Combining upper limits with a Bayesian approach *

YANG Liu(杨柳)[1,2] ZHU Kai(朱凯)[2;1)] ZHU Yong-Sheng(朱永生)[2] CAI Hao(蔡浩)[1;2)]

[1] Department of Physics, School of Physical Science and Technology, Wuhan University, 430072 Wuhan, China
[2] Institute of High Energy Physics, Chinese Academy of Science, PO Box 918(1), Beijing 100049, China

Abstract: We discuss how to determine and combine upper limits based on observed events and estimated backgrounds with a Bayesian method, when insignificant signals are observed in independent measurements. In addition to some general features deduced from the analytical formulae, systematic numerical results are obtained by a C++ program (CULBA) for low-count experiments, which can be used as a reference to combine two upper limits.

Key words: upper limit, Bayesian method, combine results

PACS: 02.50.Cw, 02.50.Tt, 02.70.Pt **DOI:** 10.1088/1674-1137/39/12/123001

1 Introduction

How to combine independent results to extract the most information appropriately is a crucial problem for experimentalists. Some statistical methods have been proposed [1], and the Bayesian method is a promising one [2]. Its basic idea is simple; the probability density function (PDF), $f(\mu|x)$, of a truth parameter μ deduced from an observational parameter x is read as [3]:

$$f(\mu|x)=\frac{f(x|\mu)f_0(\mu)}{\int f(x|\mu)f_0(\mu)\mathrm{d}\mu}, \tag{1}$$

in which $f_0(\mu)$, named the prior, is the degree of belief attributed to μ before observation, and $f(\mu|x)$, called the posterior, corresponding to the prior, is the updated likelihood that μ will produce the observed effect x.

The Bayesian method has the advantage of combining results since it provides a natural means to include additional knowledge by adding nuisance parameters, compared with other statistics methods [4]. However, an obvious weakness of this method is that its posterior depends on the choice of the prior. Even for a uninformative prior, there are different proposals such as uniform prior, Haldane prior [5], Jeffreys prior [6], reference prior [7] and so on [8]. How to select an appropriate prior is a kind of art. This situation is even worse for rare processes and small signals, although that has been discussed somewhat in depth and methods based on the spirit of the Bayesian method have been developed [9–11]. Here experimentalists face a double risk of missing a real signal or ruining physics sensitivity. In this paper, we show that, analytically and numerically, by using the first experiment result as the prior for the second one to combine two upper limits will improve this situation significantly. We then partially solve this problem.

2 Combining two probability density functions

As a starting point, let us consider a counting measurement on the number of events in a small signal. In the signal region, x events are observed, that is a Poisson random variable with average value $\lambda_\mathrm{S}+\lambda_\mathrm{B}$, where λ_S and λ_B are the expected numbers of signal and background events respectively. λ_S is the signal parameter that one wants to infer, while λ_B is a nuisance parameter, which could be estimated by background regions or theoretical predictions. From the spirit of the Bayesian method, it is natural to deduce the probability of λ_S signal [12]

$$\begin{aligned} &f(\lambda_\mathrm{S}|x,f_0(\lambda_\mathrm{S},\lambda_\mathrm{B}))\\ &=\frac{\int \mathrm{e}^{-(\lambda_\mathrm{S}+\lambda_\mathrm{B})}(\lambda_\mathrm{S}+\lambda_\mathrm{B})^x f_0(\lambda_\mathrm{S},\lambda_\mathrm{B})\mathrm{d}\lambda_\mathrm{B}}{\iint \mathrm{e}^{-(\lambda_\mathrm{S}+\lambda_\mathrm{B})}(\lambda_\mathrm{S}+\lambda_\mathrm{B})^x f_0(\lambda_\mathrm{S},\lambda_\mathrm{B})\mathrm{d}\lambda_\mathrm{S}\mathrm{d}\lambda_\mathrm{B}}. \end{aligned} \tag{2}$$

