Asymmetric Uncertainties: Sources, Treatment and Potential Dangers

G. D'Agostini
Università "La Sapienza" and INFN, Roma, Italia
(giulio.dagostini@roma1.infn.it, www.roma1.infn.it/~dagos)

Abstract

The issue of asymmetric uncertainties resulting from fits, nonlinear propagation and systematic effects is reviewed. It is shown that, in all cases, whenever a published result is given with asymmetric uncertainties, the *value* of the physical quantity of interest *is biased* with respect to what would be obtained using at best all experimental and theoretical information that contribute to evaluate the combined uncertainty. The probabilistic solution to the problem is provided both in exact and in approximated forms.

1 Introduction

We often see published results in the form

$$\text{`best value'}\ {}^{+\Delta_+}_{-\Delta_-}\,,$$

where Δ_+ and Δ_- are *usually* positive.[1] As firstly pointed out in Ref. [2] and discussed in a simpler but more comprehensive way in Ref. [3], this practice is far from being acceptable and, indeed, could bias the believed value of

[1]For examples of measurements having Δ_+ and Δ_- with all combinations of signs, see public online tables of Deep Inelastic Scattering results.[1] I want to make clear since the very beginning that it is not my intention to blame experimental or theoretical teams which have reported in the past asymmetric uncertainty, because we are all victims of a bad tradition in data analysis. At least, when asymmetric uncertainties have been given, there is some chance to correct the result, as described in Sec. 4. Since some asymmetric contributions to the global uncertainties almost unavoidably happen in complex experiments, I am more worried of collaborations that never arrive to final asymmetric uncertainties, because I must imagine they have symmetrised somehow the result but, I am afraid, without applying the proper shifts to the 'best value' to take into account asymmetric contributions, as it will be discussed in the present paper.

important physics quantities. The purpose of the present paper is, summarizing and somewhat completing the work done in the above references, to remind where asymmetric uncertainty stem from and to show why, as they are usually treated, they bias the value of physical quantities, either in the published result itself or in subsequent analyses. Once the problems are spotted, the remedy is straightforward, at least within the Bayesian framework (see e.g. [3], or [4] and [5] for recent reviews). In fact the Bayesian approach is conceptually based on the intuitive idea of probability, and formally grounded on the basic rules of probability (what are usually known as the probability 'axioms' and the 'conditional probability definition') plus logic. Within this framework many methods of 'conventional' statistics are reobtained, as approximations of general solutions, under well stated conditions of validity. Instead, in the conventional, frequentistic approach *ad hoc* formulae, prescriptions and un-needed principles are used, often without understanding what is behind these methods – before a 'principle' there is nothing!

The proposed Bayesian solutions to cure the troubles produced by the usual treatment of asymmetric uncertainties is to step up from approximated methods to the more general ones (see e.g. Ref. [3], in particular the top down approximation diagram of Fig. 2.2). In this paper we shall see, for example, how χ^2 and minus log-likelihood fit 'rules' can be derived from the Bayesian inference formulae as approximated methods and what to do when the underlying conditions do not hold. We shall encounter a similar situation regarding standard formulae to propagate uncertainty.

Some of the issues addressed here and in Refs. [2] and [3] have been recently brought to our attention by Roger Barlow [6], who proposes frequentistic ways out. Michael Schmelling had also addressed questions related to 'asymmetric errors', particularly related to the issue of weighted averages [7]. The reader is encouraged to read also these references to form his/her idea about the spotted problems and the proposed solutions.

In Sec. 2 the issue of propagation of uncertainty is briefly reviewed at an elementary level (just focusing on the sum of uncertain independent variables – i.e. no correlations considered) though taking into account asymmetry in probability density functions (p.d.f.) of the *input* quantities. In this way we understand what 'might have been done' (we are rarely in the positions to exactly know "what has been done") by the authors who publish asymmetric results and what is the danger of improper use of such a published 'best value' – *as is* – in subsequent analyses. Then, Sec. 3 we shall see in where asymmetric uncertainties stem from and what to do in order to overcome their potential troubles. This will be done in an exact way and, whenever is possible, in an approximated way. Some rules of thumb to roughly recover sensible probabilistic quantities (expected value and standard deviation) from results published with asymmetric uncertainties will be given in Sec. 4. Finally, some conclusions will be drawn.

2 Propagating uncertainty

Determining the value of a physics quantity is seldom an end in itself. In most cases the result is used, together with other experimental and theoretical quantities, to calculate the value of other quantities of interest. As it is well understood, uncertainty on the value of each ingredient is propagated into uncertainty on the final result.

If uncertainty is quantified by probability, as it is commonly done explicitly or implicitly[2] in physics, the propagation of uncertainty is performed using rules based on probability theory. If we indicate by $\boldsymbol{X}$ the set ('vector') of input quantities and by Y the final quantity, given by the function $Y = Y(\boldsymbol{X})$ of the input quantities, the most general propagation formula (see e.g. [3]) is given by (we stick to continuous variables):

$$f(y) = \int \delta[y - Y(\boldsymbol{x})] \cdot f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,, \tag{1}$$

where $f(y)$ is the p.d.f. of Y, $f(\boldsymbol{x})$ stands for the joint p.d.f. of $\boldsymbol{X}$ and δ is the Dirac delta (note the use of capital letters to name variables and small letters to indicate the values that variables may assume). The exact evaluation of Eq. (1) is often challenging, but, as discussed in Ref. [3], this formula has a nice simple interpretation that makes its Monte Carlo implementation conceptually easy.

