Including systematic uncertainties in confidence interval construction for Poisson statistics

J. Conrad, O. Botner, A. Hallgren, and C. Pérez de los Heros

Division of High Energy Physics, Uppsala University, S-75121 Uppsala, Sweden

(Received 30 January 2002; published 10 January 2003)

One way to incorporate systematic uncertainties into the calculation of confidence intervals is by integrating over probability density functions parametrizing the uncertainties. In this paper we present a development of this method which takes into account uncertainties in the prediction of background processes and uncertainties in the signal detection efficiency and background efficiency, and allows for a correlation between the signal and background detection efficiencies. We implement this method with the likelihood ratio (usually denoted as the Feldman-Cousins) approach with and without conditioning. We present studies of coverage for the likelihood ratio and Neyman ordering schemes. In particular, we present two different types of coverage tests for the case where systematic uncertainties are included. To illustrate the method we show the relative effect of including systematic uncertainties in the case of the dark matter search as performed by modern neutrino telescopes.

I. INTRODUCTION

A limit on, or a measurement of, a physical quantity at a given confidence level is usually set by comparing a number of detected events, n_o, with the number of expected events from the known background sources contributing to the physical process in question, n_b. How "compatible" these numbers are determines how much room there is for new processes, i.e., for a signal. How well the number of observed events and expected background compare strongly depends on the systematic uncertainties present in the measurement. Systematic uncertainties must, therefore, be taken into account in the limit or confidence belt calculation that is finally published.

Traditionally, confidence limits are set using a Neyman construction [1]. This is a purely frequentist method. Feldman and Cousins [2] have proposed an improved method to construct confidence intervals based on likelihood ratios, a method already known in statistics and originally described in [3]. Still, this method is based on the original Neyman construction, and needs to be extended to incorporate systematic uncertainties in the measurement. Along this line, a modification of the Neyman method that incorporates systematic uncertainties in the experimental signal efficiency has been proposed by Highland and Cousins [4]. These authors use a "semi-Bayesian" approach where an average over the probability distribution of the experimental sensitivity (and its uncertainty) is performed. By construction, the method is of limited accuracy in the limit of high relative systematic uncertainties.

Recently, an entirely frequentist approach has been proposed for the uncertainty in the background rate prediction [5]. That approach is based on a two-dimensional confidence belt construction and likelihood ratio hypothesis testing and treats the uncertainty in the background as a statistical uncertainty rather than as a systematic one.

The interest aroused recently in the high energy physics community about the many open issues on setting limits and quoting confidence levels is stressed by the organization of workshops devoted to the subject. We refer the reader to the proceedings of the recent workshops at CERN [6], Fermilab

[7], and Durham [8] for a review of the status of the field.

In this paper we extend the method of confidence belt construction proposed in [4] to include systematic uncertainties in both the signal and background efficiencies as well as theoretical uncertainties in the background prediction. The proposed method also allows us to use newer ordering schemes. A recent attempt to include systematic uncertainty in the background prediction in a similar manner has been presented in [9]. The paper is organized as follows. In Sec. II we give a short review of the confidence belt construction schemes that we will use. In Sec. III we describe how to include the systematic uncertainties; in Sec. IV we discuss how the confidence belt construction is performed and present some selected results. We compare the results of this method with other methods to include systematics in Sec. V. We introduce the tests of coverage performed in Sec. VI and present an example based on data from the Antarctic Muon and Neutrino Detector Array (AMANDA) neutrino experiment in Sec. VII.

II. THE CONSTRUCTION OF CONFIDENCE INTERVALS

The frequentist construction of confidence intervals is described in detail elsewhere [10]. Here we will give just a short review.

Let us consider a Poissonian probability density function (PDF) $p(n)_{s+b}$ for a fixed but unknown signal s in the presence of a known background with mean b. For every value of s we can find two values n_1 and n_2 such that

$$\sum_{n'=n_1}^{n_2} p(n')_{s+b} = 1-\alpha \qquad (1)$$

where $1-\alpha$ denotes the confidence level [usually quoted as a $100(1-\alpha)\%$ confidence interval]. Since we assume a Poisson distribution, the equality will generally not be satisfied exactly. A set of intervals $[n_1(s+b,\alpha), n_2(s+b,\alpha)]$ is called a *confidence belt*. Graphically, upon a measurement n_o the *confidence interval* $[s_1, s_2]$ is determined by the intersection of the vertical line drawn from the measured value n_o and the boundary of the confidence belt. This is illustrated in
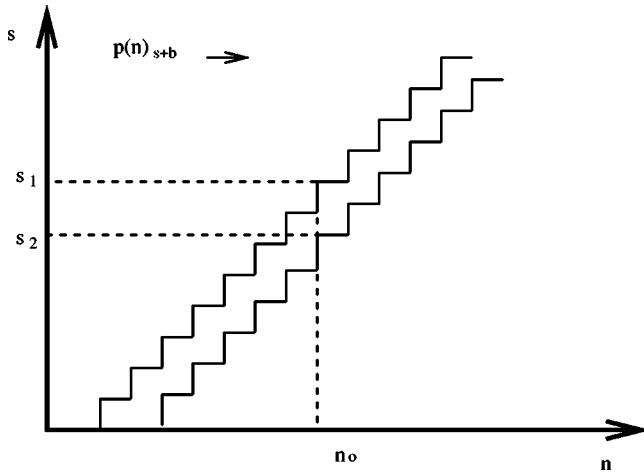
FIG. 1. Illustration of the confidence belt construction. On the x axis are the possible experimental outcomes (number of events), on the y axis the parameter of the PDF (s). In this case a Poisson PDF was assumed.

Fig. 1. The probability that the confidence interval will contain the true value s is $1-\alpha$, since this is true for all s per construction. The choice of the n_1 and n_2 to define the confidence belt is, however, not unique. An additional criterion has to be applied. The choices originally proposed by Neyman [1] are

$$\sum_{n'=0}^{n_1} p(n')_{s+b} = \sum_{n'=n_2}^{\infty} p(n')_{s+b} = \frac{1-\alpha}{2} \qquad (2)$$

for central confidence intervals, and

$$\sum_{n'=0}^{n_1} p(n')_{s+b} = 1-\alpha \qquad (3)$$

for upper confidence limits. This method presents certain drawbacks in the case of small samples and, in particular, can yield null results (in the sense that the algorithm gives no answer) in the case when no events have been observed. Also, the decision to quote a measurement (that is, a central confidence interval) or an upper limit might not be straightforward before performing an experiment.

A. Likelihood ratio ordering

To solve this problem, a modification of the Neyman method has been proposed [3,2] that is based on a more rationalized ordering scheme of the elements in the sum in Eq. (1), based on likelihood ratios. This approach automatically provides central confidence intervals when motivated and upper limits when necessary; therefore it is often denoted as the "unified approach." Instead of using the choices given in the previous section, the following ordering scheme is applied in solving Eq. (1).

