Testing the statistical compatibility of independent data sets

M. Maltoni*
Instituto de Física Corpuscular–C.S.I.C./Universitat de València Edificio Institutos de Paterna, Apt 22085, E–46071 Valencia, Spain

T. Schwetz†
Institut für Theoretische Physik, Physik Department Technische Universität München, James–Franck–Strasse, D–85748 Garching, Germany

(Received 5 May 2003; published 29 August 2003)

We discuss a goodness-of-fit method which tests the compatibility between statistically independent data sets. The method gives sensible results even in cases where the χ^2 minima of the individual data sets are very low or when several parameters are fitted to a large number of data points. In particular, it avoids the problem that a possible disagreement between data sets becomes diluted by data points which are insensitive to the crucial parameters. A formal derivation of the probability distribution function for the proposed test statistics is given, based on standard theorems of statistics. The application of the method is illustrated on data from neutrino oscillation experiments, and its complementarity to the standard goodness-of-fit is discussed.

DOI: 10.1103/PhysRevD.68.033020 PACS number(s): 14.60.Pq, 02.70.Rr

I. INTRODUCTION

The essence of any scientific progress is the comparison of theoretical predictions to experimental data. Statistics provides the scientist with so-called *goodness-of-fit* tests, which allow us to obtain well-defined probability statements about the agreement of a theory with data. By far the most popular goodness-of-fit test dates back to 1900, when K. Pearson identified the minimum of a χ^2 function as a powerful tool to evaluate the quality of the fit [1]. However, it is known that the Pearson $\chi^2_{\rm min}$ test is not very restrictive in global analyses, where data from different experiments with a large number of data points are compared to a theory depending on many parameters. The reason for this is that in such a case a given parameter is often constrained only by a small subset of the data. If the rest of the data (which can contain many data points) are reasonably fitted, a possible problem in the fit of the given parameter is completely washed out by the large amount of data points. A discussion of this problem in various contexts can be found e.g. in Refs. [2–4].

To evade this problem a modification of the original $\chi^2_{\rm min}$ test was proposed in Ref. [5] to evaluate the goodness-of-fit of neutrino oscillation data in the framework of four-neutrino models. There this method was called *parameter goodness-of-fit* (PG), and it can be applied when the global data consists of statistically independent subsets. The PG is based on parameter estimation and hence it avoids the problem of being diluted by many data points. It tests the *compatibility* of the different data sets in the framework of the given theoretical model. In this note we give a formal derivation of the probability distribution function (p.d.f.) for the test statistics of the PG, and discuss the application and interpretation of the PG on some examples. The original motivation for the PG was the analysis of neutrino oscillation data. However, the method may be very useful also in other fields of physics, especially where global fits of many parameters to data from several experiments are performed.

The outline of the paper is as follows. In Sec. II we define the PG and show that its construction is very similar to that of the standard goodness-of-fit. The formal derivation of the p.d.f. for the PG test statistics is given in Sec. III, whereas in Sec. IV a discussion of the application and interpretation of the PG is presented. In Sec. V we consider the PG in the case of correlations due to theoretical errors, and we conclude in Sec. VI.

II. GOODNESS-OF-FIT TESTS

We would like to start the discussion by citing the goodness-of-fit definition given by the Particle Data Group (see Sec. 31.3.2. of Ref. [6]): ''Often one wants to quantify the level of agreement between the data and a hypothesis without explicit reference to alternative hypotheses. This can be done by defining a *goodness-of-fit statistics*, t which is a function of the data whose value reflects in some way the level of agreement between the data and the hypothesis. [. . .] The hypothesis in question, say, H_0 will determine the p.d.f. $g(t|H_0)$ for the statistics. The goodness-of-fit is quantified by giving the p value, defined as the probability to find t in the region of equal or lesser compatibility with H_0 than the level of compatibility observed with the actual data. For example, if t is defined such that large values correspond to poor agreement with the hypothesis, then the p value would be

$$p = \int_{t_{\rm obs}}^{\infty} g(t|H_0)\,dt, \tag{1}$$

where $t_{\rm obs}$ is the value of the statistic obtained in the actual experiment.''

Let us stress that from this definition of goodness-of-fit one has complete freedom in choosing a test statistic t, as long as the correct p.d.f. for it is used.

*Electronic address: maltoni@ific.uv.es
†Electronic address: schwetz@ph.tum.de

A. The standard goodness-of-fit

Consider N random observables $\boldsymbol{\nu}=(\nu_i)$ and let $\mu_i(\boldsymbol{\theta})$ denote the expectation value for the observable ν_i, where $\boldsymbol{\theta}=(\theta_\alpha)$ are P independent parameters which we wish to estimate from the data. Assuming that the covariance matrix S is known one can construct the following χ^2 function:

$$\chi^2(\boldsymbol{\theta})=[\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta})]^T S^{-1}[\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta})] \tag{2}$$

and use its minimum $\chi^2_{\min}$ as test statistics for goodness-of-fit evaluation:

$$t(\boldsymbol{\nu})=\chi^2_{\min}. \tag{3}$$

The hypothesis we want to test determines the p.d.f. $g(t)$ for this statistics. Once the real experiments have been performed, giving the results $\boldsymbol{\nu}_{\rm obs}$, the goodness-of-fit is given by the probability of obtaining a t larger than $t_{\rm obs}$, as expressed by Eq. (1). We will refer to this procedure as *standard goodness-of-fit* (SG):

$$p_{\rm SG}=\int_{\chi^2_{\min}(\boldsymbol{\nu}_{\rm obs})}^{\infty} g(t)dt. \tag{4}$$

The great success of this method is mostly due to a very powerful theorem, which was proven over 100 years ago by K. Pearson[1] [1] and which greatly simplifies the task of calculating the integral in Eq. (4). It can be shown under quite general conditions (see e.g. Ref. [7]) that $\chi^2_{\min}$ follows a χ^2 distribution with $N-P$ degrees of freedom (d.o.f.), so that $g(t)=f_{\chi^2}(t,N-P)$. Therefore, the integral in Eq. (4) becomes

$$p_{\rm SG}={\rm CL}(\chi^2_{\min}(\boldsymbol{\nu}_{\rm obs}),N-P)\equiv\int_{\chi^2_{\min}(\boldsymbol{\nu}_{\rm obs})}^{\infty} f_{\chi^2}(t,N-P)dt, \tag{5}$$

where ${\rm CL}(\chi^2,n)$ is the *confidence level* function (see e.g., Fig. 31.1 of Ref. [6]).

In the following we propose a modification of the SG, for the case when the data can be divided into several statistically independent subsets.

