A viable method for goodness-of-fit test in maximum likelihood fit *

ZHANG Feng(张锋)[1;1)] GAO Yuan-Ning(高原宁)[2,3;2)] HUO Lei(霍雷)[1]

[1] Department of Physics, Harbin Institute of Technology, Harbin 150001, China
[2] Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, Beijing 100084, China
[3] Department of Engineering Physics, Tsinghua University, Beijing 100084, China

Abstract: A test statistic is proposed to perform the goodness-of-fit test in the unbinned maximum likelihood fit. Without using a detailed expression of the efficiency function, the test statistic is found to be strongly correlated with the maximum likelihood function if the efficiency function varies smoothly. We point out that the correlation coefficient can be estimated by the Monte Carlo technique. With the established method, two examples are given to illustrate the performance of the test statistic.

Key words: maximum likelihood fit, goodness-of-fit test, variance

PACS: 02.50.Sk, 02.50.Cw, 02.60.Cb **DOI:** 10.1088/1674-1137/35/6/014

1 Introduction

Supposing the information from an experimental observation can be characterized by a random variable, x, distributed according to a probability density function (p.d.f.) $f(x|\theta)$ with an unknown parameter θ, the unbinned maximum likelihood method is widely used to estimate the parameter from the experimental data. For N independent observations,

$$\boldsymbol{x}^{(\rm obs)}=(x_1^{(\rm obs)},x_2^{(\rm obs)},\cdots,x_N^{(\rm obs)}),$$

the log-likelihood function can be constructed as

$$\ln L(\boldsymbol{x}^{(\rm obs)}|\theta) = \sum_{i=1}^{N}\ln f(x_i^{(\rm obs)}|\theta) - N\ln\int{\rm d}x f(x|\theta). \tag{1}$$

The parameter could be estimated to be $\theta=\hat{\theta}^{(\rm obs)}$, with $\hat{\theta}^{(\rm obs)}$ maximizing the value of the log-likelihood function.

Once an estimation of the parameter is obtained, it is essential to perform a goodness-of-fit test to verify how well the p.d.f. with the obtained parameter could describe the observed data. For binned data, one can obtain a measure of the goodness-of-fit by using the χ^2 method [1, 2], after the maximum likelihood fitting is performed. However, the χ^2 method can't be used in this unbinned case. One possibility, among many tests suggested in the literature, is to use the maximum log-likelihood,

$$\hat{E}_0\equiv\ln L(\boldsymbol{x}|\hat{\theta}), \tag{2}$$

as a goodness-of-fit statistic [3], where $\hat{\theta}$ maximizes the log-likelihood function,

$$\left.\frac{\partial\ln L(\boldsymbol{x}|\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}=0, \tag{3}$$

and the observations here are known to be distributed according to the joint p.d.f.,

$$L(\boldsymbol{x}|\hat{\theta}^{(\rm obs)})=\prod_{i=1}^{N}\frac{f(x_i|\hat{\theta}^{(\rm obs)})}{\int{\rm d}x f(x|\hat{\theta}^{(\rm obs)})}. \tag{4}$$

With the condition given in Eq. (3), the parameter that maximizes the likelihood function is implicitly a function of the observations

$$\hat{\theta}=\hat{\theta}(\boldsymbol{x}). \tag{5}$$

Received 7 September 2010, Revised 22 November 2010
* Supported by National Natural Science Foundation of China (10775077, 10225522)

1) E-mail: zhangfeng8813@gmail.com
2) E-mail: gaoyn@mail.tsinghua.edu.cn

The p.d.f. of $\hat{E}_0$, with the aid of Dirac's δ-function, can be formally expressed as

$$F(\hat{E}_0)=\int \mathrm{d}\boldsymbol{x}\delta[\hat{E}_0-\ln L(\boldsymbol{x}|\hat{\theta}(\boldsymbol{x}))]L(\boldsymbol{x}|\hat{\theta}^{(\mathrm{obs})}). \quad (6)$$

Such a distribution in practice can be obtained by Monte Carlo simulations. $F(\hat{E}_0)$ is then used to define a P-value to characterize the goodness-of-fit test,

$$P\text{-value}=\int_{-\infty}^{\hat{E}_0^{(\mathrm{obs})}} \mathrm{d}\hat{E}_0 F(\hat{E}_0), \quad (7)$$

where

$$\hat{E}_0^{(\mathrm{obs})}=\ln L(\boldsymbol{x}^{(\mathrm{obs})}|\hat{\theta}^{(\mathrm{obs})}). \quad (8)$$