As it is also well known, often there is no need to go through the analytic, numerical or Monte Carlo evaluation of Eq.(1), since linearization of $Y(\boldsymbol{x})$ around the expected value of $\boldsymbol{X}$ ($\mathrm{E}[\boldsymbol{X}]$) makes the calculation of expected value and variance of Y very easy, using the well known standard propagation formulae, that for uncorrelated input quantities are

$$\mathrm{E}[Y] \approx Y(\mathrm{E}[\boldsymbol{X}]) \tag{2}$$

[2]Perhaps the reader would be surprised to learn that in the conventional statistical approach there is no room for probabilistic statements about the value of physics quantities (e.g. "the top mass is between 170 and 180 GeV with such percent probability", or "there is 95% probability that the Higgs mass is lighter than 200 GeV"), calibration constants, and so on, as discussed extensively in Ref. [8].

$$\sigma^2(Y) \;\approx\; \sum_i \left(\left.\frac{\partial Y}{\partial X_i}\right|_{\mathrm{E}[\boldsymbol{X}]} \right)^2 \sigma^2(X_i)\,. \qquad (3)$$

As far as the shape of $f(y)$, a Gaussian one is usually assumed, as a result of the central limit theorem. Holding this assumptions, E$[Y]$ and $\sigma(Y)$ is all what we need. E$[Y]$ gives the 'best value', and probability intervals, upper/lower limits and so on can be easily calculated. In particular, within the Gaussian approximation, the most believable value (*mode*), the barycenter of the p.d.f. (*expected value*) and the value that separates two adjacent 50% probability intervals (*median*) coincide. If $f(y)$ is asymmetric this is not any longer true and one needs then to clarify what 'best value' means, which could be one of the above three *position parameters*, or something else (in the Bayesian approach 'best value' stands for expected value, unless differently specified).

Anyhow, Gaussian approximation is not the main issue here and, in most real applications, characterized by several contributions to the combined uncertainty about Y, this approximation is a reasonable one, even when some of the input quantities individually contribute asymmetrically. My concerns in this paper are more related to the evaluation of E$[Y]$ and $\sigma(Y)$ when

1. instead of Eqs. (2)–(3), *ad hoc* propagation prescriptions are used in presence of asymmetric uncertainties;
2. linearization implicit in Eqs. (2)–(3) is not a good approximation.

Let us start with the first point, considering, as an easy academic example, input quantities described by the asymmetric triangular distribution shown in the left plot of Fig. 1. The value of X can range between -1 and 1, with a 'best value', in the sense of maximum probability value, of 0.5. The interval $[-0.16, +0.72]$ gives a 68.3% probability interval, and the 'result' could be reported as $X_1 = 0.50^{+0.22}_{-0.66}$. This is not a problem as long as we known what this notation means and, possibly, know the shape of $f(x)$. The problem arises when we want to make use of this result and we do not have access to $f(x)$ (as it is often the case), or we make improper use of the information [i.e. in the case we are aware of $f(x)$]. Let us assume, for simplicity, to have a second independent quantity, X_2, described exactly by the same p.d.f. and reported in the same way: $X_2 = 0.50^{+0.22}_{-0.66}$. Imagine we are now interested to the quantity $Y = X_1 + X_2$. How to report the result about Y, based on the results about Y_1 and Y_2? Here are some common, but *wrong* ways to give the result:

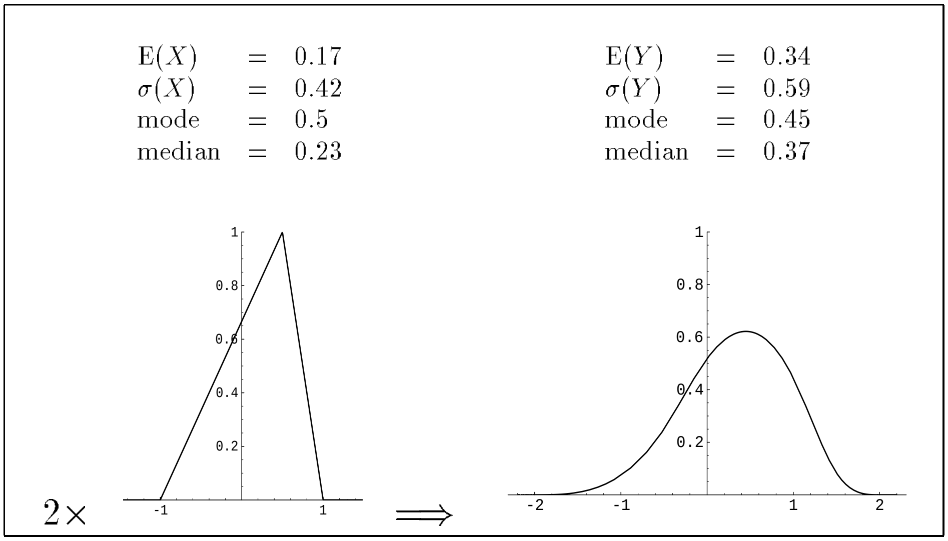

Figure 1: Distribution of the sum of two independent quantities, each described by an asymmetric triangular p.d.f. self-defined in the left plot. The resulting p.d.f. (right plot) has been calculated analytically making use of Eq.(1). This figure corresponds to Fig. 4.3 of Ref. [3].

- *asymmetric uncertainties added in quadrature*: $Y = 1.00^{+0.31}_{-0.93}$;
- *asymmetric uncertainties added linearly*: $Y = 1.00^{+0.44}_{-1.31}$.

Indeed, in this simple case we can calculate the integral (1) analytically, obtaining the curve shown in the plot on the right side of Fig. 1, where several position and shape parameters have also been reported. The 'best value' of Y, meant as expected value (i.e. the barycenter of the p.d.f.) comes out to be 0.34. Even those who like to think at the 'best value' as the value of maximum probability (density) would choose 0.45 (note that in this particular example the mode of the sum is smaller than the mode of each addend!). Instead, a 'best value' of Y of 1.00 obtained by the *ad hoc* rules, unfortunately often used in physics, corresponds neither to mode, nor to expected value or median.

The situation would have been much better if expected value and standard deviation of X_1 and X_2 had been reported (respectively 0.17 and 0.42). Indeed, these are the quantities that matter in 'error propagation', because the *theorems upon which propagation formulae rely* — exactly in the case Y is a linear combination of $\boldsymbol{X}$, or approximately in the case linearization has been performed — *speak of expected values and variances*. It is easy to verify from the numbers in Fig. 1 that exactly the correct values of $\mathrm{E}[Y] = 0.34$ and $\sigma(Y) = 0.59$ would have been obtained. Moreover, one can see that expected value, mode and median of $f(y)$ do not differ much from each other, and the shape of $f(y)$ resembles a somewhat skewed Gaussian. When Y will be combined with other quantities in a next analysis its slightly non-Gaussian shape will not matter any longer. Note that we have achieved this nice result already with only two input quantities. If we had a few more, already Y would have been much Gaussian-like. Instead, performing a bad combination of several quantities all skewed in the same side would yield 'divergent' results[3]: for $n = 10$ we get, using a quadratic combination of left and right deviations, $Y = 5.00^{+0.69}_{-2.07}$ versus the correct $Y = 1.70 \pm 1.32$.