Usually the priors of signal and background are independent, i.e. $f_0(\lambda_\mathrm{S},\lambda_\mathrm{B})=f_0(\lambda_\mathrm{S})f_0(\lambda_\mathrm{B})$. Suppose there are two experiments examining the same physics signal λ_S, with observation x_1 and x_2 respectively. As pointed out by D'Agostini [12] (section 6.3), by applying Bayes'

Received 30 January 2015, Revised 21 July 2015

∗ Supported by National Natural Science Foundation of China (NSFC) (11005115), Joint Funds of the NSFC and Chinese Academia Sinica's Big Scientific Apparatus (U1232109)

1) E-mail: zhuk@ihep.ac.cn

2) E-mail: hcai@whu.edu.cn

theorem a second time, i.e. using the posterior of the first experiment as the prior for the second one, we obtain the final posterior PDF for $\lambda_{\rm S}$, by which the final result for the $\lambda_{\rm S}$ of two experiments can be inferred. However, in general $\lambda_{\rm S}$ is related to a parameter $\mathcal{B}$, which is the quantity to be measured, by an experimental factor h via $\lambda_{\rm S}=h\mathcal{B}$. For example, in e^+e^- collision experiments, if we want to measure a branching ratio $\mathcal{B}$ for a decay $R\to f$, then $h=\mathcal{L}\epsilon$, where $\mathcal{L}$ and ϵ are the accumulated luminosity and the detection efficiency for $R\to f$ signal events, respectively. Defining h_1 and h_2 are the experimental factors of the two experiments respectively, and replacing $f_0(\lambda_{\rm S})$ in Eq. (2) with $f(\lambda_{\rm S}|x_1,f_0(\lambda_{\rm S},\lambda_{\rm B1}))$, after some derivation we get

$$f(\mathcal{B}|x_2,f_0(\lambda_{\rm S2},\lambda_{\rm B_2}))=\frac{\iint {\rm e}^{-[(h_1+h_2)\mathcal{B}+\lambda_{\rm B_1}+\lambda_{\rm B_2}]}(h_1\mathcal{B}+\lambda_{\rm B_1})^{x_1}(h_2\mathcal{B}+\lambda_{\rm B_2})^{x_2}f_0(\lambda_{\rm S1})f_0(\lambda_{\rm B_1})f_0(\lambda_{\rm B_2}){\rm d}\lambda_{\rm B_1}{\rm d}\lambda_{\rm B_2}}{\iiint {\rm e}^{-[(h_1+h_2)\mathcal{B}+\lambda_{\rm B_1}+\lambda_{\rm B_2}]}(h_1\mathcal{B}+\lambda_{\rm B_1})^{x_1}(h_2\mathcal{B}+\lambda_{\rm B_2})^{x_2}f_0(\lambda_{\rm S1})f_0(\lambda_{\rm B_1})f_0(\lambda_{\rm B_2}){\rm d}\lambda_{\rm S1}{\rm d}\lambda_{\rm B_1}{\rm d}\lambda_{\rm B_2}}. \quad (3)$$

Here $f_0(\lambda_{\rm S1})$ is only the prior of the first experiment, and it can be set to a uniform shape $f_0(\lambda_{\rm S1})=k$ (k is a constant) to indicate there is totally no knowledge before this measurement. For simplicity we normalize this formula to the number of the signal of the second experiment (renamed as $\lambda_{\rm S}$):

$$f(\lambda_{\rm S}|x_2,f_0(\lambda_{\rm S},\lambda_{\rm B_2}))=\frac{\iint {\rm e}^{-[(g+1)\lambda_{\rm S}+\lambda_{\rm B_1}+\lambda_{\rm B_2}]}(g\lambda_{\rm S}+\lambda_{\rm B_1})^{x_1}(\lambda_{\rm S}+\lambda_{\rm B_2})^{x_2}f_0(\lambda_{\rm B_1})f_0(\lambda_{\rm B_2}){\rm d}\lambda_{\rm B_1}{\rm d}\lambda_{\rm B_2}}{\iiint {\rm e}^{-[(g+1)\lambda_{\rm S}+\lambda_{\rm B_1}+\lambda_{\rm B_2}]}(g\lambda_{\rm S}+\lambda_{\rm B_1})^{x_1}(\lambda_{\rm S}+\lambda_{\rm B_2})^{x_2}f_0(\lambda_{\rm B_1})f_0(\lambda_{\rm B_2}){\rm d}\lambda_{\rm S}{\rm d}\lambda_{\rm B_1}{\rm d}\lambda_{\rm B_2}}, \quad (4)$$