In particle physics experiments, an experimental observation usually comes from one single event collected by a detector, and N such events will contribute a set of N independent observations. Generally, the p.d.f. can be expressed to be

$$f(x|\theta)=\int \mathrm{d}x' R(x,x')f_{\mathrm{t}}(x'|\theta), \quad (9)$$

where $f_{\mathrm{t}}(x'|\theta)$ denoting the p.d.f. usually comes from a theoretical prediction, and $R(x,x')$ is a function that represents the detector response. For a detector with perfect resolution on x,

$$R(x,x')=\varepsilon(x)\delta(x-x') \quad (10)$$

implies

$$f(x|\theta)=\varepsilon(x)f_{\mathrm{t}}(x|\theta), \quad (11)$$

where $\varepsilon(x)$ can be understood as the acceptance function of the detector. The log-likelihood function in Eq. (2) is simplified to be

$$\begin{aligned} E_0(\boldsymbol{x}^{(\mathrm{obs})}|\theta) &= E_1(\boldsymbol{x}^{(\mathrm{obs})})+E_2(\boldsymbol{x}^{(\mathrm{obs})}|\theta), \\ E_1(\boldsymbol{x}^{(\mathrm{obs})}) &= \sum_{i=1}^{N}\ln\varepsilon(x_i^{(\mathrm{obs})}), \\ E_2(\boldsymbol{x}^{(\mathrm{obs})}|\theta) &= \sum_{i=1}^{N}\ln f_{\mathrm{t}}(x_i^{(\mathrm{obs})}|\theta) \\ &\quad +N\ln\int \mathrm{d}x\varepsilon(x)f_{\mathrm{t}}(x|\theta). \end{aligned} \quad (12)$$

Though the detector response is supposed to be clearly known (by Monte Carlo simulations, for example), the detailed expression of $\varepsilon(x)$ is in practice very hard to obtain for x being multi-dimensional. For the E_1 term in Eq. (12) independent of the parameter θ, instead of maximizing E_0, the parameter can be estimated actually by maximizing E_2. Doing it this way, the detailed expression of $\varepsilon(x)$ is not necessary for the parameter estimation. However, without E_1, which needs detailed information about $\varepsilon(x)$, one cannot define $\hat{E}_0^{(\mathrm{obs})}$ according to Eq. (8). It is not possible to perform the goodness-of-fit test in a normal way.

The difficulty presented here is a well-known problem for the goodness-of-fit test of the unbinned maximum likelihood estimation. There is not a good solution so far for this problem. However, for most cases in practice, one may expect that the $\varepsilon(x)$ varies slowly in the allowed phase space of x. The E_1 term in Eq. (12) is then nearly constant, which suggests that E_2 should have strong correlation with E_0. The goodness-of-fit test in this case can be performed by using

$$\hat{E}_2=E_2(\boldsymbol{x}|\hat{\theta}), \quad (13)$$

as a test statistic, where $\hat{\theta}$ maximizes E_2 for a given set of observations. Similar to using $\hat{E}_0$, once the p.d.f of $\hat{E}_2$, $F_2(\hat{E}_2)$ is obtained, the P-value can then be defined to be

$$P\text{-value}=\int_{-\infty}^{\hat{E}_2^{(\mathrm{obs})}} \mathrm{d}\hat{E}_2 F_2(\hat{E}_2), \quad (14)$$

where

$$\hat{E}_2^{(\mathrm{obs})}=E_2(\boldsymbol{x}^{(\mathrm{obs})}|\hat{\theta}^{(\mathrm{obs})}).$$

The quality of such a test, of course, depends on the statistical correlation between $\hat{E}_2$ and $\hat{E}_0$. For a 100% correlation, one expects the test to be be as good as using $\hat{E}_0$. Otherwise, the statistical power of the test will be weakened.

In this article, the properties of the goodness-of-fit test based on $\hat{E}_2$ will be explored. The article is organized as follows. After this introduction, the correlation between $\hat{E}_0$ and $\hat{E}_2$ is examined in Section 2. To evaluate the test, a numerical example is discussed in Section 3 and the application in an example of multivariate analysis is given in Section 4. The last section is devoted to some discussions and a brief conclusion.