As conclusion from this section I would like to make some points:

- in case of asymmetric uncertainty on a quantity, it should be avoided to report *only* most probable value and a probability interval (be it 68.3%, 95%, or what else);

- expected value, meant as barycenter of the distribution, as well as standard deviations should *always* be reported, providing also the shape of the distribution (or its summary in terms of shape parameters, or even a parameterization of the log-likelihood function in a polynomial form, as done e.g. in Ref. [9]), if the distribution is asymmetric or non trivial.

Note that the propagation example shown here is the most elementary possible. The situation gets more complicate if also nonlinear propagation is involved (see Sec. 3.2) or when quantities are used in fits (see e.g. Sec. 12.1 of Ref. [3]).

Hoping that the reader is, at this point, at least worried about the effects of badly treated asymmetric uncertainties, let us now review the sources of asymmetric uncertainties.

[3] The reader might be curious to know what would happen in case of bad combinations of input quantities with skewness of mixed signs. Clearly there will be some compensation that lowers the risk of strong bias. As an academic exercise, let think of five independent variables each described by the triangular distribution of Fig. 1 and five others each described by a p.d.f. which is its mirror reflexed around $x = 0.5$ [$0 \le X \le 2$, $\mathrm{mode}(X) = 0.5$, $\mathrm{E}[X] = 0.83$ and $\sigma(X) = 0.42$]. The correct combination of the ten variables gives $Y = 5.00 \pm 1.33$, while adding the modes and combining quadratically left and right deviations we would get 5.00 ± 1.54.

3 Sources of asymmetric uncertainties

3.1 Non parabolic χ^2 or log-likelihood curves

The standard methods in physics to adjust theoretical parameters to experimental data are based on *maximum likelihood principle* ideas. In practice, depending on the situation, the 'minus log-likelihood' of the parameters [$\varphi(\boldsymbol{\theta};\mathrm{data}) = -\ln L(\boldsymbol{\theta};\mathrm{data})$] or the χ^2 function of the parameters [i.e. the function $\chi^2(\boldsymbol{\theta};\mathrm{data})$] are minimized. The notation used reminds that φ and χ^2 are seen as mathematical function of the parameters $\boldsymbol{\theta}$, with the data acting as 'parameters' of the functions. As it is well understood, a part from an irrelevant constant non depending on fit parameters, φ and χ^2 differ by just a factor of two when the likelihood, seen as a joint probability function or a p.d.f. of the data, is a (multivariate) Gaussian distribution of the data: $\varphi = \chi^2/2 + k$ (the constant k is often neglected, since we concentrate on the terms which depend on the fit parameters – but sometimes χ^2 and minus log-likelihood might differ by terms depending on fit parameters!). For sake of simplicity, let us take one parameter fit. Following the usual practice, we indicate the parameter by θ (though this fit parameter is just any of the input quantities $\boldsymbol{X}$ of Sec. 2).

If $\varphi(\theta)$ or $\chi^2(\theta)$ have a nice parabolic shape, the likelihood is, apart a multiplicative factor, a Gaussian function[4] of θ. In fact, as is well known from calculus, any function can be approximated to a parabola in the vicinity of its minimum. Let us see in detail the expansion of $\varphi(\theta)$ around its minimum θ_m:

$$\begin{aligned}
\varphi(\theta) &\approx \varphi(\theta_m) + \left.\frac{\partial\varphi}{\partial\theta}\right|_{\theta_m}(\theta-\theta_m) + \left.\frac{\partial^2\varphi}{\partial\theta^2}\right|_{\theta_m}\frac{1}{2}\,(\theta-\theta_m)^2 && (4)\\
&\approx \varphi(\theta_m) + \frac{1}{2}\,\frac{1}{\alpha^2}\,(\theta-\theta_m)^2\,, && (5)
\end{aligned}$$

where the second term of the r.h.s. vanishes by definition of minimum and we have indicated with α the inverse of the second derivative at the minimum. Going back to the likelihood, we get:

$$\begin{aligned}
L(\theta;\mathrm{data}) &\approx \exp\left[-\varphi(\theta_m)\right]\cdot\exp\left[-\frac{1}{2}\,\frac{1}{\alpha^2}\,(\theta-\theta_m)^2\right] && (6)\\
&\approx k\,\exp\left[-\frac{(\theta-\theta_m)^2}{2\,\alpha^2}\right], && (7)
\end{aligned}$$

[4] But not yet a probability function! The likelihood has the probabilistic meaning of a joined p.d.f. of the data given θ, and not the other way around.

apart a multiplicative factor, this is 'Gaussian' centered in θ_m with standard deviation $(\partial^2\varphi/\partial\theta^2|_{\theta_m})^{-2}$. However, although this function is mathematically a Gaussian, it does not have yet the meaning of probability density $f(\theta\,|\,\mathrm{data})$ in an inferential sense, i.e. describing our knowledge about θ in the light of the experimental data. In order to do this, we need to process the likelihood through Bayes theorem, which allows *probabilistic inversions* to be achieved using basic rules of probability theory and logic. Besides a conceptually irrelevant normalization factor (that has to be calculated at some moment) the Bayes formula is

$$f(\theta\,|\,\mathrm{data}) \;\propto\; f(\mathrm{data}\,|\,\theta)\cdot f_0(\theta)\,. \tag{8}$$

We can speak now about the "probability that θ is within a given interval" and calculate it, together with expectation of θ, standard deviation and so on.[5] If the *prior* $f_0(\theta)$ is much vaguer that what the data can teach us (via the likelihood), then it can be re-absorbed in the normalization constant, and we get:

$$f(\theta\,|\,\mathrm{data}) \;\propto\; f(\mathrm{data}\,|\,\theta) = L(\theta;\mathrm{data}) \tag{9}$$

i.e

$$\propto\; \exp\left[-\varphi(\theta;\mathrm{data})\right] \tag{10}$$

or

$$\propto\; \exp\left[-\frac{\chi^2(\theta;\mathrm{data})}{2}\right] \tag{11}$$

parabolic φ or χ^2 :

$$\rightarrow\; f(\theta\,|\,\mathrm{data}) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma_\theta}\exp\left[-\frac{(\theta-\mathrm{E}[\theta])^2}{2\,\sigma_\theta^2}\right]. \tag{12}$$

If this is the case, it is a simple exercise to show that

a) $\mathrm{E}[\theta]$ is equal to θ_m which minimizes the χ^2 or φ.

b) σ_θ can be obtained by the famous conditions $\Delta\chi^2 = 1$ or $\Delta\varphi = 1/2$, respectively, or by the second derivative around θ_m: $\sigma_\theta^{-2} = 1/2 \times (\partial^2\chi^2/\partial\theta^2)\big|_{\theta_m}$ or $\sigma_\theta^{-2} = (\partial^2\varphi/\partial\theta^2)\big|_{\theta_m}$, respectively.