B. The parameter goodness-of-fit

Consider D statistically independent sets of random observables $\boldsymbol{\nu}^r=(\nu_i^r)$ $(r=1,\ldots,D)$, each consisting of N_r observables $(i=1,\ldots,N_r)$, with $N_{\rm tot}=\Sigma_r N_r$. Now a theory depending on P parameters $\boldsymbol{\theta}=(\theta_\alpha)$ is confronted with the data. The total χ^2 is given by

$$\chi^2_{\rm tot}(\boldsymbol{\theta})=\sum_{r=1}^{D}\chi^2_r(\boldsymbol{\theta}), \tag{6}$$

where

$$\chi^2_r(\boldsymbol{\theta})=[\boldsymbol{\nu}^r-\boldsymbol{\mu}^r(\boldsymbol{\theta})]^T S_r^{-1}[\boldsymbol{\nu}^r-\boldsymbol{\mu}^r(\boldsymbol{\theta})] \tag{7}$$

is the χ^2 of the dataset r. Now we define

$$\overline{\chi^2}(\boldsymbol{\theta})=\chi^2_{\rm tot}(\boldsymbol{\theta})-\sum_{r=1}^{D}\chi^2_{r,\min}, \tag{8}$$

where $\chi^2_{r,\min}=\chi^2_r(\hat{\boldsymbol{\theta}}_r)$, and $\hat{\boldsymbol{\theta}}_r(\boldsymbol{\nu}^r)$ are the values of the parameters which minimize χ^2_r. Instead of the total χ^2 minimum we propose now to use

$$t(\boldsymbol{\nu})=\overline{\chi^2}_{\min}=\overline{\chi^2}(\tilde{\boldsymbol{\theta}}) \tag{9}$$

as test statistics for goodness-of-fit evaluation. In Eq. (9) $\overline{\chi^2}_{\min}$ is the minimum of $\overline{\chi^2}$ defined in Eq. (8), and $\tilde{\boldsymbol{\theta}}$ are the parameter values at the minimum of $\overline{\chi^2}$, or equivalently of $\chi^2_{\rm tot}$. If we now denote by $\overline{g}(t)$ the p.d.f. for this statistics, we can define the corresponding goodness-of-fit by means of Eq. (1), in complete analogy to the SG case:

$$p_{\rm PG}=\int_{\overline{\chi^2}_{\min}(\boldsymbol{\nu}_{\rm obs})}^{\infty}\overline{g}(t)dt. \tag{10}$$

This procedure was proposed in Ref. [5] with the name *parameter goodness-of-fit* (PG). Its construction is very similar to the SG, except that now $\overline{\chi^2}$ rather than χ^2 is used to define the test statistics.

In the next section we will show that also in the case of the PG the calculation of the integral appearing in Eq. (10) can be greatly simplified. Let us define

$$P_r\equiv{\rm rank}\left[\frac{\partial\boldsymbol{\mu}^r}{\partial\boldsymbol{\theta}}\right]. \tag{11}$$

This corresponds to the number of *independent* parameters (or parameter combinations), constrained by a measurement of $\boldsymbol{\mu}^r$.[2] Then under general condition $\overline{\chi^2}_{\min}$ is distributed as a χ^2 with $P_c=\Sigma_r P_r-P$ d.o.f., so that Eq. (10) reduces to

$$p_{\rm PG}={\rm CL}(\overline{\chi^2}_{\min}(\boldsymbol{\nu}_{\rm obs}),P_c). \tag{12}$$

III. THE PROBABILITY DISTRIBUTION FUNCTION OF $\overline{\chi^2}_{\min}$

In this section we derive the distribution of the test statistics for the PG. This can be done in complete analogy to the SG. Therefore, we start by reviewing the corresponding proof for the SG, see e.g. Ref. [7].

[1]Pearson uses the slightly different test statistic

$$\chi^2_{Pearson}=\sum_i\frac{[\nu_i-\mu_i(\boldsymbol{\theta})]^2}{\mu_i(\boldsymbol{\theta})}$$

and assumes that ν_i are independent. We prefer to use instead χ^2 of Eq. (2), because in this way also correlated data can be considered.

[2]If in some pathological cases P_r depends on the point in the parameter space, Eq. (11) should be evaluated at the true values of the parameters, see Sec. III B.

A. The standard goodness-of-fit

Let us start with the χ^2 defined in Eq. (2). Since the covariance matrix S is a real, positive and symmetric matrix one can always find an orthogonal matrix O and a diagonal matrix s such that $S^{-1}=O^T s^2 O$. Hence, we can write the χ^2 in the following way:

$$\chi^2(\boldsymbol{\theta})=[\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta})]^T S^{-1}[\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta})]=\mathbf{y}(\boldsymbol{\theta})^T\mathbf{y}(\boldsymbol{\theta}), \quad (13)$$

where we have defined the new variables $\mathbf{y}(\boldsymbol{\theta})=sO[\boldsymbol{\nu}-\boldsymbol{\mu}(\theta)]$. Let us denote the (unknown) true values of the parameters by $\boldsymbol{\theta}^0$ and we define

$$\mathbf{x}\equiv\mathbf{y}(\boldsymbol{\theta}^0)=sO[\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta}^0)]. \quad (14)$$

Now we assume that the x_i are normal distributed with mean 0 and the covariance matrix $\mathbf{1}_N$, which in particular implies that they are statistically independent. This assumption is obviously correct if the data ν_i are normal distributed with mean $\mu_i(\boldsymbol{\theta}^0)$ and covariance matrix S. However, it can be shown (see e.g. Refs. [7–9]) that this assumption holds for a large class of arbitrary p.d.f. for the data under quite general conditions, especially in the large sample limit, i.e. large ν_i. Under this assumption it is evident that $\chi^2(\boldsymbol{\theta}^0)=\mathbf{x}^T\mathbf{x}$ follows a χ^2 distribution with N d.o.f. According to Eq. (3) the test statistics t for the SG is given by the minimum of Eq. (13). To derive the p.d.f. for t we state the following proposition

Proposition 1. Let $\hat{\boldsymbol{\theta}}$ *be the values of the parameters which minimize Eq. (13). Then*

$$\chi^2_{\min}=\chi^2(\theta^0)-\Delta\chi^2, \quad (15)$$

with $\chi^2_{\min}=\hat{\mathbf{y}}^T\hat{\mathbf{y}}$ *and* $\hat{\mathbf{y}}\equiv\mathbf{y}(\hat{\boldsymbol{\theta}})$*, has a* χ^2 *distribution with* $N-P$ *d.o.f. and* $\Delta\chi^2$ *has a* χ^2 *distribution with* P *d.o.f. and is statistically independent of* $\chi^2_{\min}$.