2 Correlations between $\hat{E}_0$ and $\hat{E}_2$

Since the parameter $\hat{\theta}$ that maximizes E_2 also maximizes E_0, and E_1 is independent of θ, the relation

$$\hat{E}_0\left(\boldsymbol{x}|\hat{\theta}(\boldsymbol{x})\right)-\hat{E}_2\left(\boldsymbol{x}|\hat{\theta}(\boldsymbol{x})\right)=E_1(\boldsymbol{x}) \quad (15)$$

holds for a given set of N independent observations. To test the goodness-of-fit, $\boldsymbol{x}$ is set to be distributed according to the joint p.d.f given in Eq. (4). Let σ_0^2, σ_2^2 and σ_1^2 denote the variances for $\hat{E}_0$, $\hat{E}_2$ and E_1, respectively, one has

$$\sigma_0^2 = \sigma_1^2+\sigma_2^2+2\sigma_1\sigma_2\cdot\rho', \quad (16)$$

where ρ' is the correlation coefficient between $\hat{E}_1$ and $\hat{E}_2$. It is easy to express the correlation coefficient between $\hat{E}_0$ and $\hat{E}_2$ as

$$\rho \equiv \rho[\hat{E}_0, \hat{E}_2] = \frac{\sigma_2 + \sigma_1 \cdot \rho'}{\sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \cdot \rho'}}. \quad (17)$$

By assuming $\varepsilon(x)$ varying slower than $f_{\rm t}(x|\theta)$ in the phase space of x, i.e., $t \equiv \sigma_2/\sigma_1 \gg 1 > \rho'$, ρ in Eq. (17) can be expressed as

$$\rho = \frac{t+\rho'}{\sqrt{1+t^2+2t\rho'}} \approx \frac{t}{\sqrt{1+t^2}}. \quad (18)$$

So the correlation between $\hat{E}_0$ and $\hat{E}_2$ is governed by t. The statistical power of the goodness-of-fit test based on $\hat{E}_2$ should be similar to that of $\hat{E}_0$ when $t \gg 1$.

To perform the goodness-of-fit test based on $\hat{E}_2$, σ_1 is an important ingredient. An alternative approach is to take the Taylor series of E_1, i.e.,

$$E_1 = \sum_{i=1}^{N} \left[\sum_{j=1}^{\infty} (-1)^{j+1} \frac{1}{j\bar{\varepsilon}} (\varepsilon_i - \bar{\varepsilon})^j \right], \quad (19)$$

where $\bar{\varepsilon}$ is a constant denoting the average efficiency of $\varepsilon(x)$ in phase space of x, i.e.,

$$\bar{\varepsilon} = \int \varepsilon(x) f(x|\theta) {\rm d}x \quad (20)$$

and ε_i denotes $\varepsilon(x_i)$. Since the efficienc $\varepsilon_i \in [0,1]$, the Taylor series is convergent. It is reasonable to drop the higher order terms ($j>2$) in Eq. (19). Thus, E_1 reads as

$$E_1 \approx \left(\frac{1}{\bar{\varepsilon}}+1\right) \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2\bar{\varepsilon}} \sum_{i=1}^{N} \varepsilon_i^2 - N\left(1+\frac{\bar{\varepsilon}}{2}\right). \quad (21)$$

The variance of E_1 can be expressed as

$$\sigma_1^2 = \left(\frac{1}{\bar{\varepsilon}}+1\right)^2 \boldsymbol{V}\left[\sum_{i=1}^{N}\varepsilon_i\right] + \left(\frac{1}{2\bar{\varepsilon}}\right)^2 \boldsymbol{V}\left[\sum_{i=1}^{N}\varepsilon_i^2\right] - \frac{1}{\bar{\varepsilon}}\left(\frac{1}{\bar{\varepsilon}}+1\right)\sqrt{\boldsymbol{V}\left[\sum_{i=1}^{N}\varepsilon_i\right] \cdot \boldsymbol{V}\left[\sum_{i=1}^{N}\varepsilon_i^2\right]}, \quad (22)$$

where

$$\boldsymbol{V}\left[\sum_{i=1}^{N}\varepsilon_i\right] = \boldsymbol{E}\left[\sum_{i=1}^{N}\varepsilon_i^2\right] - \frac{1}{N}\left(\boldsymbol{E}\left[\sum_{i=1}^{N}\varepsilon_i\right]\right)^2, \quad (23)$$

and

$$\boldsymbol{V}\left[\sum_{i=1}^{N}\varepsilon_i^2\right] = \boldsymbol{E}\left[\sum_{i=1}^{N}\varepsilon_i^4\right] - \frac{1}{N}\left(\boldsymbol{E}\left[\sum_{i=1}^{N}\varepsilon_i^2\right]\right)^2. \quad (24)$$

