BLUE Anomalies

Michael Schmitt, Northwestern University

December 19, 2017

1 Overview

The BLUE method [1] is popular for combining measurements of a given quantity. Under certain conditions, however, the result can be anomalous (i.e., counter-intuitive): the combined value turn out to be lower than the input values or higher than the input values. We try to show how this comes about.

(Out of context.) Systematic uncertainties can often be classified as "multiplicative" or "additive." Consider the measurement of a cross section, for example:

$$\sigma = \frac{N - B}{\epsilon\, A\, L}$$

where N is the number of selected events, B is the estimated number of selected background events, ϵ is the efficiency of the selection, A is the geometric and kinematic acceptance, and L is the integrated luminosity. Systematic uncertainties for B are additive, while systematic uncertainties for ϵ, A, and L are multiplicative.

If a multiplicative uncertainty such as the luminosity uncertainty is common to two measurements, then it is common to the combined measurement. Therefore, it should not be included in the covariance matrix when making the combination. **Expand this.**

Make the distinction between uncertainties that can be reduced by repeating the measurement in exactly the same way, and those that cannot. The former are statistical, and the latter are systematic.

2 Derivation

We consider the simple example of combining two measurements of the *same* quantity:

$$x_1 \pm \sigma_{1,\mathrm{stat}} \pm \sigma_{1,\mathrm{sys}} \qquad \text{and} \qquad x_2 \pm \sigma_{2,\mathrm{stat}} \pm \sigma_{2,\mathrm{sys}}$$

where the measured values are x_1 and x_2, with statistical uncertainties $\sigma_{1,\mathrm{stat}}$ and $\sigma_{2,\mathrm{stat}}$ and systematic uncertainties $\sigma_{1,\mathrm{sys}}$ and $\sigma_{2,\mathrm{sys}}$. The BLUE method is equivalent to finding the combined value c using a least-squares method.

One must construct a covariance matrix by summing a matrix for the statistical uncertainties and another one for the systematic uncertainties. We take the measurements to be statistically independent, but with fully correlated systematic uncertainties. So

$$\begin{aligned} \mathbf{V} &= \mathbf{V}_{\mathrm{stat}} + \mathbf{V}_{\mathrm{stat}} \\ \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} &= \begin{pmatrix} \sigma_{1,\mathrm{stat}}^2 & 0 \\ 0 & \sigma_{2,\mathrm{stat}}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{1,\mathrm{sys}}^2 & \pm\sigma_{1,\mathrm{sys}}\sigma_{2,\mathrm{sys}} \\ \pm\sigma_{1,\mathrm{sys}}\sigma_{2,\mathrm{sys}} & \sigma_{2,\mathrm{sys}}^2 \end{pmatrix} \end{aligned} \tag{1}$$

where $\sigma_1^2 = \sigma_{1,\mathrm{stat}}^2 + \sigma_{1,\mathrm{sys}}^2$, $\sigma_2^2 = \sigma_{2,\mathrm{stat}}^2 + \sigma_{2,\mathrm{sys}}^2$, and

$$\rho = \frac{\pm\sigma_{1,\mathrm{sys}}\sigma_{2,\mathrm{sys}}}{\sigma_1\sigma_2} = \frac{\pm\sigma_{1,\mathrm{sys}}\sigma_{2,\mathrm{sys}}}{\sqrt{\sigma_{1,\mathrm{stat}}^2 + \sigma_{1,\mathrm{sys}}^2}\,\sqrt{\sigma_{2,\mathrm{stat}}^2 + \sigma_{2,\mathrm{sys}}^2}}.$$

The correlation coefficient ρ is close to zero if the statistical uncertainties dominate, and can be close to ± 1 if the correlated systematic uncertainties dominate.