For each n the s_{best} is found which maximizes the likelihood $\mathcal{L}(n)_{s+b}$. In the case of a simple Poissonian distribution with known background, s_{best} is given by $\max(0,n-b)$. Then for a fixed s the ratio

$$R(s,n)_{\mathcal{L}} = \frac{\mathcal{L}_{s+b}(n)}{\mathcal{L}_{s_{best}+b}(n)} \qquad (4)$$

is computed for each n, and all n's are consequently ranked according to the value of this ratio. Values of n are included in the confidence belt starting with the n with the highest rank (largest $R_{\mathcal{L}}$) and then decreasing rank until $\sum_{n=n_1}^{n_2} p(n)_{s+b} = 1-\alpha$. After the confidence belt has been constructed in this way, the confidence interval $[s_1, s_2]$ is found as described in the previous section. Note that this ordering principle is a standard method within the theory of likelihood ratio tests [3].

This approach has some undesired features as well. There is a background dependence of the upper limit in the case of fewer events observed than expected from background. This can lead to situations where measurements with higher background give a better limit, a clearly undesirable effect. Roe and Woodroofe [11] proposed a solution to this problem which we briefly describe next.

B. Conditioning

A variation of the classical method of constructing confidence belts is to use the fact that, given an observation n_o, it is known that the background cannot have been larger than n_o itself. To incorporate this knowledge into the PDF, the authors in [11] proposed the following modification:

$$q_{s+b}^{n_o}(n) = \begin{cases} \dfrac{p(n)_{s+b}}{\sum\limits_{n'=0}^{n_o} p(n')_b} & \text{if} \quad n \leqslant n_o, \\[1em] \dfrac{\sum\limits_{n'=0}^{n_o} p(n')_b\, p(n-n')_s}{\sum\limits_{n'=0}^{n_o} p(n')_b} & \text{if} \quad n > n_o. \end{cases} \qquad (5)$$

The likelihood ratio ordering can then be applied with this new PDF. Note that in this case the PDF is dependent on the number of observed events. This approach solves the background dependence of the upper limit: a limit set when no events are observed stays constant at a value of 2.44 independent of the expected background (which agrees with the result of the original likelihood ordering for no events observed and no expected background). However, this method does not satisfy all the requirements of proper coverage [12] and has problems when applied to the case of a Gaussian distribution with boundaries [13]. An extension based on a Bayesian approach with tests of coverage can be found in [14].

III. THE INCLUSION OF SYSTEMATIC UNCERTAINTIES

The way of incorporating systematic uncertainties into the confidence belt construction presented in this paper does not affect the particular ordering scheme. Instead, it takes into

account the systematic uncertainties by assuming (or if possible determining) a PDF which parametrizes our knowledge about the uncertainties and integrating over this PDF. It has been noted that averaging over systematic uncertainties in itself is a Bayesian approach [4]. Therefore the method presented is refered to as ''semi-Bayesian,'' combining classical and Bayesian elements. We return to this point in Sec. VI. Usually, uncertainties are assumed to be described by a Gaussian distribution, which we will adopt for the remainder of this paper. The implementation, however, makes it easy to use other parametrizations for the uncertainties.

We will refer in the following to the parameters with systematic uncertainties also as *nuisance parameters*.

Two examples of how the PDF is modified if systematic uncertainties are present are the following. In the case that the only uncertainty present is a theoretical uncertainty of the background process the PDF is modified to

$$q(n)_{s+b} = \frac{1}{\sqrt{2\pi}\sigma_b} \int_0^\infty p(n)_{s+b'} e^{-(b-b')^2/2\sigma_b^2} db'. \quad (6)$$

Here b is the estimated background level, and σ_b is the uncertainty in the background estimation. If, in addition to the theoretical uncertainty for background, there is the need to include the uncertainty in the signal detection efficiency the expression for $q(n)_{s+b}$ might be extended to

$$q(n)_{s+b} = \frac{1}{2\pi\sigma_b\sigma_\epsilon} \int_0^\infty \int_0^\infty p(n)_{b'+\epsilon's}$$

$$\times e^{-(b-b')^2/2\sigma_b^2} e^{-(1-\epsilon')^2/2\sigma_\epsilon^2} db' d\epsilon' \quad (7)$$

where σ_ϵ is the uncertainty in the detection efficiency expressed in *relative* terms with respect to the nominal efficiency. It is important to realize that the integration variables, here ϵ' and b', are the possible ''true'' (but unknown) values of the nuisance parameter. This indicates that this method is based on Bayesian statistics.

IV. POLE: A GENERAL ALGORITHM FOR CONFIDENCE BELT CONSTRUCTION

The integrals (6) and (7) can be solved using different methods. We note, however, that they are examples of simplified cases. The most general experimental situation involves both an uncertainty in signal efficiency as well as in the background detection efficiency, which are usually correlated, and possibly an additional theoretical uncertainty in the background process prediction. We have developed an algorithm that takes these effects into account with the proper correlations between them. The algorithm performs a Monte Carlo integration over the systematic uncertainties. It has been implemented as a FORTRAN program, POLE (Poissonian limit estimator) [15]. In the examples used in this section a Gaussian distribution of uncertainties is assumed, but the algorithm makes it easy to implement PDFs other than Gaussian (see the next section for an example of using a different distribution). For the moment the code supports a
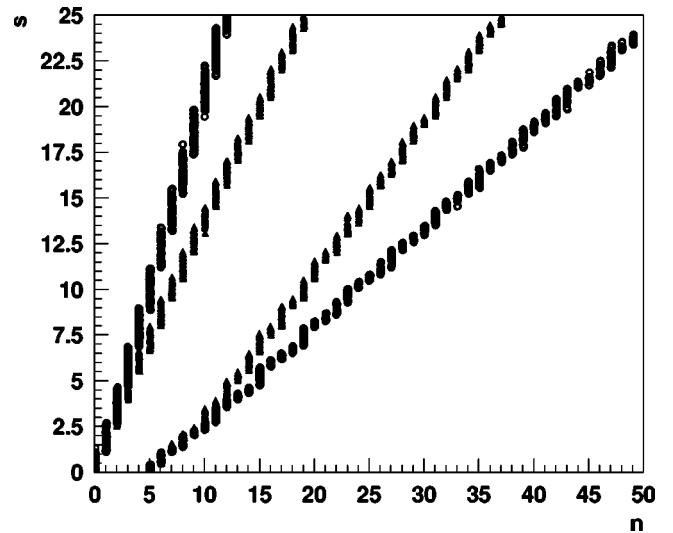

FIG. 2. 90% confidence belts obtained with POLE using the likelihood ratio ordering scheme and assuming different uncertainties in the signal efficiency. The inner band has been constructed assuming no uncertainty in the signal efficiency. The outer band represents the belt constructed with a signal efficiency uncertainty of 40%. The background expectation in this particular case was $b = 2$.

Gaussian, flat, and log-normal parametrization of the uncertainties. After determining the PDF through evaluation of the integrals, different ordering schemes can be applied for the final calculation of the confidence belt. The results presented here are mainly for the likelihood ratio ordering scheme with and without conditioning. We restrict ourselves to presenting systematic uncertainties of signal and background efficiencies separately to give a clear idea of the effect of varying a single variable at a time. Real applications usually combine those uncertainties.