[5] θ has not a probabilistic interpretation in the frequentistic approach, and therefore we cannot speak consistently, in that framework, about its probability, or determine expectation, standard deviation and so on. Most physicists do not even know of this problem and think these are irrelevant semantic quibbles. However, it is exactly this contradiction between intuitive thinking and cultural background[8] that causes wrong scientific conclusions, like those discussed in this paper.

Though in the frequentistic approach language and methods are usually more convoluted (even when the same numerical results of the Bayesian reasoning are obtained), due to the fact that probabilistic statements about physics quantities and fit parameters are not allowed in that approach, it is usually accepted that the above rules a and b are based on the parabolic behavior of the minimized functions. When this approximation does not hold, the frequentist has to replace a prescription by other prescriptions that can handle the exception.[6] The situation is simpler and clearer in the Bayesian approach, in which the above rules a and b do hold too, but only as approximations under well defined conditions. In case the underlying conditions fail we know immediately what to do:

- restart from Eq. (9) or from Eq. (11), depending on the other underlying hypotheses;
- go even one step before Eq. (9), namely to the most general Eq. (8), if priors matter (e.g. physical constraints, sensible previous knowledge, etc.).

For example, if the χ^2 description of the data was a good approximation, then $f(\theta) \propto e^{-\chi^2/2}$, properly normalized, is the solution to the problem.[7] A non parabolic, asymmetric χ^2 produces an asymmetric $f(\theta)$ (see Fig. 2),

[6] It is a matter of fact that the habit in the particle physics community of applying uncritically the $\Delta\chi^2 = 1$ or $\Delta\varphi = 1/2$ is related to the use of the software package *MINUIT*[10]. Indeed, *MINUIT* can calculate the parameter variances also from the χ^2 or φ curvature at the minimum (that relies on the same hypothesis upon which the $\Delta\chi^2 = 1$ or $\Delta\varphi = 1/2$ rules are based). But when the χ^2 or φ are no longer parabolic, the standard deviation calculated from the curvature differs from that of the $\Delta\chi^2 = 1$ or $\Delta\varphi = 1/2$ (in particular, when the minimized function is asymmetric the latter rules give two values, the (in-)famous $\Delta_\pm$ we are dealing with). People realize that the curvature at the minimum depends from the local behavior of the minimized curve, and the $\Delta\chi^2 = 1$ or $\Delta\varphi = 1/2$ rule is typically more stable. Therefore, in particle physics the latter rule has become *de facto* a standard to evaluate 'confidence intervals' at different 'levels of confidence' (depending of the value of the $\Delta\chi^2$ or $\Delta\varphi$). But, unfortunately, when those famous curves are not parabolic, numbers obtained by these rules might loose completely a probabilistic meaning. [Sorry, a frequentist would object that, indeed, these numbers do not have probabilistic meaning about θ, but they are 'confidence intervals' at such and such 'confidence level', because 'θ is a constant of unknown value', etc... Good luck!]

[7] To be precise, this approximation is valid if the parameters appear only in the argument of the exponent. In practice this means that the fitted parameters must not appear in the covariance matrix on which the χ^2 depends. As a simple example in which this approximation do not hold is that of a linear fit in which also the standard deviation σ describing the errors along the ordinate. The joint inference about line coefficients m and c and σ, having observed n points, is achieved by $f(m,c,\sigma) \propto \sigma^{-n}\, e^{-\chi^2/2}$ (see Sec. 8.2 of Ref. [3]).

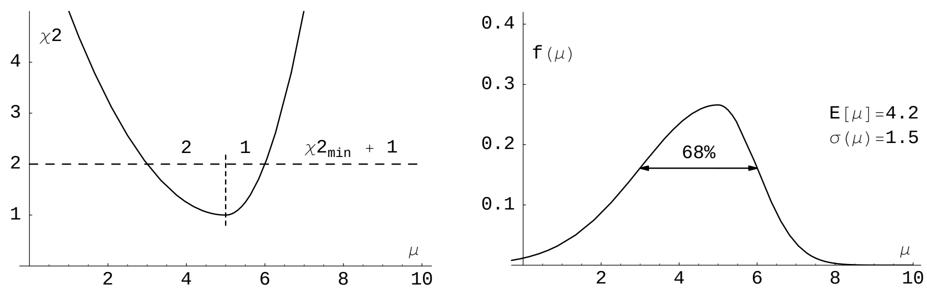

Figure 2: Example (Ref. [3]) of asymmetric χ^2 curve (left plot) with a χ^2 minimum at $\mu = 5$ (μ stands for the value of a generic physics quantity). The result based on the $\chi^2_{min} + 1$ 'prescription' is compared (plot on the right side) with the p.d.f. based on a uniform prior, i.e. $f(\mu \,|\, \mathrm{data}) \propto \exp[-\chi^2/2]$.

the mode of which corresponds, indeed, to what obtained minimizing χ^2, but expected value and standard deviation differ from what is obtained by the 'standard rule'. In particular, expected value and variance must be evaluated from their definitions:

$$\mathrm{E}[\theta] = \int \theta\, f(\theta \,|\, \mathrm{data})\, \mathrm{d}\theta \tag{13}$$

$$\sigma^2_\theta = \int (\theta - \mathrm{E}[\theta])^2\, f(\theta \,|\, \mathrm{data})\, \mathrm{d}\theta\,. \tag{14}$$

Other examples of asymmetric χ^2 curves, including the case with more than one minimum, are shown in Chapter 12 of Ref. [3], and compared with the results coming from frequentist prescriptions (but, indeed, there is not a general accepted rule to get frequentistic results – whatever they mean – when the χ^2 shape gets complicated).

