第 27 卷 第 5 期
2003 年 5 月

高 能 物 理 与 核 物 理
HIGH ENERGY PHYSICS AND NUCLEAR PHYSICS

Vol.27, No.5
May, 2003

Equivalence of Two χ^2 Forms

MO Xiao-Hu[1,2] ZHU Yong-Sheng[1]

1 (Institute of High Energy Physics, CAS, Beijing 100039, China)

2 (China Center of Advanced Science and Technology, Beijing 100080, China)

Abstract The equivalence of two χ^2 forms is proved by matrix calculation. The bias of minimization estimate is also discussed. The simplified R-value measurements are quoted to test the conclusion quantitatively.

Key words equivalence, χ^2 form, bias, R-value measurement

1 Introduction

The covariance matrix is usually used to construct the χ^2 function for correlative data which is to be minimized to acquire best estimates for parameter interested[1]. It is frequently the case that experimental data are affected by overall systematic error, often just one common normalization uncertainty. Under such case, the $n \times n$ covariance matrix $\boldsymbol{V}$ for n measurements could be constructed as follows: the diagonal elements are given by the sum of the squares of the statistical σ_{stat}, systematic point-to-point σ_{ptp}, and common normalization uncertainty σ_{norm} for each measurement. The correlation between data points i and j is contained in off-diagonal matrix element V_{ij}, which is estimated by the product of σ^i_{norm} of measurement i and σ^j_{norm} of measurement j. The expression to be minimized is then[1)]

$$\chi^2 = \eta^{\mathrm{T}} \boldsymbol{V}^{-1} \eta, \tag{1}$$

where

$$\eta = \begin{pmatrix} x_1 - k_1 \\ x_2 - k_2 \\ \vdots \\ x_n - k_n \end{pmatrix}$$

is the vector of the residuals between experimental observations x_i and theoretical estimations k_i.

Apart from matrix method, an alternative way to handle correlation problem is called factor method[3,4]. A normalization factor f is introduced and is to be fitted as a free parameter to take the correlation into account,

$$\chi^2 = \sum_i \frac{(fx_i - k_i)^2}{\sigma_i^2} + \frac{(f-1)^2}{\sigma_n^2}, \tag{2}$$

where σ_n is the relative common normalization error and $\sigma_i^2 = \sigma^2_{\text{stat}} + \sigma^2_{\text{ptp}}$.

The relation between these two methods was first discussed by D'Agostini[5]. However in Ref. [5], the equivalence of two χ^2 forms is proved only for two measurements case. In this paper, the equivalence proof has been extended to multi-measurement case. In addition, some experiment results are quoted to test the equivalence quantitatively.

2 Equivalence proof

2.1 Covariance matrix construction

In order to get analytical results, following D'Agostini's approach, we only consider the constant fitting case, that is, the theoretical expectation k_i is a constant, denoted

Received 24 September 2002

* Supported by National Natural Science Foundation of China (19991483) and 100 Talents Programme of CAS (U-25)

1) In this paper, chi-square minimization is adopted to obtain the best estimated value. For experimental data minimization, the MINUIT package is used. The detail explanation about chi-square minimization technique and MINUIT package can be found in Ref. [2].

371—376

as k. Assume there are n measurement values x_i, and one scale factor f, which affects all n measurements. As an example, x_i could be the cross section at energy i, and f denotes the overall normalization factor in the determination of luminosities at all n energies. Then, x_i will be related with the luminosity L_i of e^+e^- collider at energy i by the equation

$$x_i = \frac{N_i}{L_i} \quad (N_i: \text{events number}),$$

and therefore be related with the overall normalization factor f. We further assume that f has the expectation equal to 1, and its relative error is known as σ_f; and x_i has uncertainty of σ_i, which is known[1)]. In this case, considering the correlations between n measurements, the covariance matrix is then

$$V_x = \left(\begin{array}{cccc|c} \sigma_1^2 & 0 & \cdots & 0 & \\ 0 & \sigma_2^2 & \cdots & 0 & \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_n^2 & \\ \hline & & 0 & & \sigma_f^2 \end{array}\right).$$