The confidence belt constructions have been performed using steps of 0.05 in signal expectation and performing the construction up to a maximal signal expectation of 50 and a maximal number of detected events of 100. Including systematic uncertainties generally leads to a widening of the confidence belt. Figure 2 shows an example of a likelihood ratio confidence belt construction with and without uncertainty in the signal efficiency, where a background expectation $b = 2$ has been assumed.

Examples for some resulting intervals are given in Tables I and II. Different combinations of the number of observed events n_0 and the expected background b are given for different uncertainties in the signal and background efficiency.

The width of the interval for two particular examples of observed events and expected background as a function of signal efficiency uncertainty and background uncertainty is shown in Fig. 3. Note that for low background expectation the uncertainties in the background can be neglected (see also Table II). Figures 4 and 5 give more extended information on the resulting intervals.

An interesting case arises when there are significantly fewer events observed than expected from background and there is an uncertainty in the signal efficiency. In this case, the width of the confidence interval does not increase (see

TABLE I. Examples of likelihood ratio 90% confidence intervals including systematic uncertainty in the signal efficiency and assuming no uncertainty in the background prediction.

n_0	b	Signal efficiency uncertainty (%)	Likelihood ratio interval	Likelihood ratio interval with conditioning
2	2	0	0: 3.90	0: 4.00
		0.2	0: 3.95	0: 4.34
		0.3	0: 4.10	0: 4.75
		0.4	0: 4.65	0: 5.35
3	2	0	0: 5.40	0: 5.30
		0.2	0: 5.70	0: 5.65
		0.3	0: 5.95	0: 6.20
		0.4	0: 6.80	0: 7.10
4	2	0	0: 6.60	0: 6.60
		0.2	0: 7.10	0: 7.30
		0.3	0: 7.75	0: 7.85
		0.4	0: 8.95	0: 9.15
5	2	0	0.40: 7.95	0.50: 8.05
		0.2	0.40: 8.60	0.50: 8.60
		0.3	0.40: 9.55	0.50: 9.65
		0.4	0.40:11.15	0.50:11.20
6	2	0	1.10: 9.45	1.10: 9.45
		0.2	1.05,10.05	1.05:10.10
		0.3	1.05:11.50	1.05:11.50
		0.4	1.05:13.35	1.05:13.35

TABLE II. Examples of likelihood ratio 90% confidence intervals including systematic uncertainty in the background expectation and assuming no uncertainty in the signal efficiency.

n_0	b	Background uncertainty	Likelihood ratio interval	Likelihood ratio interval with conditioning
2	2	0	0: 3.90	0: 4.00
		0.2	0: 3.95	0: 4.10
		0.3	0: 3.95	0: 4.25
		0.4	0: 3.95	0: 4.35
3	2	0	0: 5.40	0: 5.30
		0.2	0: 5.45	0: 5.35
		0.3	0: 5.45	0: 5.45
		0.4	0: 5.50	0: 5.55
4	2	0	0: 6.60	0: 6.60
		0.2	0: 6.95	0: 6.65
		0.3	0: 6.95	0: 6.80
		0.4	0: 6.95	0: 6.80
5	2	0	0.40: 7.95	0.50: 8.05
		0.2	0.35: 7.95	0.50: 8.10
		0.3	0.30: 8.00	0.50: 8.10
		0.4	0.20: 8.20	0.45: 8.15
6	2	0.	1.10: 9.45	1.10: 9.45
		0.2	1.05: 9.45	1.10: 9.50
		0.3	1.00: 9.50	1.05: 9.50
		0.4	0.95: 9.50	1.00: 9.50

Table III). Note that if we use conditioning the effect disappears. The same cannot be observed in the case where we consider only an increasing background uncertainty (see Table IV).

Negative values of the nuisance parameters: Using a log-normal distribution

In experimental situations where the systematic uncertainties are high, a problem might arise due to the fact that sampling from a Gaussian PDF allows negative values. POLE deals with these cases by truncating the Gaussian distribution and renormalizing the part above zero.

We examine the effect of truncating the Gauss distribution by calculating the confidence interval for different values of the truncation point (see Fig. 6). Considering a Gaussian distribution centered on 1 with $\sigma = 40\%$, a truncation at zero removes only 0.7%. Figure 6 therefore indicates that effects on the confidence interval due to the truncation are negligible for all cases considered in this paper.

A PDF for the nuisance parameters extending to negative values or which falls off to zero discontinuously is certainly
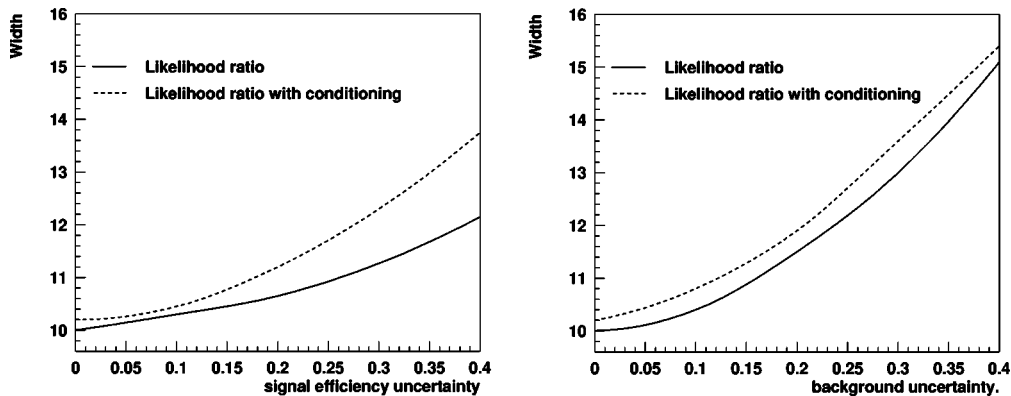