Unfortunately, it is not easy to translate numbers obtained by *ad hoc* rules into probabilistic results, because the dependence on the actual shape of the χ^2 or φ curve can be not trivial. Anyhow, some *rules of thumb* can be given in next-to-simple situations where the χ^2 or φ has only one minimum and the χ^2 or φ curve looks like a 'skewed parabola', like in Fig. 2:

- the 68% 'confidence interval' obtained by the $\Delta\chi^2 = 1$, or $\Delta\varphi = 1/2$ rule still provides a 68% probability interval for θ.

- the standard deviation obtained using Eq. (14) is approximately equal to the average between the Δ_+ and Δ_- values obtained by the $\Delta\chi^2 =$

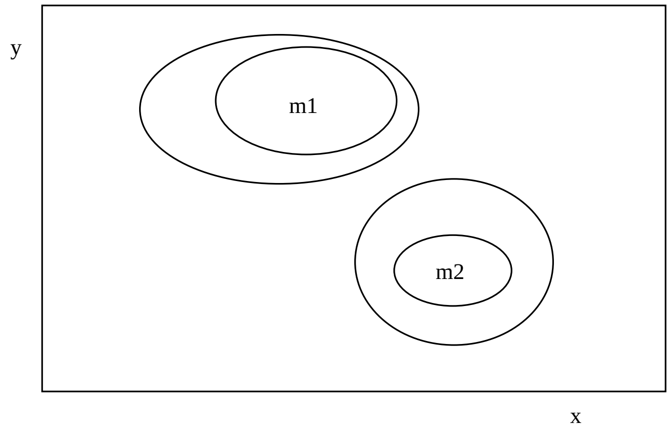

Figure 3: Example of two-dimensional multi-spots "68% CL" and "95% CL" contours obtained slicing the χ^2 or the minus log-likelihood curve at some *magic levels*. What do they mean?

1, or $\Delta\varphi = 1/2$ rule:

$$\sigma_\theta \approx \frac{\Delta_+ + \Delta_-}{2}\,; \tag{15}$$

- the expected value is equal to the mode (θ_m, coinciding with the maximum likelihood or minimum χ^2 value) plus a *shift*:

$$\mathrm{E}[\theta] \approx \theta_m + \mathcal{O}(\Delta_+ - \Delta_-)\,. \tag{16}$$

  [This latter rule is particularly rough because $\mathrm{E}[\theta]$ is more sensitive than σ_θ on the exact shape of χ^2 or φ curve. Equation (16) has to be taken only to get an idea of the order of magnitude of the effect. For example, in the case depicted in Fig 2 the shift is 80% of $(\Delta_+ - \Delta_-)$.]

The remarks about misuse of $\Delta\chi^2 = 1$ and $\Delta\varphi = 1/2$ rules can be extended to cases where several parameters are involved. I do not want to go into details (in the Bayesian approach there is nothing deeper than studying $k\,e^{-\chi^2/2}$ or $k\,e^{-\varphi}$ in function of several parameters.[8]), but I just want to get the reader worried about the meaning of contour plots of the kind shown in Fig. 3.

[8] See footnote 7 concerning a possible pitfall in the use of $k\,e^{-\chi^2/2}$.

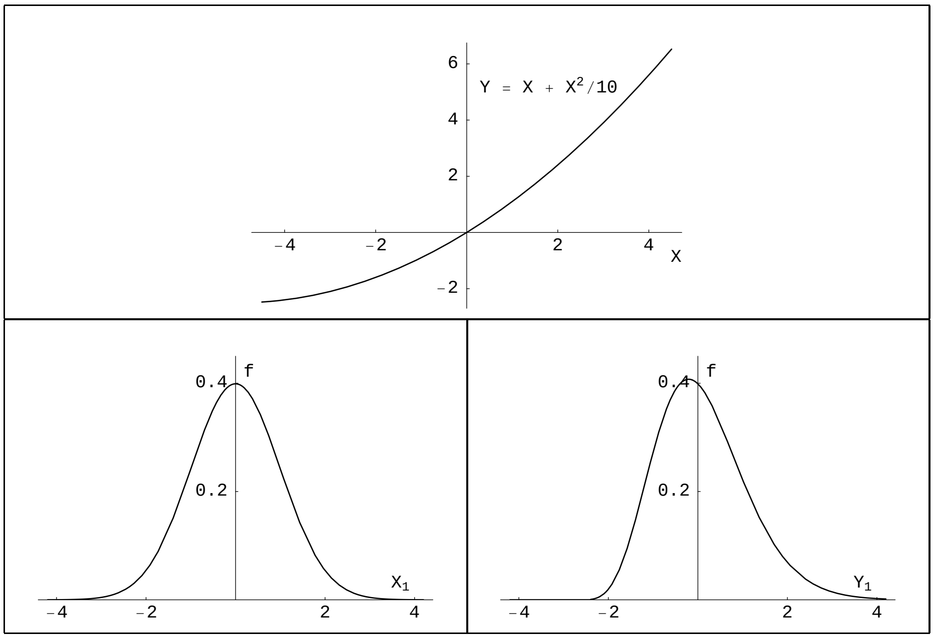

Figure 4: Propagation of a Gaussian distribution under a nonlinear transformation. $f(Y_i)$ were obtained analytically using Eq.(1) (part of Fig. 12.2 of Ref.[3]).

3.2 Nonlinear propagation

Another source of asymmetric uncertainties is nonlinear dependence of the output quantity Y on some of the input $\boldsymbol{X}$ in a region a few standard deviations around $\mathrm{E}(\boldsymbol{X})$. This problem has been studied with great detail in Ref. [2], also taking into account correlations on input and output quantities, and somewhat summarized in Ref. [3]. Let us recall here only the most relevant outcomes, in the simplest case of only one output quantity Y and neglecting correlations.

Figure 4 shows a non linear dependence between X and Y and how a Gaussian distribution has been distorted by the transformation [$f(y)$ has been evaluated analytically using Eq.(1)]. As a result of the nonlinear transformation, mode, mean, median and standard deviation are transformed in non trivial ways (in the example of Fig. 4 mode moves left and expected value right). In the general case the complete calculations should be performed, either analytically, or numerically or by Monte Carlo. Fortunately, as it has been shown in Ref. [2], second order expansion is often enough to take into account small deviations from linearity. The resulting formulae are still compact and depend on location and shape parameters of the initial distributions.