The corrected measurement values are

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} fX_1 \\ fX_2 \\ \vdots \\ fX_n \end{pmatrix}$$

and the covariance matrix for Y is obtained from the error propagation

$$V_Y = MV_XM^{\mathrm{T}}, \text{ or } (V_Y)_{mn} = \sum_{ij} M_{mi}(V_X)_{ij}M^{\mathrm{T}}_{jn},$$

where $M_{mi} = \partial Y_m/\partial X_i|_{x_i}$. The elements of V_Y are given

$$V_{Y_{mn}} = \sum_{ij}\left.\frac{\partial Y_m}{\partial X_i}\right|_{x_i}\left.\frac{\partial Y_n}{\partial X_j}\right|_{x_j} V_{X_{ij}},$$

and explicitly[2)], we have

$$V_Y = \begin{pmatrix} \sigma_1^2 + x_1^2\sigma_f^2 & x_1x_2\sigma_f^2 & \cdots & x_1x_n\sigma_f^2 \\ x_2x_1\sigma_f^2 & \sigma_2^2 + x_2^2\sigma_f^2 & \cdots & x_2x_n\sigma_f^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1\sigma_f^2 & x_nx_2\sigma_f^2 & \cdots & \sigma_n^2 + x_n^2\sigma_f^2 \end{pmatrix}. \quad (3)$$

Correspondingly, the χ^2 reads:

$$\chi^2_{\mathrm{m}} = \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - k)\cdot(V^{-1})_{ij}\cdot(x_j - k), \quad (4)$$

where subscript m indicates the matrix method and for simplification, the V_Y's subscript Y has been suppressed. Notice $x_i\sigma_f = \sigma^i_{\mathrm{norm}}$, that is the normalization uncertainty for points i, the V here is just the covariance matrix used in Eq.(1).

2.2 Expectation and variance from covariance matrix method

The minimization of χ^2_{m} expressed by Eq. (4) leads to the best estimates for expectation and variance:

$$\hat{k}_{\mathrm{m}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} x_iV^*_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} V^*_{ij}} \left(\text{from } \frac{\partial\chi^2_{\mathrm{m}}}{\partial k} = 0\right), \text{and } \sigma^2_{\hat{k}_{\mathrm{m}}} = \frac{2}{\frac{\partial^2\chi^2_{\mathrm{m}}}{\partial k\,\partial k}}, \quad (5)$$

where V^*_{ij} is the cofactor corresponding to V_{ij}. For clearness, most of symbol conventions and complex formulae of matrix are relegated to the appendix. Here, the key issue lies in the calculation of the inverse matrix. According to Formula (A.1), together with Eqs. (A.4) and (A.5), we finally transform the calculation involving inverse-matrix into the summation of cofactors of the adjoint matrix, as shown in Eq.(5). Notice

$$x_i = \frac{(\delta_{ij}\sigma_i^2 + \sigma_f^2x_ix_j) - \delta_{ij}\sigma_i^2}{\sigma_f^2x_j} = \frac{V_{ij} - \delta_{ij}\sigma_i^2}{\sigma_f^2x_j},$$

combining Formulae (A.2) and (A.9), it can be worked out:

$$\hat{k}_{\mathrm{m}} = \frac{\bar{x}}{1 + \frac{\sigma_f^2\sigma_s^2}{2}\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(x_i - x_j)^2}{\sigma_i^2\sigma_j^2}} \quad (6)$$

where $\bar{x}$ is the weighted average defined as

$$\bar{x} = \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\Big/\sum_{i=1}^{n}\frac{1}{\sigma_i^2},$$

and σ_s^2 is defined as

$$\frac{1}{\sigma_s^2} = \sum_{i=1}^{n}\frac{1}{\sigma_i^2} \quad \text{or} \quad \sigma_s^2 = 1\Big/\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right).$$

Notice Formulae (A.6) and (A.7), it can be obtained

1) σ_i could contain the the statistical and systematic contribution, that is $\sigma_i^2 = \sigma^2_{\mathrm{stat}} + \sigma^2_{\mathrm{ptp}}$.