FIG. 3. Example of the dependence of the likelihood ratio confidence interval width on the systematic uncertainties, with and without conditioning, as obtained with POLE. The left plot shows the width as a function of the uncertainty in signal efficiency assuming no additional uncertainty in background expectation. The right plot shows the width as a function of the background uncertainty. We have used $n_o = 17$ and a background of 15 in constructing the plots.
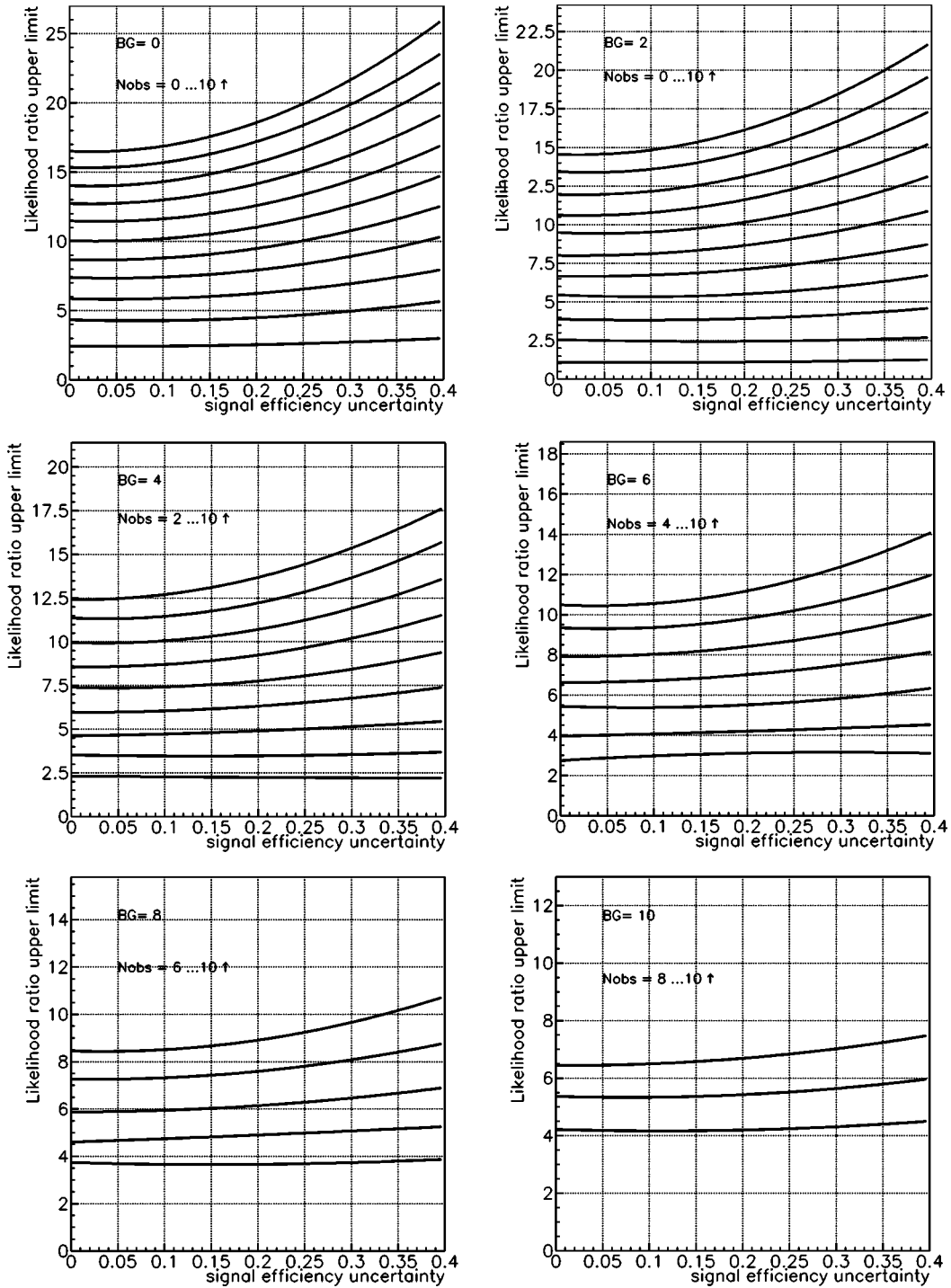
FIG. 4. Likelihood ratio upper limit as a function of signal efficiency uncertainty for expected backgrounds of 0, 2, 4, 6, 8, and 10. Cases with number of observed events significantly fewer than expected background have been omitted.

undesired from a conceptual point of view. We therefore test the behavior of the confidence interval if we replace the Gaussian distribution with a log-normal distribution, which in the general form is given by

$$q(x)_{\mu,\sigma} = \frac{1}{\sqrt{2\pi}x\sigma} e^{-(\ln x - \mu)^2/2\sigma^2}. \tag{8}$$

We require the mean of the log-normal distribution to be the

nominal value of the nuisance parameter and use the Gaussian standard deviation as before (the variance of the log-normal distribution will then be approximately the same). The confidence intervals for Neyman and likelihood ratio ordering under these assumptions are shown for one particular example of the number of observed events and expected background as a function of signal efficiency in Fig. 7. The differences between using a Gaussian distribution and using a log-normal distribution are generally small, in our example
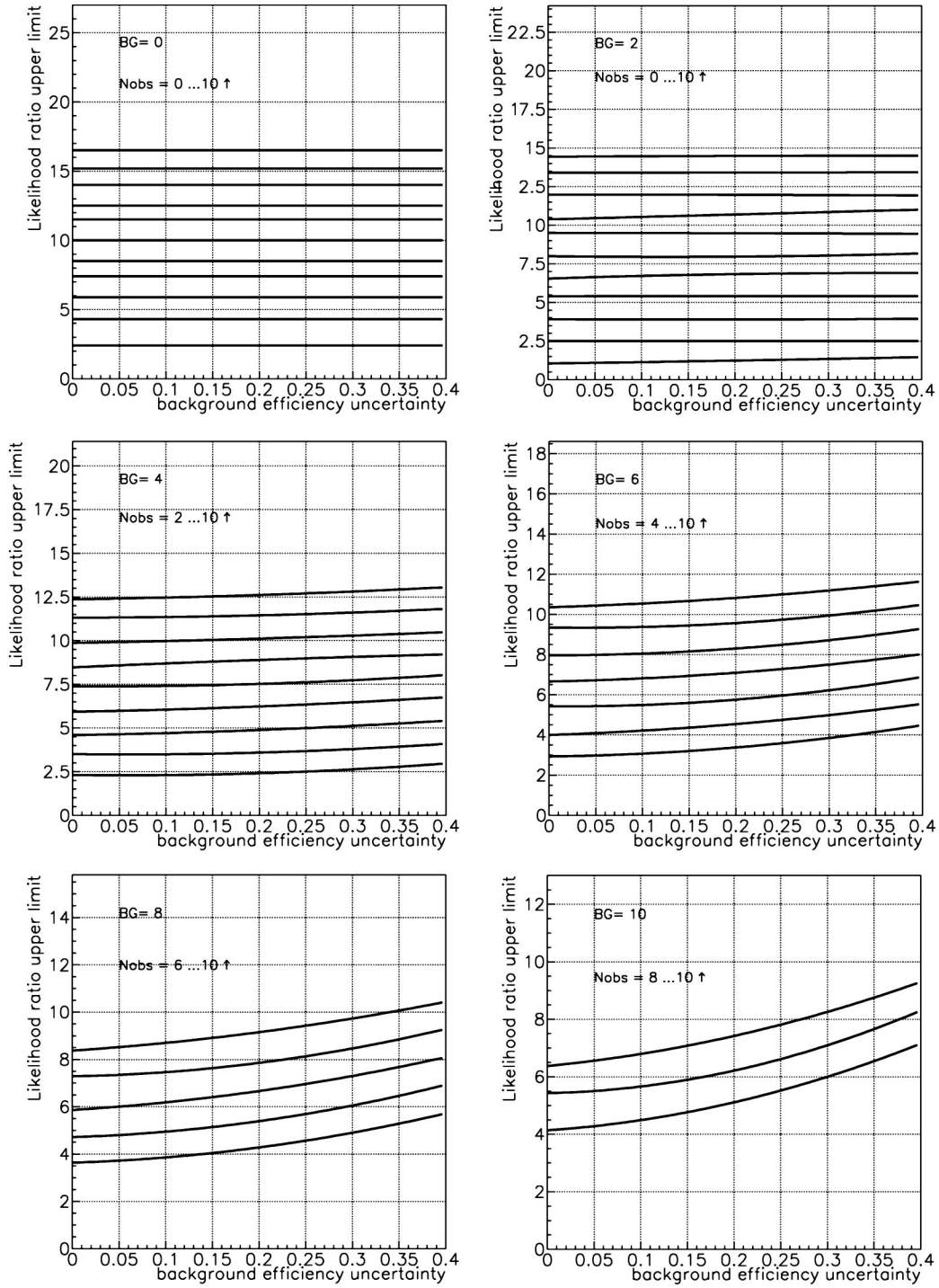
FIG. 5. Likelihood ratio upper limit as a function of background efficiency uncertainty for expected backgrounds of 0, 2, 4, 6, 8, and 10. Cases with number of observed events significantly fewer than expected background have been omitted.

less than $\sim 2\%$. The use of a log-normal distribution is implemented as an option in POLE.