Second order propagation formulae depend on first and second derivatives. In practical cases (especially as far as the contribution from systematic effects are concerned) the derivatives are obtained numerically[9] as

$$\left.\frac{\partial Y}{\partial X}\right|_{\mathrm{E}[X]} \approx \frac{1}{2}\left(\frac{\Delta_+}{\sigma(X)}+\frac{\Delta_-}{\sigma(X)}\right)=\frac{\Delta_+ + \Delta_-}{2\,\sigma(X)}\,, \tag{17}$$

$$\left.\frac{\partial^2 Y}{\partial X^2}\right|_{\mathrm{E}[X]} \approx \frac{1}{\sigma(X)}\left(\frac{\Delta_+}{\sigma(X)}-\frac{\Delta_-}{\sigma(X)}\right)=\frac{\Delta_+ - \Delta_-}{\sigma^2(X)}\,, \tag{18}$$

where Δ_- and Δ_+ now stand for the left and right deviations of Y when the *input variable X varies by one standard deviation* around $\mathrm{E}[X]$. Second order propagation formulae are conveniently given in Ref. [2] in terms of the $\Delta_\pm$ deviations[10]. For Y that depends only on a single input X we get:

$$\mathrm{E}(Y) \approx Y(\mathrm{E}[X])+\delta\,, \tag{21}$$

$$\sigma^2(Y) \approx \overline{\Delta}^2+2\,\overline{\Delta}\cdot\delta\cdot S(X)+\delta^2\cdot[\mathcal{K}(X)-1]\,, \tag{22}$$

where δ is the semi-difference of the two deviations and $\overline{\Delta}$ is their average:

$$\delta = \frac{\Delta_+ - \Delta_-}{2} \tag{23}$$

$$\overline{\Delta} = \frac{\Delta_+ + \Delta_-}{2}\,, \tag{24}$$

while $\mathcal{S}(X)$ and $\mathcal{K}(X)$ stand for skewness and kurtosis of the input variable.[11]

[9] Note that sometimes people do not get asymmetric uncertainty, not because the propagation is approximately linear, but because asymmetry is hidden by the standard propagation formula! Therefore also in this case the approximation might produce a bias in the result (for example, the second order formula of the expected value of the ratio of two quantities is known to experts[12]). The merit of numerical derivatives is that at least it shows the asymmetries.

[10] In terms of analytically calculated derivatives, δ and $\overline{\Delta}$ are given by

$$\delta = \frac{1}{2}\left.\frac{\partial^2 Y}{\partial X^2}\right|_{\mathrm{E}[X]}\sigma^2(X) \tag{19}$$

$$\overline{\Delta} = \left.\frac{\partial Y}{\partial X}\right|_{\mathrm{E}[X]}\sigma(X)\,. \tag{20}$$

[11] After what we have seen in Sec. 2 we should not forget that the input quantities

For many input quantities we have

$$\mathrm{E}(Y) \approx Y(\mathrm{E}[X]) + \sum_i \delta_i\,, \tag{25}$$

$$\sigma^2(Y) \approx \sum_i \sigma^2_{X_i}(Y)\,, \tag{26}$$

where $\sigma^2_{X_i}(Y)$ stands for each individual contribution to Eq. (22). The expression of the variance gets simplified when all input quantities are Gaussian (a Gaussian has skewness equal 0 and kurtosis equal 3):

$$\sigma^2(Y) \approx \sum_i \overline{\Delta}_i^2 + 2\sum_i \delta_i^2\,, \tag{27}$$

and, as long as δ_i^2 are much smaller that $\overline{\Delta}_i^2$, we get the convenient *approximated formulae*[13]

$$\mathrm{E}(Y) \approx Y(\mathrm{E}[\boldsymbol{X}]) + \sum_i \delta_i\,, \tag{28}$$

$$\sigma^2(Y) \approx \sum_i \overline{\Delta}_i^2 \tag{29}$$

valid also for other symmetric input p.d.f.'s (the kurtosis is about 2 to 3 in typical distribution and its exact value is irrelevant if the condition $\sum_i \delta_i^2 \ll \sum_i \overline{\Delta}_i^2$ holds). The resulting practical rules (28)–(29) are quite simple:

- the *expected value* of Y *is shifted* by the sum of the individual shifts, each given by half of the semi-difference of the deviations $\Delta_\pm$;
- each input quantity contributes (in quadrature) to the combined standard uncertainty with a term which is approximately the average between the deviations $\Delta_\pm$.

Moreover, if there are many contributions to the uncertainty, the final uncertainty will be symmetric and approximately Gaussian, thanks to the central limit theorem.

---

could have non trivial shapes. Since skewness and kurtosis are related to 3rd and 4th moment of the distribution, Eq. (22) makes use up to the 4th moment and is definitely better that the usual propagation formula, that uses only second moments. In Ref. [2] approximated formulae are given also for skewness and kurtosis of the output variable, from which it is possible to reconstruct $f(y)$ taking into account up to 4-th order moment of the distribution.

3.3 Uncertainty due to systematics

Finally, and this is often the case that we see in publications, asymmetric uncertainty results from systematic effects. The Bayesian approach offers a natural and clear way to treat systematics – and I smile at the many attempts[12] of 'squaring the circle' using frequentistic prescriptions... – simply because probabilistic concepts are consistently applied to all *influence quantities* that can have an effect on the quantity of interest and whose value is not precisely known. Therefore we can treat them using probabilistic methods. This was also recognized by metrologic organizations[15].

Indeed, there is no need to treat systematic effects in a special way. They are treated as any of the many input quantities $\boldsymbol{X}$ discussed in Sec. 3.2, and, in fact, their asymmetric contributions come frequently from their nonlinear influence on the quantity of interest. The only word of caution, on which I would like to insist, is to use expected value and standard deviation for each systematic effect. In fact, sometimes the uncertainty about the value of the influence quantities that contribute to systematics is intrinsically asymmetric.

I also would like to comment shortly on results where either of the $\Delta_\pm$ is negative, for example $1.0^{+0.5}_{+0.3}$ (see e.g. Ref. [1] to have an idea of the variety of signs of $\Delta_\pm$). This means that that the we are in proximity of a minimum (or a maximum if Δ_+ were negative) of the function $Y = Y(X_i)$. It can be shown [2, 3] that Eqs. (21)-(22) hold for this case too.[13]

For further details about meaning and treatment of uncertainties due systematics and their relations to ISO *Type B* uncertainties[15], see Refs. [2] and [3].

