some models as being inappropriate for their use) so there is a spread. There are also overall shifts between the pdf treatment and the $\ln L$ treatment, which is not arbitrary. If you use the wrong treatment this will give wrong results, by an amount which is somewhat larger than the inescapable differences owing to the choice of model.

5.3. A final result

A typical final result is presented in the form $x^{+\sigma_1^+ \ +\sigma_2^+}_{-\sigma_1^- \ -\sigma_2^-}$ where "the first error is statistical and the second is systematic". The first is thus, presumably, an error obtained from the likelihood to some final fit, and the second from some pdf-based uncertainty obtained from combining many separate sources. In such cases one may still want to present a combined result: $x^{+\sigma^+}_{-\sigma^-}$.

As a pedagogical example we consider the determination of the total strength (in Becquerel, or counts per second) of an isotropic point source using a counter of intrinsic efficiency η. If n counts are measured

[2] It is regrettable to use the same symbol for cross-sections and for standard deviations, but the use is entrenched.

in time t then the strength R is obtained from n through a conversion factor A

$$R = An = \frac{4\pi}{S}\frac{1}{\eta t}n, \tag{14}$$

where S is the solid angle presented by the counter to the source.

We suppose that η and S are well known, but that there is a Gaussian error $\sigma_t = 0.1s$ on the calibration of the time measurement t which is systematic, i.e. described by a pdf, and gives an asymmetric error on $1/t$.

Considering an example where $A = 100, \eta = 1, n = 50$ and $t = 1$ s, we would write $R = 5000^{+741+556}_{-674-455}$ where the first, 'statistical', error comes from the Poisson errors on $n = 50$ and the second 'systematic' from the error on $1/t$.

If we, wrongly, treat the pdf error as if it were a likelihood error and combine the two errors this gives a combined result of 5000^{+906}_{-828} (using the linear-sigma model: other models give near-identical values). If we, wrongly, treat the ln$L$ error as if it were a pdf and combine them, this gives 5000^{+909}_{-831} (using the dimidiated model: other models give near-identical values. There is little difference between the two outcomes.

As a validation we performed 10^7 toy experiments, generating values of n from a Poisson distribution of mean 50 and values of t from a Gaussian distribution of mean 1 and standard deviation 0.1, and applying the two prescriptions. The pdf method gave a coverage of $69.074 \pm 0.015\%$, only slightly above the nominal 68.269%, and the lnL method was even closer with a coverage of $68.387 \pm 0.015\%$

It thus appears justifiable to combine (with caution) asymmetric 'systematic' and 'statistical' errors on a final result, treating them both either as one or the other.

6. Conclusions

We trust we have convinced any reader of this paper that asymmetric errors are not to be taken up lightly: if differences appear between the σ^+ and σ^- values in some part of an analysis, they will introduce considerable complication both technically and conceptually. If these differences are small and can legitimately be ignored then our advice would be to do so, rather than to retain them out of misplaced diligence. If the differences are large then the best course would be to report the full likelihood, at least for the parameter(s) of interest, or perhaps to provide future meta-analysts with some fuller description of the behavior of the likelihood function.

If asymmetric errors cannot be avoided then they can be handled by the methods given. You need to be clear whether they are likelihood errors or pdf errors, and whether you are combining results or combining errors. You also need to choose models from among those given here, or ones of your own, always taking more than one model to check the robustness of the result. Experience should show which model(s) are appropriate for a particular type of problem.

This treatment assumes that the errors being considered are independent, and that the final result is a single quantity. We hope to consider its extension to correlated inputs and outputs in a future publication.

CRediT authorship contribution statement

Roger Barlow: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Alessandra Rosalba Brazzale:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Igor Volobouev:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank our colleagues for many discussions and suggestions, (including a PHYSTAT informal review), especially the PHYSTAT organizers: Sara Algeri, Olaf Behnke, Lydia Brenner, Louis Lyons (who suggested some of the examples) and Nick Wardle. This paper arose out of a workshop at the Banff Center for Arts and Creativity in Alberta, Canada, and we are very happy to thank the Banff International Research Station (BIRS) and its staff for their organization and hospitality. We are grateful to Hans Dembinski for helpful comments.

ARB acknowledges that the publication was produced with funding from the Italian Ministry of University and Research under the Call for Proposals related to the scrolling of the final rankings of the PRIN 2022 call (Project title "Latent variable models and dimensionality reduction methods for complex data", Project No. 20224CRB9E, CUP C53C24000730006, PI Prof. Paolo Giordani, Grant Assignment Decree No. 1401 adopted on 18.9.2024 by MUR).

IV acknowledges partial support by the US Department of Energy grant DE-SC0015592.

Appendix A. Modeling pdf errors

We present the details of possible models for near-Gaussian pdf functions. The list is not exhaustive, and others, such as the Generalized Extreme Value distribution, can also be suggested [10]. Their motivations, and their technical challenges, vary widely. There is no universal 'best' model and the choice is ultimately down to the user. Use of more than one model in any application is strongly recommended as the differing results will give a feeling for the robustness of the whole procedure.

A.1. The dimidiated Gaussian

Given the results of an OPAT analysis one can hypothesize that the dependence of R on ν can be described by two straight lines, as shown in the left of Fig. 18. This can be expressed as

$$R = \begin{cases} M + \sigma^+\nu, & \nu \geq 0 \\ M + \sigma^-\nu, & \nu \leq 0 \end{cases}. \tag{15}$$

The pdf for the variable ν is a unit Gaussian, so for R the pdf is a dimidiated (or split) Gaussian with two equal halves, as shown on the right of (15). This is sometimes referred to as a bifurcated Gaussian, but that is certainly inaccurate. The dimidiated Gaussian for R is the probability transform of the unit Gaussian for ν, under the model that they are related by Eq. (15)

Although the model is clearly unrealistic, it has some nice simple features. The quantile parameters provide the parameters of the pdf directly, as given by Eq. (16)

$$p(x) = \begin{cases} \frac{1}{\sigma^-\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-M}{\sigma^-}\right)^2} & \text{for } x < M \\ \frac{1}{\sigma^+\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-M}{\sigma^+}\right)^2} & \text{for } x > M \end{cases}. \tag{16}$$

The moments are given by

$$\begin{aligned} \mu &= M + \frac{\sigma^+ - \sigma^-}{\sqrt{2\pi}}, \\ V &= \frac{1}{2}(\sigma^{+2} + \sigma^{-2}) - \frac{1}{2\pi}(\sigma^+ - \sigma^-)^2, \\ \gamma &= \frac{1}{\sqrt{2\pi}}\left[2(\sigma^{+3} - \sigma^{-3}) - \frac{3}{2}(\sigma^+ - \sigma^-)(\sigma^{+2} + \sigma^{-2}) + \frac{1}{\pi}(\sigma^+ - \sigma^-)^3\right]. \end{aligned} \tag{17}$$

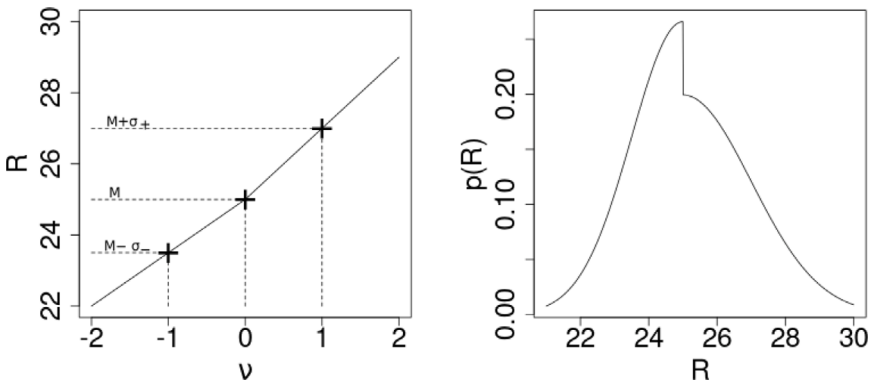

Fig. 18. The Dimidiated Gaussian. The left hand plot shows the results of a simple OPAT analysis that gave a result $25.0^{+2.0}_{-1.5}$, and the proposed approximation. The right hand plot shows the resulting pdf projected on the vertical axis that would come from a Gaussian distribution for ν on the horizontal axis.

To determine the quantile parameters from the moments these equations have to be solved numerically. If we write $D = \sigma^+ - \sigma^-$ and $S = \sigma^{+2} + \sigma^{-2}$, the equations

$$S = 2V + \frac{D^2}{\pi}, \qquad (18)$$
$$D = \frac{2}{3S}\left[\sqrt{2\pi}\gamma - D^3(\frac{1}{\pi} - 1)\right]$$

can be solved as a cubic for D, using Cardano's formula. Having thus determined S and D the quantile parameters are given by

$$M = \mu - \frac{D}{\sqrt{2\pi}}, \qquad \sigma^+ = \frac{1}{2}(\sqrt{2S - D^2} + D), \qquad \sigma^- = \frac{1}{2}(\sqrt{2S - D^2} - D). \qquad (19)$$

Even though the model can handle arbitrarily large asymmetries, in the sense of $\frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}$, there is a limit in terms of moments. If one of the two half-Gaussians has zero width, then

$$\frac{|\gamma|}{\sqrt{V^3}} = \frac{\pi + 1}{\sqrt{(\pi - 1)^3}} \approx 1.641, \qquad (20)$$

so asymmetries, in the $\frac{\gamma}{V^{3/2}}$ sense, larger than this cannot be handled. But such cases are extreme.

A.1.1. Convolution of two dimidiated Gaussians

For the dimidiated Gaussian model, the convolution of two pdfs can be done analytically. If $z = x+y$ then its pdf is $p_z(z) = \int p_x(x)p_y(z-y)\,dx$. The integral covers the top left and bottom right quadrants of the (x, y) plane and either the bottom left quadrant (if z is negative) or the top right (if z is positive). Writing, for clarity:

$$\sigma_+^2 = \sigma_x^{+2} + \sigma_y^{+2}, \qquad \sigma_\pm^x = \sigma_x^{+2} + \sigma_y^{-2}, \qquad \sigma_\mp^2 = \sigma_x^{-2} + \sigma_y^{+2}, \qquad \sigma_-^2 = \sigma_x^{-2} + \sigma_y^{-2}, \qquad (21)$$

the Gaussian integrals give, for $z > 0$,

$$p_z(z) = \phi\left(\frac{z}{\sigma_\mp}\right)\Phi\left(\frac{-z\sigma_x^-}{\sigma_y^+\sigma_\mp}\right) + \phi\left(\frac{z}{\sigma_+}\right)\left[\Phi\left(\frac{z\sigma_y^+}{\sigma_x^+\sigma_+}\right) - \Phi\left(\frac{-z\sigma_x^+}{\sigma_y^+\sigma_+}\right)\right] + \phi\left(\frac{z}{\sigma_\pm}\right)\left[1 - \Phi\left(\frac{z\sigma_y^-}{\sigma_x^+\sigma_\pm}\right)\right], \qquad (22)$$

where ϕ is the Gaussian pdf and Φ is the corresponding cumulative density function. For $z < 0$ the expression is

$$p_z(z) = \phi(\frac{z}{\sigma_\mp})\Phi\left(\frac{z\sigma_y^+}{\sigma_x^-\sigma_\mp}\right) + \phi(\frac{z}{\sigma_-})\left[\Phi\left(\frac{-z\sigma_x^-}{\sigma_y^-\sigma_-}\right) - \Phi\left(\frac{z\sigma_y^-}{\sigma_x^-\sigma_-}\right)\right] + \phi\left(\frac{z}{\sigma_\pm}\right)\left[1 - \Phi\left(\frac{-z\sigma_x^+}{\sigma_y^-\sigma_\pm}\right)\right]. \qquad (23)$$

An example is shown in Fig. 19. Two quantities x and y, both with large positive skewness, $\sigma^+ = 1.5, \sigma^- = 0.5$, are combined. The black curve shows the convolution of the two dimidiated Gaussians, using the above equations: if two dimidiated Gaussians are combined this is the 'correct' pdf. The red curve shows the curve from the 'wrong method' of combining positive and negative sigmas separately and agreement is poor. The green curve shows the dimidiated Gaussian whose first 3 moments match that of the convolution, and gives much better agreement. Note that the central value (the median) has shifted, also that the asymmetry is reduced by the combination, in accordance with the Central Limit Theorem: the green curve has $\sigma^+ = 1.93, \sigma^- = 0.97$.