With

$$\varepsilon^{(n)} \equiv \int (\varepsilon(x))^n f_{\rm t}(x|\theta) {\rm d}x \quad (25)$$

denoting the nth-order moment of $\varepsilon(x)$ under $f_{\rm t}(x|\theta)$, it is easy to obtain the expectation of $\sum_{i=1}^{N}\varepsilon_i^j$ ($j = 1,2,\cdots$) under $f(x|\theta)$

$$\boldsymbol{E}\left[\sum_{i=1}^{N}\varepsilon_i^j\right] = \frac{N\varepsilon^{(j+1)}}{\varepsilon^{(1)}} \quad (26)$$

by the definition.

Though the detailed expression of $\varepsilon(x)$ is unknown, some useful information could be extracted from the Monte Carlo events generated. In fact, the moments of $\varepsilon(x)$ can be obtained too. For this purpose, a full detector simulation and the event selection procedure can be performed on the events generated according to $f_{\rm t}(x|\theta)$. The acceptance efficiency is $\varepsilon^{(1)}$. If the process could be done n times, the nth-order moment of $\varepsilon(x)$ under $f_{\rm t}(x|\theta)$, $\varepsilon^{(n)}$, can be obtained.

However, performing the process n times is generally complicated in practice. An iterative computation method can be used to calculate $\varepsilon^{(j)}$ ($j>2$) if $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are known. The first order moment, $\varepsilon^{(1)}$, is the acceptance efficiency, which is easy to obtain. If the process mentioned above is implemented again, the second order moment, $\varepsilon^{(2)}$, is given.

In the large sample limit, $\hat{E}_1$ under $f_{\rm t}(x|\theta)$ approximatively approaches a Gaussian distribution with mean

$$\mu' \equiv N\varepsilon^{(1)}, \quad (27)$$

and variance

$$\sigma'^2 \equiv N\left(\varepsilon^{(2)} - \left(\varepsilon^{(1)}\right)^2\right). \quad (28)$$

Then one can obtain

$$N\varepsilon^{(n)} = \Gamma^{(n)} - \sum_{n_1+n_2+\cdots+n_N=n} \frac{n!}{\prod_i n_i!} \prod_i \varepsilon^{(n_i)}, \quad (29)$$

where n_i is the exponent of ε_i, and $n_i < n$. $\Gamma^{(n)}$ is the nth-order moment of Gaussian distribution $N(\mu', \sigma')$.

With $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ obtained by Monte Carlo, the expectation values of $\sum_{i=1}^{N}\varepsilon_i^j$ are straightforward and the variance of $\hat{E}_1$, σ_1^2, can be calculated too.

3 An illustrative example

We consider a simple univariate case in experiment. In the decay $e^+e^- \to \mu^+\mu^-$, the angular distribution of μ^+ follows

$$f(\cos\theta;\alpha) = 1 + \alpha\cos^2\theta, \quad (30)$$

where θ is the angle between e^+ and μ^+ and $\alpha = 1.0$ predicted by theory.

The efficiency of the detector, $\varepsilon(\cos\theta)$, is assumed to be described by a quadratic function,

$$\varepsilon(\cos\theta) = -\lambda\cos^2\theta + \lambda', \tag{31}$$

where λ, λ' are smoothing parameters and $0 < \lambda < \lambda' < 1$. The efficiency function $\varepsilon(\cos\theta)$ implies the detector preferring events with $|\cos\theta|$ close to zero and the maximum efficiency is λ' in the phase space.

Monte Carlo simulation is used to evaluate the test proposed in Section 2. For the purpose, events are generated according to Eq. (30), and Eq. (31) is used to perform the selection process on the events generated. The input values of the parameters are fixed as

$$\begin{aligned} &\alpha = 1.0, \\ &\lambda = 0.01, 0.04, 0.07, 0.10, 0.13, 0.16, \\ &\lambda' = 0.3. \end{aligned} \tag{32}$$

5000 experiments are simulated by Monte Carlo and for each experiment $N = 10000$ events are generated. Those events are used to estimate the parameter α with maximum likelihood estimators E_0 and E_2. Maximum values of the log-likelihood function, $\hat{E}_0$ and $\hat{E}_2$ are recorded.