A rigorous proof of this proposition is somewhat intricate and can be found e.g. in Ref. [7]. In the following we give an outline of the proof dispensing with mathematical details for the sake of clarity.

The $\hat{\boldsymbol{\theta}}$ are obtained by solving the equations

$$\frac{\partial\chi^2}{\partial\theta_\alpha}=2\mathbf{y}^T\frac{\partial\mathbf{y}}{\partial\theta_\alpha}=0. \quad (16)$$

It can be proved (see e.g. Ref. [7]) under very general conditions that Eq. (16) has a unique solution $\hat{\boldsymbol{\theta}}$ which converges to the true values $\boldsymbol{\theta}^0$ in the large sample limit. In this sense it is a good approximation[3] to write

$$\hat{\mathbf{y}}\approx\mathbf{x}+B(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0), \quad (17)$$

where we have defined the rectangular $N\times P$ matrix B by

$$B\equiv\left.\frac{\partial\mathbf{y}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^0}. \quad (18)$$

[3]Note that Eq. (17) is exact if $\mathbf{y}$ depends linearly on the parameters $\boldsymbol{\theta}$.

Without loss of generality we assume that[4] $\mathrm{rank}[B]=P$. From Eq. (17) we obtain

$$\left.\frac{\partial\mathbf{y}}{\partial\boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}}\approx\left.\frac{\partial\mathbf{y}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^0}=B. \quad (19)$$

Using this last relation in Eq. (16) we find that $\hat{\mathbf{y}}$ fulfils $\hat{\mathbf{y}}^T B=0$. Multiplying Eq. (17) from the left side by B^T this leads to

$$B^T\mathbf{x}=-B^TB(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0). \quad (20)$$

Using Eqs. (17) and (20) we obtain

$$\hat{\mathbf{y}}^T\hat{\mathbf{y}}=\mathbf{x}^T\mathbf{x}-(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)^TB^TB(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0). \quad (21)$$

The symmetric $P\times P$ matrix B^TB can be written as $B^TB=Rb^2R^T$ with the orthogonal matrix R and the diagonal matrix b, and Eq. (20) implies $b^{-1}R^TB^T\mathbf{x}=-bR^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)$. Defining the $N\times P$ matrix

$$H\equiv BRb^{-1} \quad (22)$$

we find $(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)^TB^TB(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)=\mathbf{x}^THH^T\mathbf{x}$, and Eq. (21) becomes

$$\hat{\mathbf{y}}^T\hat{\mathbf{y}}=\mathbf{x}^T(\mathbf{1}_N-HH^T)\mathbf{x}. \quad (23)$$

Note that the matrix H obeys the orthogonality relation $H^TH=\mathbf{1}_P$, showing that the P column vectors of length N in H are orthogonal. We can add $N-P$ columns to the matrix H completing it to an orthogonal $N\times N$ matrix: $V=(H,K)$. Here K is an $N\times(N-P)$ matrix with $K^TK=\mathbf{1}_{(N-P)}$, $H^TK=0$ and the completeness relation

$$VV^T=HH^T+KK^T=\mathbf{1}_N. \quad (24)$$

Now we transform to the new variables

$$\mathbf{x}'=V^T\mathbf{x}, \quad \mathbf{x}'=\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}=\begin{pmatrix}H^T\mathbf{x}\\ K^T\mathbf{x}\end{pmatrix}, \quad (25)$$

where $\mathbf{v}=H^T\mathbf{x}$ is a vector of length P and $\mathbf{w}=K^T\mathbf{x}$ is a vector of length $N-P$. In general, if the covariance matrix of the random variables $\mathbf{x}$ is S, then the covariance matrix S' of $\mathbf{x}'=V^T\mathbf{x}$ is given by $S'=V^TSV$. Hence, since in the present case x_i are normal distributed with mean 0 and covariance matrix $\mathbf{1}_N$ the same is true for x_i'. In particular also $\mathbf{v}$ and $\mathbf{w}$ are statistically independent. Using Eqs. (23) and (24) we deduce

[4]If $\mathrm{rank}[B]=P'<P$ some of the parameters θ_α are not independent. In this case one can perform a change of variables and choose a new sets of parameters θ'_β, such that $\chi^2(\boldsymbol{\theta}')$ depends only on the first P' of them. The remaining parameters are not relevant for the problem and can be eliminated from the very beginning. When repeating the construction in the new set of variables, the number of parameters will be equal to the rank of B.

$$\hat{\mathbf{y}}^T\hat{\mathbf{y}}=\mathbf{x}^T(\mathbf{1}_N-HH^T)\mathbf{x}=\mathbf{x}^TKK^T\mathbf{x}=\mathbf{w}^T\mathbf{w} \tag{26}$$

proving that $\chi^2_{\rm min}=\hat{\mathbf{y}}^T\hat{\mathbf{y}}$ has a χ^2 distribution with $N-P$ d.o.f. Finally, we obtain

$$\Delta\chi^2=\chi^2(\theta^0)-\chi^2_{\rm min}=\mathbf{x}^T\mathbf{x}-\hat{\mathbf{y}}^T\hat{\mathbf{y}}=\mathbf{x}^THH^T\mathbf{x}=\mathbf{v}^T\mathbf{v}, \tag{27}$$

showing that $\Delta\chi^2$ has a χ^2 distribution with P d.o.f. and is statistically independent of $\chi^2_{\rm min}$. □

B. The parameter goodness-of-fit

Moving now to the PG we generalize in an obvious way the formalism of the previous section by attaching and index r for the data set to each quantity. We have

$$\chi^2_{\rm tot}(\boldsymbol{\theta})=\sum_r\ \mathbf{y}_r^T(\boldsymbol{\theta})\mathbf{y}_r(\boldsymbol{\theta}),\quad \chi^2_{\rm tot}(\boldsymbol{\theta}^0)=\sum_r\ \mathbf{x}_r^T\mathbf{x}_r, \tag{28}$$

and

$$\overline{\chi^2}(\boldsymbol{\theta})\equiv\chi^2_{\rm tot}(\boldsymbol{\theta})-\sum_r\ \chi^2_{\rm r,min}=\sum_r\ [\mathbf{y}_r^T(\boldsymbol{\theta})\mathbf{y}_r(\boldsymbol{\theta})-\hat{\mathbf{y}}_r^T\hat{\mathbf{y}}_r]. \tag{29}$$

