A study of the methods for signal significance estimation in ground-based gamma-ray detectors

Yuncheng Nan [*]

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, CAS, Beijing 100049, China.
University of Chinese Academy of sciences, Beijing 100049, China.
E-mail: nanyc@ihep.ac.cn

Songzhan Chen [†]

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, CAS, Beijing 100049, China.

Searching for new astrophysical sources is an important physical goal of current and nearly further ground-based gamma-ray detectors, such as HAWC, LHAASO and CTA. The gamma-ray signal appears as an excess number of events. The significance of the excess is estimated using statistical method and is used to determine whether to accept the signal or not. Up to now, four main statistical methods, including the classical Li-Ma method, the weighted method, the likelihood method and the chi-squared method, have been used in the majority of ground-based extensive air shower (EAS) experiments, such as Tibet ASγ, MILAGRO, ARGO-YBJ and HAWC. While these methods are not well tested for the accuracy in previous works. In this work, we develop a standard to test the accuracy of different methods basing on different experimental conditions. We find some problems for some methods, which have been already used in experiment data analysis. Finally, we obtain the limitation of different methods and provide application conditions for different methods, respectively. Additionally, we also test the superiority of different statistical methods and provide our best recommendations for different physical analyses.

35th International Cosmic Ray Conference - ICRC2017
10-20 July, 2017
Bexco, Busan, Korea

---

1. Introduction

Very high energy (VHE) gamma-rays are a powerful probe for astrophysics and fundamental physics under extreme conditions. Thanks to the success of ground-based detectors, about two hundred VHE gamma-ray sources have been detected in the past three decades [1]. Several categories of VHE gamma-ray emitters have been firmly established: active galactic nuclei (AGNs), pulsar wind nebulae (PWNs), supernova remnants (SNRs), X-ray binaries (XBs), and starburst galaxies. Searching for new VHE sources is still an important physical goal of the planned ground-based gamma-ray detectors, e.g. LHAASO and CTA, which are expected to boost the number of VHE sources by a factor of 10.

The gamma-ray emission from a source would appear as an excess number of events coming from the directions of candidate over the cosmic ray background. To quantify the confidence level of the observed excess as a source, the statistical significance should be estimated in a hypothesis test. The method to estimate the statistical significance is a crucial issue for ground-based detector to search for new sources. A better method could improve the detector sensitivity. e.g. Miguel et al.[1] has adopted the Gaussian weighting technique to improve the sensitivity by 10% compared with the traditional optimal binned analysis. While a wrong method may underestimate the detector sensitivity or lead to fake source, e.g., Li& Ma [2] found the incorrectness of early methods and then proposed the famous Li-Ma formulae. Besides excess signals, the significance of deficit signal are also need to be estimated sometimes, e.g., the Sun and Moon shadow.

Up to now, four main methods have been adopted by the majority of ground-based EAS experiments. The classical Li-Ma method is widely used in many experiments, including ASγ, ARGO-YBJ, and so on. According to [2], the equation (9) can only be used when the number of background is large enough, and the optimal equation (17) also limit the observed counts($N_{off} \gtrsim 10, N_{on} \gtrsim 10$) and the time ratio($\alpha = 0.1 - 10$). To improve the sensitivity, weighted methods are adopted by MILAGRO and ARGO-YBJ experiments. MILAGRO adopted a extension of the Li-Ma formula (17) [3], while, ARGO-YBJ adopted a extension of the Li-Ma formula (9) [4] to calculate the significance. Recently, HAWC adopted the likelihood method to calculate the significance [5], and ARGO-YBJ also proposed the chi-squared method. These methods have been used for data analysis, however, they don't suffer strict evaluation as that in [2]. For example, the uncertainty of the estimated background is ignored in the likelihood method in [5], which is not strict in the statistics and will affect the correctness more or less. In this paper, we will adopt Monte Carlo method to test the correctness and also the superiority of these methods.

The paper is organized as follows. In Section 2 a whole process of testing the significance estimation is built. The results are reported in Section 3. In Section 4 some problems arising from the use of these methods are pointed out and the best recommendations for different physical analyses are also showed in this section. A conclusion is given in Section 5.

2. Method

2.1 Building the sky map

In the observation, the detector points in the direction of a suspected source for a certain time

---

[1]http://tevcat.uchicago.edu/ (Version: 3.400, as of 2017 July).

t_{on} and counts N_{on} photos, and then it turns for background measurement for a time interval t_{off} and counts N_{off} photos. The quantity α is the ratio of the on-source time t_{on} to the off-source time t_{off}[2]. On the basis, the sky map with $15^o \times 15^o$ size and $0.1^o \times 0.1^o$ bin size in celestial coordinates(right ascension and declination) is built.