where $g\equiv h_1/h_2$ is the normalization factor, that represents the ratio between the experimental factors of the two experiments. $g\ll 1$ or $g\gg 1$ means one experiment is much more sensitive than the other one. Due to our study, in this kind of situation, the final result will be dominated only by the more sensitive experiment. So what we really care about is the situation with $g\approx 1$. We will set $g=1$ in the following derivation and calculation, and discuss $g\neq 1$ later.

To illustrate Eq. (4) further, we assume we know the background very well then simplify $f_0(\lambda_{\rm B})=\delta(\lambda_{\rm B}-m_{\rm B})$ (δ is the Dirac delta function and $m_{\rm B}$ is the expected background). Then Eq. (4) can be rewritten as:

$$f(\lambda_{\rm S}|x_2,f_0(\lambda_{\rm S},\lambda_{\rm B2}))=\frac{{\rm e}^{-(2\lambda_{\rm S}+m_{\rm B1}+m_{\rm B2})}(\lambda_{\rm S}+m_{\rm B1})^{x_1}(\lambda_{\rm S}+m_{\rm B2})^{x_2}}{\int {\rm e}^{-(2\lambda_{\rm S}+m_{\rm B1}+m_{\rm B2})}(\lambda_{\rm S}+m_{\rm B1})^{x_1}(\lambda_{\rm S}+m_{\rm B2})^{x_2}{\rm d}\lambda_{\rm S}}\propto {\rm e}^{-(2\lambda_{\rm S}+m_{\rm B1}+m_{\rm B2})}(\lambda_{\rm S}+m_{\rm B1})^{x_1}(\lambda_{\rm S}+m_{\rm B2})^{x_2}. \quad (5)$$

From Eq. (5), it is easy to see in this simplified case that the posterior of the combined results of two measurements is just proportional to the product of the posterior of each one. We can then infer that a prior distribution with a more pronounced peak at $\lambda_{\rm S}=0$ will produce a more stringent posterior if combined with the same secondary experiment. This means that a more accurate measurement will play a dominant role in the combined result. Furthermore, from Eq. (3) or (5) it is obvious that if we switch the sequence of any two experiments, i.e. change the prior candidate for the other, the final result does not change. This feature is natural and intuitive, and is advantageous compared with the Serialization method [13], whose results depend on the sequence of the combined experiments.

3 Numerical illustration

Now let us consider a more practical situation, in which the estimated backgrounds are assumed to satisfy a Gaussian distribution whose mean value $m_{\rm B}$ is the same as that in the previous δ function while its standard deviation $\sigma_{\rm B}$ depends on the uncertainty of the estimation method. Eq. (4) will be transformed into a more complex function, so we composed a program, named CULBA, based on C++ and ROOT's [14] built-in functions to implement the integration in Eq. (2) and later calculations as well as plotting. To simplify plots and discussion, we set $g=1$ during the numerical illustration, without losing the general features of the results as discussed in Section 2. Here we choose the numerical method instead of analytical expressions such as that used in Ref. [15], because the numerical method will make our program more flexible to handle more types of priors in case the exact formulae are missing or very complex. To illustrate the functions of this program and its basic idea, we suppose there are three independent measurements, I, II and III, and list the data sets of the numbers of observed (x) and backgrounds ($\lambda_{\rm B}$) in the signal region respectively in Table 1. When the observed number of events x is not significantly larger than the expected backgrounds, just the upper limits of these measurements are determined. Figure 1 shows the posterior and their cumulative distributions with the uniform prior of the three data sets; the upper limits are determined at 90% credible level (C.L.). As we mentioned before, we can use one result as an in-

put prior to calculate the posterior of the other one, as shown with Eq. (4); we refer to this as a "transfer prior".