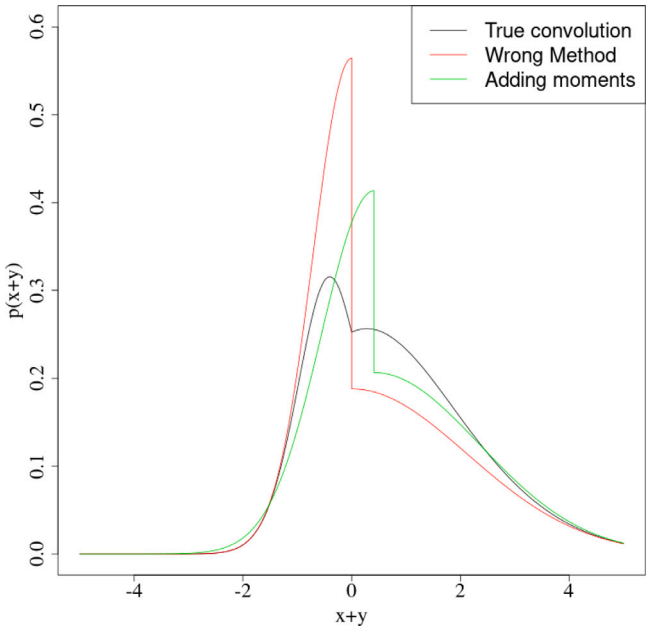

Fig. 19. Combination of errors using the dimidiated Gaussian. Here x and y both have positive skewness: $\sigma^+ = 1.5$ and $\sigma^- = 0.5$.

A.2. The distorted Gaussian

A more sophisticated procedure to handle the results of an OPAT analysis is to draw a parabola through the three points, as in Fig. 20,

$$R = M + a\nu + b\nu^2. \qquad (24)$$

The parameters a and b are given very simply from the quantile parameters: $a = \frac{1}{2}(\sigma^+ + \sigma^-)$ and $b = \frac{1}{2}(\sigma^+ - \sigma^-)$. The pdf is again the probability transform of the unit Gaussian, but using a single quadratic transform rather than two piecewise linear ones. Although this is more plausible than the two straight lines of the dimidiated Gaussian, care

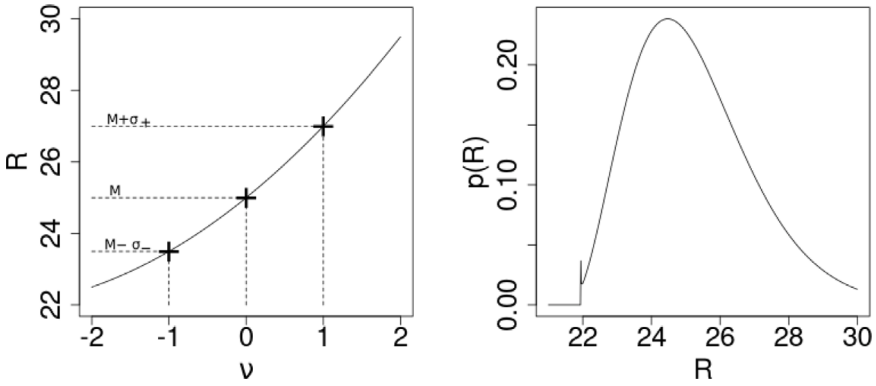

Fig. 20. The Distorted Gaussian. The left hand plot shows the results of a simple OPAT analysis that gave a result $25.0^{+2.0}_{-1.5}$, as before, and the proposed approximation. The right hand plot shows the pdf projected on the vertical axis that would come from a Gaussian on the horizontal axis.

needs to be taken to include both arms of the parabola in evaluating the pdf. The support for $p(R)$ is limited by the minimum (or, for a negative skewness, the maximum) of the parabola. Where this occurs, there is a Jacobian spike, and this appears in Fig. 20 around $R = 22$.

If ν is distributed according to the standard Gaussian with $\mu = 0$ and $\sigma = 1$ and $x(\nu) = x_0 + a\nu + b\nu^2$, then the pdf for x is $p(x) = \frac{\phi(\nu)}{|x'(\nu)|} = \frac{\phi(\nu)}{|a+2b\nu|}$, where a and b are as given in Eq. (24), and $\phi(\nu)$ is the unit Gaussian. As this is a quadratic, there are two values of ν for a given x: $\nu = \frac{\sqrt{a^2+4bx}-a}{2b}$ and $\frac{-\sqrt{a^2+4bx}-a}{2b}$. Except for small asymmetries and/or small deviations, where the contribution from the second can be neglected, the pdf for x is the sum of the two.

The moments are given by

$$\mu = M + b, \qquad V = a^2 + 2b^2, \qquad \gamma = 2b(3a^2 + 4b^2). \tag{25}$$

Again, the parameters can be determined from the moments numerically,

$$b = \frac{\gamma}{V - 4b^2}, \qquad a = \sqrt{V - 4b^2}, \qquad M = \mu - b. \tag{26}$$

From these, if desired, the other parameters are just given by $\sigma^+ = a + b, \sigma^- = a - b$.

Because of the second arm of the parabola, the parameters M, σ^+, σ^- do not give the exact quantiles for R. Because of this ambiguity the distorted Gaussian is treated slightly differently in the software packages: the R implementation uses the parameters described here and the Python/C++ implementation uses the parameters giving the exact quantiles.

A.3. The "railway" Gaussian

The railway Gaussian density attempts to mitigate the drawbacks of the dimidiated Gaussian (the discontinuity in the middle) and of the distorted Gaussian (the Jacobian spike). This density is also based on a coordinate transformation. The curve drawn through the OPAT analysis points is a parabola in the $[-1, 1]$ region which smoothly transitions into straight lines on both left and right sides of the region. These transition regions are highly reminiscent of the track transition curves in railroad tracks, and this similarity gives the density its name. The coordinate transformation is defined by the equation

$$R = f(\nu) = \begin{cases} L(\nu, -1 - h_l) & \text{for } \nu \le -1 - h_l \\ T(\nu, -1, -h_l) & \text{for } -1 - h_l \le \nu \le -1 \\ M + a\nu + b\nu^2 & \text{for } -1 \le \nu \le 1 \\ T(\nu, 1, h_r) & \text{for } 1 \le \nu \le 1 + h_r \\ L(\nu, 1 + h_r) & \text{for } \nu \ge 1 + h_r \end{cases}, \tag{27}$$

where $L(\nu, \nu_0)$ is a straight line defined by its value and its derivative at ν_0, while $T(\nu, \nu_0, h)$ is a cubic transition curve. The coefficients of the cubic are determined from the value of the cubic and of its first and second derivatives at ν_0 as well as from the requirement that the second derivative decays linearly to 0 at $\nu_0 + h$. The explicit expression is

$$T(\nu, \nu_0, h) = \left[\frac{f''(\nu_0)}{2}\left(1 - \frac{\nu - \nu_0}{3h}\right)(\nu - \nu_0) + f'(\nu_0)\right](\nu - \nu_0) + f(\nu_0). \tag{28}$$

According to Eq. (27), for the right transition region $[1, 1 + h_r]$ with $\nu_0 = 1$, we must set $f(\nu_0) = M + 1 + b$, $f'(\nu_0) = a + 2b$, $f''(\nu_0) = 2b$. The transition curve in the $[-1 - h_l, -1]$ left transition region is similarly parameterized by the function and its derivatives at $\nu_0 = -1$. An example railway coordinate transformation and the corresponding density is shown in Fig. 21.

In comparison with the dimidiated and distorted Gaussian distributions, the railway Gaussian has two additional parameters, h_l and h_r (the widths of the left and right transition regions, respectively). A reasonable default choice for them is given by $h_l = \left|\frac{f'(-1)}{f''(-1)}\right|$ and $h_r = \left|\frac{f'(1)}{f''(1)}\right|$.

The R code (Appendix D) evaluates the moments of the railway Gaussian algebraically, whereas in the python/C++ code (Appendix C) the moments are evaluated numerically and, for numerical reasons, the values of h used are restricted to the range $[0.1, 10.0]$.

Note that, in general, the railway coordinate transformation is allowed to be non-monotonic (for example, when σ^+ and σ^- have different signs). In this case, similar to the distorted Gaussian, the railway Gaussian will have a semi-infinite support and will exhibit a Jacobian spike.

A.4. The double cubic Gaussian

To construct the double cubic Gaussian, the transformation generating the railway Gaussian is simplified by eschewing the parabolic section in the middle. The natural choice $h_l = h_r = 1$ then leads to

$$R = M + \begin{cases} \frac{1}{4}(5\sigma^- - \sigma^+)(\nu + 1) - \sigma^- & \text{for } \nu \le -1 \\ \nu\left[\frac{1}{4}\left(\nu^2 + 3\nu + 2\right)(\sigma^+ - \sigma^-) + \sigma^-\right] & \text{for } -1 \le \nu \le 0 \\ \nu\left[\frac{1}{4}\left(\nu^2 - 3\nu + 2\right)(\sigma^- - \sigma^+) + \sigma^+\right] & \text{for } 0 \le \nu \le 1 \\ \frac{1}{4}(5\sigma^+ - \sigma^-)(\nu - 1) + \sigma^+ & \text{for } \nu \ge 1 \end{cases}. \tag{29}$$

The dependence of R on ν is continuous together with its first two derivatives, while the third derivative exhibits discontinuities at $\nu = -1, 0, 1$. An example double cubic coordinate transformation and the corresponding density are shown in Fig. 22.

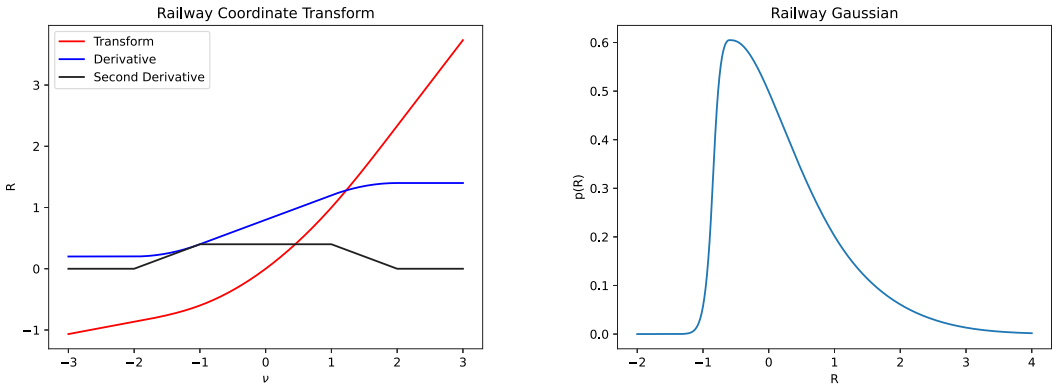

Fig. 21. The Railway Gaussian. The panel on the left illustrates the coordinate transformation used to construct a railway Gaussian with parameters $M = 0$, $\sigma^+ = 1$, $\sigma^- = 0.6$, $h_l = h_r = 1$. The panel on the right shows the corresponding density.

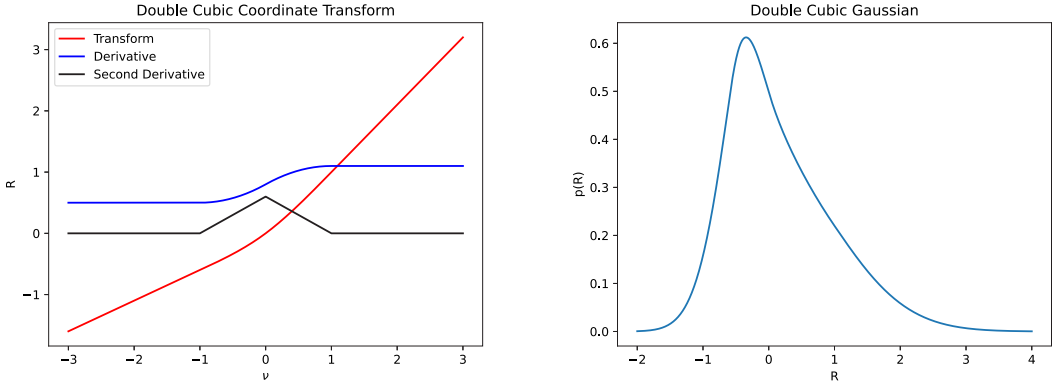

Fig. 22. The Double Cubic Gaussian. The panel on the left illustrates the coordinate transformation used to construct a double cubic Gaussian with parameters $M = 0$, $\sigma^+ = 1$, $\sigma^- = 0.6$. The panel on the right shows the corresponding density.