2) The details of correlated matrix construction can be found in Ref.[5], where two frequently happened cases, the offset and the normalization cases, have been studied.

$$\sigma^2_{\hat{k}_m} = \frac{\sigma_s^2 + \sigma_f^2 \overline{x}^2}{1 + \frac{\sigma_f^2\sigma_s^2}{2} \cdot \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(x_i - x_j)^2}{\sigma_i^2\sigma_j^2}}.$$

Call r the ratio between $\hat{k}_m$ and $\overline{x}$, then

$$r = \frac{\hat{k}_m}{\overline{x}} = \frac{1}{1 + \sigma_f^2 \cdot \sum_{i=1}^{n}\frac{x_i(x_i - \overline{x})}{\sigma_i^2}}.$$

Here the definition of $\overline{x}$ has been adopted to transform the denominator of Eq. (7) into that of Eq. (8). Using identity $\sum_{i=1}^{n}\frac{\overline{x}_i(x_i - \overline{x})}{\sigma_i^2} \equiv 0$, r could be further written as

$$r = \frac{1}{1 + \sigma_f^2 \cdot \sum_{i=1}^{n}\frac{(x_i - \overline{x})^2}{\sigma_i^2}}.$$

Notice $\sum_{i=1}^{n}\frac{(x_i - \overline{x})^2}{\sigma_i^2} \approx \chi^2(n-1)$, so

$$\langle r \rangle = \frac{1}{1 + \sigma_f^2 \cdot (n-1)}. \qquad (9)$$

This is just the formula firstly given by D' Agostini in Ref. [5]. It can be seen that if there are few data points and normalization error σ_f is so small that $r \approx 1$, the most probable value of the physical quantity equals to the weighted average $\overline{x}$; otherwise the Eq. (9) shows a bias on the χ^2-estimated result when, for a non-vanishing σ_f, a large number of data points are fitted. In particular, the fit produces on average a bias larger than the normalization error itself if $\sigma_f > 1/(n-1)$.

2.3 Expectation and variance from factor method

Now we turn to factor method. A scale factor f is introduced to take into consideration of the correlation. Chi-square has the comparatively simplified form:

$$\chi_f^2 = \sum_{i=1}^{n}\frac{(fx_i - k)^2}{\sigma_i^2} + \frac{(f-1)^2}{\sigma_f^2}, \qquad (10)$$

where subscript f indicates the factor method. The minimization calculation is simple for factor method. From

$$\begin{cases}\frac{\partial \chi_f^2}{\partial k} = 0, \\ \frac{\partial \chi_f^2}{\partial f} = 0,\end{cases}$$

it can be obtained

$$\begin{cases}\hat{k}_f = \hat{f} \cdot \overline{x}, \\ \hat{f} = \frac{1}{1 + \frac{\sigma_f^2\sigma_s^2}{2} \cdot \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(x_i - x_j)^2}{\sigma_i^2\sigma_j^2}}.\end{cases}$$

The inverse of covariance matrix is

$$V_f^{-1} = \frac{1}{2}\begin{pmatrix}\frac{\partial^2\chi_f^2}{\partial k\,\partial k} & \frac{\partial^2\chi_f^2}{\partial f\,\partial k} \\ \frac{\partial^2\chi_f^2}{\partial k\,\partial f} & \frac{\partial^2\chi_f^2}{\partial f\,\partial f}\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^{n}\frac{1}{\sigma_i^2} & -\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} \\ -\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} & \frac{1}{\sigma_f^2} + \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\end{pmatrix},$$

therefore

$$V_f = \frac{1}{D_{V_f^{-1}}} \cdot \begin{pmatrix}\frac{1}{\sigma_f^2} + \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{1}{\sigma_i^2}\end{pmatrix},$$

where

$$D_{V_f^{-1}} = |V_f^{-1}| = \frac{1}{\sigma_f^2} \cdot \sum_{i=1}^{n}\frac{1}{\sigma_i^2} + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{x_i^2 - x_ix_j}{\sigma_i^2\sigma_j^2}.$$