To gain the sample of observations (N_{on}, N_b) in each bin to bulid the "N_{on} map" and "N_b map", the Toy Monte Carlo simulation is applied. First, the expected value of the background number $\langle N_b \rangle$ in each bin is spread considering the photon probability distributions of the events. The angular resolution σ is 1° in this work. Moreover, the "N_{on} map" is divided into three kinds according to different sources. If only the background events were detected, we gain the number of events N_{on} in each bin from the Poisson distribution with expectation $\langle N_b \rangle$. When the excess source was detected, two orthogonal normal distributions which the mean value is -0.8^o and -0.1^o separately is added to the map. For the deficit source, "N_{on} map" is gained through the "N_{on} map" with only the background events minus 1 in the corresponding positions of the source events. Figure 1 provides the "N_{on} map" with only the background and the "N_b map".
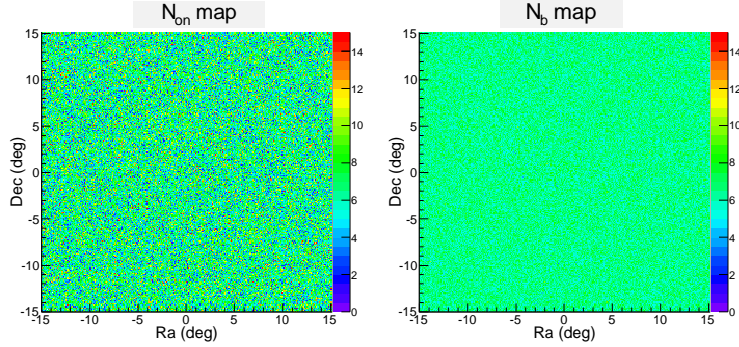

Figure 1: The "N_{on} map" with only the background and the "N_b map".

The smoothing procedure is applied on the "N_{on} map" and the "N_b map" to obtain the significance map. The point spread function(PSF) which is a Gaussian shape requires the bins in the maps to be "integrated" over a circular area. Each bin is filled with the content of all the surrounding bins whose center is closer than 1.58σ. Then we obtain the smoothed "N_{on} map" and the smoothed "N_b map". The statistical significance in each bin is computed in the next section.

2.2 Significance estimation

2.2.1 The classical Li-Ma method

The classical Li-Ma method proposes three formulae to evaluate the statistical significance of an observational result. The Li-Ma formula (5) is given by simply defining the significance S as a radio of the number of signal events to its standard deviation:

$$S = \frac{N_{on} - \alpha N_{off}}{\sqrt{N_{on} + \alpha^2 N_{off}}}, \tag{2.1}$$

Assuming that the observed signal is due only to the background events, the Li-Ma formula (9) is proposed:

$$S = \frac{N_{on} - \alpha N_{off}}{\sqrt{\alpha(N_{on} + N_{off})}}, \tag{2.2}$$

2

The Li-Ma formula (17) evaluates the significance by using the method of statistical hypotheses test. The likelihood function is given by

$$L = P(N_{on}, N_{off} | < N_s >, < N_b >), \tag{2.3}$$

where P is the probability density function(PDF) of N_{on} and N_{off}. Both of them satisfy the Poisson distribution. The methods of the estimate of the unknown parameter $\langle N_s \rangle$ and $\langle N_b \rangle$ has been mentioned in the first two formulae. Then the likelihood ratio test is performed. The significance is expressed as follows:

$$S = \sqrt{2} \left\{ N_{on} \ln \left[\frac{1+\alpha}{\alpha} \left(\frac{N_{on}}{N_{on} + N_{off}} \right) \right] + N_{off} \ln \left[(1+\alpha) \left(\frac{N_{off}}{N_{on} + N_{off}} \right) \right] \right\}^{1/2}. \tag{2.4}$$

2.2.2 The weighted method

Only the Gaussian-shaped PSF be used as the weight here. That is,

$$\omega(r) = \frac{1}{2\pi\sigma^2} e^{-r^2/(2\sigma^2)}, \tag{2.5}$$

where r is the space angle to the position of the possible source and the σ is the angular resolution. The events in a circular area centered on each bin whin an angular radius of 2σ are summed after weighting. Then the significances can be obtained by substituting the weighted N_{on} and N_{off} into the Li-Ma formulae[3, 4].