$\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ can be obtained in an experiment simulated by Monte Carlo. Then the variance of E_1, σ_1^2, can be estimated using the method proposed in Section 2. On the other hand, the distribution of E_1 simulated from the 5000 Monte Carlo experiments can also give a "true" σ_1^2. Comparing σ_1 estimated to σ_1 simulated from the 5000 Monte Carlo experiments for different smoothing parameter λ is shown in Fig. 1. It is obvious that the estimation of σ_1^2 is good. Besides, $\sigma_2^2 = 22$ can also be obtained from the distribution of $\hat{E}_2$ simulated by the 5000 Monte Carlo experiments.

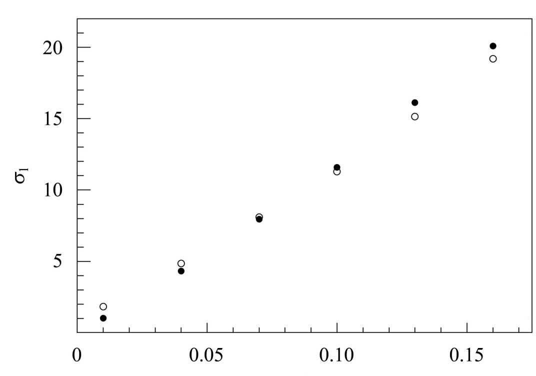

Fig. 1. Comparison of σ_1 estimated with the "true" σ_1 for different smoothing parameter λ.

The distributions of $\hat{E}_0$, E_1 and $\hat{E}_2$ for $\lambda = 0.04$ are shown in Fig. 2, and the scatter-plots of $\hat{E}_2$ versus $\hat{E}_0$ for $\lambda = 0.04$ are shown in Fig. 3. Here, $t = 21.9/4.3 = 5$ and $\rho = 0.98$. It is obvious that $\hat{E}_2$ can be taken as a goodness-of-fit test.

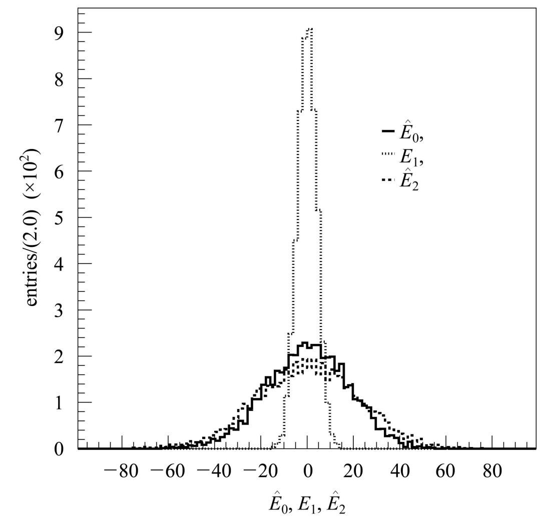

Fig. 2. The distributions of $\hat{E}_0$, E_1 and $\hat{E}_2$ simulated by Monte Carlo.

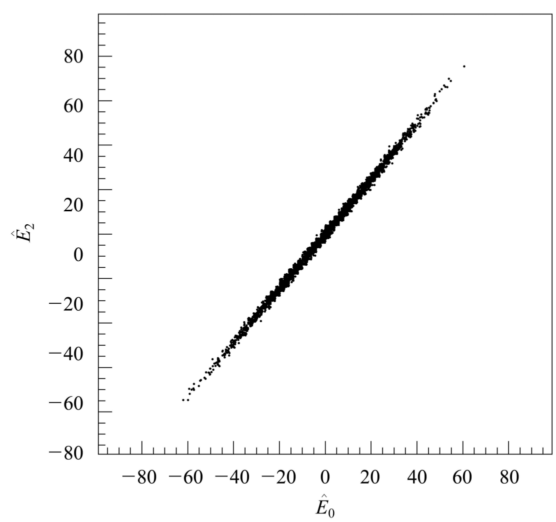

Fig. 3. The scatter-plots of $\hat{E}_2$ versus $\hat{E}_0$.