A.5. The symmetric beta Gaussian

The symmetric beta Gaussian distribution is constructed using the following transformation:

$$R(v) = A \int_0^v dy \int_0^y b(x, p, h) dx + kv + M, \tag{30}$$

where

$$b(x, p, h) = \frac{1}{A} \frac{d^2 R(v)}{dv^2} = \begin{cases} 0 & \text{for } |x| \geq h \\ \left[1 - \left(\frac{x}{h}\right)^2\right]^p & \text{for } |x| < h \end{cases} \tag{31}$$

is a shifted and scaled (as well as unnormalized) density of the symmetric beta distribution. h is a positive width parameter. In order to maintain the continuity of the transformation second derivative and to enable efficient numerical calculation of the cumulants, the power parameter p is expected to be a small positive integer (the C++/Python software is restricted to $p \in \{1, \ldots, 20\}$; it is not yet implemented in the R software). For given values of p, h, and M, parameters A and k are chosen to satisfy the conditions $R(-1) = M - \sigma^-$ and $R(1) = M + \sigma^+$, while the condition $R(0) = M$ is satisfied by Eq. (30) automatically.

The transformation defined by Eq. (30) is continuous together with its $p+1$ leading derivatives. This is reflected in the smoothness of the R distribution whose density will possess at least p continuous derivatives at all inner points of its support. (The Jacobian spike, if present, lies at the support boundary.) In the limit $h \to 0$ the distribution of R tends to the dimidiated Gaussian, and in the limit $h \to \infty$ it becomes the distorted Gaussian. Practically usable values of h are between 0.1 and 10.0 or so; values outside this range can lead to numerical instabilities in the software.

Examples of the transformations generated by Eq. (30), together with the corresponding densities, are illustrated in Fig. 23 for a number of different p and h settings.

A.6. The quantile variable width Gaussian

The quantile variable width (QVW) Gaussian distribution is a special case of the simple Q-normal distribution proposed in [18]. It is based on the following idea. Suppose, $Q(z)$ is the quantile function (i.e., the inverse cumulative distribution function) of the standard normal. Then $\mu + \sigma Q(z)$ is the quantile function of the Gaussian distribution with mean μ and standard deviation σ. We can also define a quantile function which looks like $\mu + \sigma(z) Q(z)$. In a wider sense, the distribution defined by this quantile function can be interpreted as a Gaussian with variable width $\sigma(z)$. The width in this case depends on the cumulative probability rather than on the coordinate. The QVW Gaussian distribution is employing the following simple parameterization:

$$\sigma(z) = \sigma_0 \left[1 + a\left(z - \frac{1}{2}\right)\right], \tag{32}$$

where a is a shape parameter governing the asymmetry. The complete quantile function is then

$$q(z) = \mu_0 + \sigma_0 \left[1 + a\left(z - \frac{1}{2}\right)\right] Q(z). \tag{33}$$

The density that corresponds to the cdf value z is, of course, $\left(\frac{dq(z)}{dz}\right)^{-1}$. The z value corresponding to a given density argument x has to be determined by solving the equation $q(z) = x$ numerically.

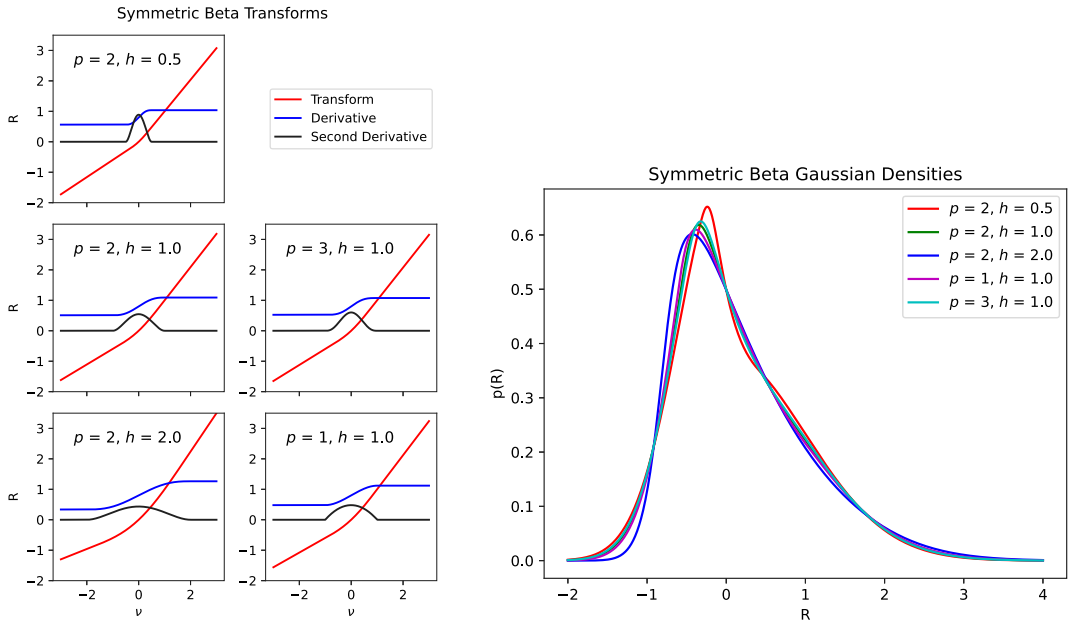

Fig. 23. The Symmetric Beta Gaussian. The panel on the left illustrates the coordinate transformations used to construct densities shown on the right. In all cases, $M = 0$, $\sigma^+ = 1$, $\sigma^- = 0.6$.

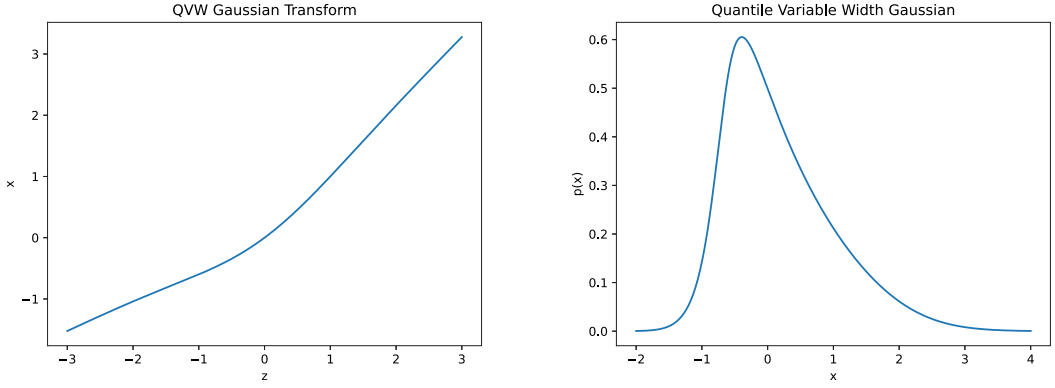

Fig. 24. The QVW Gaussian. The panel on the right illustrates the density which corresponds to the asymmetric uncertainty result $0^{+1.0}_{-0.6}$. The panel on the left shows the corresponding coordinate transformation given by $x = q(\Phi(z))$.

The first three moments of the QVW Gaussian distribution are

$$\begin{aligned}\mu &= \mu_0 + a\sigma_0 M_{11}\\ \sigma^2 &= \sigma_0^2\left[1 + a^2\left(M_{22} - M_{11}^2 - \frac{1}{4}\right)\right]\\ \gamma &= \frac{a\sigma_0^3}{4}\left[8a^2M_{11}^3 - 3M_{11}(4 + a^2(4M_{22} - 1)) + 3(a-2)^2M_{31}\right.\\ &\quad \left. - 6(a-2)aM_{32} + 4a^2M_{33}\right],\end{aligned} \tag{34}$$

where

$$M_{kn} = \int_{-\infty}^{\infty} z^k \Phi(z)^n \phi(z) dz. \tag{35}$$

Here, $\phi(z)$ is the standard normal density, and $\Phi(z)$ is the standard normal cdf. In particular, $M_{11} = \frac{1}{2\sqrt{\pi}}$, $M_{22} = \frac{\sqrt{3}+2\pi}{6\pi}$, $M_{31} = M_{32} = \frac{5}{4\sqrt{\pi}}$. Multiple precision numerical evaluation of M_{33} gives 0.675106426094598067428 4983.

An example QVW Gaussian density and the corresponding transformation of the standard normal are shown in Fig. 24.

A.7. The fechner distribution

The probability density function of the Fechner (also known as the split-normal) distribution is given by

$$p(x) = \begin{cases} \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_1+\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma_1}\right)^2\right) & \text{for } x \le m \\ \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_1+\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma_2}\right)^2\right) & \text{for } x \ge m \end{cases}, \tag{36}$$

where m is the distribution mode, while non-negative parameters σ_1 and σ_2 define characteristic distribution widths below and above the mode, respectively.

The first three moments of the Fechner distribution are given by

$$\begin{aligned}\mu &= m + \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1)\\ \sigma^2 &= \left(1 - \frac{2}{\pi}\right)(\sigma_2 - \sigma_1)^2 + \sigma_1\sigma_2\\ \gamma &= \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1)\left[\left(\frac{4}{\pi} - 1\right)(\sigma_2 - \sigma_1)^2 + \sigma_1\sigma_2\right]\end{aligned} \tag{37}$$

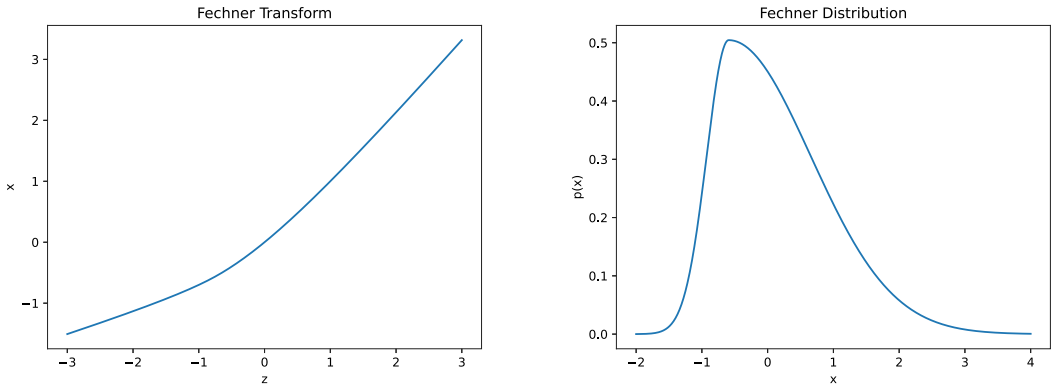

Fig. 25. The Fechner distribution. The panel on the right illustrates the probability density function which corresponds to the asymmetric uncertainty result $0^{+1.0}_{-0.7}$. (The result $0^{+1.0}_{-0.6}$ for which a number of other distributions are plotted in this Appendix cannot be represented by Fechner.) The panel on the left shows the corresponding coordinate transformation given by $x = q(\Phi(z))$.

The largest possible normalized skewness, $\gamma/\sigma^{3/2}$, is reached for $\sigma_1 = 0$ and $\sigma_2 > 0$. It corresponds to the skewness of the half-Gaussian, $\sqrt{2}(4-\pi)(\pi-2)^{-3/2}$. This limits the asymmetry that can be represented by the Fechner distribution, $\frac{\sigma^+-\sigma^-}{\sigma^++\sigma^-}$, to about 0.21564027.

Construction of the Fechner distribution from moments or from quantiles has to be performed numerically. An example Fechner probability density function and the corresponding transformation of the standard normal are shown in Fig. 25.

A.8. The edgeworth expansion

The Edgeworth expansion [16] is appropriate with an asymmetry ascribable to the fact that the sample size is finite and the Central Limit Theorem has not had a chance to make the distribution accurately Gaussian. This is not a probability transform so there is no equivalent to Figs. 18 or 20. The pdf (taking only the first term after the Gaussian) is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \left[1 + \frac{\gamma}{6\sigma^3} He_3\left(\frac{x-\mu}{\sigma}\right)\right], \tag{38}$$

where $He_3(z) = z^3 - 3z$ is the third order Hermite polynomial. The Edgeworth model has the nice feature that the moments provide the function parameters for Eq. (38), with $V = \sigma^2$.

The cumulative equivalent of this $p(x)$, the cumulative distribution function, is

$$F(z) = \Phi(z) - \frac{\gamma}{6\sigma^3}\phi(z)(z^2-1), \tag{39}$$

where $z \equiv (x-\mu)/\sigma$ and $\phi(z)$ and $\Phi(z)$ are the Gaussian density and distribution functions. The second term vanishes at the 1-sigma points, $z = \pm 1$, so the width of the 68% central interval is the same for all values of γ. As μ is also independent of γ, this means that the choice of parameters $\mu^{+\sigma^+}_{-\sigma^-}$ cannot be used with this form, as mentioned in .