This covariance matrix is the central element for the χ^2 calculation:

$$\chi^2 = \begin{pmatrix} c - x_1 & c - x_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} c - x_1 \\ c - x_2 \end{pmatrix}. \tag{2}$$

where c is the single free parameter and represents the combined value for the measured quantity. The inverse of the covariance matrix is

$$\mathbf{V}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

A little bit of algebra leads to

$$\chi^2 = \frac{1}{1-\rho^2} \left[\frac{(c-x_1)^2}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2}(c-x_1)(c-x_2) + \frac{(c-x_2)^2}{\sigma_2^2} \right].$$

If $\rho = 0$, meaning that the two measurements are completely independent, then the expression for χ^2 is transparent:

$$\chi^2 = \frac{(c-x_1)^2}{\sigma_1^2} + \frac{(c-x_2)^2}{\sigma_2^2}.$$

The minimization of χ^2 with respect to c leads to

$$c = \frac{\frac{x_1}{\sigma_1^2} - 2\rho\frac{(x_1+x_2)/2}{\sigma_1\sigma_2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}} = \frac{x_1(\sigma_2^2 - \rho\sigma_1\sigma_2) + x_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_2^2 - 2\rho\sigma_1\sigma_2 + \sigma_1^2}. \tag{3}$$

Again, with $\rho = 0$, the result is familiar:

$$c = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

which is just the weighted average of x_1 and x_2.

In the special case that $\sigma_1 = \sigma_2$, $c = (x_1 + x_2)/2$ independently of ρ.

The uncertainty of c is related to the curvature of χ^2 at its minimum: the greater the curvature, the tighter the constraint on c. In fact, if σ_c is the uncertainty for c, then

$$\frac{1}{\sigma_c^2} = \frac{1}{2} \left. \frac{\partial^2\chi^2}{\partial c^2} \right|_{\text{min}} = \frac{1}{1-\rho^2} \left(\frac{1}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2} \right)$$

or

$$\sigma_c = \frac{\sqrt{1-\rho^2}\,\sigma_1\sigma_2}{\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} = \left(\frac{\sigma_1\sigma_2}{\sigma_1 + \sigma_2} \right) \frac{\sqrt{1-\rho^2}}{\sqrt{1 - 2(1+\rho)\frac{\sigma_1\sigma_2}{(\sigma_1+\sigma_2)^2}}}$$

where it may help to recall that $0 \leq \sigma_1\sigma_2/(\sigma_1+\sigma_2)^2 \leq 1/4$. When $\sigma_1 = \sigma_2$ then

$$\sigma_c = \frac{\sigma_1}{\sqrt{2}} \sqrt{\frac{1+\rho}{1-\rho}}$$

which gives $\sigma_c \to 0$ when $\rho \to -1$, i.e., if two measurements are perfectly anti-correlated, then their combined value has no uncertainty – basically, the fluctuations cancel each other out. When $\rho \to 1$, however, $\sigma_c \to \infty$, showing that two measurements that are perfectly correlated must give the same value, which implies $\sigma_1 = \sigma_2 = 0$ (i.e., there are no fluctuations). These two extremes are obviously artificial and perhaps even paradoxical, indicating that values of ρ close to ± 1 have to be handled with extreme care.

2.1 Algebraic approach

The question is: what happens if ρ is definitely not zero?

We first simply carry out the algebraic solution. Without loss of generality, we take $x_2 = x_1 + \Delta$ with $\Delta > 0$. Furthermore, we can set $x_1 = 0$. Then,

$$c = \left(\frac{\frac{1}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2}}{\frac{1}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}} \right) \Delta.$$

The two anomalous cases would be $c < 0$ or $c > \Delta$. Since we have chosen $\Delta > 0$, we can have $c < 0$ only if

$$\frac{1}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2} < 0 \qquad \text{or} \qquad \rho > \frac{\sigma_1}{\sigma_2}$$

which is possible if $\sigma_1 < \sigma_2$. We can only have $c > \Delta$ if

$$\frac{1}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2} > \frac{1}{\sigma_2^2} - \frac{2}{\sigma_1\sigma_2} + \frac{1}{\sigma_1^2} \qquad \text{or} \qquad \rho > \frac{\sigma_2}{\sigma_1}$$

which is possible if $\sigma_2 < \sigma_1$.