2.2.3 The likelihood method

The significance is gained through the method of statistical hypotheses test. The likelihood function is given by

$$L = \prod_i^{ROI} \frac{\lambda_i^{N_i} e^{-\lambda_i}}{N_i!}, \tag{2.6}$$

where N_i and λ_i are the observed and expected signals in the ith bin. The λ is the number of background events N_b plus N_s which is the number of the signal in a bin. The area of the region of interest(ROI) is rounded with the radius of 3σ. Using the maximum likelihood technique the parameter $\langle N_s \rangle$ is obtained by using the MINUIT package. The likelihood ratio test is applied and the significance is obtained.

Besides,we call the $0.1^o \times 0.1^o$ bins "2D bin". In order to gain more data and information, the bins in the ROI is regrouped. The ROI is divided according to the distance from the space angle to the central bin. The width of the annular bin is 0.1^o and we call it "1D bin".

2.2.4 The chi-squared method

In the statistical hypotheses test, the chi-square value is expressed as follows:

$$\chi^2 = \sum_i^{ROI} \left(\frac{N_i - \lambda_i}{\sigma_i} \right)^2, \tag{2.7}$$

where the σ_i is same as the standard deviation of the N_s in the equation (2.2). The N_{on_i}, λ_i and the ROI are same as the values mentioned in the likelihood method. Using the minimum chi-squared technique the parameter $\langle N_s \rangle$ is also obtained by using the MINUIT package. The significance in each bin is evaluated with chi-square difference test.

3

2.3 Inspection standard and comparison method

If the null hypothesis $\langle N_s \rangle = 0$ is true, in other words, if all the counts come from the background, the Significance of the observed result follow a standard normal distribution. Based on this theory, the relative error between the distribution of significance of the Toy Monte Carlo samples and the standard normal distribution is applied to test the accuracy of the estimation methods of the significance. If the relative error is less than 20%, the method is judged to be accurate enough. For different α and $\langle N_b \rangle$, approximately 5×10^7 independent samples are generated to test the accuracy of the first two methods and 1.058×10^7 independent samples for the last two methods.

Ensuring that those methods are accurate, a positive source or a deficit source is added to the "N_{on} map" to test the superiority of different methods. The significance map is gained from the "N_{on} map" and "N_b map". We compared the significance of the center of the source in different methods. Each procedure of sampling and the evaluating was repeated 10^4 times.
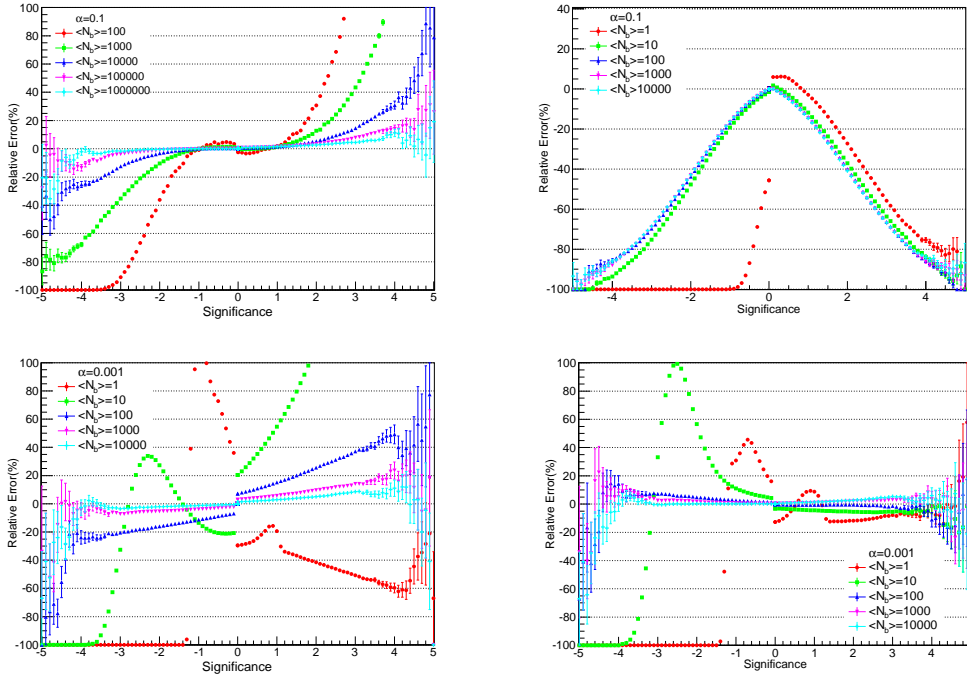
3. Results

Figure 2: The relative errors as a function of the significance. Upper part of the figure: the samples are generated from the equation (2.2) in the classical Li-Ma method and the weighted equation (2.4) (left and right plot, respectively). Lower part: the "2D bin" and the "1D bin" of the likelihood method are used(left and right plot, respectively).