V. COMPARISON WITH THE χ^2 METHOD

Since the ratio of the likelihoods is asymptotically χ^2 distributed the approximation

$$\Delta \chi^2 = -2 \times \min_{\epsilon} \frac{\mathcal{L}(n_o, s, \epsilon)}{\mathcal{L}(n_o, s_{best}, \epsilon_{best})} \qquad (9)$$

is often used; see, e.g., [16].

Here for a given observation n_0 the $\Delta \chi^2$ is calculated as a function of s and a cut on $\Delta \chi^2$ is performed to obtain the confidence interval (e.g., $\Delta \chi^2 = 2.71$ corresponds to the 90%

TABLE III. Likelihood ratio confidence intervals with systematic uncertainty in the signal efficiency and no uncertainty in the background expectation. Here two examples are shown where there are fewer events observed than the expected background. The interval does not increase with increasing uncertainty if there are significantly fewer events observed than the expected background. However, if the expected background and number of observed events are comparable, the interval becomes larger. When conditioning is applied, it grows larger in all cases.

n_0	b	Signal efficiency uncertainty	Likelihood ratio interval	Likelihood ratio interval with conditioning
2	6	0	0: 1.55	0: 3.15
		0.2	0: 1.55	0: 3.35
		0.4	0: 1.45	0: 4.00
4	6	0	0: 2.85	0: 4.30
		0.2	0: 3.20	0: 4.60
		0.4	0: 3.35	0: 5.35

confidence level for one degree of freedom).

Figure 8 illustrates the effect of including uncertainties on the resulting confidence intervals for the χ^2 approximation as compared to the method proposed here. Generally, the χ^2 approximation gives more conservative results than the POLE method. Since—as we will see in the following sections—using the POLE method leads to some overcoverage, this is clearly undesirable.

VI. TESTS OF COVERAGE

From a frequentist point of view, an algorithm is said to have the correct *coverage* if, given a confidence level $1-\alpha$ and a large number of repeated identical experiments, it provides correct answers in a fraction $1-\alpha$ of the cases, independent of the value of s. To test the coverage of the algorithm proposed in this paper, we perform the construction described in the previous sections for a large number of simulated experiments, where we predefine the *true* signal and background and then determine n_o by random sampling from a Poisson distribution. We then calculate how often the

TABLE IV. Likelihood ratio confidence intervals with systematic uncertainty in the background expectation and no uncertainty in the signal efficiency. Here two examples are shown where there are fewer events observed than expected background. The confidence interval becomes larger with increasing uncertainty in the background expectation.

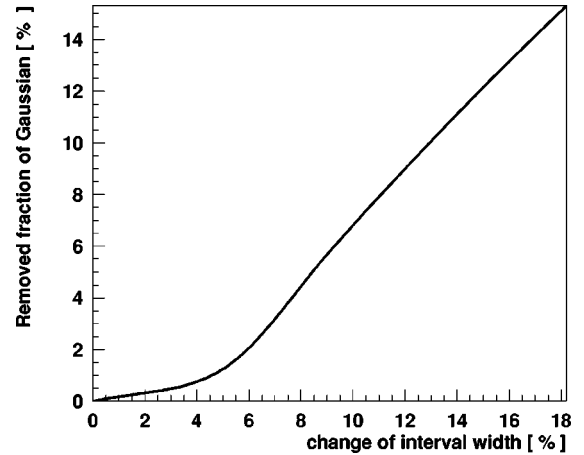| n_0	b	Background uncertainty (%)	Likelihood ratio interval	Likelihood ratio interval with conditioning
2	6	0	0: 1.55	0: 3.15
		0.2	0: 1.55	0: 3.50
		0.4	0: 2.64	0: 3.85
4	6	0	0: 2.85	0: 4.30
		0.2	0: 3.25	0: 4.55
		0.4	0: 4.60	0: 5.55

FIG. 6. The relative change of the likelihood ratio interval width as a function of fraction of Gaussian removed. In this example, $n_o=4$ and $b=4$ have been assumed.

obtained confidence interval does not contain the predefined s. We define the *coverage ratio*

$$R = \frac{n_{false}}{n_{tot}}. \qquad (10)$$

Here, n_{false} denotes the number of simulated experiments in which the result of the algorithm does not contain the predefined s, and n_{tot} denotes the number of simulated experiments performed. If we choose $1-\alpha$ to be 0.9, perfect coverage would mean $R=0.1$, independent of the signal expectation assumption. A value of R smaller than 0.1 means that the method overcovers. Expected coverage has been studied mostly in the context of Bayesian intervals, small numbers of events, or including conditioning [18–20] and without taking into account systematic uncertainties. Very recently, a study was presented considering coverage including systematic uncertainties [17].

In the next sections, in addition to presenting coverage tests done for higher signal expectations without uncertain-
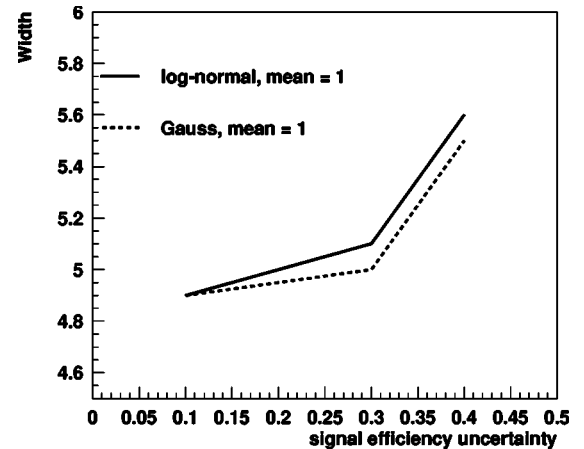