Proposition 2. Let $\tilde{\boldsymbol{\theta}}$ *be the values of the parameters which minimize* $\overline{\chi^2}(\boldsymbol{\theta})$*, or equivalently* $\chi^2_{\rm tot}(\boldsymbol{\theta})$*. Then* $\overline{\chi^2}_{\rm min}=\overline{\chi^2}(\tilde{\boldsymbol{\theta}})$ *follows a* χ^2 *distribution with* P_c *d.o.f., with*

$$P_c\equiv\mathcal{P}-P,\quad \mathcal{P}\equiv\sum_{r=1}^{D}P_r,\quad P_r\equiv{\rm rank}[B_r]\quad \text{and}$$

$$B_r\equiv\left.\frac{\partial\mathbf{y}_r}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^0}. \tag{30}$$

The matrices B_r are of order $N_r\times P$. Since a given data set r may depend only on some of the P parameters, or on some combination of them, in general one has to consider the possibility of $P_r\leqslant P$.[5] This means that the symmetric $P\times P$ matrix $B_r^TB_r$ can be writen as $R_rb_r^Tb_rR_r^T$, where R_r is an orthogonal matrix and b_r is a $P_r\times P$ "diagonal" matrix, such that the diagonal $P\times P$ matrix $b_r^Tb_r$ will have P_r nonzero entries. Let us now define the $P\times P_r$ "diagonal" matrix b_r^{-1} in such a way that $(b_r^{-1})_{ii}\equiv 1/(b_r)_{ii}$ for each of the P_r nonvanishing entries of b_r, and all other elements are zero. In analogy to Eq. (22) we introduce now the matrices

$$H_r\equiv B_rR_rb_r^{-1}, \tag{31}$$

which are of order $N_r\times P_r$. To prove Proposition 2 we define the vectors of length $\mathcal{P}$

$$\mathbf{Y}(\boldsymbol{\theta})\equiv\begin{pmatrix}H_1^T\mathbf{y}_1(\boldsymbol{\theta})\\ \vdots\\ H_D^T\mathbf{y}_D(\boldsymbol{\theta})\end{pmatrix},\quad \mathbf{X}\equiv\begin{pmatrix}H_1^T\mathbf{x}_1\\ \vdots\\ H_D^T\mathbf{x}_D\end{pmatrix}=\begin{pmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_D\end{pmatrix}. \tag{32}$$

In the first part of the proof we show that $\overline{\chi^2}_{\rm min}=\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}$ with $\tilde{\mathbf{Y}}\equiv\mathbf{Y}(\tilde{\boldsymbol{\theta}})$. With arguments similar to the ones leading to Eq. (21) we find

$$\sum_r\ \tilde{\mathbf{y}}_r^T\,\tilde{\mathbf{y}}_r=\sum_r\ \mathbf{x}_r^T\mathbf{x}_r-(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)^T\sum_r\ B_r^TB_r(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0). \tag{33}$$

Using further Eq. (23) for each r we obtain

$$\begin{aligned}\overline{\chi^2}_{\rm min}&=\sum_r\ \tilde{\mathbf{y}}_r^T\,\tilde{\mathbf{y}}_r-\sum_r\ \hat{\mathbf{y}}_r^T\hat{\mathbf{y}}_r\\ &=\sum_r\ \mathbf{x}_r^TH_rH_r^T\mathbf{x}_r-(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)^T\sum_r\ B_r^TB_r(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0).\end{aligned} \tag{34}$$

On the other hand we can use that the minimum values $\tilde{\boldsymbol{\theta}}$ are converging to the true values $\boldsymbol{\theta}^0$ in the large sample limit and write $\tilde{\mathbf{Y}}\approx\mathbf{X}+\mathcal{B}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)$, where we have defined the $\mathcal{P}\times P$ matrix

$$\mathcal{B}\equiv\left.\frac{\partial\mathbf{Y}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^0}=\begin{pmatrix}H_1^TB_1\\ \vdots\\ H_D^TB_D\end{pmatrix}. \tag{35}$$

Without loss of generality we assume that ${\rm rank}[\mathcal{B}]=P$. Again, with arguments similar to those leading to Eq. (21) we derive

$$\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}=\mathbf{X}^T\mathbf{X}-(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)^T\mathcal{B}^T\mathcal{B}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0). \tag{36}$$

Using Eq. (31) it is easy to show that $\mathcal{B}^T\mathcal{B}=\Sigma_rB_r^TB_r$, and by comparing Eqs. (36) and (34) we can readily verify the relation $\overline{\chi^2}_{\rm min}=\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}$.

To complete the proof we identify $\mathbf{Y}\leftrightarrow\mathbf{y}$ and $\mathbf{X}\leftrightarrow\mathbf{x}$ and proceed in perfect analogy to the proof of Proposition 1 given in Sec. III A. In particular, from the arguments presented there it follows that the elements of $\mathbf{v}_r$ are P_r independent Gaussian variables with mean 0 and variance 1. Since the D data sets are assumed to be statistically independent the vector $\mathbf{X}$ contains $\mathcal{P}$ independent Gaussian variables with mean 0 and variance 1. In analogy to the matrices H,K of Sec. III A we now obtain the $\mathcal{P}\times P$ matrix $\mathcal{H}$ and the $\mathcal{P}\times P_c$ matrix $\mathcal{K}$, which fulfil $\mathcal{H}\mathcal{H}^T+\mathcal{K}\mathcal{K}^T=\mathbf{1}_{\mathcal{P}}$, and Eq. (36) becomes

$$\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}=\mathbf{X}^T(\mathbf{1}_{\mathcal{P}}-\mathcal{H}\mathcal{H}^T)\mathbf{X}. \tag{37}$$

In analogy to the vector $\mathbf{w}$ from Eq. (25) we now define $\mathbf{W}\equiv\mathcal{K}^T\mathbf{X}$, containing $P_c=\mathcal{P}-P$ independent Gaussian variables with mean 0 and variance 1, and Eq. (37) gives

[5]Note that the definition of P_r in Eq. (30) is equivalent to that given in Eq. (11).

$$\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}=\mathbf{X}^T\mathcal{K}\mathcal{K}^T\mathbf{X}=\mathbf{W}^T\mathbf{W}. \tag{38}$$

From Eq. (38) it is evident that $\overline{\chi^2_{\min}}=\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}$ follows a χ^2 distribution with P_c d.o.f. □

Let us conclude this section by noting that both Proposition 1 and 2 are *exact* if the data are multinormally distributed and the theoretical predictions $\boldsymbol{\mu},\boldsymbol{\mu}^r$ depend linearly on the parameters $\boldsymbol{\theta}$. If these requirements are not fulfilled, simplified expressions (5) and (12) are valid only approximately, and to calculate the SG and the PG one should in principle use general formulas (4) and (10) instead. However, we want to stress that under rather general conditions $\chi^2_{\min}$ and $\overline{\chi^2_{\min}}$ will be distributed as a χ^2 *in the large sample limit* (i.e. for large $\boldsymbol{\nu}$ and $\boldsymbol{\nu}^r$, respectively), so that even in the general case Eqs. (5) and (12) can still be used.