What happens at the special values for ρ? For example, suppose $\sigma_1 < \sigma_2$ and $\rho = \sigma_1/\sigma_2$. The χ^2 function collapses into a transparent form,

$$\chi^2 = \frac{c^2}{\sigma_1^2} + \frac{\Delta^2}{\sigma_2^2 - \sigma_1^2} = \frac{(c - x_1)^2}{\sigma_1^2} + \frac{(x_2 - x_1)^2}{\sigma_2^2 - \sigma_1^2}$$

and by Eq (3), $c = x_1$, and of course $\sigma_c = \sigma_1$. Apparently the measurement x_2 is ignored. The covariance matrix has the special form

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_2^2 \end{pmatrix}$$

where the off-diagonal elements are the same as the smaller of the two variances. In the other case with $\rho = \sigma_2/\sigma_1 \leq 1$, $c = x_2$ and $\sigma_c = \sigma_2$. It is surprising that in these specific special cases, the best result is to take the measurement with the smaller uncertainty and to ignore the other.

After some algebraic manipulations, we can show that

$$\chi^2_{\min} = \frac{\Delta^2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \tag{4}$$

where $\Delta = x_2 - x_1 > 0$. There is only one degree of freedom, and although the probability distribution for χ^2 peaks at zero and is highly skewed, we will have $E[\chi^2] = 1$ for the expectation value. Taking $\chi^2 = 1$, $(x_2 - x_1)^2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2$. In the extreme case of completely anticorrelated measurements,

$$x_2 - x_1 = \sigma_1 + \sigma_2,$$

which suggests that there is relatively more room for differences $x_2 - x_1$. If, however, the two measurements are completely correlated, then

$$x_2 - x_1 = |\sigma_2 - \sigma_1|$$

which means that only rather close values of the measurements lead to a good value of χ^2, and in the case $\sigma_2 \to \sigma_1$, a finite difference $\Delta > 0$ always leads to a bad χ^2. Although these two extreme cases are artificial (note, for example, that the covariance matrix is singular when $\rho = \pm 1$), they are intuitive. For anticorrelated measurements, one value fluctuating upward is paired with another that fluctuates downward, hence their difference should be relatively large. For strongly correlated measurements, however, one value fluctuating upward should be accompanied by the other also fluctuating upward, leading to a small difference. In the special case where the size of the fluctuations is the same ($\sigma_2 = \sigma_1$) and the correlation is perfect ($\rho = 1$), they must have the same value, and if they do not, then something very improbable has occurred ($\chi^2 \to \infty$).

2.2 Geometric approach

We recall two very basic results from geometry / analysis.
1 The sum of two parabolas is another parabola.

$$\begin{aligned} y &= (x-\alpha)^2 + \lambda(x-\beta)^2 \\ &= (1+\lambda)x^2 - 2(\alpha+\lambda\beta)x + (\alpha^2 + \lambda\beta^2) \\ &= (1+\lambda)\left[x - \left(\frac{\alpha+\lambda\beta}{1+\lambda}\right)\right]^2 + \frac{\lambda}{1+\lambda}(\alpha-\beta)^2. \end{aligned}$$

Here, $\lambda > 0$ and plays the role of relative weight. If $\lambda = 1$ so that the two parabolas have the same curvature, then the new minimum is just the arithmetic average of the two initial minima:

$$y = 2\left[x - \frac{1}{2}(\alpha+\beta)\right]^2 + \frac{1}{2}(\alpha-\beta)^2.$$

It is interesting to note that the minimum value of y is greater than the minimum of both parabolas.
2 Adding a line to a parabola shifts the parabola.

$$\begin{aligned} y &= (x-\alpha)^2 - 2\gamma x \\ &= \big(x - (\alpha+\gamma)\big)^2 - \gamma(2\alpha+\gamma). \end{aligned}$$

If $\gamma > 0$ then the minimum of the parabola shifts to the right and to lower values.

The formula given in Eq. (2) for χ^2 as a function of c is the sum of two parabolas and a straight line:

$$\chi^2 = \left(1 - 2\rho\frac{\sigma_1}{\sigma_2}\right)\frac{(c-x_1)^2}{\sigma_1^2} + \left(1 - 2\rho\frac{\sigma_2}{\sigma_1}\right)\frac{(c-x_2)^2}{\sigma_2^2} - \frac{2\rho}{\sigma_1\sigma_2}(x_1+x_2)c + \frac{2\rho}{\sigma_1\sigma_2}\Big((x_1-x_2)^2 + x_1x_2\Big).$$

The first two terms are parabolas, the third term is a straight line, and the fourth term is a constant independent of c. So, based on point **1**, we have a parabola, and based on point **2**, the minimum of that parabola will be shifted to the right for $\rho > 0$ and to the left for $\rho < 0$. Can that shift place the minimum of χ^2 outside the range $[x_1, x_2]$? The answer depends on the size of ρ.