If the quantile parameters $M^{+\sigma^+}_{-\sigma^-}$ are given, the moments are

$$\begin{aligned} \mu &= M + \frac{1}{2}(\sigma^+ - \sigma^-) \\ \sqrt{V} \equiv \sigma &= \frac{1}{2}(\sigma^+ + \sigma^-) \\ \gamma &= \frac{6\sigma^3}{(z^2-1)G(z)}\left(\Phi(z) - \frac{1}{2}\right) \text{ with } z = \frac{M-\mu}{\sigma}. \end{aligned} \tag{40}$$

The reverse process again needs a numerical solution. Starting from $z = 0$ one can iterate

$$z = \Phi^{-1}\left[\frac{1}{2} + \frac{\gamma}{6\sigma^3}(z^2-1)G(z)\right], \tag{41}$$

where Φ^{-1} is the inverse of the Gaussian cdf. This seems to work better than Newton's method, thanks to the constantly changing gradient. From z one obtains $M = \sigma z + \mu$, and thence $\sigma^+ = \sigma + \mu - M$ and $\sigma^- = \sigma - \mu + M$.

The Edgeworth form, as one soon discovers, has the disadvantage that unless the asymmetry is quite small, the pdf goes negative, which is clearly disallowed.

A.9. The skew normal

Azzalini [19] suggests using a distribution whose density can be elegantly written as

$$p(z) = 2\phi(z)\Phi(\alpha z). \tag{42}$$

In addition to the asymmetry parameter α one introduces a scale parameter ω and a location parameter ξ, so $z \equiv \frac{x-\xi}{\omega}$. Writing for convenience $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$, the moments are given by

$$\begin{aligned} \mu &= \xi + \omega\delta\sqrt{\frac{2}{\pi}} \\ V &= \omega^2\left(1 - \frac{2\delta^2}{\pi}\right) \\ \gamma &= \left(\frac{4-\pi}{2}\right)(\omega\delta\sqrt{2/\pi})^3. \end{aligned} \tag{43}$$

To convert in the reverse direction

$$\begin{aligned} \omega\delta &= \sqrt{\frac{\pi}{2}}\sqrt{[3]}\frac{2\gamma}{4-\pi} \\ \omega &= \sqrt{V + \frac{2(\omega\delta)^2}{\pi}} \\ \xi &= \mu - \omega\delta\sqrt{\frac{2}{\pi}}. \end{aligned} \tag{44}$$

For the quantile parameters we need the cumulative distribution, which is not so elegant,

$$F(z) = \Phi(z) - 2T(z,\alpha), \tag{45}$$

where $T(z,\alpha)$ is Owen's T function

$$T(z,\alpha) = \frac{1}{2\pi}\int_0^\alpha \frac{e^{-\frac{1}{2}z^2(1+x^2)}}{1+x^2}\,dx. \tag{46}$$

From a value of α one can determine the appropriate values of z to get the required quantiles $F(z) = 0.16$, 0.5, and 0.84, iterating using Newton's method, and then use ξ and ω to convert these to $M^{+\sigma^+}_{-\sigma^-}$.

To convert in the other direction, when the distribution is specified by $M^{+\sigma^+}_{-\sigma^-}$, the parameter α can be found numerically by adjusting it to get the desired value of the asymmetry $\frac{\sigma^+-\sigma^-}{\sigma^++\sigma^-}$. ω is then found by scaling to get the desired value of $\sigma^+ + \sigma^-$, and ξ from the desired value for M. If the distribution is specified by $\mu^{+\sigma^+}_{-\sigma^-}$ then to obtain the other

forms is not impossible (as it is for the Edgeworth distribution), but is so impractical as to be essentially useless.

This distribution does not describe large asymmetries, like the Edgeworth distribution, but for a different reason. In the large α limit it becomes a half Gaussian, and higher asymmetries cannot be accommodated.

A.10. The johnson system

The combination of Johnson distributions S_B (bounded), S_U (unbounded), together with their limiting cases (Gaussian and log normal), form a very flexible system of four-parameter densities [20–22]. It turns out that there is a Johnson distribution for every possible combination of mean, standard deviation, skewness, and kurtosis. It is in fact convenient to think of the set as if it was a single distribution whose density is given by $j(R,\mu,\sigma,s,\kappa)$. (In this parameterization, skewness s and kurtosis κ are normalized.)

Four parameters are one too many to represent a central value with two asymmetric uncertainties. There is, however, a rather natural way to restrict the system to three parameters as follows: $j_r(R,\mu,\sigma,s) = j(R,\mu,\sigma,s,\hat{\kappa})$, where the kurtosis parameter $\hat{\kappa}$ is chosen for every s according to the maximum entropy principle: $\hat{\kappa} = \underset{\kappa}{\text{argmax}}[S(s,\kappa)]$. The standardized entropy $S(s,\kappa)$ is given by

$$S(s,\kappa) = -\int_{-\infty}^{\infty} j(u,0,1,s,\kappa)\ln j(u,0,1,s,\kappa)\,du.$$

The entropy does not depend on μ. As κ is normalized, $\hat{\kappa}$ does not depend on σ either.

It turns out that, for $|s| > 0$, the maximum entropy principle always selects the S_U distribution (and the Gaussian for $s = 0$). Therefore, the S_U distribution will be discussed here in more detail. Its density is given by

$$p(x) = \frac{\delta}{\lambda\sqrt{2\pi\left[1+\left(\frac{x-\xi}{\lambda}\right)^2\right]}}\, e^{-\frac{1}{2}\left[\gamma+\delta\sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right]^2}, \tag{47}$$

where the set of parameters $\{\xi,\lambda,\gamma,\delta\}$ is in one-to-one correspondence to the set $\{\mu,\sigma,s,\kappa\}$. Variable z distributed according to the standard normal can be obtained from Johnson's S_U variate x by the transformation $z = \gamma + \delta\sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)$. Alternatively,

$$x = \xi + \lambda\sinh\left(\frac{z-\gamma}{\delta}\right) \tag{48}$$

transforms the standard normal into Johnson's S_U.

With the notation $\Omega = \gamma/\delta$ and $\omega = \exp(\delta^{-2})$, the moments of the S_U distribution are given by [21]

$$\begin{aligned}
\mu &= \xi - \lambda\sqrt{\omega}\sinh\Omega \\
\sigma^2 &= \frac{\lambda^2}{2}(\omega-1)(\omega\cosh 2\Omega+1) \\
s^2 &= \frac{\omega(\omega-1)\left[\omega(\omega+2)\sinh 3\Omega + 3\sinh\Omega\right]^2}{2(\omega\cosh 2\Omega+1)^3} \\
\kappa &= \frac{\omega^2(\omega^4+2\omega^3+3\omega^2-3)\cosh 4\Omega + 4\omega^2(\omega+2)\cosh 2\Omega + 3(2\omega+1)}{2(\omega\cosh 2\Omega+1)^2}.
\end{aligned} \tag{49}$$

The sign of the skewness s is opposite to that of γ. The mapping from the parameter set $\{\mu,\sigma,s,\kappa\}$ back to $\{\xi,\lambda,\gamma,\delta\}$ can be efficiently performed numerically, according to the algorithm presented in [23]. The calculation of $\hat{\kappa}$ for the given s also has to be performed numerically, as well as the determination of μ, σ, and s from quantiles.

An example S_U density and the corresponding transformation are shown in Fig. 26.

A.11. The log normal

The log normal distribution is a special three-parameter member of the Johnson system. For a fixed skewness, the kurtosis of the log normal is larger than the kurtosis of all S_B distributions, but smaller than the kurtosis of all S_U. It is thus the limiting distribution for S_U with the smallest possible kurtosis and for S_B with the largest possible one.

A variable distributed according to the log normal can be obtained from the standard normal variate z by applying the transformation

$$x = \xi + \exp\left(\frac{z-\gamma}{\delta}\right). \tag{50}$$

This also means that the variable $z = \gamma + \delta\ln(x-\xi)$ is distributed according to the standard normal if x is distributed according to the log normal. With this parameterization,[3] for positive values of skewness, the log normal density is given by

$$p(x) = \begin{cases} 0, & x \le \xi \\ \frac{\delta}{\sqrt{2\pi}(x-\xi)}\, e^{-\frac{1}{2}[\gamma+\delta\ln(x-\xi)]^2}, & x > \xi \end{cases}. \tag{51}$$

Distributions with negative values of skewness can be obtained by utilizing $p(-x)$ and adjusting the parameters ξ and γ as necessary.

With the notation $\omega = \exp(\delta^{-2})$, the moments of the log normal distribution (with normalized s and κ) are given by [21]

$$\begin{aligned}
\mu &= \xi + \sqrt{\omega}\exp\left(-\frac{\gamma}{\delta}\right) \\
\sigma^2 &= \omega(\omega-1)\exp\left(-\frac{2\gamma}{\delta}\right) \\
s^2 &= (\omega-1)(\omega+2)^2 \\
\kappa &= \omega^4+2\omega^3+3\omega^2-3
\end{aligned}. \tag{52}$$

The mapping from the parameter set $\{\mu,\sigma,s\}$ back to $\{\xi,\gamma,\delta\}$ can be performed by first solving the equation $s^2 = (\omega-1)(\omega+2)^2$ for ω, calculating δ, and then finding all other parameters. See [21] for details. The mapping from the quantiles to the distribution parameters has to be constructed numerically.

An example log normal density and the corresponding transformation are shown in Fig. 27.

Appendix B. Modeling likelihood errors

Here we describe several models for likelihood curves that are approximately parabolic. There are many other possible forms (for example, the 'rotated parabola'), and both implementations of software provided are flexible so that more models (including models with more parameters) can be added without needing to re-write all the code.

B.1. The cubic

The obvious extension to a parabola is the cubic

$$\ln L(a) = -\frac{1}{2}\left(\alpha a^2 + \beta a^3\right) \tag{53}$$

with

$$\alpha = \frac{\sigma^{+3}+\sigma^{-3}}{\sigma^{+2}\sigma^{-2}(\sigma^{+}+\sigma^{-})} \quad \text{and} \quad \beta = \frac{\sigma^{-2}-\sigma^{+2}}{\sigma^{+2}\sigma^{-2}(\sigma^{+}+\sigma^{-})}. \tag{54}$$

If σ^+ and σ^- differ by less than a factor of 2, this gives curves which behave sensibly in the range $[\hat{a}-\sigma^-, \hat{a}+\sigma^+]$, but outside that range the cubic term gives an unwanted turning point. (If the asymmetry is large, the curve will have a minimum inside the range in addition to the maximum at $\hat{a} = 0$.) Also the curve does not go to $-\infty$ at both large positive and negative a. So although the cubic is the obvious extension of the quadratic, it often gives problems.

[3] The parameterization described here is consistent with [21], but differs from the standard one.

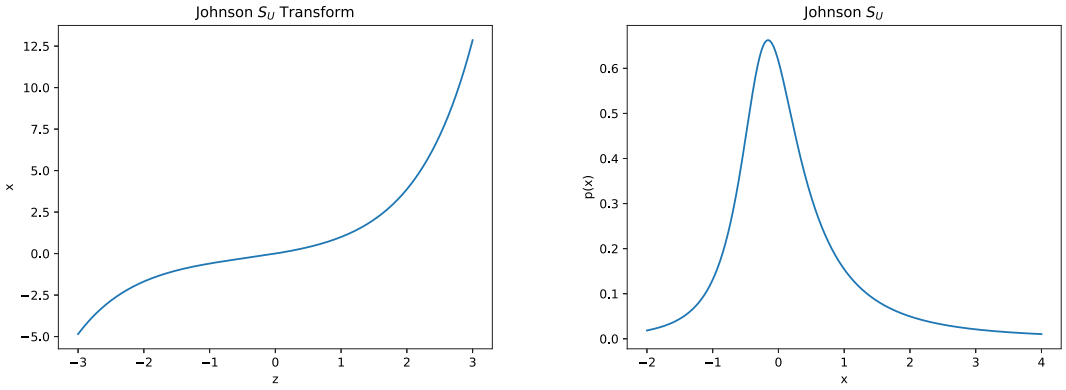

Fig. 26. The Johnson distribution. The panel on the right illustrates the Johnson S_U density which corresponds to the asymmetric uncertainty result $0^{+1.0}_{-0.6}$. The panel on the left shows the matching coordinate transformation calculated according to Eq. (48). As expected, this transformation maps -1 to -0.6, 0 to 0, and 1 into 1.