In Figure 2, the equation (2.2) obviously overestimates the statistical significance for the positive excess and underestimates for the deficit when the $\langle N_b \rangle < 1 \times 10^5$. The weighted equation (2.4) is a poor estimator of the significance both for the excess and the deficit sources. In the left plot of the lower part, the "2D bin" in the likelihood method can't be applied when the value of

$\langle N_b \rangle$ is small. To expand the scope of the use of the likelihood method, we proposed the "1D bin". The "1D bin" increases the number of background events $\langle N_b \rangle$ in each source bin. Under the same experimental conditions, the likelihood method with the "1D bin" is available for smaller $\langle N_b \rangle$ both for the excess and the deficit sources.

Table 1: Application conditions for different methods for α=0.001, 0.01 and 0.1 respectively.

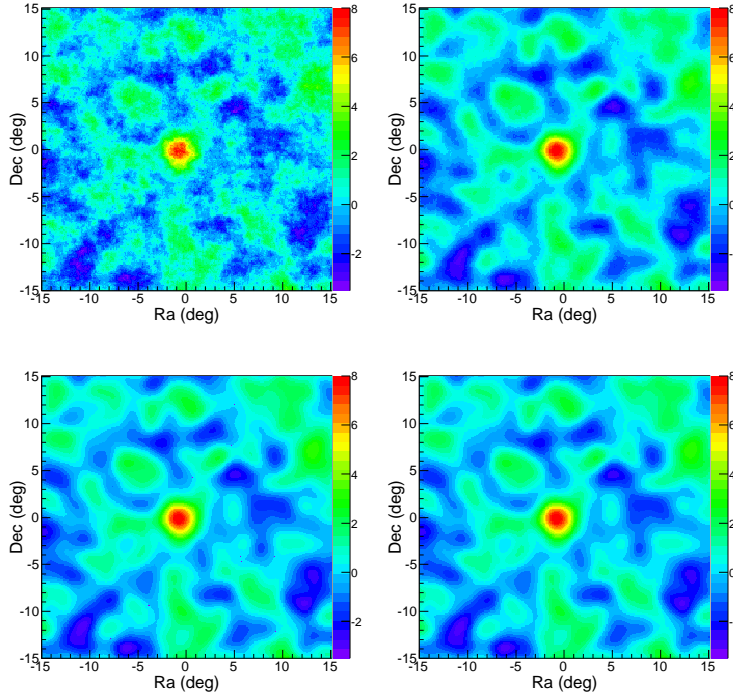| Method		Range of $\langle N_b \rangle$(S<0)	Range of $\langle N_b \rangle$(S>0)
The classical Li-Ma method	Eq.(2.1)	$\geq$100000	$\geq$10000
	Eq.(2.2)	$\geq$10000	$\geq$10000
	Eq.(2.4)	$\geq$100	$\geq$100,10,1
The weighted method	Eq.(2.1)	$\geq$100000	>10000
	Eq.(2.2)	>10000	>10000
	Eq.(2.4)	$\times$	$\times$
The likelihood method	"2D bin"	$\geq 1000, 100000, \times$	$\geq 1000, 100000, \times$
	"1D bin"	$\geq 100, 100, \times$	$\geq 1, 100, \times$
The chi-squared method		>10000	>10000

Figure 3: Significance maps of the same positive source in different estimation methods. Upper part of the figure: source is estimated through the classical Li-Ma method and the weighted method(left and right plot, respectively). Lower part: the likelihood method and the chi-squared method are used(left and right plot, respectively).

After discussing all the methods basing on different experimental conditions, the application

conditions are shown in the Table 1. It is noteworthy that the equation (2.4) in the classical Li-Ma method and the situation of the "1D bin" in the likelihood method ($\alpha \neq 0.1$) have wider ranges of application than other methods. On the contrary, the weighted equation (2.4) is unavailable. Similarly, the likelihood method can't be applied when the $\alpha = 0.1$. Other methods have relatively narrow ranges as showed in the Table 1.

By following the methods described above, we can obtain the maps of the positive source represented in the Figure3. The number of the $\langle N_b \rangle$ is 10^5 and the $\alpha = 0.001$. Except for the map which estimated by the classical Li-Ma method, all the maps have clearer and more complete contours of the source. For further comparison of those methods, the increases of the significance compared with the classical Li-Ma method are shown in Table 2. The statistical significances calculated by the weighted method, the likelihood method and the chi-squared method are about 10% bigger than the result from the classical Li-Ma method.