FIG. 7. Likelihood ratio confidence interval width as a function of signal efficiency uncertainty for a Gaussian and a log-normal distribution with mean at 1. In this example, $n_o=4$ and $b=4$ have been assumed and the Gaussian was truncated at zero.
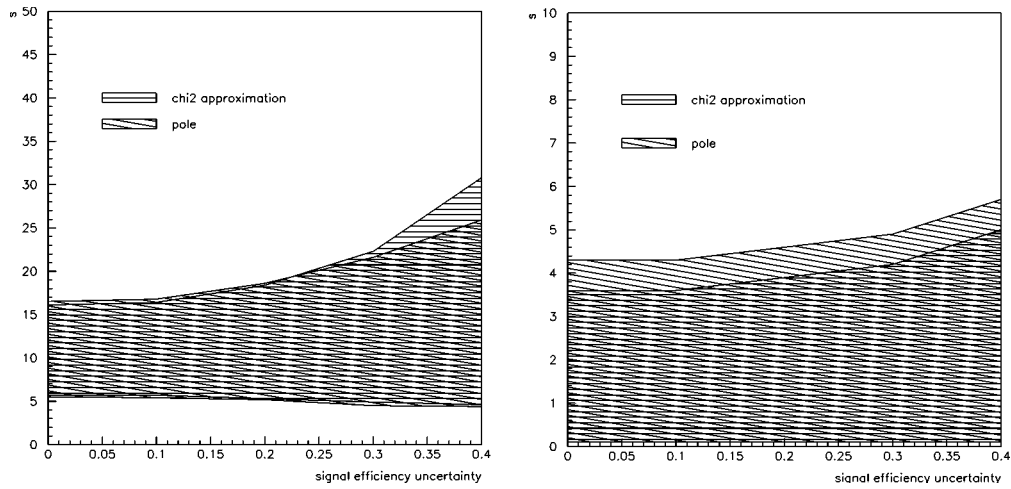
FIG. 8. The likelihood ratio confidence interval as calculated by POLE and using the χ^2 approximation. Number of observed events is 10 (left panel) and 1 (right panel). The background was assumed to be zero in both cases.

ties, we will focus on the coverage of the methods if systematic uncertainties are included.

A. Coverage without systematic uncertainties

We show an example of a plot of the coverage ratio (here using steps of 0.1 in signal space) in Fig. 9 for Neyman and likelihood ratio ordering. Both methods seem to overcover for almost all cases (which is expected because of the discreteness of the Poisson distribution). There is no signal expectation dependence of the coverage ratio except for the "seesaw" behavior, which again reflects the discreteness of the Poisson distribution.

B. Coverage with systematic uncertainties

Introducing systematic uncertainties in the calculation of confidence intervals and tests of coverage leads to the question of what is meant by a *repeated* experiment. If we adhere to the traditional definition in which an experiment is repeated with fixed parameters such as efficiency or background rate, the algorithm presented here will inevitably

yield overcoverage. Figure 10 shows the mean coverage ratio (mean here taken over six different signal expectation assumptions) as a function of different systematic uncertainties. The overcoverage will not only increase with increasing uncertainties but also be dependent on the signal expectation (Fig. 11).

1. Bayesian coverage

The overcoverage described in the previous section is a consequence of the fact that efficiencies and background are not random variables (there is a true but unknown fixed efficiency and background rate) but they are treated as random variables in the construction of the confidence belt [Eqs. (6) and (7)]. Thus, while in the construction we are using a PDF which is a convolution of a Poisson distribution with a Gaussian distribution, repeated measurements (with parameters fixed) will produce a Poisson distribution. However, one has to keep in mind that the distribution obtained in this way is not the *underlying* one. To infer from the measured Poisson distribution the underlying one, the signal efficiency and the background have to be taken into account. In particu-
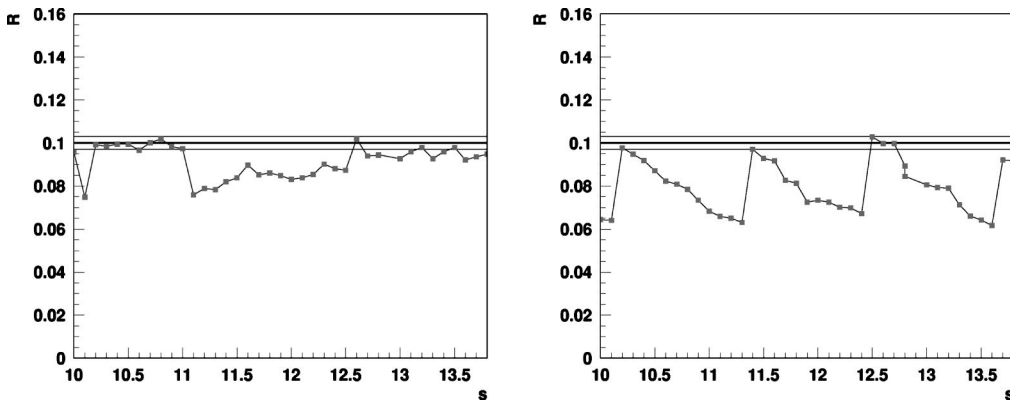

FIG. 9. Coverage ratio as a function of different signal expectation assumptions. Left plot: Likelihood ratio ordering. Right plot: Neyman ordering. The thick line gives the line of perfect coverage, the thinner lines denote the measurement precision of this ratio that we can achieve with 10000 simulated experiments (taken as 1σ of a binomial distribution). A constant background expectation of $b=10$ has been assumed.
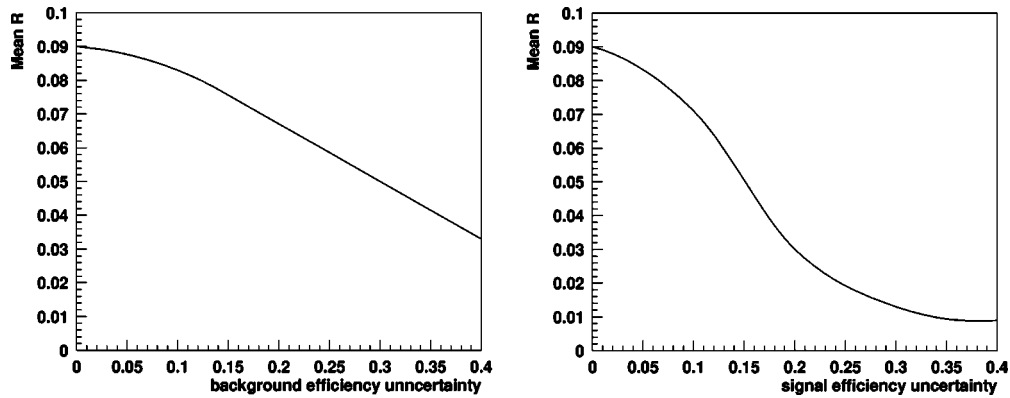
FIG. 10. Mean coverage ratio as a function of background uncertainty (left plot) and signal efficiency uncertainty (right plot). Here the mean is taken over six signal expectation assumptions between 12 and 42. The background expectation was taken to be constant $b = 12$. All other uncertainties than the one displayed were assumed to be zero.

lar, if these parameters are uncertain, there will not be a single underlying Poisson distribution, but a set of distributions that are weighted with the probabilities of the possible different efficiencies and backgrounds. In a way, we thus give different hypotheses different weights. To take this into account we modify the coverage test described in the previous section. Instead of drawing a measurement from Poisson distributions with predefined signal expectation and background, we draw the signal expectation and background prediction used in each simulated experiment from Gaussian distributions centered on the predefined true signal and background, where the width of the Gaussian is the associated systematic uncertainty. The measurement is then produced by taking these new values as means for the final Poisson distributions.