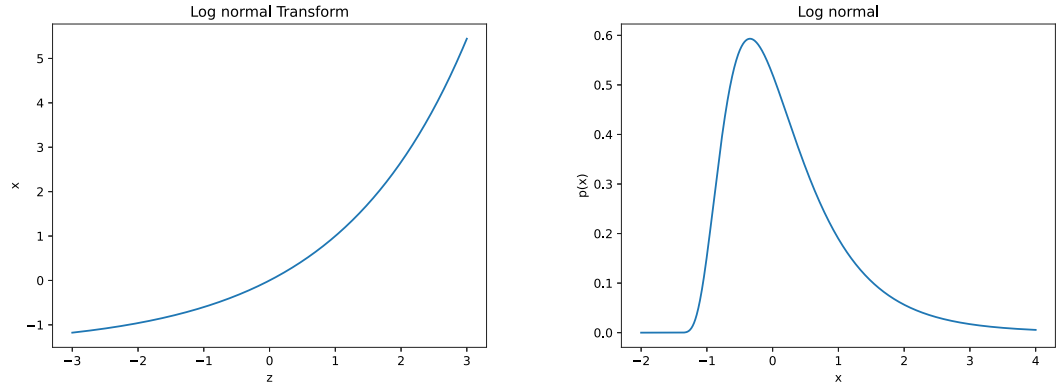

Fig. 27. The Log Normal distribution. The panel on the right illustrates the density which corresponds to the asymmetric uncertainty result $0^{+1.0}_{-0.6}$. The panel on the left shows the matching coordinate transformation calculated according to Eq. (50).

B.2. The broken parabola

Another obvious simple solution which gives $\Delta \ln L = -\frac{1}{2}$ at $a = \sigma^+$ and $a = -\sigma^-$ is just

$$\ln L(a) = g(a) = \begin{cases} -\frac{1}{2}\frac{a^2}{\sigma^{-2}} & \text{for } a \le 0 \\ -\frac{1}{2}\frac{a^2}{\sigma^{+2}} & \text{for } a \ge 0 \end{cases}. \tag{55}$$

(Here, we explicitly assume $\hat{a} = 0$.) Unfortunately, this solution suffers from the unphysical discontinuity of the second derivative at $a = 0$.

B.3. The symmetrized parabola

The symmetrized parabola model is constructed in a Bayesian manner. It starts by assuming a flat prior for the parameter and then treats Eq. (55) as a logarithm of the (unnormalized) parameter probability density. The resulting distribution is Fechner with density given by Eq. (36) in which $m = \hat{a}$, $\sigma_1 = \sigma^-$, and $\sigma_2 = \sigma^+$. This distribution is then exchanged for the Gaussian distribution with the same mean and standard deviation as the Fechner where the moments are calculated according to Eq. (37). The symmetrized parabola is the logarithm of this Gaussian density.

Note that the peak position of the symmetrized parabola is no longer at $\hat{a}$; it is shifted to the right if $\sigma^+ > \sigma^-$ or to the left in case $\sigma^+ < \sigma^-$.

B.4. The constrained quartic

A quartic curve can be constrained to give only one maximum by making the second derivative a perfect square:

$$\begin{aligned} (\ln L)''(a) &= -\frac{1}{2}(\alpha + \beta a)^2 \qquad \text{which yields} \\ \ln L(a) &= -\frac{1}{2}\left(\frac{\alpha^2 a^2}{2} + \frac{\alpha\beta a^3}{3} + \frac{\beta^2 a^4}{12}\right). \end{aligned} \tag{56}$$

The parameters are given by

$$\beta = \frac{1}{\sigma^+\sigma^-}\sqrt{\frac{12(\sigma^- + \sigma^+)^2 \pm 24\sqrt{4\sigma^+\sigma^{-3} + 4\sigma^-\sigma^{+3} - 2\sigma^{+4} - 2\sigma^{-4}}}{3\sigma^{-2} + 2\sigma^+\sigma^- + 3\sigma^{+2}}}. \tag{57}$$

The $\pm$ term should be taken as negative to give a small quartic term. In very asymmetric cases, where σ^+ and σ^- differ by more than a factor of $(1 + \sqrt[4]{12} + \sqrt{3})/2 \approx 2.29663$, the inner root is negative, indicating that there is no solution in the desired form.

Having found β, α is found from the two $\Delta \ln L = -\frac{1}{2}$ points. These specify that both

$$\frac{\alpha^2\sigma^{+2}}{2} + \frac{\alpha\beta\sigma^{+3}}{3} + \frac{\beta^2\sigma^{+4}}{12} = 1 \qquad \text{and} \qquad \frac{\alpha^2\sigma^{-2}}{2} - \frac{\alpha\beta\sigma^{-3}}{3} + \frac{\beta^2\sigma^{-4}}{12} = 1 \tag{58}$$

which have respective solutions

$$\alpha = -\frac{\beta\sigma^+}{3} \pm \frac{\sqrt{72 - 2\beta^2\sigma^{+4}}}{6\sigma^+} \quad \text{and} \quad \alpha = \frac{\beta\sigma^-}{3} \pm \frac{\sqrt{72 - 2\beta^2\sigma^{-4}}}{6\sigma^-}. \tag{59}$$

The two pairs of results have one solution in common, which is the one to be taken. (The roots in Eq. (59) do not go negative.)

This gives better behavior than the cubic for large a, but is not always so satisfactory in the central $[\hat{a} - \sigma^-, \hat{a} + \sigma^+]$ region.

B.5. The molded quartic

To address the discontinuity problem of the broken parabola given by Eq. (55), one can make a quartic polynomial $q(a)$which looks as close as possible to the broken parabola $g(a)$ on the interval $[-\sigma^-, \sigma^+]$ but, of course, has continuous derivatives at 0. The *molded quartic* $\ln L(a) \equiv q(a)$ is constructed by minimizing $\int_{-\sigma^-}^{\sigma^+}(q(a) - g(a))^2 da$ subject to the constraints $q(-\sigma^-) = q(\sigma^+) = -1/2$, $q(0) = 0$, $q'(0) = 0$, and $q''(0) < 0$. One can imagine that, in this construction, $g(a)$ serves as a mold into which $q(a)$ is stamped.

The solution of this constrained minimization problem is $q(a) = -\frac{1}{2}(\alpha a^4 + \beta a^3 + \gamma a^2)$, where the parameters are given by

$$\begin{aligned}
\alpha &= 3(\sigma^- - \sigma^+)^2(5\sigma^{-6} + 8\sigma^{-5}\sigma^+ + 5\sigma^{-4}\sigma^{+2} + 8\sigma^{-3}\sigma^{+3} \\
&\quad + 5\sigma^{-2}\sigma^{+4} + 8\sigma^-\sigma^{+5} + 5\sigma^{+6})/\eta \\
\beta &= (\sigma^- - \sigma^+)[25(\sigma^{-8} + \sigma^{+8}) + 14(\sigma^{-7}\sigma^+ - \sigma^{-6}\sigma^{+2} + \sigma^{-5}\sigma^{+3} \\
&\quad - \sigma^{-4}\sigma^{+4} + \sigma^{-3}\sigma^{+5} - \sigma^{-2}\sigma^{+6} + \sigma^-\sigma^{+7})]/\eta \\
\gamma &= (10\sigma^{-10} - 5\sigma^{-9}\sigma^+ + 30\sigma^{-7}\sigma^{+3} - 6\sigma^{-6}\sigma^{+4} + 6\sigma^{-5}\sigma^{+5} - 6\sigma^{-4}\sigma^{+6} \\
&\quad + 30\sigma^{-3}\sigma^{+7} - 5\sigma^-\sigma^{+9} + 10\sigma^{+10})/\eta
\end{aligned} \tag{60}$$

Here, $\eta = 2\sigma^{-2}\sigma^{+2}(\sigma^- + \sigma^+)^4(5\sigma^{-4} - 10\sigma^{-3}\sigma^+ + 12\sigma^{-2}\sigma^{+2} - 10\sigma^-\sigma^{+3} + 5\sigma^{+4})$.

With these parameters, the constraint $q''(0) < 0$ (equivalently, $\gamma > 0$) is satisfied automatically for all $\sigma^- > 0$ and $\sigma^+ > 0$. The requirement that $q(a)$ has a single extremum (the maximum at 0) leads to the condition that σ^+ and σ^- must not differ by more than a factor of about 3.40804.

B.6. The matched quintic

The matched quintic log likelihood model assumes that, outside of the interval $[-\sigma^-, \sigma^+]$, the log likelihood behaves as a second order polynomial with the second derivative $-(1/\sigma^+)^2$ for $a > \sigma^+$ and $-(1/\sigma^-)^2$ for $a < -\sigma^-$. Inside the interval $[-\sigma^-, \sigma^+]$ the polynomial takes the form

$$\ln L(a) = -\frac{1}{2}(\alpha a^5 + \beta a^4 + \gamma a^3 + \delta a^2). \tag{61}$$

The coefficients are determined from the condition $\Delta \ln L(a) = -\frac{1}{2}$ at the interval boundaries and from the requirement that the second derivative remains continuous at $-\sigma^-$ and σ^+. This leads to

$$\begin{aligned}
\alpha &= -10(\sigma^- - \sigma^+)/\eta \\
\beta &= -18(\sigma^- - \sigma^+)^2/\eta \\
\gamma &= 45\sigma^-\sigma^+(\sigma^- - \sigma^+)/\eta \\
\delta &= (8\sigma^{-4} + 19\sigma^{-3}\sigma^+ - 19\sigma^{-2}\sigma^{+2} + 19\sigma^-\sigma^{+3} + 8\sigma^{+4})/\eta
\end{aligned} \tag{62}$$

Here, $\eta = \sigma^{-2}\sigma^{+2}(8\sigma^{-2} + 19\sigma^-\sigma^+ + 8\sigma^{+2})$.

This log likelihood model behaves similarly to Eq. (55) for large values of $|a|$ but avoids the discontinuity in the second derivative. The third derivative is still discontinuous at $a = -\sigma^-$ and $a = \sigma^+$.

The requirement that $\ln L(a)$ has a single extremum (the maximum at 0) leads to the condition that σ^+ and σ^- must not differ by more than a factor of about 2.426419.

B.7. The interpolated 7th degree polynomial

Another way of avoiding the second derivative problem of Eq. (55) consists in assuming that the log likelihood curve should take the form of Eq. (55) for $a < -\sigma^-$ and $a > \sigma^+$, while on the interval $[-\sigma^-, \sigma^+]$ the curve is constructed by matching the values of the broken parabola $g(a)$ and its derivatives $g'(a)$, and $g''(a)$ at the interval boundaries: $a = -\sigma^-$ and $a = \sigma^+$. Together with the conditions $\ln L(0) = 0$ and $\frac{d}{da}\ln L(a)\Big|_{a=0} = 0$, we have eight equations that can be satisfied by a 7th degree polynomial. This polynomial takes the form

$$\ln L(a) = -\frac{1}{2}(\alpha a^7 + \beta a^6 + \gamma a^5 + \delta a^4 + \epsilon a^3 + \zeta a^2) \tag{63}$$

with the coefficients given by

$$\begin{aligned}
\alpha &= 6(\sigma^- - \sigma^+)/\eta \\
\beta &= 15(\sigma^- - \sigma^+)^2/\eta \\
\gamma &= 10(\sigma^- - \sigma^+)(\sigma^{-2} - 4\sigma^-\sigma^+ + \sigma^{+2})/\eta \\
\delta &= -30\sigma^-\sigma^+(\sigma^- - \sigma^+)^2/\eta \\
\epsilon &= 30\sigma^{-2}\sigma^{+2}(\sigma^- - \sigma^+)/\eta \\
\zeta &= (\sigma^{-6} + 4\sigma^{-5}\sigma^+ + 6\sigma^{-4}\sigma^{+2} - 6\sigma^{-3}\sigma^{+3} + 6\sigma^{-2}\sigma^{+4} + 4\sigma^-\sigma^{+5} + \sigma^{+6})/\eta
\end{aligned} \tag{64}$$

Here, $\eta = \sigma^{-2}\sigma^{+2}(\sigma^- + \sigma^+)^4$.

The requirement that $\ln L(a)$ has a single extremum (the maximum at 0) leads to the condition that σ^+ and σ^- must not differ by more than a factor of about 2.744405.

In combination with Eq. (55) for $a < -\sigma^-$ and $a > \sigma^+$, this model has a very reasonable behavior for large values of $|a|$ but can be somewhat "wavy" on the interval $[-\sigma^-, \sigma^+]$. It also has a discontinuous third derivative at $a = -\sigma^-$ and $a = \sigma^+$.