IV. EXAMPLES AND DISCUSSION

In this section we illustrate the application of the PG on some examples. In Sec. IV A we show that in the simple case of two measurements of a single parameter the PG is identical to the intuitive method of considering the difference of the two measurements, and in Sec. IV B we show the consistency of the PG and the SG in the case of independent data points. In Sec. IV C we discuss the application of the PG to neutrino oscillation data in the framework of a sterile neutrino scheme. This problem was the original motivation to introduce the PG in Ref. [5]. In Sec. IV D we add some general remarks on the PG.

A. The determination of one parameter by two experiments

Let us consider two data sets observing the data points $\boldsymbol{\nu}^1=(\nu_i^1)$ $(i=1,\ldots,N_1)$ and $\boldsymbol{\nu}^2=(\nu_i^2)$ $(i=1,\ldots,N_2)$. Further, we assume that the expectation values for both data sets can be calculated from a theory depending on one parameter η: $\boldsymbol{\mu}^r(\eta)$ $(r=1,2)$, and all ν_i^r are independent and normal distributed around the expectation values with variance σ_i^r. Then we have the following χ^2 functions for the two data sets $r=1,2$:

$$\chi^2_r(\eta)=\sum_{i=1}^{N_r}\left(\frac{\nu_i^r-\mu_i^r(\eta)}{\sigma_i^r}\right)^2=\chi^2_{r,\min}+\left(\frac{\hat{\eta}_r-\eta}{\hat{\sigma}_r}\right)^2, \tag{39}$$

where $\hat{\eta}_r=\hat{\eta}_r(\boldsymbol{\nu}^r)$ is the value of the parameter at the χ^2 minimum of data set r. Now one may ask the question whether the results of the two experiments are consistent. More precisely, we are interested in the probability of obtaining $\hat{\eta}_1$ and $\hat{\eta}_2$ under the assumption that both result from the *same* true value η^0.

A standard method (see e.g. Ref. [8] Sec. 14.3) to answer this question is to consider the variable

$$z=\frac{\hat{\eta}_1-\hat{\eta}_2}{\sqrt{\hat{\sigma}_1^2+\hat{\sigma}_2^2}}. \tag{40}$$

If the theory is correct z is normal distributed with mean 0 and variance 1. Hence we can answer the question raised above by citing the probability of obtaining $|z|\geqslant|z_{\rm obs}|$:

$$p=1-\int_{-|z_{\rm obs}|}^{|z_{\rm obs}|}f_N(z;0,1)dz, \tag{41}$$

where f_N denotes the normal distribution.

If the PG is applied to this problem, one obtains from Eq. (39)

$$\overline{\chi^2}(\eta)=\left(\frac{\hat{\eta}_1-\eta}{\hat{\sigma}_1}\right)^2+\left(\frac{\hat{\eta}_2-\eta}{\hat{\sigma}_2}\right)^2, \tag{42}$$

and after some simple algebra one finds $\overline{\chi^2_{\min}}=z^2$, where z is given in Eq. (40). Obviously, applying Eq. (12) to calculate the p value according to the PG with the relevant number of d.o.f. $P_c=2-1=1$ leads to the same result as Eq. (41).

Hence, we arrive at the conclusion that in this simple case of testing the compatibility of two measurements for the mean of a Gaussian, the PG is identical to the intuitive method of testing whether the difference of the two values is consistent with zero.

B. Consistency of PG and SG for independent data points

As a further example of the consistency of the PG method we consider the case of N statistically independent data points ν_i. Let us denote by σ_i the standard deviation of the observation ν_i $(i=1,\ldots,N)$, and the corresponding theoretical prediction by $\mu_i(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of P parameters. For simplicity, we assume that each of the μ_i depends at least on one parameter. Then the χ^2 is given by

$$\chi^2(\boldsymbol{\theta})=\sum_{i=1}^{N}\chi_i^2(\boldsymbol{\theta}),$$

where

$$\chi_i^2(\boldsymbol{\theta})=\frac{[\nu_i-\mu_i(\boldsymbol{\theta})]^2}{\sigma_i^2}, \tag{43}$$

and from the SG construction (see Sec. III A) we know that $\chi^2_{\min}$ follows a χ^2 distribution with $N-P$ degrees of freedom. On the other hand, if we consider each single data point as an independent data set and we apply the PG construction, we easily see that $\chi^2_{i,\min}=0$ for each i. This implies $\overline{\chi^2}(\boldsymbol{\theta})=\chi^2(\boldsymbol{\theta})$, and in particular $\overline{\chi^2_{\min}}=\chi^2_{\min}$. Therefore, for the specific case considered here one expects that SG and PG are identical.

To show that this is really the case let us first note that each matrix $\partial\mu_i/\partial\boldsymbol{\theta}$ consists just of *a single line*, and therefore it obviously has rank one. Hence, Eq. (11) gives $P_i=1$ for each i. This reflects the fact that from the measurement of a single observable we cannot derive *independent* bounds on P parameters, but only *a single combination* of them is constrained. Therefore, the number of d.o.f. relevant for the calculation of the PG is given by $P_c=\Sigma_{i=1}^N P_i-P$

$=N-P$, which is exactly the number of d.o.f. relevant for the SG. Hence, we have shown that in the considered case the two methods are equivalent and consistent.

C. Application to neutrino oscillation data

In this section we use real data from neutrino oscillation experiments to discuss the application of the PG and to compare it to the SG. We consider the so-called (2+2) neutrino mass scheme, where a fourth (sterile) neutrino is introduced in addition to the three standard model neutrinos. In general this model is characterized by nine parameters: three neutrino mass-squared differences $\Delta m^2_{\rm sol}$, $\Delta m^2_{\rm atm}$, $\Delta m^2_{\rm LSND}$ and six mixing parameters $\theta_{\rm sol}$, $\theta_{\rm atm}$, $\theta_{\rm LSND}$, d_μ, η_s, η_e. An interested reader can find precise definitions of the parameters, applied approximations, an extensive discussion of physics aspects, and references in Refs. [3,5,10]. Here we are interested mainly in the statistical aspects of the analysis, and therefore we consider a simplified scenario.

