Theoretical errors

Francesco Giacosa
ITP, J. W. Goethe University

February 1, 2013

Abstract

How to evaluate theoretical errors out of a χ^2 analysis

1 Hesse matrix

Consider a theoretical approach which dependens on N unknown parameters x_1, x_2, ..., x_N. We can determine them by searching the minimum of

$$\chi^2[x_1, ...x_N] = \left(\frac{A_1[x_1,...,x_N] - A_1^{\exp}}{\delta A_1}\right)^2 + \tag{1}$$

$$\left(\frac{A_2[x_1,...,x_N] - A_2^{\exp}}{\delta A_2}\right)^2 + ... + \left(\frac{A_q[x_1,...,x_N] - A_q^{\exp}}{\delta A_q}\right)^2 \tag{2}$$

$$= \sum_{i=1}^{Q} \left(\frac{A_i[x_1,...,x_N] - A_i^{\exp}}{\delta A_i}\right)^2 \tag{3}$$

where $A_i[x_1, ..., x_N]$ are the theoretical expressions ($Q \geqslant N$) for which there are experimental measurements δA_i.

Let the point $\{x_1^{\min}, ..., x_N^{\min}\}$ be the minimum of χ^2 realized at the point $\{x_1^{\min}, ..., x_N^{\min}\}$. We introduce the new variables y_i :

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_1 - x_1^{\min} \\ \vdots \\ x_N - x_N^{\min} \end{pmatrix} = \vec{x} - \vec{x}^{\min}. \tag{4}$$

In therms of the variables y_i the minimum is obviously realized at $P = \{0, ...0\}$.

Let us now write the Taylor expansion of χ^2 around $y_i = 0$:

$$\chi^2[y] = \chi_0^2 + y^t H y \tag{5}$$

where the Hesse matrix H is given by:

$$H_{ij} = \frac{1}{2}\left(\frac{\partial^2 \chi^2[y_1, ...y_N]}{\partial y_i \partial y_j}\right)_P = .\frac{1}{2}\left(\frac{\partial^2 \chi^2[x_1, ...y_N]}{\partial x_i \partial x_j}\right)_P \tag{6}$$

Note, we have used the fact the the first derivative vanishes: $\left(\frac{\partial \chi^2[y_1,\ldots y_n]}{\partial y_i}\right)_P =$ 0.

2 Diagonalization

We now introduce new variables z_i in the following way:

$$\vec{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_N \end{pmatrix} = B \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = B\vec{y} = B(\vec{x} - \vec{x}^{\min}) \tag{7}$$

with $B \subset SO(N)$ in such a way that

$$BHB^t = \lambda = diag\{\lambda_1, ..., \lambda_N\} \tag{8}$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of the matrix H. Note, the matrix B can be written as

$$B = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}_P \tag{9}$$

and

$$B^t = \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} & \cdots \\ \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}_{z_i=0} \tag{10}$$

The quantity χ^2 expressed in terms of z_i takes the simple form:

$$\begin{aligned} \chi^2[z] &= \chi_0^2 + z^t BHB^t z = \chi_0^2 + z^t \lambda z = \\ &= \chi_0^2 + z_1^2\lambda_1 + ... + z_N^2\lambda_N. \end{aligned} \tag{11}$$

The errors on the quantities z_i are now easily given by

$$\delta z_i = \frac{1}{\sqrt{\lambda_i}} \tag{12}$$

which correspond to a variation of ± 1 for the χ^2 :

$$\chi^2[0, 0, ..., z_i = \pm\frac{1}{\sqrt{\lambda_i}}, 0, ...0] = \chi_0^2 + 1. \tag{13}$$

If we shift at the same time all the quantities z_i we obtain

$$\chi^2[\pm\frac{1}{\sqrt{\lambda_1}}..., z_i = \pm\frac{1}{\sqrt{\lambda_i}}, ..., \pm\frac{1}{\sqrt{\lambda_N}}] = \chi_0^2 + N. \tag{14}$$

This is the maximum variation allowed.

3 Error calculation for a generic function F

Now, all the theoretical quantities we would like to determine are -most probably-function of the original parameters x_i. Let us consider, for instance, some function

$$F = F[x_1, ..., x_N] \tag{15}$$

which we want to determine. The central value is easy:

$$F_0 = F[x_1^{\min}, ..., x_N^{\min}]. \tag{16}$$

The problem is how to calculate the error δF. In fact, we know the errors only in terms of the variables z_i, thus we should re-express F as function $F = F[z]$ and evaluate

$$\delta F = \sqrt{\left(\frac{\partial F}{\partial z_1}\delta z_1\right)^2 + ... + \left(\frac{\partial F}{\partial z_n}\delta z_n\right)^2} \tag{17}$$

Instead of doing it explicitly we notice that

$$\frac{\partial F}{\partial z_1} = \frac{\partial F}{\partial x_1}\frac{\partial x_1}{\partial z_1} + ... + \frac{\partial F}{\partial x_N}\frac{\partial x_N}{\partial z_1} \tag{18}$$

where the derivatives $\frac{\partial x_1}{\partial z_1}$, ..., $\frac{\partial x_n}{\partial z_1}$ are the first column of the matrix B^t.