B.8. The double quartic

This curve is constructed as follows:

$$\ln L(a) = \begin{cases} P_2^-(a) & \text{for } a \le -\sigma^- \\ P_4^-(a) & \text{for } -\sigma^- \le a \le 0 \\ P_4^+(a) & \text{for } 0 \le a \le \sigma^+ \\ P_2^+(a) & \text{for } \sigma^+ \le a \end{cases}, \tag{65}$$

where $P_2^-(a)$ and $P_2^+(a)$ are quadratic polynomials and $P_4^-(a)$ and $P_4^+(a)$ are quartics defined by the following conditions:

$$\begin{aligned}
P_2^-(-\sigma^-) = P_4^-(-\sigma^-) &= P_2^+(\sigma^+) = P_4^+(\sigma^+) = -\frac{1}{2}, \\
\frac{d}{da}P_2^-(a)\Big|_{a=-\sigma^-} &= \frac{d}{da}P_4^-(a)\Big|_{a=-\sigma^-}, \\
\frac{d^2}{da^2}P_2^-(a)\Big|_{a=-\sigma^-} &= \frac{d^2}{da^2}P_4^-(a)\Big|_{a=-\sigma^-} = -\frac{1}{\sigma^{-2}}, \\
\frac{d}{da}P_2^+(a)\Big|_{a=\sigma^+} &= \frac{d}{da}P_4^+(a)\Big|_{a=\sigma^+}, \\
\frac{d^2}{da^2}P_2^+(a)\Big|_{a=\sigma^+} &= \frac{d^2}{da^2}P_4^+(a)\Big|_{a=\sigma^+} = -\frac{1}{\sigma^{+2}},
\end{aligned} \tag{66}$$

$$\begin{aligned}
P_4^-(0) &= P_4^+(0) = 0 \\
\frac{d}{da}P_4^-(0)\Big|_{a=0} &= \frac{d}{da}P_4^+(0)\Big|_{a=0} = 0, \\
\frac{d^2}{da^2}P_4^-(0)\Big|_{a=0} &= \frac{d^2}{da^2}P_4^+(0)\Big|_{a=0} = -\frac{1}{\sigma_0^2}.
\end{aligned} \tag{67}$$

σ_0 is a somewhat arbitrary parameter describing the second differential at the peak. When it is selected, these conditions define the polynomial coefficients uniquely. The resulting curve is continuous together with its first two derivatives. Two choices of σ_0 have been investigated.

The *simple double quartic* is obtained by setting $\sigma_0 = \sqrt{\sigma^-\sigma^+}$. For this choice of σ_0, the requirement that $\ln L(a)$ has a single extremum

(the maximum at 0) leads to the condition that σ^+ and σ^- must differ by less than a factor of $\frac{68}{11}$.

The *molded double quartic* minimizes the "molding" functional $\frac{1}{\sigma^-}\int_{-\sigma^-}^{0}[\ln L(a)-g(a)]^2\,da+\frac{1}{\sigma^+}\int_{0}^{\sigma^+}[\ln L(a)-g(a)]^2\,da$, with $g(a)$ defined by Eq. (55). This expression is minimized by setting $\sigma_0=\sqrt{\frac{\sigma^{-4}+\sigma^{+4}}{\sigma^{-2}+\sigma^{+2}}}$. This choice results in a curve exhibiting reasonable behavior even for very large asymmetries.

B.9. The double quintic

This curve is constructed as follows:

$$\ln L(a)=\begin{cases}P_2^-(a) & \text{for } a\le-\sigma^-\\ P_5^-(a) & \text{for } -\sigma^-\le a\le 0\\ P_5^+(a) & \text{for } 0\le a\le\sigma^+\\ P_2^+(a) & \text{for } \sigma^+\le a\end{cases}, \tag{68}$$

where $P_2^-(a)$ and $P_2^+(a)$ are quadratic polynomials and $P_5^-(a)$ and $P_5^+(a)$ are quintics defined by the following conditions:

$$\begin{aligned}P_2^-(-\sigma^-)=P_5^-(-\sigma^-)&=P_2^+(\sigma^+)=P_5^+(\sigma^+)=-\frac{1}{2},\\ \frac{d}{da}P_2^-(a)\Big|_{a=-\sigma^-}&=\frac{d}{da}P_5^-(a)\Big|_{a=-\sigma^-}=\frac{1}{\sigma^-},\\ \frac{d^2}{da^2}P_2^-(a)\Big|_{a=-\sigma^-}&=\frac{d^2}{da^2}P_5^-(a)\Big|_{a=-\sigma^-}=-\frac{1}{\sigma^{-2}},\\ \frac{d}{da}P_2^+(a)\Big|_{a=\sigma^+}&=\frac{d}{da}P_5^+(a)\Big|_{a=\sigma^+}=-\frac{1}{\sigma^+},\\ \frac{d^2}{da^2}P_2^+(a)\Big|_{a=\sigma^+}&=\frac{d^2}{da^2}P_5^+(a)\Big|_{a=\sigma^+}=-\frac{1}{\sigma^{+2}},\\ P_5^-(0)&=P_5^+(0)=0,\\ \frac{d}{da}P_5^-(0)\Big|_{a=0}&=\frac{d}{da}P_5^+(0)\Big|_{a=0}=0,\\ \frac{d^2}{da^2}P_5^-(0)\Big|_{a=0}&=\frac{d^2}{da^2}P_5^+(0)\Big|_{a=0}=-\frac{1}{\sigma_0^2}.\end{aligned} \tag{69}$$

The resulting curve is continuous together with its first two derivatives. It coincides with Eq. (55) outside of the interval $[-\sigma^-,\sigma^+]$.

Again, when the value of the parameter σ_0 is selected, these conditions define the polynomial coefficients uniquely. Two choices of σ_0 have been investigated.

The *simple double quintic* is obtained by setting $\sigma_0=\sqrt{\sigma^-\sigma^+}$. For this choice of σ_0, the requirement that $\ln L(a)$ has a single extremum (the maximum at 0) leads to the condition that σ^+ and σ^- must differ by less than a factor of 13.5.

The *molded double quintic* minimizes the "molding" functional $\frac{1}{\sigma^-}\int_{-\sigma^-}^{0}[\ln L(a)-g(a)]^2\,da+\frac{1}{\sigma^+}\int_{0}^{\sigma^+}[\ln L(a)-g(a)]^2\,da$, with $g(a)$ defined by Eq. (55). This expression is minimized by setting $\sigma_0=\sqrt{\frac{\sigma^{-4}+\sigma^{+4}}{\sigma^{-2}+\sigma^{+2}}}$. This choice results in a curve exhibiting reasonable behavior even for very large asymmetries.

B.10. The conservative spline

This curve is constructed as follows:

$$\ln L(a)=\begin{cases}P_2^-(a) & \text{for } a\le a_{\text{ left}}\\ \alpha a^3+\beta a^2 & \text{for } a_{\text{ left}}\le a\le a_{\text{ right}}\\ P_2^+(a) & \text{for } a\ge a_{\text{ right}}\end{cases}, \tag{70}$$

where $P_2^-(a)$ and $P_2^+(a)$ are quadratic polynomials. The values, derivatives, and second derivatives of these polynomials are matched to the values, derivatives, and second derivatives of the central cubic at the points $a_{\text{ left}}$ (with $a_{\text{ left}}$ restricted to the interval $[-\sigma^-,\,0]$) and $a_{\text{ right}}$ (with a $_{\text{right}}$ in $[0,\,\sigma^+]$), so that the whole curve is continuous together with its first two derivatives. The parameters α and β are chosen so that the curve satisfies the condition $\ln L(-\sigma^-)=\ln L(\sigma^+)=-\frac{1}{2}$. Assuming, for the moment, that $\sigma^+<\sigma^-$, the parameters $a_{\text{ left}}$ and $a_{\text{ right}}$ are set in such a way that the magnitude of the second derivative of this curve does not exceed $\frac{\kappa}{\sigma^{+2}}$, and equals to $\frac{\kappa}{\sigma^{+2}}$ for $a\ge a_{\text{ right}}$, and does not become less than $\frac{1}{\kappa\sigma^{-2}}$, and equals to $\frac{1}{\kappa\sigma^{-2}}$ for $a\le a_{\text{ left}}$, where $\kappa\ge 1$ is some value specified by the user. Larger values of κ result in a smoother second derivative (smaller $|\alpha|$ and larger interval $[a_{\text{ left}},\,a_{\text{ right}}]$), while the choice $\kappa=1$ results in $a_{\text{ left}}=a_{\text{ right}}=0$ thus reducing Eq. (70) to Eq. (55).

This log-likelihood construction mitigates the discontinuity of the second derivative afflicting Eq. (55) while maintaining user-controlled lower and upper bounds on the Fisher information of the log-likelihood in the parameter space. If a linear log-likelihood curve is added to this one, the uncertainties of the result determined from the $\Delta\ln L(a)=-\frac{1}{2}$ equation will be bounded by the range $[\min(\sigma^-,\sigma^+)/\sqrt{\kappa},\ \sqrt{\kappa}\max(\sigma^-,\sigma^+)]$ no matter how large is the slope of the added curve.

B.11. The log logistic-beta

The density of the logistic-beta distribution (a.k.a. the type IV generalized logistic) is given by [24]:

$$f(x,\alpha,\beta)=\frac{\sigma^\alpha(x)\sigma^\beta(x)}{B(\alpha,\beta)}, \tag{71}$$

where $B(\alpha,\beta)$ is the beta function, $\sigma(x)=1/(1+e^{-x})$ is the logistic function, and $\alpha,\beta>0$ are the shape parameters. After adding a scale parameter and shifting the distribution so that its mode is at 0, the logarithm of this density can be parameterized as

$$\begin{aligned}\ln L(a)=gc^2\Big((A-1)\ln\Big[\frac{1}{2}\Big\{A\Big(e^{\frac{a}{c}}-1\Big)+e^{\frac{a}{c}}+1\Big\}\Big]\\ -(A+1)\ln\Big[\frac{1}{2}e^{-\frac{a}{c}}\Big\{A\Big(e^{\frac{a}{c}}-1\Big)+e^{\frac{a}{c}}+1\Big\}\Big]\Big).\end{aligned} \tag{72}$$

Here, the logarithm of the logistic-beta density (with the $x\to a$ replacement) has been reparameterized in terms of the new parameters $g>0$, $c>0$, and $|A|<1$, and a constant term has been added so that $\ln L(0)=0$. The role of the new parameters can be appreciated as follows:

- A is the asymmetry parameter. The limiting behavior of $\ln L(a)$ for $a\to-\infty$ is an asymptote with the positive slope $gc(A+1)$. For $a\to+\infty$, the asymptote has the negative slope $gc(A-1)$. In case $A=0$, these slopes are the same in magnitude and $\ln L(a)$ becomes an even function.
- g is the parameter regulating the magnitude of the $\frac{d^2\ln L(a)}{da^2}$ maximum. Indeed, it can be shown that $\ln L(a)$ is strictly concave, while $\left|\frac{d^2\ln L(a)}{da^2}\right|$ is maximized at $a=c\ln\left(\frac{1-A}{1+A}\right)$. At that point, $\frac{d^2\ln L(a)}{da^2}=-\frac{g}{2}$.
- c is the shape parameter regulating how quickly the asymptotic behavior is approached. For small values of c the curve looks like two straight lines joined by a narrow transition section. For large values of c the curve looks more parabolic.

Eq. (72) satisfies the conditions $\ln L(0)=0$ and $\frac{d\ln L(a)}{da}\Big|_{a=0}=0$ automatically. It also needs to satisfy the conditions $\ln L(-\sigma^-)=\ln L(\sigma^+)=-\frac{1}{2}$. While with three parameters A, c, and g the latter conditions can be satisfied in a multitude of ways, we choose the parameterization that minimizes g, i.e., A and c are chosen in such a way that $|\ln L(\sigma^+)/g|$ is maximized subject to the constraint $\ln L(-\sigma^-)=\ln L(\sigma^+)$. This results in a conservative log-likelihood model with limited slopes and the second derivative exceeding $\frac{1}{\min(\sigma^-,\sigma^+)^2}$ in magnitude only on a short interval.

B.12. The logarithmic

One can also use a logarithmic form

$$\ln L(a)=-\frac{1}{2}\left(\frac{\ln(1+\gamma a)}{\ln\beta}\right)^2 \tag{73}$$

with

$$\beta = \frac{\sigma^+}{\sigma^-} \quad \text{and} \quad \gamma = \frac{\sigma^+ - \sigma^-}{\sigma^+\sigma^-}. \tag{74}$$

This is easy to implement, and has some motivation as it describes a parabola which has been modified by the expansion/contraction of a at a constant rate. Its unpleasant features are that it is undefined for a beyond some point in the direction of the smaller error, as $1+\gamma a$ goes negative, and it does not give a parabola in the $\sigma^+ = \sigma^-$ limit.