We do not include LSND, KARMEN and all the experiments sensitive to $\Delta m^2_{\rm LSND}$ and the corresponding mixing angle $\theta_{\rm LSND}$. Hence, we are left with three datasets from solar, atmospheric and reactor neutrino experiments. The solar dataset includes the current global solar neutrino data from the SNO, Super-Kamiokande, Gallium and Chlorine experiments, making a total of 81 data points, whereas the atmospheric data sample includes 65 data points from the Super-Kamiokande and MACRO experiments (for details of the solar and atmospheric analysis see Ref. [10]). In the reactor data set we include only the data from the KamLAND and the CHOOZ experiments, leading to a total of $13+14=27$ data points [11,12]. In general the reactor experiments (especially CHOOZ) depend in addition to $\Delta m^2_{\rm sol}$ and $\theta_{\rm sol}$ also on $\Delta m^2_{\rm atm}$ and a further mixing parameter η_e. However, we adopt here the approximation $\eta_e=1$, which is very well justified in the (2+2) scheme [3]. This implies that the dependence on $\Delta m^2_{\rm atm}$ disappears and we are left with the parameters $\Delta m^2_{\rm sol}$ and $\theta_{\rm sol}$ for both reactor experiments, KamLAND as well as CHOOZ.

Under these approximations the experimental datasets we are using are described only by the six parameters $\Delta m^2_{\rm sol}$, $\Delta m^2_{\rm atm}$, $\theta_{\rm sol}$, $\theta_{\rm atm}$, η_s, d_μ. The parameter structure is illustrated in Fig. 1. This simplified analysis serves well for discussing the statistical aspects of the problem; a more general treatment including a detailed discussion of the physics is given in Refs. [3,5]. In Table I we summarize the parameter

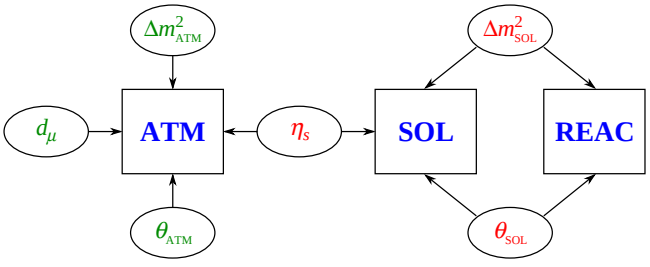

FIG. 1. Parameter structure of the three data sets from reactor, solar and atmospheric neutrino experiments.

TABLE I. Parameter dependence, total number of data points, $\chi^2_{\rm min}$ and the corresponding SG for the three data sets.

Dataset	Parameters	N	$\chi^2_{\rm min}$/d.o.f.	SG
Reactor	$\Delta m^2_{\rm sol}, \theta_{\rm sol}$	27	11.5/25	99%
Solar	$\Delta m^2_{\rm sol}, \theta_{\rm sol}, \eta_s$	81	65.8/78	84%
Atmospheric	$\Delta m^2_{\rm atm}, \theta_{\rm atm}, \eta_s, d_\mu$	65	38.4/61	99%

dependence, the number of data points, the minimum values of the χ^2 functions and the resulting SG. We observe that all the datasets analyzed alone give a very good fit. Let us remark that especially in the case of reactor and atmospheric data the SG is suspicious high. This may indicate that the errors have been estimated very conservatively.

In Table II we show the results of an SG and PG analysis for various combinations of the three datasets. In the first three lines in the table only two out of the three datasets are combined. By combining solar and atmospheric neutrino data we find a $\chi^2_{\rm min}$ of 126.7. With the quite large number of d.o.f. of 140 this gives an excellent SG of 78.3%. If, however, the PG is applied we obtain a goodness-of-fit of only 3.54×10^{-6}. The reason for this very bad fit can be understood from Figs. 1 and 2. From Fig. 1 one finds that solar and atmospheric data are coupled by the parameter η_s. In Fig. 2 $\Delta\chi^2$ is shown for both sets as a function of this parameter. We find that there is indeed significant disagreement between the two datasets:[6] solar data prefers values of η_s close to 0, whereas atmospheric data prefers values close to one. There are two reasons why this strong disagreement does not show up in the SG. First, since the SG of both datasets alone is very good, there is much room to ''hide'' some problems in the combined analysis. Second, because of the large number of data points many of them actually might not be sensitive to the parameter η_s, where the disagreement becomes manifest. Hence, the problem in the combined fit becomes diluted due to the large number of data points. We conclude that the PG is very sensitive to disagreement of the data sets, even in cases where the individual χ^2 minima are very low, and when the number of data points is large.

In the reactor + solar analysis one finds complete agreement between the two data sets for the SG as well as for the PG. This reflects the fact that the determination of the parameters $\theta_{\rm sol}$ and $\Delta m^2_{\rm sol}$ from reactor and solar neutrino experiment are in excellent agreement [11]. Finally, in the case of the combined analysis of reactor and atmospheric data the PG cannot be applied. In our approximation these datasets have no parameter in common as one can see in Fig. 1. Hence, it makes no sense to test their compatibility, or even to combine them at all.

In the lower part of Table II we show the results from combining all three data sets. By comparing these results

[6]The physical reason for this is that both datasets strongly disfavor oscillations into sterile neutrinos. Since it is a generic prediction of the (2+2) scheme that the sterile neutrino must show up either in solar or in atmospheric neutrino oscillations the model is ruled out by the PG test [5].

TABLE II. Comparison of SG and PG for various combinations of the datasets from solar (sol), atmospheric (stm) and reactor (react) neutrino experiments.

Datasets	$N_{\rm tot}$	$\chi^2_{\rm tot,min}$/d.o.f.	SG	$\Sigma_r P_r$	P	$\overline{\chi}^2_{\rm min}/P_c$	PG
Sol + atm	146	126.7/140	78.3%	3+4	6	21.5/1	3.54×10^{-6}
React + sol	108	77.4/105	98.0%	2+3	3	0.13/2	93.5%
React + atm	92	49.9/86	99.9%	2+4	6	0.0/0	
KamL + sol + atm	159	132.7/153	88.1%	2+3+4	6	21.7/3	7.53×10^{-5}
React + sol + atm	173	138.2/167	95.0%	2+3+4	6	21.7/3	7.53×10^{-5}

with that from the solar + atmospheric analysis one can appreciate the advantage of the PG. If we add only the 13 data points from KamLAND to the solar and atmospheric samples we observe that the SG improves from 78.3% to 88.1%, whereas if both reactor experiments are included we obtain an SG of 95.0%. This demonstrates that the SG strongly depends on the number of data points, especially the 14 data points from CHOOZ contain nearly no relevant information, since the best fit values of $\Delta m^2_{\rm sol}$ and $\theta_{\rm sol}$ are in the no-oscillation regime for CHOOZ implying that χ^2 is flat. Moreover, since reactor data are not sensitive to the parameter η_s (see Fig. 1) the disagreement between solar and atmospheric data becomes even more diluted by the additional reactor data points. This clearly illustrates that the SG can be drastically improved by adding data which contain no information on the relevant parameters. Also the PG improves slightly by adding reactor data, reflecting the good agreement between solar and reactor data. However, the resulting PG is still very small due to the disagreement between solar and atmospheric data in the model under consideration. Moreover, the PG is completely unaffected by the addition of the CHOOZ data, because χ^2 of CHOOZ is flat in the relevant parameter region, and the PG is sensitive only to the parameter dependence of the datasets.