Upper limit results of each single measurement and the combined ones obtained by our method are listed in Tables 1 and 2, respectively. From these, two further intuitive features of the upper limit combination based on the Bayesian method are deduced: 1) a more stringent final result is expected when the results of two measurements are combined, especially if these two measurements are at the same precision level; 2) if one measurement is much more precise than the other, then the final combined result depends dominantly on the more precise one. For comparison we use an "experimentally practical" method, which is widely applied in high energy physics analyses such as Refs. [16, 17], to calculate upper limits and combine any two results. All the results based on the these methods are listed in Tables 1 and 2 for comparison, and it turns out the "experimental practical" method provides similar results to ours for both single and combined upper limits.

Further than this simple illustration, a systematic study with different combinations of general conditions of experiments are implemented by using CULBA. For

Table 1. The three data sets and upper limits (UL) at 90% credible level based on Bayesian method with uniform prior (uni). For comparison, the "experimentally practical" method (exp) is also applied to calculate the upper limits at 90% credible level.

Mea.	I $(x/\lambda_B/\sigma_B)$	II $(x/\lambda_B/\sigma_B)$	III$(x/\lambda_B/\sigma_B)$
	16/16/4	9/9/3	1/4/2
UL(uni)	10.38	8.05	3.09
UL(exp)	10.56	8.24	2.99

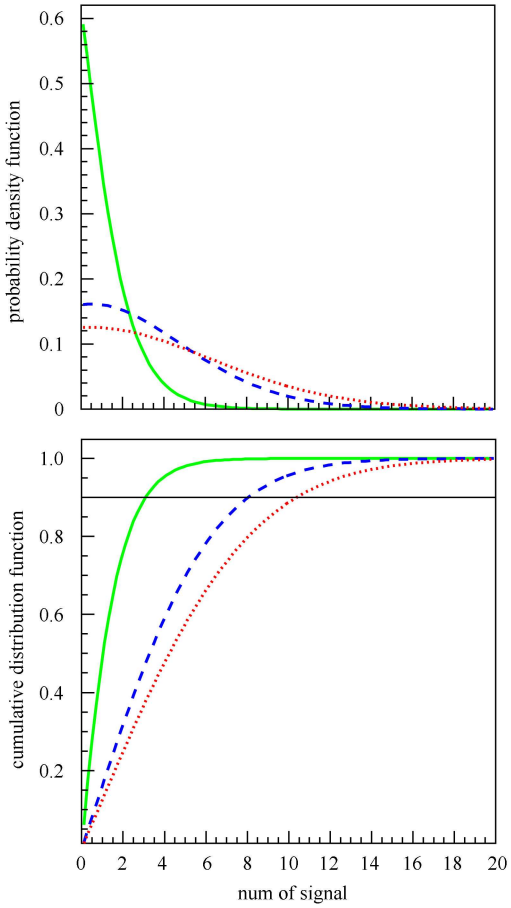

Fig. 1. (color online) PDF and cumulative distribution functions of sets I, II and III, where the uniform prior is applied. The dotted, dashed and solid lines correspond to sets I, II and III respectively. The input values of these sets are shown in Table 1.