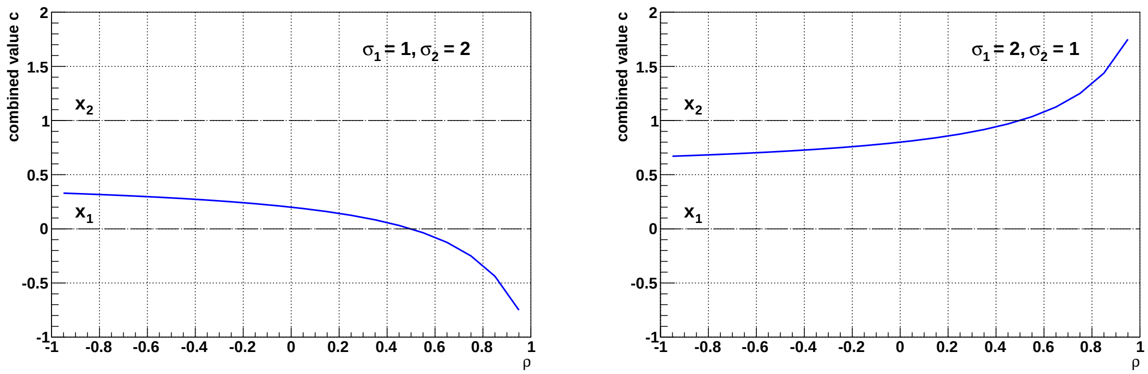


Figure 1: Least-squares solution c as a function of the correlation coefficient ρ. Naively one expects the solution to fall between the two input values (indicated by the long-dash lines), but when ρ is large enough the solution lies outside that range.

2.3 Graphical approach

The two previous sections indicate that ρ is the crucial quantity. If $\rho = 0$ then there can be no anomaly. Anomalies arise only if ρ is large enough and positive. Also, no anomalies arise if $\sigma_1 = \sigma_2$ or if $x_2 = x_1$. For illustration we choose $\sigma_2 = 2\sigma_1$ and plot c as a function of ρ. For concreteness, we set $x_1 = 0$, $x_2 = 1$, and $\sigma_1 = 1$. Notice that if the two measurements were uncorrelated then in this scenario they would be statistically compatible. The solution given in Eq. (3) is plotted in Fig. 1 (left) which shows that c diverges outside the range $[x_1, x_2]$ for $\rho > 0.5$. The other case, with $\sigma_1 = 2\sigma_2$ and $\sigma_2 = 1$ is shown in Fig. 1 (right).

An important feature of Fig. 1 is that the "deviation" of c outside the interval $[x_1, x_2]$ is determined by which of the two measurements is more precise. When $\sigma_1 < \sigma_2$, then c deviates toward and beyond x_1 as ρ increases, and when $\sigma_2 < \sigma_1$, then it deviates toward and beyond x_2.

Figure 2 shows the variation of the uncertainty σ_c and χ^2_{min} as a function of ρ, for the same input values as in Fig. 1 (left) ($x_1 = 0$ and $x_2 = 1$). The uncertainty is largest when $\rho = \sigma_1/\sigma_2$. The value of χ^2_{min} is lowest for $\rho = -1$ and rises monotonically with ρ, as one expects from Eq. (4). The fact that $\chi^2_{\text{min}} < 1$ for all ρ indicates that the input values are not incompatible even if highly correlated.

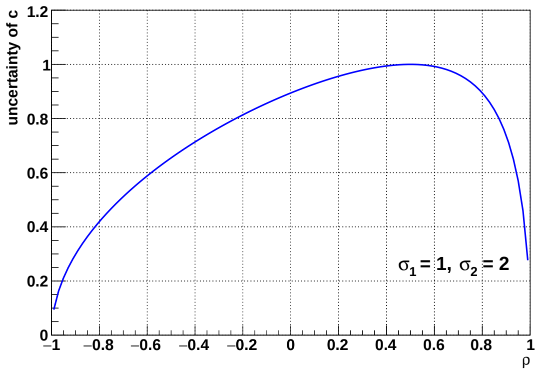
3 Analysis

The anomalies occur when ρ is large and positive. In the limit $\rho \to 1$, one expects $x_2 = x_1$, so any difference $\Delta = x_2 - x_1 \neq 0$ violates the expectations about x_2 and x_1.