Finally, we mention that in view of the analyses shown in Table II the meaning of P_c, the number of d.o.f. for PG becomes clear. It corresponds to the number of parameters coupling the datasets. Solar and atmospheric data are coupled only by η_s, hence $P_c=1$, whereas reactor and solar data are coupled by $\theta_{\rm sol}$ and $\Delta m^2_{\rm sol}$ and $P_c=2$. Atmospheric data has no parameter in common with reactor data, therefore $P_c=0$. In the combination of reactor + solar + atmospheric datasets, the three parameters η_s, $\theta_{\rm sol}$, $\Delta m^2_{\rm sol}$ provide the coupling and $P_c=3$.

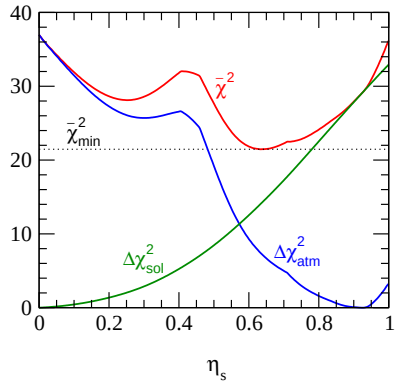

FIG. 2. PG for solar and atmospheric neutrino data in the (2 +2) scheme.

D. General remarks on the PG

(1) Using the relation $\overline{\chi}^2_{\rm min}=\Sigma_r\Delta\chi^2_r(\tilde{\boldsymbol{\theta}})$ one can obtain more insight into the quality of the fit by considering the contribution of each dataset to $\overline{\chi}^2_{\rm min}$. If the PG is poor it is possible to identify the data sets leading to the problems in the fit by looking at the individual values of $\Delta\chi^2_r(\tilde{\boldsymbol{\theta}})$. In this sense the PG is similar to the so-called ''pull approach'' discussed in Ref. [13] in relation with solar neutrino analysis.

(2) One should keep in mind that the PG is completely insensitive to the goodness-of-fit of the *individual* data sets. Because of the subtraction of the $\chi^2_{r,\rm min}$ in Eq. (8) all the information on the quality of the fit of the datasets alone is lost. One may benefit from this property if the SG of the individual datasets is very good (see the example in Sec. IV C). On the other hand, if e.g. one dataset gives a bad fit on its own this will not show up in the PG. Only the *compatibility* of the datasets is tested, irrespective of their individual SGs.

(3) The PG might be also useful if one is interested whether a dataset consisting of very few data points is in agreement with a large data sample.[7] The SG of the combined analysis will be completely dominated by the large sample and the information contained in the small data sample may be drowned out by the large number of data points. In such a case the PG can give valuable information on the compatibility of the two sets, because it is not diluted by the number of data points in each set and it is sensitive only to the parameter dependence of the sets.

V. CORRELATIONS DUE TO THEORETICAL UNCERTAINTIES

One of the limitations of the PG is that it can be applied only if the datasets are *statistically independent*. In many physically interesting situations (for example, different solar neutrino experiments) this is not the case since *theoretical uncertainties* introduce correlations between the results of

[7]For example, one could think of a combination of the 19 neutrino events from the Super Nova 1987A with the high statistics global solar neutrino data.

different—and otherwise independent—experiments. However, in such a case one can take advantage of the so-called pull approach, which, as demonstrated in Ref. [13], is equivalent to the usual covariance method. In that paper it was shown that if correlations due to theoretical errors exist, it is possible to account for them by introducing new parameters ξ_a and adding penalty functions to χ^2. In this way it is possible to get rid of unwanted correlations and the PG can be applied. The correlation parameters ξ_a should be treated in the same way as the parameters $\boldsymbol{\theta}$ of the theoretical model.

In this section we illustrate this procedure by considering a generic experiment with an uncertainty on the normalization of the predicted number of events. Let the experiment observe some energy spectrum which is divided into N bins. The theoretical prediction for the bin i is denoted by $\mu_i(\boldsymbol{\theta})$ depending on P parameters $\boldsymbol{\theta}$. In praxis often $\mu_i(\boldsymbol{\theta})$ is not known exactly. Let us consider the case of a fully correlated relative error $\sigma_{\rm th}$. A common method to treat such an error is to add statistical and theoretical errors in quadrature, leading to the correlation matrix

$$S_{ij}(\boldsymbol{\theta}) = \delta_{ij}\sigma_{i,\rm stat}^2 + \sigma_{\rm th}^2 \mu_i(\boldsymbol{\theta})\mu_j(\boldsymbol{\theta}), \tag{44}$$

where $\sigma_{i,\rm stat}$ is the statistical error in the bin i. In the case of neutrino oscillation experiments such a correlated error results e.g. from an uncertainty of the initial flux normalization or of the fiducial detector volume. The χ^2 is given by

$$\chi^2(\boldsymbol{\theta}) = \sum_{i,j=1}^{N} [\nu_i - \mu_i(\boldsymbol{\theta})] S_{ij}^{-1}(\boldsymbol{\theta}) [\nu_j - \mu_j(\boldsymbol{\theta})], \tag{45}$$

where ν_i are the observations. As shown in Ref. [13], instead of Eq. (45) we can equivalently use

$$\chi^2(\boldsymbol{\theta},\xi) = \sum_{i=1}^{N} \left(\frac{\nu_i - \xi\mu_i(\boldsymbol{\theta})}{\sigma_{i,\rm stat}} \right)^2 + \left(\frac{\xi - 1}{\sigma_{\rm th}} \right)^2, \tag{46}$$

and minimize with respect to the new parameter ξ.