Table 2: The increases of the significance compared with the classical Li-Ma method.

Source	The weighted method	The likelihood method	The chi-squared method
S>0	$(9.76 \pm 0.30)\%$	$(12.91 \pm 0.31)\%$	$(11.33 \pm 0.30)\%$
S<0	$(9.97 \pm 0.30)\%$	$(12.25 \pm 0.30)\%$	$(10.40 \pm 0.30)\%$

4. Discussion

When the α=0.1, the applicable range of the $\langle N_b \rangle$ in equation (2.4) in the classical Li-Ma method is not less than 1 for the positive source. The results are in agreement with the results from Li and Ma[2]. For the deficit source, the applicable condition of the $\langle N_b \rangle$ in equation (2.4) is given explicitly ($\langle N_b \rangle \geq 100$). For all the three formulas, the α is used in a wider range which is as low as 0.001 compared with that from Li and Ma.

The weighting technique is used in the equation (2.4) leads to the unavailability of the method the MILAGRO experiment applied. The normalization technique can be added to improve the limitation of the application. For another two formulas in the weighted method, approximately 10% improvement in the significance over the classical Li-Ma method.

The likelihood method proposed by HAWC experiment can't be used when the $\alpha = 0.1$. In this method, the error of the estimated background $\langle N_b \rangle$ is neglect in this method. When the α is big relatively, the difference between the estimated background $\langle N_b \rangle$ and the truth value of $\langle N_b \rangle$ becomes larger. This makes the method incorrect. Besides, "1D bin" is proposed to effectively improve the applicable range of $\langle N_b \rangle$. The same question also appeared in the chi-squared method. The method can't be used when the $\langle N_b \rangle$ is too small ($\langle N_b \rangle$>10000) due to the neglect of the error of $\langle N_b \rangle$. Both the likelihood and the chi-squared method have approximately 10% increase in significance compared with the classical Li-Ma method.

Owing to the different time scales of the gamma-ray and the deficit sources, the values of $\langle N_b \rangle$ are different. Based on the data analysis results of the ARGO-YBJ experiment, all the values of $\langle N_b \rangle$ are in a range of 1 to 10^8. In order to choose the more accurate and superior method to estimate the significance in different physical analyses, the best recommendations are proposed as follows: (1) For the sources which have the small value of $\langle N_b \rangle$, such as the GRBs, the classical

Li-Ma method is more accurate, fast and convenient. The likelihood method is another choice because of its' slightly higher significance, but it is restricted by the application conditions($\alpha<0.01$, $\langle N_b \rangle>100$). (2) If the $\langle N_b \rangle$ of the sources is more than 10^4, a faster and more superior method is the weighted method. The likelihood method and the chi-squared method can also be used according to their application conditions. (3)When the expansion of the source can't be ignored, conversely, the likelihood method and the chi-squared method are the good estimator of the significance.

5. Conclusion

All the methods of estimating the statistical significance of the signals have been tested. The application conditions of each method are presented explicitly. The weighted equation (2.4) used by MILAGRO experiment can't be applied directly. And the likelihood method adopted by HAWC experiment is accurate only when α is far less than one. Besides, the equation (2.4) in the classical Li-Ma method and the improved "1D bin" likelihood method has a wider applicable range of $\langle N_b \rangle$. The comparisons between the four methods are also presented explicitly within the limitation of accuracy. The weighted method, the likelihood method, and the chi-squared method have an about 10% increase in significance from the classical Li-Ma method. Based on the application conditions and the comparison among methods, we provide our best recommendations for different physical analyses to choose the methods. The classical Li-Ma method and the weighted method are the best choices when the value of $\langle N_b \rangle$ is small and large separately. The likelihood method and the chi-squared method are used when considering the expansion of the source.

References

[1] M. F. Morales, D. A. Williams, T. Deyoung, *An advanced analysis technique for transient searches in wide-field gamma-ray observatories*, Astropart. Phys..

[2] T. P. Li, Y. Q. Ma, *Analysis methods for results in gamma-ray astronomy*, ApJ.

[3] A. A. Abdo, B. T. Allen, T. Aune et al, *Observation and spectral measurements of the crab nebula with MILAGRO*, ApJ.

[4] B. Bartoli, P. Bernardini, X. J. Bi et al, *TeV gamma-ray survey of the northern sky using the ARGO-YBJ detector*, ApJ.

[5] H. Zhou, C. M. Hui, P. Huentemeyer, *TeV observations of the galactic plane with HAWC and joint analysis of GeV data from Fermi*, in proceedings of the 34th ICRC.

PoS(ICRC2017)879