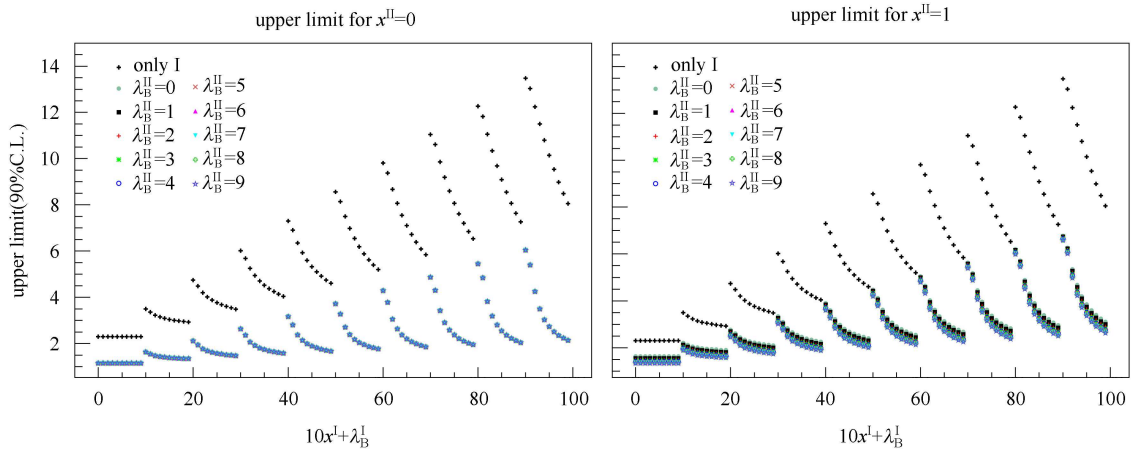

Fig. 2. (color online) Upper limits at 90% C.L. with $x^{II}=0$ and $x^{II}=1$. In each plot, $10\times10\times10=1000$ different combined conditions are considered, where the observation x^{I} and expected background λ_B^{I} of experiment I, and the expected background λ_B^{II} of experiment II are varied from 0 to 9 respectively. The upper limit of a single experiment is also provided for comparison.

two experiments I and II, we vary their observations x and expected backgrounds $\lambda_{\rm B}$ from 0 to 9 with a step 1 respectively, while the uncertainty of the background is taken as 1 when $\lambda_{\rm B}=0$ or $\sqrt{\lambda_{\rm B}}$ when $\lambda_{\rm B}\geqslant 1$ for simplicity. Experiment I uses the uniform prior, then its posterior PDF is used as the prior input to Experiment II. The combined upper limits at 90% C.L. for the $(10\times10)^2=10000$ different combinations are calculated, and they are shown in Figs. 2–6 classified by the observation $x^{\rm II}$ of experiment II. In each plot, the y-axis is the upper limit at 90% C.L. and the x-axis is the experimental condition type, i.e. $10x^{\rm I}+\lambda_{\rm B}^{\rm I}$. The upper limits for each single experiment I are also drawn in these plots for comparison.

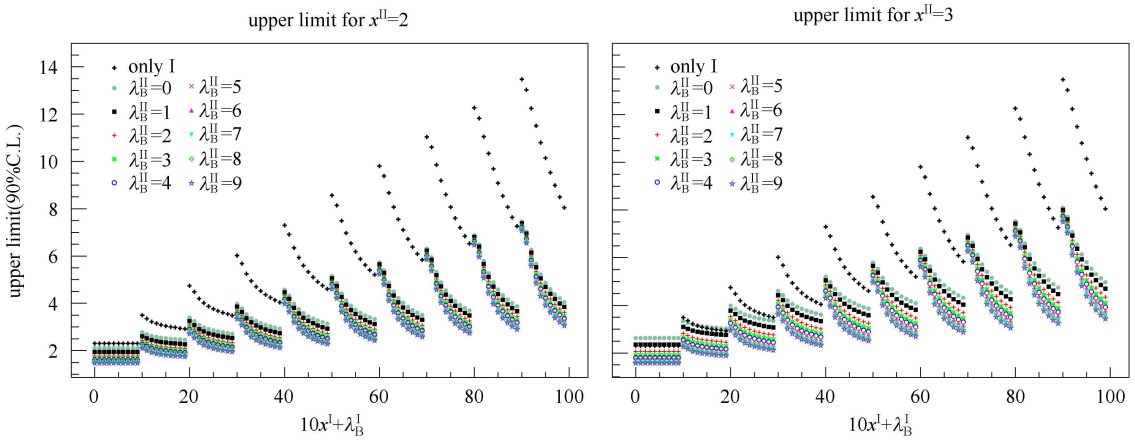

Fig. 3. (color online) Upper limits at 90% C.L. with $x^{\rm II}$=2 and $x^{\rm II}$=3.