From V_f, the variance of $\hat{k}_f$ reads

$$\sigma^2_{\hat{k}_f} = \frac{\sigma_s^2 + \sigma_f^2\overline{x}^2}{1 + \frac{\sigma_f^2\sigma_s^2}{2} \cdot \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(x_i - x_j)^2}{\sigma_i^2\sigma_j^2}}.$$

Comparing Eqs. (6) and (7) with (11) and (12), the exactly same analytical results of two methods show their equivalence directly.

3 Experiment testing

R, the ratio of the hadron production cross section via single photon annihilation to the lowest order point-like QED $\mu^+\mu^-$ cross section $\sigma_{pt} = 4\pi\alpha^2/3s$, is a fundamental quantity in e^+e^- interaction. It is calculated in the quark-parton model as $R = 3\sum_q Q_q^2$, where Q_q is the quark electric charge, and the summation runs over all the produced flavors. Neglecting the lowest order QCD correction and the electro-weak effect, in the energy region without any resonances, R is a constant, which equals to 11/3 within the region from 22 GeV to 37 GeV.

In experiment, the ratio of R is calculated according to[1)]

1) The R value measurement at BESII has been described in Refs. [6,7], where the detailed calculation about experiment R value could be found.

$$R = \frac{(N - N_{bg})}{L(1+\delta)\cdot\sigma_{pt}}, \quad (13)$$

where N is the number of multi-hadronic events detected, N_{bg} is the estimated number of background events, L is the integrated luminosity, $\epsilon(1+\delta)$ is the acceptance for the multi-hadronic events with radiative effect included and $(1+\delta)$ is the radiative correction factor due to higher order QED processes up to order α^3. All the quantities on the right side of Eq.(13), except σ_{pt}, contain possible systematic errors. According to the error analysis of the first paper of Ref.[4], the total systematic error is ±3%, where the overall normalizaiton error contributes ±2.4% and ±1.8% is due to the point-to-point error. The relevant R values and corresponding errors are listed in Table 1.

Table 1. Values for R[4]. The errors quoted include the statistical and point-to-point systematic errors.

ord.	E_{cm}/GeV	R value	Error ΔR
1	22.00	4.11	0.13
2	25.01	4.24	0.29
3	27.66	3.85	0.48
4	29.93	3.55	0.40
5	30.38	3.85	0.19
6	31.29	3.83	0.28
7	33.89	4.16	0.10
8	34.50	3.93	0.20
9	35.01	3.93	0.10
10	35.45	3.93	0.18
11	36.38	3.71	0.21
Fitting value		R	ΔR
χ^2 method		3.97475	0.1079
Matrix method		3.97464	0.1079
Ratio R_{χ^2}/R_{matrix}		1.00003	

We fit the average R value from above experiment values by two methods, the concrete expressions read:

$$\chi_f^2 = \sum_i \frac{(fR_{exp}^i - R)^2}{(\Delta R_{exp}^i)^2} + \frac{(f-1)^2}{\sigma_f^2},$$

$$\chi_m^2 = \sum_{ij}(R_{exp}^i - R)(V_m^{-1})_{ij}(R_{exp}^j - R),$$

where σ_f is the overall error of normalization factor f, which equals to 2.4%. In the last equation above, matrix V_m has the similar form as V in Eq.(3), but with the following substitute

$$\sigma_i = \Delta R_{exp}^i, \quad \text{and} \quad x_i = R_{exp}^i.$$

Fig.1 shows the fitting result. The last row of Table 1 gives the ratio of two fitted R values. The almost-one ratio value indicates the equivalence clearly.