4 Application in multivariate condition

We consider a complicated multivariate case in experiment. For the cascade decay $e^+e^- \to J/\psi$, $J/\psi \to \Lambda\bar{\Lambda}$, $\Lambda \to p\pi^-$, $\bar{\Lambda} \to \bar{p}\pi^-$, the full angular distribution reads [4]

$$\frac{d\sigma}{d\Omega d\Omega_1 d\Omega_2} \propto \mathcal{F}(\Theta, \Phi; \theta_1, \phi_1; \theta_2, \phi_2)$$

$$\equiv \sum_{M\lambda_p\lambda_{\bar{p}}} |\beta_{\lambda_p}|^2 |\bar{\beta}_{\lambda_{\bar{p}}}|^2 \sum_{\lambda_1\lambda_2\lambda_1'\lambda_2'} \alpha_{\lambda_1\lambda_2}\alpha^*_{\lambda_1'\lambda_2'}$$

$$\times\ \mathrm{e}^{\mathrm{i}[(\lambda_1-\lambda_1')\phi_1+(\lambda_2-\lambda_2')\phi_2]} d^1_{M\lambda_1-\lambda_2}(\Theta) d^1_{M\lambda_1'-\lambda_2'}(\Theta)$$

$$\times\ d^{\frac{1}{2}}_{\lambda_1\lambda_\mathrm{p}}(\theta_1) d^{\frac{1}{2}}_{\lambda_1'\lambda_\mathrm{p}}(\theta_1) d^{\frac{1}{2}}_{\lambda_2\lambda_{\bar{\mathrm{p}}}}(\theta_2) d^{\frac{1}{2}}_{\lambda_2'\lambda_{\bar{\mathrm{p}}}}(\theta_2), \qquad (33)$$

where M, λ_1, λ_2, λ_p, $\lambda_{\bar{p}}$ denote respectively the helicities of J/ψ, Λ, $\bar{\Lambda}$, p and $\bar{\mathrm{p}}$. Θ, Φ are the emission angles of Λ versus e^+ beam direction $\hat{z}$. $\theta_1,\phi_1(\theta_2,\phi_2)$, measured in $\Lambda(\bar{\Lambda})$ rest frame, are polar and azimuth angles of $\mathrm{p}(\bar{\mathrm{p}})$ versus $\Lambda(\bar{\Lambda})$ momentum direction. The explicit expression of $d^{\mathrm{J}}_{\mathrm{mm'}}(\beta)$ can be found, for example, in Ref. [5].

$\alpha_{\lambda_1\lambda_2}$ and $\beta_{\lambda_\mathrm{p}}$, $\bar{\beta}_{\lambda_{\bar{\mathrm{p}}}}$ are complex numbers. With CPT and C invariance in the decay process, the free parameter $\alpha_{\lambda_1\lambda_2}$ can be re-parameterized as

$$\alpha_{\frac{1}{2}\frac{1}{2}}=\mathrm{e}^{\mathrm{i}\frac{\eta}{2}}, \quad \alpha_{-\frac{1}{2}-\frac{1}{2}}=\mathrm{e}^{-\mathrm{i}\frac{\eta}{2}},$$

$$\alpha_{\frac{1}{2}-\frac{1}{2}}=\alpha_{-\frac{1}{2}\frac{1}{2}}=\sqrt{\frac{2(1+\alpha)}{1-\alpha}}, \qquad (34)$$

where η, α are real, and $\alpha=0.65\pm0.11$ (stat) $\pm$ 0.03 (syst) has been obtained by experimental measurement [6]. $\beta_{\lambda_\mathrm{p}}$ and $\bar{\beta}_{\lambda_{\bar{\mathrm{p}}}}$ can be re-parameterized as

$$\beta_{\frac{1}{2}}=\sqrt{1+\beta}, \quad \beta_{-\frac{1}{2}}=\sqrt{1-\beta},$$

$$\bar{\beta}_{\frac{1}{2}}=\sqrt{1-\bar{\beta}}, \quad \bar{\beta}_{-\frac{1}{2}}=\sqrt{1+\bar{\beta}}, \qquad (35)$$

with β, $\bar{\beta}$ being real and $\beta\approx\bar{\beta}=0.642\pm0.013$ [5].