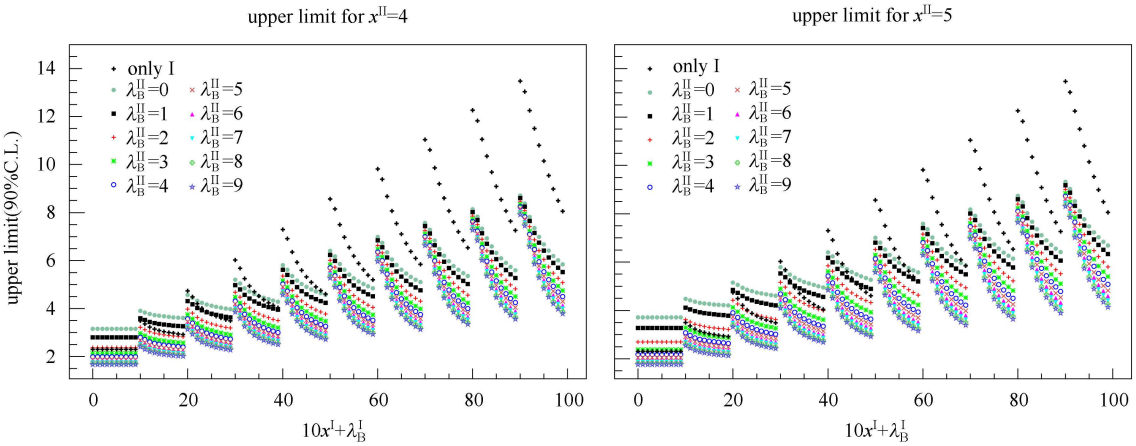

Fig. 4. (color online) Upper limits at 90% C.L. with $x^{\rm II}$=4 and $x^{\rm II}$=5.

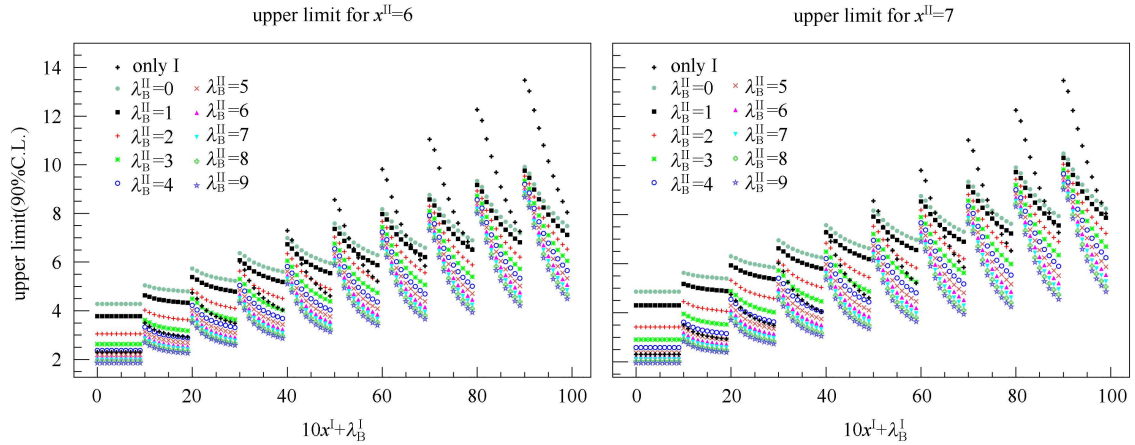

Fig. 5. (color online) Upper limits at 90% C.L. with $x^{\rm II}$=6 and $x^{\rm II}$=7.