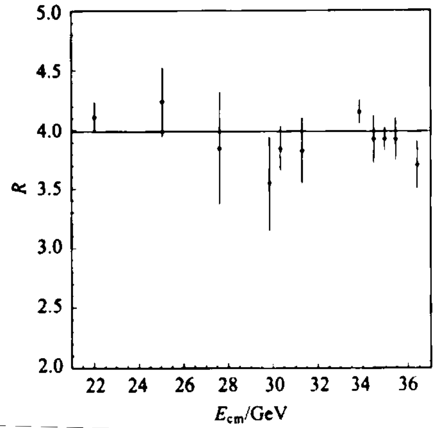

Fig.1. The R value, error bars include the statistical and point-to-point systematic errors. The solid line represents the best fitted R value, data points taken from Ref.[4].

4 Summary

Two χ^2 forms have been constructed to handle correlated data fitting. The equivalence of these two forms has been proved strictly by matrix calculation, and tested quantitatively by a typical simplified R value measurement experiment.

However Eq.(9) shows the existence of a bias for two χ^2 forms, and the deviation from the weighted average $\bar{x}$ may be considerably obvious, if the fitted points are too many, or the error normalization factor is rather large. This kind of bias often produces unexpected results and must be avoided in actual experiment[8,9]. Therefore it is essential to develop a χ^2 form which could deal with correlated data fitting without bias. We will discuss the issue in another paper[10].

Thanks should be given to Prof. C. Z. Yuan, who recommend the useful reference paper on correlation treatment, which becomes the start point of the study of this paper.

References

1 CELLO Collab., Behrend H J et al. Phys. Lett., 1987, **B182**:400—411

2 CERN Library, MINUIT Reference Manual version 92.1 (March 1992)

3 TASSO Collab., Brandelik R et al. Phys. Lett., 1982, **B113**:499—508; TASSO Collab., Brandelik R et al. Z. Phys., 1980, **C4**:87—93

4 JADE Collab., Bartel W et al. Phys. Lett., 1983, **B129**:145—152

5 D'Agostini G. Nucl. Instr. Meth., 1994, **A346**:306—311

6 BAI J Z et al. Phys. Rev. Lett., 2000, **84**:594

7 BAI J Z et al. Phys. Rev. Lett., 2002, **88**:101802

8 Swartz M L. Phys. Rev., 1996, **D53**:5268—5282

9 Takeuchi T. Prog. Thero. Phys. Suppl., 1996, **123**:247—264

10 MO X H, ZHU Y S. Different χ^2 Forms for Correlated Data and Their Properties. submitted to HEP & NP
(莫晓虎,朱永生. 关联实验数据的四种 χ^2 形式及其性质. 高能物理与核物理,待发表)

11 Greub Wener. Linear Algebra. Fourth Edition. Springer-Verlag New York Inc., 1981; Peking University. Advanced Algebra. Second Edition. Beijing: Advanced Education Publishing House, 1988. (in Chinese)
(北京大学. 高等代数. 第二版. 北京:高等教育出版社 1998.)

Appendix: Matrix Formulae

Some necessary matrix formulae are collected in this appendix, as to the knowledge of matrix, see Ref.[11].

For a square matrix $\boldsymbol{A}$, its element is denoted as A_{ij} and determinant as $D_{\boldsymbol{A}} = |\boldsymbol{A}|$. The inverse matrix of $\boldsymbol{A}$, or $\boldsymbol{A}^{-1}$, can be calculated by

$$\boldsymbol{A}^{-1} = \frac{\boldsymbol{A}^{*}}{D_{\boldsymbol{A}}}, \tag{A.1}$$

where $\boldsymbol{A}^{*}$ is the adjoint matrix of $\boldsymbol{A}$, whose element is denoted as A^{*}_{ij}, which is also the cofactor corresponding to A_{ij}. One useful formula involving A_{ij} and A^{*}_{ij} is

$$\sum_{j=1}^{n} A_{ij}A^{*}_{ij} = \delta_{ii}D_{\boldsymbol{A}}, \text{ or } \sum_{i=1}^{n} A_{ii}A^{*}_{ii} = \delta_{ii}D_{\boldsymbol{A}}.$$