In order to gain some insight into this problem, we generalize the formulation of χ^2 to two independent parameters, c_1 and c_2, in place of the single parameter c:

$$\chi^2(c_1, c_2) = \begin{pmatrix} c_1 - x_1 & c_2 - x_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} c_1 - x_1 \\ c_2 - x_2 \end{pmatrix}. \tag{5}$$

If $\rho = 0$ then the solution is obvious: $c_1 = x_1$ and $c_2 = x_2$. When $\rho \neq 0$, the solution is the same.



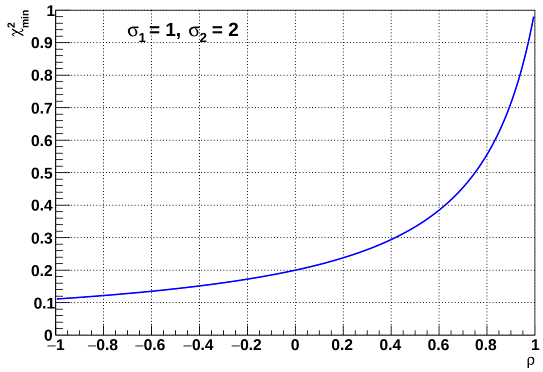


Figure 2: Variation of the uncertainty σ_c (left) and $\chi^2_{\rm min}$ (right) as a function of ρ. The numerical values are the same as in Fig. 1.

Figure 3 shows contours of χ^2 in the (c_1, c_2) plane, for $\rho = 0.1$, which results in a "normal" solution, and $\rho = 0.9$, which results in an "anomalous" solution. (We still have $x_1 = 0$, $x_2 = 1$, $\sigma_1 = 1$, and $\sigma_2 = 2\sigma_1$.) The thin diagonal line shows the constraint $c_1 = c_2$ and the thickened part of the diagonal line shows the intuitive expectation $x_1 < c < x_2$. The dot shows the position of the minimum of χ^2 and the star shows the solution Eq. (3) obtained with the constraint $c_1 = c_2$. For the case $\rho = 0.1$ the star falls in the thick part of the diagonal line, while for $\rho = 0.9$ it lies outside.

4 Remedy

It is not clear that a remedy is needed. As Lyons et al. point out, if x_2 is highly correlated with x_1, then if x_1 fluctuates above the true value $x_{\rm true}$, then so will x_2. Consequently, one should take x_1 and x_2 to perform an *extrapolation* rather than an *interpolation*, which is precisely what this method provides (see alo Fig. 3).

Reference [1] was concerned primarily with statistical correlations, however. Correlations caused by common systematic uncertainties are fundamentally different – see the earlier discussion on additive versus multiplicative systematic uncertainties. **MORE**

5 MORE WORK TO DO

Things I should still investigate:

1. Point out that correlations between measurements of the same quantity are what we examine here. Correlations between measurements of different quantities are also important but do not tend to display the anomalous behavior described here. (See Valassi.)

2. Plot the distribution of χ^2.

3. Show that the variance of c is minimized and is smaller than taking the straight average.

4. Point out that the exact values of the correlations are important and that one should check the sensitivity of the central value and its uncertainty under mild variations of the correlation coefficients.

5. Despite the fact that c is unbiased, it is still unreasonable to have c outside the range $[x_1, x_2]$ for the sake of a luminosity uncertainty. Can we enforce $c = (x_1 + x_2)/2$ as far as luminosity and similar uncertainties are concerned? What is the impact on the *bias* and on the *variance*?

   This is obvious. The results above show that c is pushed outside the interval $[x_1, x_2]$ in the direction of whichever measurements has the smaller uncertainty. As far as luminosity is concerned, however, the uncertainty on either of the measurements is irrelevant and there is no reason to suppose that $c < x_1$ is better than $c > x_2$, or vice-versa.

   So... it is important to check whether one of the experiments gives better information about a nuisance parameter such as luminosity than the other experiment.

   This suggests that uncertainties on measured quantities and on nuisance parameters *somehow* need to be separated.