We now present two ways to write δF.

Way 1 to write δF in a compact form:

We introduce the vector

$$\vec{f} = \left(\vec{\nabla} F\right)_P = \begin{pmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_N} \end{pmatrix}_P . \tag{19}$$

Then:

$$\begin{aligned} \frac{\partial F}{\partial z_1} &= \frac{\partial F}{\partial x_1}\frac{\partial x_1}{\partial z_1} + ... + \frac{\partial F}{\partial x_N}\frac{\partial x_N}{\partial z_1} & (20) \\ &= \left(\vec{f}^t B^t\right)_1 = \left(B\vec{f}\right)_1^t = \left(B\vec{f}\right)_1 . & (21) \end{aligned}$$

Ergo:

$$\delta F = \sqrt{\frac{1}{\lambda_1}\left(B\vec{f}\right)_1^2 + ... + \frac{1}{\lambda_N}\left(B\vec{f}\right)_N^2} . \tag{22}$$

Way 2 to write δF in a compact form:

It is useful to introduce the matrix D

$$D = \begin{pmatrix} \frac{\partial F}{\partial x_1}\frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_1}\frac{\partial F}{\partial x_2} & ... \\ \frac{\partial F}{\partial x_2}\frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2}\frac{\partial F}{\partial x_2} & ... \\ ... & ... & ... \end{pmatrix}_{x=x^{\min}} \tag{23}$$

and then the matrix product

$$D \cdot B^t \cdot \lambda^{-1/2}. \tag{24}$$

We then pick up only the diagonal elements of the latter matrix and form the vector

$$\alpha = \begin{pmatrix} \left(D \cdot B^t \cdot \lambda^{-1/2}\right)_{11} \\ \vdots \\ \left(D \cdot B^t \cdot \lambda^{-1/2}\right)_{NN} \end{pmatrix}. \tag{25}$$

Now, the error of the function F is given by

$$\delta F = \sqrt{\alpha^t \alpha}. \tag{26}$$

When changing the function F with another function G all goes as before, provided that one recalculates the matrix D as

$$D = \begin{pmatrix} \frac{\partial G}{\partial x_1} & \frac{\partial G}{\partial x_2} & ... \\ \frac{\partial G}{\partial x_1} & \frac{\partial G}{\partial x_2} & ... \\ ... & ... & ... \end{pmatrix}_{x=x^{\min}} \tag{27}$$

4 Error on the original parameters x_i

As a last step we consider a special case. We suppose that the function $G = x_1$. Using the way 2 described above, the matrix D takes the form

$$D = \begin{pmatrix} 1 & 0 & ... \\ 1 & 0 & ... \\ ... & ... & ... \end{pmatrix} \tag{28}$$

One obtains the error on the original parameter x_1:

$$\delta x_1 = \sqrt{\left(\frac{\partial x_1}{\partial z_1}\frac{1}{\sqrt{\lambda_1}}\right)^2 + \left(\frac{\partial x_1}{\partial z_2}\frac{1}{\sqrt{\lambda_2}}\right)^2 + ...} \tag{29}$$

Note that the latter can be obtained also in another way. To this end we recall that

$$BHB^t = \lambda. \tag{30}$$

Ergo:

$$B(B^t \lambda^{-1} B) = 1 \tag{31}$$

out of which we identify the inverse of the Hesse matrix:

$$H^{-1} = B^t \lambda^{-1} B. \tag{32}$$

Now, let us consider the element H^{-1}_{11}. With simple passages we obtain that

$$H^{-1}_{11} = \left(\frac{\partial x_1}{\partial z_1}\right)^2 \frac{1}{\lambda_1} + \left(\frac{\partial x_1}{\partial z_2}\right)^2 \frac{1}{\lambda_2} + ... \tag{33}$$

and therefore

$$\delta x_1 = \sqrt{H_{11}^{-1}}. \tag{34}$$

The error on the original parameter x_1 is the first element of the inverse Hesse matrix. More in general

$$\delta x_i = \sqrt{H_{ii}^{-1}} \tag{35}$$

i.e., the diagonal of the inverse Hesse matrix gives us the squared errors of the original parameters.

Note that, in virtue of the obtained δx_i, one would be tempted to calculate the error of a function F as

$$\delta F_{naive} = \sqrt{\left(\frac{\partial F}{\partial x_1}\delta x_1\right)^2 + ... + \left(\frac{\partial F}{\partial x_N}\delta x_N\right)^2}. \tag{36}$$

However, the variables x_i do not diagonalize the Hesse matrix and correlations are not eliminated. One has therefore

$$\delta F_{naive} \geq \delta F. \tag{37}$$

The evaluation of δF_{naive} is therefore an overestimate of the real error. In fact, we cannot vary the parameters x_i as we want in the hypercube defined by $x_i \pm \delta x_i$ because in this way we neglect the correlations and we may reach value of χ^2 which are much larger than the 'limit' $\chi_0^2 + N$ evaluated above.