According to matrix properties, it can be shown that

$$\begin{vmatrix} a_{11}+\delta & a_{12}+\delta & \cdots & a_{1n}+\delta \\ a_{21}+\delta & a_{22}+\delta & \cdots & a_{2n}+\delta \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+\delta & a_{n2}+\delta & \cdots & a_{nn}+\delta \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \delta\cdot\sum_{i=1}^{n}\sum_{j=1}^{n} A^{*}_{ij}.$$

Let $\delta = 1$, then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A^{*}_{ij} = \begin{vmatrix} a_{11}+1 & a_{12}+1 & \cdots & a_{1n}+1 \\ a_{21}+1 & a_{22}+1 & \cdots & a_{2n}+1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+1 & a_{n2}+1 & \cdots & a_{nn}+1 \end{vmatrix} - \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{A.4}$$

With mathematical induction, the following equation is proved[1)]

$$\begin{vmatrix} x_1+\delta & a_1b_2+\delta & \cdots & a_1b_n+\delta \\ a_2b_1+\delta & x_2+\delta & \cdots & a_2b_n+\delta \\ \vdots & \vdots & \ddots & \vdots \\ a_nb_1+\delta & a_nb_2+\delta & \cdots & x_n+\delta \end{vmatrix} = \left\{1+\sum_{i=1}^{n}\frac{a_ib_i+\delta}{x_i-a_ib_i}+\delta\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{a_ib_i-a_ib_j}{(x_i-a_ib_i)(x_j-a_jb_j)}\right\}\cdot\left[\prod_{i=1}^{n}(x_i-a_ib_i)\right]. \tag{A.5}$$

Therefore, for some special matrices, such as the matrix given in Eq. (A.5), the right side of Eq. (A.4) can be calculated explicitly. Notice the Eq. (A.1), the inverse of matrix could be handled on some extent. And this is the mathematical basis of equivalence proof. For the special error matrix given in Eq. (3), there are following formulae,

$$D_V = \left\{1+\sigma_f^2\cdot\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right\}\cdot\left[\prod_{i=1}^{n}\sigma_i^2\right],$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} V^{*}_{ij} = \left\{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}+\sigma_f^2\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{x_i^2-x_ix_j}{\sigma_i^2\sigma_j^2}\right\}\cdot\left[\prod_{i=1}^{n}\sigma_i^2\right], \tag{A.7}$$

1) The equation has been proved completely by X.H.Mo.

$$= \frac{D_V}{\sigma_k^2} - \frac{\sigma_f^2 x_k^2}{\sigma_k^4} \cdot \left[\prod_{i=1}^{n} \sigma_i^2 \right], \tag{A.8}$$

or

$$= D_V - \frac{\sigma_f^2 x_k^2}{\sigma_k^2} \cdot \left[\prod_{i=1}^{n} \sigma_i^2 \right]. \tag{A.9}$$

In fact, the determinant $|V|$ is just a special form of that given in Eq. (A.5).

两种 χ^2 形式的等价性*

莫晓虎[1,2] 朱永生[1]

1(中国科学院高能物理研究所 北京 100039)

2(中国高等科学技术中心 北京 100080)

摘要 通过矩阵计算证明了两种 χ^2 形式的等价性,同时讨论了最小化估计值的有偏性. 利用简化的 R 值测量定量地检验了等价性的结论.

关键词 等价性 χ^2 形式 有偏性 R 值测量

2002-09-24 收稿

* 国家自然科学基金(19991483)、中国科学院重大项目(KJ95T-03)、中国科学院百人计划(U-25)资助