6. Bring out the connection between the compatibility of measurement values x_1 and x_2 as reflected in their difference $x_2 - x_1$.

7. What happens if a systematic uncertainty is **not** Gaussian distributed? Do a simulation with the luminosity distributed Gaussian, triangle, and box. Perhaps consider other models for the pdf (thinking of theoretical uncertainties for example).

8. Investigate Jens Erler's method.

9. Improve the discussion of additive versus multiplicative uncertainties.

10. What would be a Bayesian approach to this problem? What is the posterior for c? What happens if we enforce $x_1 < c < x_2$ using the prior?

11. Is there a way to couch this problem in the language of risk or cost analysis?

12. Find a way to connect the bias (e.g. from taking the arithematic mean) and the variance – discuss the MSE in this problem.

13. How would (Shannon) entropy enter this problem?

14. Is there a way to apply machine learning to this problem? One could train the algorithm (e.g. ANN) easily and then apply it to the data. Pay attention to bias. The training should differentiate between random measurement errors and correlated systematic errors/uncertainties.

References

[1] L. Lyons, D. Gibaut and P. Clifford, "How to Combine Correlated Estimates of a Single Physical Quantity," Nucl. Instrum. Meth. A **270**, 110 (1988). doi:10.1016/0168-9002(88)90018-6

[2] A. Valassi, "Combining correlated measurements of several different physical quantities," Nucl. Instrum. Meth. A **500**, 391 (2003). doi:10.1016/S0168-9002(03)00329-2

[3] R. Nisius, "On the combination of correlated estimates of a physics observable," Eur. Phys. J. C **74**, no. 8, 3004 (2014) doi:10.1140/epjc/s10052-014-3004-2

[4] J. Erler, "On the Combination Procedure of Correlated Errors," Eur. Phys. J. C **75**, no. 9, 453 (2015) doi:10.1140/epjc/s10052-015-3688-y

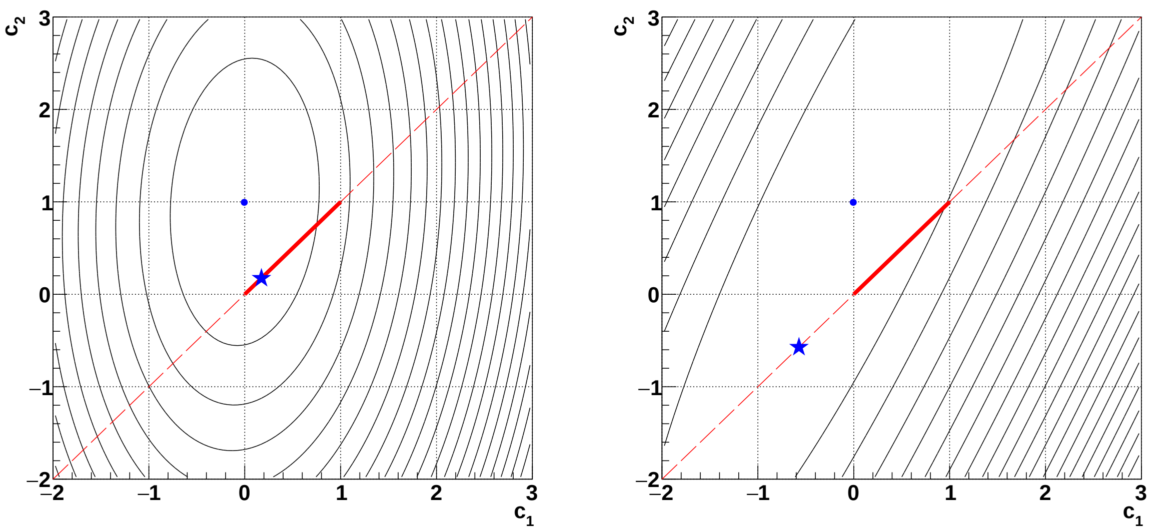


Figure 3: Contours of χ^2 in the (c_1, c_2) plane. The thin dashed diagonal line represents the case $c_1 = c_2$, with the solid thick part demarcating the expected range $x_1 < c < x_2$. The blue dot shows the minimum of χ^2 and the star shows the solution Eq. (3) subject to the constraint $c_1 = c_2$.