B.13. The generalized Poisson

Starting from the Poisson likelihood, $\ln L(a) = -a + N\ln a - \ln N!$, which has a positive skewness, one can generalize to

$$\ln L(a) = -\alpha(a+\beta) + \mathcal{N}\ln(a+\beta) + constant, \tag{75}$$

where $\mathcal{N}$ is a continuous parameter which produces the skewness, and α and β are scale and location constants. To get the maximum in the right place requires $\mathcal{N} = \alpha\beta$ and, adjusting the constant to make $\ln L(0) = 0$, this becomes

$$\ln L = -\alpha a + \mathcal{N}\ln\left(1 + \frac{\alpha a}{\mathcal{N}}\right). \tag{76}$$

Writing $\gamma = \alpha/\mathcal{N}$ the equations at σ^- and σ^+ lead to

$$\frac{1-\gamma\sigma^-}{1+\gamma\sigma^+} = e^{-\gamma(\sigma^+ + \sigma^-)}. \tag{77}$$

This has to be solved numerically for γ. The solution lies between 0 and $1/\sigma^-$, and can be found by repeatedly taking the midpoint.[4] $\mathcal{N}$ is then found from

$$\mathcal{N} = \frac{1}{2(\gamma\sigma^+ - \ln(1+\gamma\sigma^+))}. \tag{78}$$

This does fairly well, though the need for a numerical solution makes it inelegant. If the curve to be fitted has a negative skewness ($\sigma^- > \sigma^+$), then the sign of a has to be flipped.

B.14. The linear sigma

Bartlett suggests (and justifies) that the likelihood function is described by a Gaussian whose width varies as a function of the parameter [13–15]

$$\ln L(a) = -\frac{1}{2}\left(\frac{a-\hat{a}}{\sigma(a)}\right)^2. \tag{79}$$

This does not include the $-\ln\sigma$ term from the denominator of the Gaussian, however omitting this term improves the accuracy of $\Delta\ln L = -\frac{1}{2}$ errors [13].

If we suppose that the variation of σ is linear, at any rate in the region of interest

$$\sigma(a) = \sigma + \sigma'(a-\hat{a}), \tag{80}$$

$$\ln L(a) = -\frac{1}{2}\left(\frac{a-\hat{a}}{\sigma + \sigma'(a-\hat{a})}\right)^2, \tag{81}$$

then the requirement that this go through the two $-\frac{1}{2}$ points gives

$$\sigma = \frac{2\sigma^+\sigma^-}{\sigma^+ + \sigma^-} \quad \text{and} \quad \sigma' = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-}. \tag{82}$$

This form does well in most cases, and the parameters are easy to find.

[4] Attempts to use more sophisticated algorithms were unsuccessful.

B.15. The linear variance

As an alternative to Eq. (81), we could equally plausibly assume that the variance has a linear variation

$$V(a) = V + V'(a-\hat{a}) \tag{83}$$

giving

$$\ln L(a) = -\frac{1}{2}\frac{(a-\hat{a})^2}{V + V'(a-\hat{a})}. \tag{84}$$

The parameters are again easy to find:

$$V = \sigma^+\sigma^- \qquad V' = \sigma^+ - \sigma^-, \tag{85}$$

and this also does very well. Non-physical results can occur in principle when $V + V'(a-\hat{a})$ goes negative, but only for large deviation and large asymmetries and this does not present a problem in practice.

B.16. The linear sigma in the log space

Eq. (79) works for any positive, monotonic function $\sigma(a)$ as long as $\sigma(\hat{a}-\sigma^-) = \sigma^-$ and $\sigma(\hat{a}+\sigma^+) = \sigma^+$. As the values of σ are supposed to be positive, it seems natural to interpolate these values in the log space. However, simple linear interpolation of $\ln\sigma$ as a function of a is not going to work: for large values of $|a|$ the exponent of a linear function will grow faster than $|a-\hat{a}|$, and the magnitude of $\ln L(a)$ will decrease, creating at least one additional unphysical extremum.

Instead, it can be useful to perform linear interpolation of $\ln\sigma$ as a function of some variable $u = Q(a)$, where $Q(a)$ is a monotonic transform that maps the range of a from $[-\infty,\infty]$ to some compact interval $[u_{\min}, u_{\max}]$. Without loss of generality, this interval can be chosen to be $[0,1]$. In this case, $Q(a)$ becomes a cumulative density function of some distribution supported on $[-\infty,\infty]$.

Here, we choose $Q(a)$ to be the cdf of the Fechner distribution whose density is given by Eq. (36) in which we set $m = \hat{a}$, $\sigma_1 = \sigma^-$, and $\sigma_2 = \sigma^+$. The equation for $\ln\sigma$ is then

$$\ln\sigma(u) = \alpha u + \beta, \tag{86}$$

where the coefficients α and β are chosen to satisfy the conditions

$$\begin{aligned} \ln\sigma(Q(\hat{a}-\sigma^-)) &= \ln\sigma^- \\ \ln\sigma(Q(\hat{a}+\sigma^+)) &= \ln\sigma^+. \end{aligned} \tag{87}$$

Then, of course, $\sigma(a) = e^{\alpha Q(a)+\beta}$.

The requirement that $\ln L(a)$ has a single extremum (the maximum at 0) leads to the condition that σ^+ and σ^- must not differ by more than a factor of about 5.338453.

B.17. The double cubic sigma in the log space

Note that Eq. (79) results in a curve with a continuous second derivative if $\sigma(a)$ is continuous with its first and second derivatives everywhere except the point $a = \hat{a}$. At that point it is sufficient for $\sigma(a)$ just to be continuous, while the continuity of the derivatives is not required (as long as the first and the second derivative are finite on both sides of $\hat{a}$). This allows us to construct a useful log likelihood model as follows:

$$\ln\sigma(a) = \begin{cases} \ln\sigma^- & \text{for } a \le \hat{a}-\sigma^- \\ P_3^-(a) & \text{for } \hat{a}-\sigma^- \le a \le \hat{a} \\ P_3^+(a) & \text{for } \hat{a} \le a \le \hat{a}+\sigma^+ \\ \ln\sigma^+ & \text{for } \hat{a}+\sigma^+ \le a \end{cases}, \tag{88}$$

where $P_3^-(a)$ and $P_3^+(a)$ are cubic polynomials satisfying the following conditions:

$$\begin{aligned} P_3^-(\hat{a}-\sigma^-) &= \ln\sigma^-, \\ \frac{d}{da}P_3^-(a)\Big|_{a=\hat{a}-\sigma^-} &= 0, \end{aligned}$$

$$\begin{aligned}
\left.\frac{d^2}{da^2}P_3^-(a)\right|_{a=\hat{a}-\sigma^-} &= 0,\\
P_3^-(\hat{a}) = P_3^+(\hat{a}) &= \ln\sigma_0,\\
P_3^+(\hat{a}+\sigma^+) &= \ln\sigma^+.\\
\left.\frac{d}{da}P_3^+(a)\right|_{a=\hat{a}+\sigma^+} &= 0,\\
\left.\frac{d^2}{da^2}P_3^+(a)\right|_{a=\hat{a}+\sigma^+} &= 0.
\end{aligned} \tag{89}$$

When the value of the overall parameter σ_0 is selected, these conditions define the polynomial coefficients uniquely. We choose σ_0 by numerically minimizing the following "molding" functional: $\frac{1}{\sigma^-}\int_{\hat{a}-\sigma^-}^{\hat{a}}\{\ln L(a)-g(a-\hat{a})\}^2\,da + \frac{1}{\sigma^+}\int_{\hat{a}}^{\hat{a}+\sigma^+}\{\ln L(a)-g(a-\hat{a})\}^2\,da$, with $g(a)$ defined by Eq. (55). The resulting curve has two continuous derivatives and exhibits reasonable behavior even for very asymmetric errors.

B.18. The quintic sigma in the log space

This model is defined by the following dependence of σ on the parameter:

$$\ln\sigma(a) = \begin{cases} \ln\sigma^- & \text{for } a \le \hat{a}-\sigma^- \\ P_5(a) & \text{for } \hat{a}-\sigma^- \le a \le \hat{a}+\sigma^+ \\ \ln\sigma^+ & \text{for } \hat{a}+\sigma^+ \le a \end{cases}. \tag{90}$$

The coefficients of the quintic polynomial $P_5(a)$ are chosen in such a way that the $\ln\sigma(a)$ function remains continuous together with its first and second derivatives at $a=\hat{a}-\sigma^-$ and $a=\hat{a}+\sigma^+$.

The requirement that $\ln L(a)$ has a single extremum (the maximum at 0) leads to the condition that σ^+ and σ^- must not differ by more than a factor of about 4.107184572.

B.19. The PDG method

In combining results with asymmetric errors the Particle Data Group [25] uses σ^+ for $a > \hat{a}+\sigma^+$, σ^- for $a < \hat{a}-\sigma^-$, and a linearly varying sigma, as given by Eq. (81), for the intermediate region. Note that the PDG log likelihood has discontinuous derivatives at $a=\hat{a}-\sigma^-$ and $a=\hat{a}+\sigma^+$.

B.20. The edgeworth expansion

The Edgeworth expansion of Eq. (38) gives

$$\ln L(a) = -\frac{1}{2}\left(\frac{a-a_0}{\sigma}\right)^2 + \ln\left(1+\alpha He_3\left(\frac{a-a_0}{\sigma}\right)\right) \tag{91}$$

with $He_3(z) = z^3 - 3z$.

Note that the peak of the distribution is not at a_0, and the peak value is not 0. The parameter α can be found numerically from the asymmetry $(\sigma^+-\sigma^-)/(\sigma^++\sigma^-)$ as that is independent of scale and location. Once this is found, the values of σ and a_0 are just found by scaling and shifting.

B.21. The skew normal

The skew normal pdf, Eq. (42), takes the likelihood form

$$\ln L(a) = -\frac{1}{2}\left(\frac{a-\xi}{\omega}\right)^2 + \ln\Phi\left(\alpha\left[\frac{a-\xi}{\omega}\right]\right). \tag{92}$$

Again the location parameter, here ξ, is not the peak position, and the log likelihood at the peak is not zero but must be determined. Given parameters ξ, ω, α the log likelihood can be mapped out numerically to find the peak and the $\Delta\ln L = -\frac{1}{2}$ errors. For the reverse procedure α is first found from the asymmetry, $\frac{\sigma^+-\sigma^-}{\sigma^++\sigma^-}$, which is independent of ω and ξ, then the scale ω from $(\sigma^+ + \sigma^-)$, and finally the location ξ from the peak position.

Appendix C. Implementation in C++/Python

Open source C++/Python code for modeling asymmetric errors is available on GitHub. The code is split into two packages, one implemented in C++ and the other mostly in Python. The reason for dual language implementation is that C++ is better suited for number crunching while Python environment allows for rapid development of small programs and provides convenient plotting facilities.

The first package, "ase", contains a C++ library implementing probability distributions described in Appendix A and log-likelihood curves described in Appendix B, as well as a number of requisite numerically intensive algorithms (minimization, root finding, special functions, quadratures, convolutions, etc.). The URL for this package is https://github.com/igvgit/AsymmetricErrors

The second package, "asepy", contains a Python API generated by SWIG [26] for the C++ classes and functions from the first package. It also includes a number of additional high-level utilities and programs for combining errors and results, as well as for plotting distributions and log-likelihoods. This package comes with its own documentation and a sizeable collection of examples. The URL for this package is https://github.com/igvgit/AsymmetricErrorsPy

Appendix D. Implementation in R

The implementation in R is available as a package, which can be installed, from within an R session, by `install.packages`("https://barlow.web.cern.ch/programs/AsymmetricErrors.tar.gz"). Alternatively, if preferred, the file can be downloaded and installed with R `CMD install` *downloadfilename*. (The case required for `install` appears to vary between systems.) Once installed, when required it can be loaded and attached whenever needed by `library(AsymmetricErrors)`.

The package comprises a set of functions that use S3 classes to distinguish between different models. Not all models have been implemented (yet). A list of likelihood models can be found by `methods(lnL)`.

There are 4 functions to handle likelihoods:

`getlnLpars(`*spec*`,` *type*`)`

> where *spec* is a named vector containing `val`, the position of the peak, and the positive and negative errors `sp` and `sm`. *type* is the model to be used. It returns a named vector with the appropriate model parameters added to *spec*, with class *type*. If there is no set of parameters that can be found for a model to match the required specification then a `warning` is raised and a `NULL` is returned. The `warning` can be suppressed (if appropriate) by a standard R `suppressWarning` call, and robust code will check for any `NULL` returns.

`lnL(p, a)`

> evaluates the log likelihood at `a` (which can be a vector) according to the model and parameters `p`. It returns a vector of the same length as `a`.