The efficiency of the detector for a track with momentum $\boldsymbol{p}$ and polar angle θ in the laboratory frame, $\varepsilon(\boldsymbol{p},\cos\theta)$, is assumed to be described by a quadratic function,

$$\varepsilon(\boldsymbol{p},\cos\theta)=-(\epsilon_0-\epsilon_1\cdot|\boldsymbol{p}|)\cos^2\theta+\epsilon_2, \qquad (36)$$

where ϵ_0, ϵ_1, ϵ_2 are smoothing parameters. There are four particles in the final states of the decay process. Thus the efficiency of the detector for the decay process is

$$\varepsilon = \varepsilon(\boldsymbol{p}_\mathrm{p},\cos\theta_\mathrm{p})\varepsilon(\boldsymbol{p}_{\bar{\mathrm{p}}},\cos\theta_{\bar{\mathrm{p}}})$$

$$\times\varepsilon(\boldsymbol{p}_{\pi^+},\cos\theta_{\pi^+})\varepsilon(\boldsymbol{p}_{\pi^-},\cos\theta_{\pi^-}), \qquad (37)$$

where $\boldsymbol{p}_{\mathrm{p},\bar{\mathrm{p}},\pi^+,\pi^-}$ and $\theta_{\mathrm{p},\bar{\mathrm{p}},\pi^+,\pi^-}$ denote respectively the momenta and polar angles of $\mathrm{p},\bar{\mathrm{p}},\pi^+,\pi^-$ in the laboratory frame, respectively.

Events are generated according to Eq. (33), and Eq. (37) is used to perform the selection process on events generated. The input values of the parameters are fixed as

$$\alpha=0.65,\ \beta=\bar{\beta}=0.64,\ \eta=0.0,$$

$$\epsilon_0=0.04,\ \epsilon_1=0.04,\ \epsilon_2=0.8. \qquad (38)$$

1000 experiments are simulated by Monte Carlo and for each experiment, $N=10000$ events are generated. Those events are used to estimate the parameter η with maximum likelihood estimators E_0 and E_2. The scatter-plot of $\hat{E}_2$ versus $\hat{E}_0$ is shown in Fig. 4. Here, $t=34/1.2=28$ and $\rho=1.0$. It is obvious that $\hat{E}_2$ can be taken as a goodness-of-fit test.

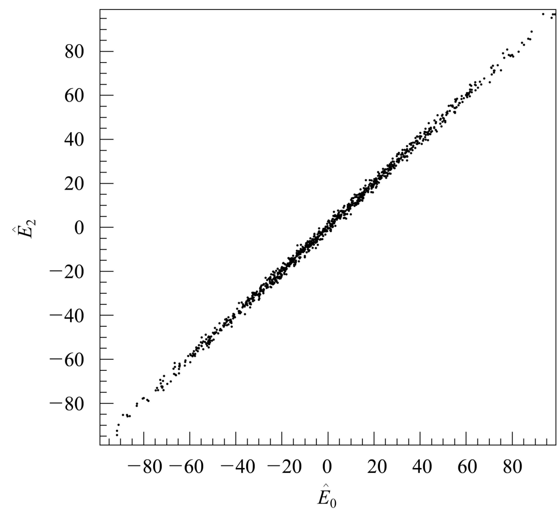

Fig. 4. The scatter-plot of $\hat{E}_2$ versus $\hat{E}_0$.

5 Discussions and conclusion

A test statistic is proposed to perform the goodness-of-fit test in the unbinned maximum likelihood fit. Without an explicit expression of the efficiency function, the test statistic is found to be strongly correlated with the maximum likelihood function if we can assume the efficiency of detector varying slower than pdf in the phase space. We point out that the correlation coefficient can be estimated by the Monte Carlo technique.

References

1 Hoel P G. Ann. Math. Stat., 1938, **9**: 158
2 Baker S, Cousins R D. Nucl. Instrum. Methods, 1984, **221**: 437
3 Cowen G. Statistical Data Analysis. Oxford: Clarendon Press, 1998. 89
4 ZHANG Feng, GAO Yuan-Ning, HUO Lei. Phys. Lett. B, 2009, **681**: 237
5 Amsler C et al. (Particle Data Group). Phys. Lett. B, 2008, **667**: 1
6 Ablikim M et al. (BES collaboration). Phys. Lett. B, 2006, **632**: 181