4 Some rules of thumb to unfold probabilistic sensible information from results published with asymmetric uncertainties

Having understood what one should have done to obtain expected value and standard deviation in the situations in which people are used to report asym-

[12] It has been studied by psychologists how sometimes our efforts to solve a problem are the analogous with the moves along elements of a group structure (in the mathematical sense). There is no way to reach a solution until we not break out of this kind of trapping psychological or cultural cages.[14]

[13] In this special case there should be no doubt that a shift should be applied to the best value, since moving X_i by $\pm\sigma(X_i)$ around its expected value $\mathrm{E}[X_i]$ the final quantity Y only moves in one side of $Y(\mathrm{E}[X_i])$.

metric uncertainties, we might attempt to recover those quantities from the published result. It is possible to do it exactly only if we know the detailed contributions to the uncertainty, namely the χ^2 or log-likelihood functions of the so called 'statistical terms' and the pairs $\{\Delta_{+_i}, \Delta_{-_i}\}$, together to the probabilistic model, for each 'systematic term'. However, these pieces of information are usually unavailable. But we can still make some *guesses*, based on some rough assumptions, lacking other information:

- asymmetric uncertainties in the 'statistical part' are due to asymmetric χ^2 or log-likelihood: $\rightarrow$ apply corrections given by Eqs. (15)–(16);
- asymmetric uncertainties in the 'systematic part' comes from nonlinear propagation: $\rightarrow$ apply corrections given by Eqs. (28)–(29).

As a numerical example, imagine we read the following result (in arbitrary units):

$$Y = 6.0\,^{+1.0}_{-2.0}\,^{+0.3}_{-0.9}\,, \tag{30}$$

(that somebody would summary as $6.0\,^{+1.0}_{-2.2}$!). The only certainty we have, seeing two asymmetric uncertainties with the same sign of skewness, is that *the result is definitively biased.* Let us try to make our estimate of the bias and calculate the corrected result (that, not withstanding all uncertainties about uncertainties, will be closer to the 'truth' than the published one):

1. the first contribution gives roughly [see. Eqs. (15)–(16)]:

$$\delta_1 \approx -1.0 \tag{31}$$
$$\sigma_1 \approx 1.5\,; \tag{32}$$

2. for the second contribution we have [see. Eqs. (24)–(24), (28)–(29)]:

$$\delta_2 \approx -0.31 \tag{33}$$
$$\sigma_2 \approx 0.62\,. \tag{34}$$

Our guessed best result would then become[14]

$$Y \approx 4.69 \pm 1.5 \pm 0.62 = 4.69 \pm 1.62 \tag{35}$$
$$\approx 4.7 \pm 1.6\,. \tag{36}$$

[14] The ISO Guide [15] recommends to give the result using the standard deviation within parenthesis, instead of using the $\pm xx$ notation. In this example we would have $Y \approx 4.69\,(1.5)\,(0.62) = 4.69\,(1.62) \Rightarrow Y \approx 4.7\,(1.6)$. Personally, I do not think this is a very important issue as long as we know what the quantity xx means. Anyhow, I understand the ISO rational, and perhaps the proposed notation could help to make a break with the 'confidence intervals'.

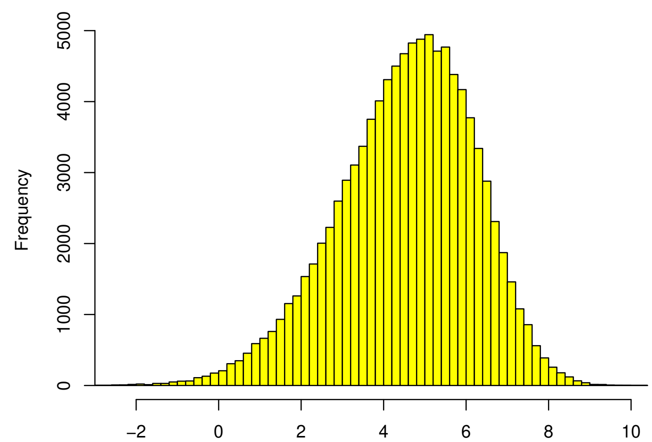

Figure 5: Monte Carlo estimate of the shape of the p.d.f. of the sum of three independent variables, one described by the p.d.f. of Fig. 2 and the other two by the triangular distribution of Fig. 1.

(The exceeding number of digits in the intermediate steps are just to make numerical comparison with the correct result that will be given in a while.)

If we had the chance to learn that the result of Eq. (30) was due to the asymmetric χ^2 fit of Fig. 2 plus two systematic corrections, each described by the triangular distribution of Fig. 1, then we could calculate expectation and variance exactly:

$$\mathrm{E}(Y) = 4.2 + 2 \times 0.17 = 4.54 \tag{37}$$
$$\sigma^2(Y) = 1.5^2 + 2 \times 0.42^2 = 1.61^2\,, \tag{38}$$

i.e. $Y = 4.54 \pm 1.61$, quite different from Eq. (30) and close to the result corrected by rule of thumb formulae. Indeed, knowing exactly the ingredients, we can evaluate $f(y)$ from Eq.(1) as

$$f(y) = \int \delta(y - x_1 - x_2 - x_3)\, f_1(x_1)\, f_2(x_2)\, f_3(x_3)\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3\,, \tag{39}$$

although by Monte Carlo. The result is given in Fig. 5, from which we can evaluate a mean value of 4.54 and a standard deviation of 1.65 in perfect agreement with the figures given in Eqs. (37)–(38).[15] As we can see from

[15]The slight difference between the standard deviations comes from rounding, since $\sigma(\mu) = 1.5$ of Fig. 2 is the rounded value of 1.54. Replacing 1.5 by 1.54 in Eq. (38), we get exactly the Monte Carlo value of 1.65.

the figure, also those who like to think at 'best value' in term of most probable value have to realize once more that *the most probable value of a sum is not necessarily equal to the sum of most probable values of the addends* (and analogous statements for all combinations of uncertainties[16]). In the distribution of Fig. 5, the mode of the distribution is around 5. [Note that expected value and variance are equal to those given by Eqs. (37)–(38, since in the case of a linear combination they can be obtained exactly.] Other statistical quantities that can be extracted by the distribution are the median, equal to 4.67, and some 'quantiles' (values at which the cumulative distribution reaches a given percent of the maximum – the median being the 50% quantile). Interesting quantiles are the 15.85%, 25%, 75% and 84.15%, for which the Monte Carlo gives the following values of Y: 2.88, 3.49, 5.72 and 6.18. From these values we can calculate the *central* 50% and 68.3% intervals,[17] which are [3.49, 5.72] and [2.88, 6.18], respectively. Again, the information provided by Eq. (30) is far from any reasonable way to provide the uncertainty about Y, given the information on each component.