`combinelnLresults(r)`

> takes a list of parameter vectors (which must all be of the same model) and gives the combined result as a parameter vector.

`combinelnLerrors(r)`

> takes a list of parameter vectors, which must all be of the same model, and gives the combined result as a parameter vector. (The central value of the result, which is set to zero, is meaningless.)

These use named vectors as parameters and return values. The indexing of parameters within these vectors is not standardized: access should always be by name.

Pdfs are handled similarly. A list of models can be found by `methods(Pdf)`. (Notice the uppercase P, which avoids confusion with existing R graphics-related functions.) They are a little more complicated (but this is hidden from the user) as the specification in `getPdfpars` can be given either as the moments, `mu, V, gamma` or by the quantiles `M, sp, sm` (with `M` the median and `M-sp` and `M+sp` the 68% central interval). The function will take whichever set it is given and evaluate the model parameters. It will then use those to calculate the other set and include these in the returned vector. This may be inefficient as they will not always be needed, but it makes the handling easier: any parameter vector produced by `getPdfpars` will contain both the moments and the quantile parameters, as well as the specific model parameters.

`getPdfpars(`*spec*`, `*type*`)`

where *spec* is a named vector containing either the moments `mu, V, gamma`, or the quantile parameters `M, sp, sm`. *type* is the model to be used. It returns a named vector with the appropriate model parameters and the alternative specifiers added to *spec*, with class *type*. Again, if there is no set of parameters for the required specification then a `warning` is raised and `NULL` is returned.

`Pdf(p, a)`

evaluates the pdf at `a` according to the model and parameters `p`.

`combinePdferrors(r)`

takes a list of parameter values (which must all be of the same model), adds their moments to get the total μ, V, γ and returns the parameter set that give this combined result.

`combinePdfresults(r)`

takes a list of parameter values (which must all be of the same model) and gives the combined result.

`getflipPdfpars(`*spec, type, direction*`)`

returns the model parameters for a 'flipped' OPAT result. `direction` is + 1 (the default) when both deviations are positive and −1 when both are negative (the given values of both `sp` and `sm` are always positive). Only implemented for the dimidiated model.

All these functions have an associated `help` giving further details.

Some classes (such as `edgeworth` and `azzalini`) are used both for the likelihoods and for the pdfs, but the two are distinct, with no connection. The function `Title(p)`, which returns the name of the model of the parameters in printable form, with spaces and upper-case letters, may be useful.

Examples. We give some simple illustrative examples. They are very basic, just to illustrate simple uses, and do not contain the checks and elaboration required for 'good practice'. They assume that the library has been installed on your system and loaded and attached in this R session by `library(AsymmetricErrors)`.

- Likelihood errors

1. To combine two results for the Higgs width, using the linear variance model:

```
ATLAS <- getlnLpars(c(val=4.5,sp=3.3,
sm=2.5),"linearvariance")
CMS <- getlnLpars(c(val=3.2,sp=2.4,sm=
1.7),"linearvariance")
p <- combinelnLresults(list(ATLAS,CMS))
print(p)
```

2. To plot the lnL curves for the result $5.3^{+1.2}_{-0.9}$ using all available models:

```
result <- c(val=5.2,sp=1.2,sm=0.9)
a <- seq(0,10,.01)
plot(0,0,type='n',xlim=range(a),ylim=
c(-5,0))
for (t in methods(lnL)) {
 p <- getlnLpars(result,substring(t,5))
 lines(a,lnL(p,a))
}
```

The substring call is to remove the leading `lnL.` from the model names. The program should be elaborated to give titles, colors, and a legend, as desired.

3. The near end of an object is 100^{+3}_{-4} km away, and the far end is 200^{+4}_{-5} km away. To calculate the error on the length of the object:

```
near <- getlnLpars(c(val=100,sp=3,sm=4),
"linearsigma")
far <- getlnLpars(c(val=200,sp=4,sm=5),
"linearsigma")
print(combinelnLerrors(list(near,far)))
```

4. The combination of 3 results shown in Fig. 13 can be obtained by:

```
t <- "linearvariance"
v1 <- getlnLpars(c(sp=0.7,sm=0.5,
val=1.9),t)
v2 <- getlnLpars(c(sp=0.6,sm=0.8,
val=2.4),t)
v3 <- getlnLpars(c(sp=0.5,sm=0.4,
val=3.1),t)
a <- seq(1,4,.01)
plot(a,lnL(v1,a),type='l',ylab="lnL")
lines(range(a),-.5*c(1,1))
lines(a,lnL(v2,a))
lines(a,lnL(v3,a))
res <- combinelnLresults(list(v1,v2,v3))
lines(a,lnL(res,a),col='red')
print(res)
```

- Pdf Errors

1. The left hand plot curves of Fig. 4 were produced by the code:

```
x <- seq(0,10,.01)
p1 <- getPdfpars(c(M=5,sp=1.1,sm=0.9),
"dimidiated")
p2 <- getPdfpars(c(M=5,sp=1.1,sm=0.9),
"distorted")
plot(x,Pdf(p1,x),type='l',lwd=2,col=
'red',ylab="P(x)")
lines(x,Pdf(p2,x),col='green',lwd=2)
legend("topright",col=c("red","green"),
lwd=2,
    legend=c("Dimidiated","Distorted"))
```

and the right hand plot is the same, with `sp = 0.85, sm = 1.15` in place of `sp = 1.1, sm = 0.9`.

2. If a table of OPAT-style systematic errors has been saved on a file `errors.txt` as bare numbers, one set per line, for σ^+ and σ^-, then to find the total, using the dimidiated model:

```
df <- read.table("errors.txt")
N <- dim(df)[1]
errorlist <- list()
for(i in 1:N) errorlist[[i]] <-
 getPdfpars(c(M=0,sp=df[i,1],sm=
 df[i,2]),"dimidiated")
 print(combinePdferrors(errorlist))
```

3. To find the first three moments for a pdf specified as $5.0^{+1.1}_{-0.9}$ using all available models:

```
models <- methods(Pdf)
for(m in models){
m <- substring(m,5)
p <- getPdfpars(c(M=5,sp=1.1,sm=0.9),m)
print(paste("model",m,"mean ",p['mu'],
  "variance ",p['V'],"skewness ",
  p['gamma']))
}
```

(The `substring` call is to remove the `Pdf.` at the start of each name.)

4. To combine two measurements $12.34^{+0.56}_{-0.78}$ and $12.43^{+0.65}_{-0.87}$ using the railway Gaussian:

```
p1=getPdfpars(c(M=12.34,sp=0.56,sm=
0.78),"railway")
p2=getPdfpars(c(M=12.43,sp=0.65,sm=
0.87),"railway")
print(combinePdfresults(list(p1,p2)))
```

- Likelihood and Pdf Errors

1. The combined results of Example 5.3 were produced by the code:

```
poissonerrors <- function(x){
  L0 <- x*log(x)-x
  g <- function(y) { return (-y+x*log(y)-
  L0 +.5)}
  hi <- uniroot(g,c(x,x+2*sqrt(x)),
  tol=1.e-8)
  lo <- uniroot(g,c(x,max(0.0001,
  x-2*sqrt(x))),tol=1.e-8)
  return(c(sp=hi$root-x,sm=x-lo$root))
  }
phi <- function(x){
 # seek Delta ln L = -1/2 upwards
  xx <- x[1]
  L0 <- x[2]
  g <- function(y){ return(-y+xx*log(y)-
  L0+.5)}
  return(uniroot(g,c(xx,xx+2*sqrt(xx)),
  tol=1.e-8)$root-xx)
  }
plo <- function(x){
# seek Delta ln L = -1/2 downwards
  xx <- x[1]
  L0 <- x[2]
  g <- function(y){ return(-y+xx*log(y)
  -L0+.5)}
  return(xx-uniroot(g,c(xx,max(0.0001,
  xx-2*sqrt(xx))),tol=1.e-8)$root)
  }
model1 <- "distorted"
model2 <- "linearvariance"
t <- 1 # recording time of 1 second
s <- as.numeric(readline("Error on time in
sec - suggest 0.1 "))
sp <- 1/(t-s)-1/t
sm <- 1/t-1/(t+s)
rt <- 1/t
print(paste("result for time factor ",rt,
"+ ",sp, "- ",sm))

A <- 100 # acceptance factor

n <- as.numeric(readline("Expected number
of events - suggest 50 "))
pp <- poissonerrors(n)
print(paste("Number with poisson errors
is ",n,"+",pp['sp'],"-",pp['sm']))
r <- n*A*rt # the result
sp1 <- as.numeric(pp['sp']*A*rt)
sm1 <- as.numeric(pp['sm']*A*rt)
sp2 <- n*A*sp
sm2 <- n*A*sm
print(paste("Quoted result ", r, "+",sp1,"-",
sm1," (stat) +",sp2,"-",sm2," (syst)"))

p1p <- getPdfpars(c(M=r,sp=sp1,sm=sm1),
model1)
p2p <- getPdfpars(c(M=r,sp=sp2,sm=sm2),
model1)
p3p <- combinePdferrors(list(p1p,p2p))
print("Combine (wrongly) as two pdfs");
print(p3p)
p1L <- getlnLpars(c(val=r,sp=sp1,sm=sm1),
model2)
p2L <- getlnLpars(c(val=r,sp=sp2,sm=sm2),
model2)
p3L <- combinelnLerrors(list(p1L,p2L))
print("combine(wrongly) as two
likelihoods"); print(p3L)
```

Data availability

The code is available, as described in the paper.

References

[1] S. Manzoni, Higgs boson mass and width measurement with the ATLAS detector, in: Proc. EPS-HEP 2023, Venice, 2023.

[2] F. Errico, Higgs boson properties (mass/width) at CMS, in: Proc. EPS-HEP 2023, Venice, 2023.

[3] M. Frank, The latest three-flavor neutrino oscillation results from NOvA, in: Proc. EPS-HEP 2023, Venice, 2023.

[4] M. Reif, Recent belle II results on hadronic b decays, in: Proc. EPS-HEP 2023, Venice, 2023.

[5] V. Chobanova, Time-dependent cp violation measurements in b decays, in: Proc. EPS-HEP 2023, Venice, 2023.

[6] M. Schmelling, Averaging measurements with hidden correlations and asymmetric errors, 2000, arXiv:hep-ex/0006004v1.
[7] R.J. Barlow, Asymmetric systematic errors, 2003, arXiv:physics/0306138v1.
[8] R.J. Barlow, Asymmetric statistical errors, 2004, arXiv:physics/0406120v1.
[9] G. d'Agostini, Asymmetric uncertainties: Sources, treatment and potential dangers, 2004, arXiv:physis/0403086.v2.
[10] C.Merkat.A. Possolo, O. Biodnar, Asymmetrical uncertainties, Metrologia 56 (2019) 045009.
[11] L. Demortier, L. Lyons, Everything you always wanted to know about pulls, 2008, CDF Note, CDF/ANAL/PUBLIC/5776 (v3).
[12] S. Navas, et al., (Particle data group) review of particle properties, Phys. Rev. D 110 (2024) 030001.
[13] R.J. Barlow, A note on $\Delta \ln l = -\frac{1}{2}$, 2004, arXiv/physics/0403046.
[14] M.S. Bartlett, On the statistical estimation of mean lifetimes, Phil. Mag. 44 (1953) 244.
[15] M.S. Bartlett, Estimation of mean lifetimes from multiple plate cloud chamber tracks, Phil. Mag. 44 (1953) 1407.
[16] P. Hall, The Bootstrap and Edgeworth Expansion, Springer, 1992.
[17] P. Hall, Theoretical comparison of bootstrap confidence intervals, Ann. Stat. 16 (1988) 927.
[18] T.W. Keelin, B.W. Powley, Quantile-parameterized distributions, Decis. Anal. 8 (2011) 206.
[19] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Stat. 12 (1985) 181.
[20] N.L. Johnson, Systems of frequency curves generated by methods of translation, Biometrika 36 (1949) 149.
[21] W.P. Elderton, N.L. Johnson, Systems of Frequency Curves, Cambridge University Press, 1969.
[22] G.J. Hahn, S.S. Shapiro, Statistical Models in Engineering, Wiley, 1994.
[23] R. Hill I.D. Hill, R.L. Holder, Algorithm AS 99: fitting johnson curves by moments, Appl. Stat. 25 (1976) 180.
[24] S. Kotz N.L. Johnson, N. Balakrishnan, Continuous Univariate Distributions, vol. 2, Wiley, 1995.
[25] Wei-Ming Jao, Private communication, 2004.
[26] Simplified Wrapper and Interface Generator. https://www.swig.org.