On the other hand, if ξ is considered as an additional parameter, on the same footing as $\boldsymbol{\theta}$, all the data points are formally uncorrelated and it is straightforward to apply the PG. Subtracting the minimum of the first term in Eq. (46) with respect to $\boldsymbol{\theta}$ and ξ one obtains

$$\overline{\chi^2}(\boldsymbol{\theta},\xi) = \Delta\chi^2(\boldsymbol{\theta},\xi) + \left(\frac{\xi - 1}{\sigma_{\rm th}} \right)^2. \tag{47}$$

The external information on the parameter ξ represented by the second term in Eq. (47) is considered as an additional dataset. Evaluating the minimum of Eq. (47) for 1 d.o.f. is a convenient method to test if the best fit point of the model is in agreement with the constraint on the overall normalization. In particular one can identify whether a problem in the fit comes from the spectral shape (first term) or the total rate (second term).

Moreover, one may like to divide the data into two parts, set I consisting of bins $1, \ldots, n$ and and set II consisting of bins $n, \ldots, N$, and test whether these datasets are compatible. Equation 46 can be written as

$$\chi^2(\boldsymbol{\theta},\xi) = \sum_{i=1}^{n} \left(\frac{\nu_i - \xi\mu_i(\boldsymbol{\theta})}{\sigma_{i,\rm stat}} \right)^2 + \sum_{i=n}^{N} \left(\frac{\nu_i - \xi\mu_i(\boldsymbol{\theta})}{\sigma_{i,\rm stat}} \right)^2 + \left(\frac{\xi - 1}{\sigma_{\rm th}} \right)^2, \tag{48}$$

and subtracting the minima of the first two terms gives the $\overline{\chi^2}$ relevant for the PG:

$$\overline{\chi^2}(\boldsymbol{\theta},\xi) = \Delta\chi^2_{\rm I}(\boldsymbol{\theta},\xi) + \Delta\chi^2_{\rm II}(\boldsymbol{\theta},\xi) + \left(\frac{\xi - 1}{\sigma_{\rm th}} \right)^2. \tag{49}$$

Assuming that the datasets I and II both depend on all P parameters $\boldsymbol{\theta}$, the minimum of this $\overline{\chi^2}$ has to be evaluated for $P+2$ d.o.f. to obtain the PG. This procedure tests whether the data sets I and II are consistent with each other *and* the constraint on the overall normalization. By considering the relative contributions of the three terms in Eq. (49) it is possible to identify potential problems in the fit. For example, one may test whether a bad fit is dominated only by a small subset of the data, e.g. a few bins at the low or high end of the spectrum. Alternatively, the two datasets I and II can come from two different experiments correlated by a common normalization error, e.g. two detectors observing events from the same beam.

It is straightforward to apply the method sketched in this section also in more complicated situations. For example, if there are several sources of theoretical errors leading to more complicated correlations the pull approach can also be applied by introducing a parameter ξ_a for each theoretical error [13]. In a similar way one can treat the case when the compatibility of several experiments should be tested, which are correlated by common theoretical uncertainties. (Consider e.g. the various solar neutrino experiments, which are correlated due to the uncertainties on the solar neutrino flux predictions.)

VI. CONCLUSIONS

In this note we have discussed a goodness-of-fit method which was proposed in Ref. [5]. The so-called *parameter goodness-of-fit* can be applied when the global data consists of several statistically independent subsets. Its construction and application are very similar to the standard goodness-of-fit. We gave a formal derivation of the probability distribution function of the proposed test statistics, based on standard theorems of statistics, and illustrated the application of the PG on some examples. We have shown that in the simple case of two datasets determining the mean of a Gaussian, the PG is identical to the intuitive method of testing whether the difference of the two measurements is consistent with zero. Furthermore, we have compared the standard goodness-of-fit and the PG by using real data from neutrino oscillation experiments, which have been the original motivation for the PG. In addition we have illustrated that the so-called pull approach allows us to apply the PG also in cases where the datasets are correlated due to theoretical uncertainties.

The proposed method tests the compatibility of different

datasets, and it gives sensible results even in cases where the errors are estimated very conservatively and/or the total number of data points is very large. In particular, it avoids the problem that a possible disagreement between datasets becomes diluted by data points which are insensitive to the problem in the fit. The PG can also be very useful when a set consisting of a rather small number of data points is combined with a very large data sample.

To conclude, we believe that physicists should keep an open mind when choosing a statistical method for analyzing experimental data. In many cases much more information can be extracted from data if the optimal statistical tool is used. We think that the method discussed in this note may be useful in several fields of physics, especially where global analyses of large amount of data are performed.

ACKNOWLEDGMENTS

We thank C. Giunti and P. Huber for very useful discussions. Furthermore, we would like to thank many participants of the NOON 2003 workshop in Kanazawa, Japan, for various comments on the PG. M.M. was supported by the Spanish Grant No. BFM2002-00345, by the European Commission RTN network HPRN-CT-2000-00148, by the ESF Neutrino Astrophysics Network N. 86, and by the Marie Curie Contract No. HPMF-CT-2000-01008. T.S. was supported by the ''Sonderforschungsbereich 375-95 für Astro-Teilchenphysik'' der Deutschen Forschungsgemeinschaft.

[1] K. Pearson, Philos. Mag. V **50**, 157 (1900).
[2] J.C. Collins and J. Pumplin, hep-ph/0105207.
[3] M. Maltoni, T. Schwetz, and J.W.F. Valle, Phys. Rev. D **65**, 093004 (2002).
[4] P. Creminelli, G. Signorelli, and A. Strumia, J. High Energy Phys. **05**, 052 (2001); M.V. Garzelli and C. Giunti, *ibid.* **12**, 017 (2001).
[5] M. Maltoni, T. Schwetz, M.A. Tòrtola, and J.W.F. Valle, Nucl. Phys. **B643**, 321 (2002).
[6] Particle Data Goup, K. Hagiwara *et al.*, Phys. Rev. D **66**, 010001 (2002).
[7] H. Cramér, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, 1946).
[8] A.G. Frodesen, O. Skjeggestad, and H. Tofte, *Probability and Statistics in Particle Physics* (Universitatsforlaget, Bergen, 1979).
[9] B.P. Roe, *Probability and Statistics in Experimental Physics* (Springer-Verlag, New York, 1992).
[10] M. Maltoni, T. Schwetz, M.A. Tòrtola, and J.W.F. Valle, Phys. Rev. D **67**, 013011 (2003).
[11] M. Maltoni, T. Schwetz, and J.W.F. Valle, Phys. Rev. D **67**, 093003 (2003).
[12] M. Apollonio *et al.*, Eur. Phys. J. C **27**, 331 (2003).
[13] G.L. Fogli, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, Phys. Rev. D **66**, 053010 (2002).