Since, by using this approach, we give different weights to different hypotheses, we call this modified coverage test *Bayesian coverage*. In this way, the PDFs used in the construction and in the coverage test are consistent with each other, and the algorithm should, per construction, give the correct coverage (except for discreteness effects). In particular, the coverage defined in this way should be independent of the magnitude of the uncertainties present in the experi-

ment. Figure 12 shows the mean Bayesian coverage for different uncertainties in the signal efficiency together with the frequentist coverage. The mean is here taken over the 29 points in signal expectation space that were tested. As expected the Bayesian coverage ratio is nearly constant.

Thus, if we loosen the criteria on the definition of "repeated experiment," allowing the "true" (unknown) efficiencies to vary for each experiment repetition, the method has the desired statistical property of correct coverage.

2. Remark on the choice of ensemble

In the previous subsection we consider an ensemble in which the true value of the nuisance parameter, corresponding to ϵ' in Eq. (7) is varied in each of the members of the ensemble. We find, as expected, that the POLE method satisfies the requirement of correct coverage with respect to this ensemble.

However, it can be argued that this ensemble does not describe the situation usually encountered in experimental physics. The systematic uncertainty is a *measurement* uncertainty, i.e., it is not the true value of the nuisance parameter
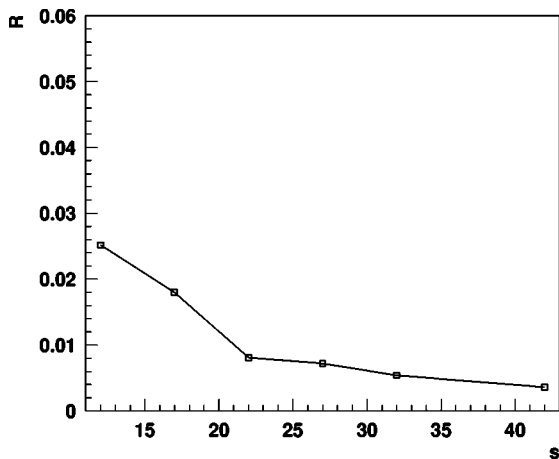

FIG. 11. Signal expectation dependence of the coverage ratio, here for the case where signal efficiency uncertainty is 30%.
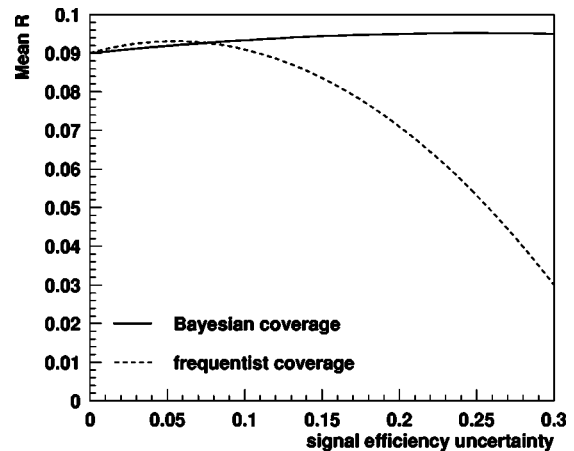

FIG. 12. Bayesian mean coverage ratio as a function of the uncertainty in signal efficiency. The mean is here taken over 29 signal expectation values. For comparison the frequentist result using the same signal expectation assumptions has been included.
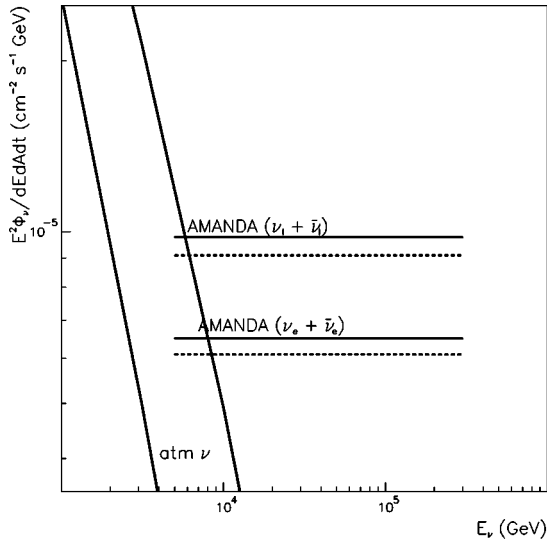
FIG. 13. Limit on the flux from cosmic neutrinos of all flavors and electron neutrinos as presented in [22]. Signal efficiency uncertainties of 25% and 30% in backround prediction lead to an increase of the upper limit by about 10%.

FIG. 15. Effect of including uncertainties on a generic WIMP limit. The solid line represents the limit calculation without any experimental systematic uncertainties. The dashed lines represent (from bottom to top) a 20% uncertainty in the background expectation only, a 10%, 20%, 30%, and 40% uncertainty on the signal efficiency on top of the 20% background uncertainty. Additionally, a 30% uncertainty in the theoretical atmospheric neutrino background expectation has been assumed in all cases. The absolute scale is arbitrary.

that changes in each experiment but the *measured* one. Studies for a few cases indicate that with respect to such an ensemble the POLE method leads to moderate overcoverage [17].

VII. LIMITS ON HIGH ENERGY COSMIC NEUTRINOS

In experimental situations where the systematic uncertainties are negligible small, limits or central confidence inter-
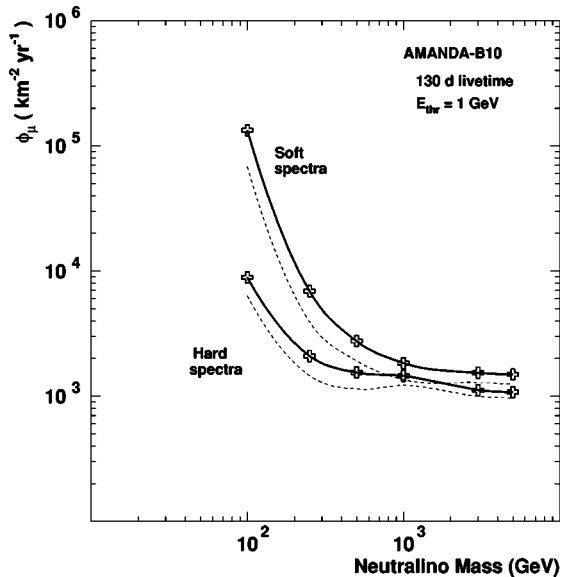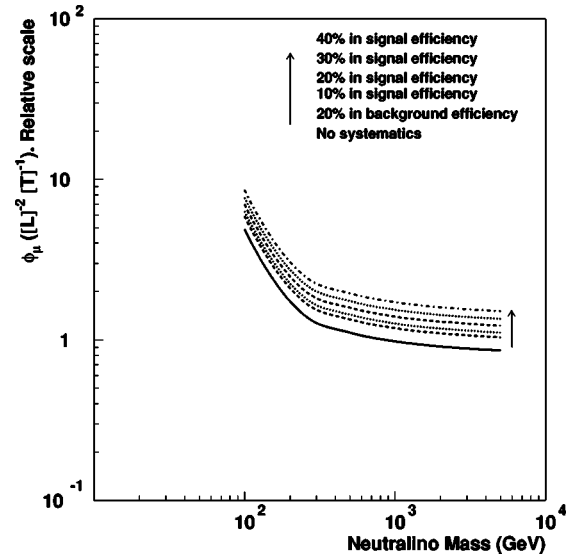