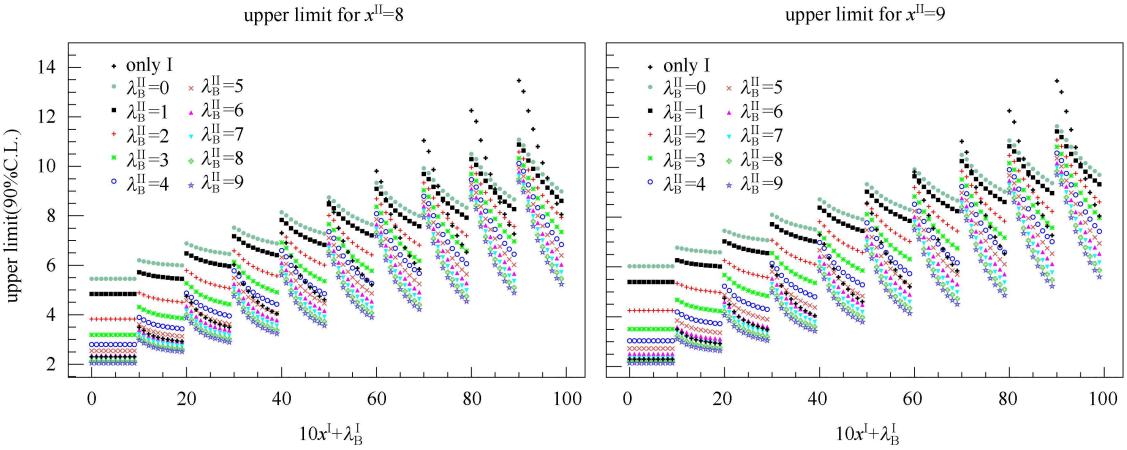

Fig. 6. (color online) Upper limits at 90% C.L. with x^{II}=8 and x^{II}=9.

Table 2. The three data sets and combined upper limits (UL) at 90% credible level based on the Bayesian method with transfer prior (trans) are presented. For comparison, the "experimentally practical" method (exp) is also applied to calculate the upper limits at 90% credible level.

Mea.	I $(x/\lambda_B/\sigma_B)$	II $(x/\lambda_B/\sigma_B)$	III$(x/\lambda_B/\sigma_B)$
	16/16/4	9/9/3	1/4/2
Com.	I+II	II+III	III+I
UL(tra)	6.28	2.77	2.88
UL(exp)	6.98	2.89	2.96

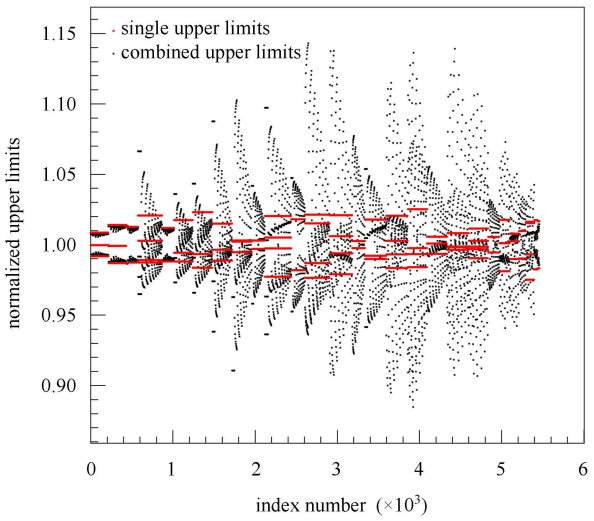

Fig. 7. (color online) Comparison of the upper limits of single experiments I (represented by red solid circles) and their combinations with experiment II (represented by grey stars) after grouping. The results are normalized to the average of each group.