[16] Discussing this issues with several persons I have realized, with my great surprise, that this misconception is deeply rooted and strenuously defended by many colleagues, even by data analysis experts (they constantly reply "yes, but..."). This attitude is probably one of the consequences of being anchored to what I call un-needed principles (namely maximum likelihood, in this case), such that even the digits resulting from these principles are taken with a kind of religious respect and it seems blasphemous to touch them.

[17] I give the central 68.3% interval with some reluctance, because I know by experience that in many minds the short circuit

$$\text{“68\% probability interval”} \longleftrightarrow \text{“sigma”}$$

is almost unavoidable (I have known physicists convinced – and who even taught it! – that the standard deviation only 'makes sense for the Gaussian' and that it was defined via the '68% rule'). For this reason, recently I have started to appreciate thinking in terms of 50% probability intervals, also because they force people to reason in terms of better perceived fifty-to-fifty bets. I find these kind of bets very enlighting to show why practically all standard ways (including Bayesian ones!) fail to report upper/lower *confidence limits* in *frontier case situations* characterized by *open likelihoods* (see chapter 12 in Ref.[3]). I like to ask "please use your method and give me a 50% C.L. upper/lower limit", and then, when I have got it, "are you really 50% confident that the value is below that limit and 50% confident that it is above it? Would you equally bet on either side of that limit?". And the supporters of 'objective' methods are immediately at loss. (At least those who use Bayesian formulae realize that there must be some problem with the choice of priors.)

Besides the lucky case[18] of this numerical example (which was not constructed on purpose, but just recycling some material from Ref. [3]), it seems reasonable that even results roughly corrected by rule of thumb formulae are already better than those published directly with asymmetric result.[19] But the accurate analysis can only be done by the authors who know the details of the individual contribution to the uncertainty.

5 Conclusions

Asymmetric uncertainties do exist and there is *no way to remove them artificially*. If they are not properly treated, i.e. using prescriptions that do not have a theoretical ground but are more or less rooted in the physics community, the published result is biased. Instead, if they are properly treated using probability theory, in most cases of interest the final result is practically symmetric and approximately Gaussian, with expected value and standard deviations which take into account the several shifts due to individual asymmetric contributions. Note that some of the simplified methods to make statistical analyses had a *raison d'être* many years ago, when the computation was a serious limitation. Now it is not any longer a problem to evaluate, analytically or numerically, integrals of the kind of those appearing e.g. in Eqs.(1), (13) and (14).

In the case the final uncertainty remains asymmetric, the authors should provide detailed information about the 'shape of the uncertainty', giving also most probable value, probability intervals, and so on. But the best estimate of the *expected value and standard deviation should be always given* (see also the *ISO Guide* [15]).

To conclude, I would like to leave the final word to my preferred quotation with whom I like to end seminars and courses on probability theory applied to the evaluation and the expression of uncertainty in measurements:

[18] In the example here we have been lucky because an over-correction of the first contribution was compensated by an under-correction of the second contribution. Note also that the hypothesis about the nonlinear propagation was not correct, because we had, instead, a linear propagation of asymmetric p.d.f.'s. Anyhow the overall shift calculated by the guessed hypothesis is comparable to that calculable knowing the details of the analysis (and, in any case, using in subsequent analyses the roughly corrected result is definitely better than sticking to the published 'best value').

[19] Note that even if we were told that Y was $6.0^{+1.0}_{-2.2}$, without further information, we could still try to apply some shift to the result, obtaining 4.8 ± 1.6 or 5.4 ± 1.6 depending on some guesses about the source of the asymmetry. In any case, either results are better than $6.0^{+1.0}_{-2.2}$!

"Although this Guide *provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty, and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value."*[15]

It is a pleasure to thank *Superfaber* (Fabrizio Fabbri in *hepnames*) for helpful discussions on the subject and for his *super*vision of the manuscript.

References

[1] H1 Collaboration, www-h1.desy.de/psfiles/figures/d97-042.tab_correlations.txt;
ZEUS Collaboration, www-zeus.desy.de/physics/sfew/PUBLIC/sfew_results/published/nc-9900/table-2.dat.

[2] G. D'Agostini and M. Raso, CERN-EP/2000-026, February 2000 [hep-ex/0002056].

[3] G. D'Agostini, *Bayesian Reasoning in Data Analysis: A Critical Introduction*, World Scientific, 2003 (book info, including list of contents and hypertexted bibliography can be found at www.roma1.infn.it/~dagos/WSPC/).

[4] G. D'Agostini, *Rep. Prog. Phys.* **66** (2003) 1383.

[5] V. Dose, *Rep. Prog. Phys.* **66** (2003) 1421.

[6] R. Barlow, physics/0306138 , physics/0401042, physics/0403046.

[7] M. Schmelling, hep-ex/0006004.

[8] G. D'Agostini, Proc. XVIII International Workshop on Maximum Entropy and Bayesian Methods, Garching (Germany), July 1998, V. Dose et al. eds., Kluwer Academic Publishers, Dordrecht, 1999, 157–170 [physics/9811046].

[9] ZEUS Collaboration, J. Breitweg et al., *Eur. Phys. J.* **C14** (2000) 239.

[10] F. James and M. Roos, *Comp. Phys. Comm.* **10** (1975) 343.

[11] R.S. Thorne et al., hep-ph/9912439.

[12] See, e.g., A.M. Mood, D.C. Boes and F.A. Graybill, *Introduction to the theory of statistics*, McGraw-Hill, 1974.

[13] G. D'Agostini, DESY-95-242 and Roma1 N. 1070, hep-ph/9512295.

[14] P. Waltzlawick, J.H. Weakland and R. Fisch, *Change: principles of problem formation and problem resolution*, W.W. Norton, 1974.

[15] International Organization for Standardization (ISO), *Guide to the expression of uncertainty in measurement*, Geneva, Switzerland, 1993.