FIG. 14. Comparison of the effect of systematics on the limit on the neutrino-induced muon flux from the center of the Earth from neutralino annihilation. The solid line represents the current limit set by the AMANDA Collaboration, the dashed line is the limit without including the systematic uncertainties present in their analysis (figure taken from [23], where more details of the analysis can be found).

vals can be calculated without evaluating the effects of the former. In more general situations, including systematic uncertainties in the calculation of the limits is essential, since it is a way of incorporating the real sensitivity of a given experiment to the quantity being measured.

In this section we will consider two real examples taken from published results of the AMANDA neutrino telescope, where the program POLE was used to include the systematics in the final results. The AMANDA Collaboration has recently published 90% confidence limits on the diffuse flux of cosmic electron neutrinos and of neutrinos of all flavors in the energy range between 5 TeV and 300 TeV [22]. The analysis revealed zero events with an expected background from atmospheric neutrinos of 0.01 events. The systematic uncertainty in the signal efficiency for this analysis was $\sim 25\%$, determined from Monte Carlo studies. On top of that, the current theoretical systematic uncertainty in the atmospheric neutrino flux prediction in the energy range relevant to the analysis is about 30% [21], which has to be taken into account as well. The effect of including the systematic uncertainties is shown in Fig. 13. The effect is to worsen the limits by about 10%.

Another example worth noticing is the results published by the same collaboration on searches for supersymmetric dark matter in the form of weakly interacting massive particles (WIMPs) [23]. In this analysis the uncertainties in signal efficiency range from 10% to 25% depending on the assumed signal spectrum and, additionally, an uncertainty in the background detection efficiency, estimated to be 20%, has to be taken into account in this case. A further complication arises in including these uncertainties in the calcula-

tion of the final limits since the efficiencies in signal and background detection are correlated. Moreover, the mentioned theoretical uncertainty on the overall normalization of the atmospheric neutrino flux has to be added. In Fig. 14 we show the limits to the muon flux from the center of the earth as a function of WIMP mass. The full lines show the limits for two diferent assumptions on the signal spectra, and the dashed lines the corresponding limits without including systematic uncertainties in the calculations.

For the purpose of illustration, we show in Fig. 15 the effect of including systematic uncertainties in the limit calculation for five different values (dashed lines), from bottom to top a 20% uncertainty in background expectation only, and a 10%, 20%, 30%, and 40% uncertainty in signal efficiency (on top of the mentioned 20% background uncertainty). The absolute scale of the plot is arbitrary since we are just interested in showing the relative effect of the inclusion of systematic uncertainties in the limit calculation, with respect to the no-systematics case (full line). The figure shows the importance of correctly evaluating systematics and including them in the final result, since the effects can be important.

VIII. CONCLUSIONS

In this note we present a Monte Carlo algorithm for introducing systematic uncertainties in the evaluation of classical confidence intervals which allows us to include uncertainties in the background prediction, in the background detection efficiency, and in the signal detection efficiency, and correlations between them, by integrating over the (assumed) PDFs of these parameters. We apply the method for a Poisson process with a background under the assumption of a Gaussian PDF describing the uncertainties. We present re-sults where the construction has been performed using likelihood ratio ordering with and without conditioning.

Generally, the introduction of systematic uncertainties leads to an increase in the confidence interval width. However, an interesting result is that likelihood ratio (as well as Neyman) confidence intervals which take into account the systematic uncertainty in the signal efficiency do not become larger with larger uncertainty, in the case that significantly fewer events have been observed than expected background. With respect to an ensemble with strictly identical experiments, introducing systematic uncertainties in the presented manner inevitably leads to overcoverage, increasing with the magnitude of th systematic uncertainties. However, we show that, with respect to an ensemble where the systematic uncertainties are taken into account in the coverage test by varying their *true* assumed values in each pseudoexperiment, the method presented here provides overcoverage only on the level already present due to the discreteness of the Poisson distribution. Both ensembles are ideal ensembles and have to be seen as approximations to the ensemble encountered in experimental physics. In summary, the algorithm presented here provides a practical and flexible way to quantitatively take into account systematic uncertainties present in experimental situations in the calculation of confidence intervals.

[1] J. Neyman, Philos. Trans. R. Soc. London **767**, 333 (1937).

[2] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).

[3] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics, Vol. 2, Classical Inference and Relationship* (Oxford University Press, New York, 1991).

[4] R. D. Cousins and V. L. Highland, Nucl. Instrum. Methods Phys. Res. A **320**, 331 (1992).

[5] W. A. Rolke and A. M. Lopez, Nucl. Instrum. Methods Phys. Res. A **458**, 745 (2001).

[6] Proceedings of Workshop on Confidence Limits, Geneva, Switzerland, 2000, edited by F. James and L. Lyons (CERN Yellow Report No. 2000-005, Geneva, 2000).

[7] Proceedings of the 2nd Workshop on Confidence Limits, Fermilab, Chicago, 2000, http://conferences.fnal.gov/cl2k/

[8] Proceedings of the Conference on Advanced Statistical Techniques in Particle Physics, Durham, 2002, edited by M. R. Whalley and L. Lyons (IPPP/02/38, DCPT/02/78).

[9] C. Giunti, Phys. Rev. D **59**, 113009 (1999).

[10] W. Eadie *et al.*, *Statistical Methods in Experimental Physics* (North-Holland, Amsterdam, 1982).

[11] B. P. Roe and M. B. Woodroofe, Phys. Rev. D **60**, 053009 (1999).

[12] G. Zech, hep-ex/0106023.

[13] R. D. Cousins, Phys. Rev. D **62**, 098301 (2000).

[14] B. P. Roe and M. B. Woodroofe, Phys. Rev. D **63**, 013009 (2001).

[15] http://www3.tsl.uu.se/~conrad/pole.html

[16] G. J. Feldman, in Ref. [7].

[17] J. Conrad, O. Botner, A. Hallgren, and C. P. de los Heros, in Ref. [8].

[18] I. Narsky, hep-ex/0005019.

[19] M. Mandelkern and J. Schultz, J. High Energy Phys. **11**, 036 (2000).

[20] O. Helene, Phys. Rev. D **60**, 037901 (1999).

[21] T. Gaisser *et al.*, in Proceeding of the XVII International Cosmic Ray Conference (ICRC), Hamburg, Germany, 2001, available from http://www.copernicus.org/icrc

[22] J. Ahrens *et al.*, astro-ph/0206487.

[23] J. Ahrens *et al.*, Phys. Rev. D **66**, 032006 (2002).