4 Discussion

From the numerical results, all four expected features mentioned in Section 2, i.e. "switchable", "transitive", "improvable", and "dominant", are observed. But we should notice that sometimes the combined upper limit will be larger than the result from a single measurement. That happens only if the observation is large and background is small, for example x=5 while λ_B=0. This situation just indicates a significant signal is observed and an upper limit claim is not proper anymore. A Bayesian method is available to deal with these conditions and is able to provide combined mean values instead of upper limits, but that is out of the scope of this paper so is not discussed here. We also notice not only the relative but the absolute numbers of signal and backgrounds are meaningful, because they will lead to different PDF shapes then different combination results. Another interesting topic is the dependence of likelihood shapes, i.e. when some individual experiments with different observations and backgrounds give similar upper limits, will their combinations with another experiment give similar results too? To study this, we divided the experiments I into different groups with respect to their single upper limits by requiring the difference between any two experiments to be less than 5% inside each group. For instance, assuming there are seven experimental conditions with upper limits of 1.00, 1.02, 2.01, 2.05, 4.23, 4.37, and 4.44 respectively, we divide them into three groups (1.00, 1.02), (2.01, 2.05), (4.23, 4.37, 4.44) by requiring the difference between any two experiments in one group to be less than 5%. Then within each group, we combined experiments I with experiments II to obtain the combined upper limits. After that, all the single and combined upper limits are normalized to the average of the group they belong to respectively. The results are shown in Fig. 7 for comparison, where the results from all groups are included. Figure 7 shows that when the difference of individual experiments is limited to 5%, mostly the difference of combined results are within 10% while the largest difference reaches 25%. That means in

a situation where only the upper limits are known but without the detailed information for the signal and background, we can still use this method to get a reasonable combined upper limit with a larger uncertainty. This conclusion is also suitable for the situation with the normalization factor $g\neq1$ but $g\approx1$. Here, the upper limit of the first experiment should be normalized to the second one by considering the experimental factors such as luminosities and efficiencies, then an approximate combined result can be obtained with this method.

All the relevant numerical results can be used as a reference to combine two experimental results appropriately. They are saved to a plain text file and uploaded to the arxiv server as "other formats" [18].

Notice that we only discussed combining two measurements in this article, but with the Bayesian method it is simple and easy to expand to any number of measurements with a combination chain, where each result in the previous step will be used as an input prior for the next step.

References

1 Proceedings of Workshop on Confidence Limits. Edited by James F, Lyons L, Perrin Y. CERN, Geneva, Switzerland, 17-18 January 2000. http://cds.cern.ch/record/411537?ln=en
2 D'Agostini G. Rept. Prog. Phys., 2003, **66**: 1383–1420
3 Bayes R T. Phil. Trans. Roy. Soc. Lond, 1764, **53**: 370–418
4 Olive K A et al. (Particle Data Group collaboration). Chin. Phys. C, 2014, **38**: 090001
5 Haldane J B S. Mathematical Proceedings of the Cambridge Philosophical Society, 1932, **28**: 55–61
6 Jeffreys H. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1946, **186**(1007): 453–461
7 Bernardo J M. J. Royal Statistical Society B, 1979, **41**: 113–147; Reprinted in Bayesian Inference 1 (Tiao G C, Polson N G, eds). Oxford: Edward Elgar, 229–263
8 Jaynes E T. IEEE Trans. on Systems Science and Cybernetics, 1968, **4**: 227
9 Narsky I V. Nucl. Instrum. Methods A, 2000, **450**: 444 [hep-ex/9904025]
10 Escoubes B, Unamuno S De, Helene O. Nucl. Instrum. Methods A, 1987, **257**: 346
11 ZHU Y. Nucl. Instrum. Methods A, 2007, **578**: 322
12 D'Agostini G. Bayesian reasoning in data analysis: A critical introduction. New Jersey, USA: World Scientific, 2003. 167
13 Yellin S. arXiv:1105.2928 [physics.data-an]
14 Rene Brun, Fons Rademakers. Nucl. Instrum. Methods. in Phys. Res. A, 1997, **389**: 81–86. http://root.cern.ch/
15 Loparco F, Mazziotta M N. Nucl. Instrum. Methods A, 2011, **646**: 167
16 Ablikim M et al. (BESIII collaboration). Phys. Rev. D, 2014, **90**(11): 112014
17 Ablikim M et al. (BESIII collaboration). Phys. Rev. D, 2015, **91**(3): 032002
18 YANG L, ZHU K, ZHU Y S, CAI H. arXiv:1307.3804 [physics.data-an]