ELSEVIER

Nuclear Instruments and Methods in Physics Research A 434 (1999) 435–443

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

www.elsevier.nl/locate/nima

Confidence level computation for combining searches with small statistics

Thomas Junk*

Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S 5B6

Received 8 February 1999; accepted 11 April 1999

Abstract

This article describes an efficient procedure for computing approximate confidence levels for searches for new particles where the expected signal and background levels are small enough to require the use of Poisson statistics. The results of many independent searches for the same particle may be combined easily, regardless of the discriminating variables which may be measured for the candidate events. The effects of systematic uncertainty in the signal and background models are incorporated in the confidence levels. The procedure described allows efficient computation of expected confidence levels.

PACS: 06.20.Dk

Keywords: Confidence level computation Poisson statistical

1. Introduction

The problem of combining the results of several independent searches for a new particle and producing a confidence level (CL) has become very important at the LEP collider in its high-energy phase of running. Typically, both the expected number of signal events and the expected number of background events are small, and few candidate events are observed in the data for any particular search analysis. The ability to exclude the presence of a possible signal at a desired CL is often improved significantly by combining the results of several searches, particularly if the sensitivity is limited by the collected luminosity, and not by a kinematic boundary. In addition, sophisticated search analyses may provide information about the observed candidates, such as one or more reconstructed masses or other experimental information relating to the expected features of the signal. These variables provide better discrimination of signal from background, and also help to indicate which signal hypothesis is preferred among many. Sometimes no such information is available, and these search analyses must be combined with other types of analyses for an optimal CL. Binning the search results of the analyses in their discriminant variables and treating each bin as a statistically independent counting search provides a simple, uniform representation of the data well suited for combination.

*CERN/EP, CH-1211 Geneva 23, Switzerland. Tel.: +41-22-767-6042; fax: +41-22-767-9330.

E-mail address: tom.junk@cern.ch (T. Junk)

0168-9002/99/$ - see front matter
PII: S0168-9002(99)00498-2

Often, as is the case with searches for MSSM Higgs bosons at LEP2, a broad range of model parameters which affect the production of signal events must be considered and exclusion limits placed for all possible values of these parameters. The expected experimental signatures of the new particles in general vary with the model parameters which govern their production and decay, and the combination of complementary channels provides the best exclusion for all values of the parameters. A rapid procedure for computing confidence levels is therefore necessary in order to explore fully the possibilities of the model.

This article describes an efficient, approximate method of computing combined exclusion confidence levels in these cases, allowing also for the possibility of uncertainty in the estimated signal and background.

2. Modified frequentist confidence levels

For the case of n independent counting search analyses, one may define a test statistic X which discriminates signal-like outcomes from background-like ones. An optimal choice for the test statistic is the likelihood ratio [1–3]. If the estimated signal in the ith channel is s_i, the estimated background is b_i, and the number of observed candidates is d_i, then the likelihood ratio can be written as

$$X = \prod_{i=1}^{n} X_i \tag{1}$$

with

$$X_i = \frac{e^{-(s_i+b_i)}(s_i+b_i)^{d_i}}{d_i!} \Big/ \frac{e^{-b_i} b_i^{d_i}}{d_i!}. \tag{2}$$

This test statistic has the properties that the joint test statistic for the outcome of two channels is the product of the test statistics of the two channels separately, and that it increases monotonically in each channel with the number of candidates d_i.

The confidence level for excluding the possibility of simultaneous presence of new particle production and background (the $s+b$ hypothesis), is

$$\mathrm{CL}_{s+b} = P_{s+b}(X \le X_{\mathrm{obs}}) \tag{3}$$

i.e., the probability, assuming the presence of both signal and background at their hypothesized levels, that the test statistic would be less than or equal to that observed in the data. This probability is the sum of Poisson probabilities

$$P_{s+b}(X \le X_{\mathrm{obs}}) = \sum_{X(\{d_i'\}) \le X(\{d_i\})} \prod_{i=1}^{n} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{d_i'}}{d_i'!} \tag{4}$$

where $X(\{d_i\})$ is the test statistic computed for the observed set of candidates in each channel $\{d_i\}$, and the sum runs over all possible final outcomes $\{d_i'\}$ which have test statistics less than or equal to the observed one.

The confidence level $(1 - \mathrm{CL}_{s+b})$ may be used to quote exclusion limits although it has the disturbing property that if too few candidates are observed to account for the estimated background, then any signal, and even the background itself, may be excluded at a high confidence level. It, nonetheless, provides exclusion of the signal at exactly the confidence level computed. Because the candidates counts are integers, only a discrete set of confidence levels is possible for a fixed set of s_i and b_i.

A typical limit computation, however, involves also computing the confidence level for the background alone,

$$\mathrm{CL}_b = P_b(X \le X_{\mathrm{obs}}), \tag{5}$$

where the probability sum assumes the presence only of the background. This confidence level has been suggested to quantify the confidence of a potential discovery, as it expresses the probability that background processes would give fewer than or equal to the number of candidates observed. Then the Modified Frequentist confidence level CL_s is computed as the ratio

$$\mathrm{CL}_s = \mathrm{CL}_{s+b}/\mathrm{CL}_b. \tag{6}$$

This confidence level is a natural extension of the common single-channel $\mathrm{CL} = 1 - \mathrm{CL}_s$ [4–7], and for the case of a single counting channel is identical to it.

The task of computing confidence levels for experimental searches with one or more discriminating variables measured for each event reduces to the case of combining counting-only searches by

binning each search analyses' results in the measured variables. Each bin of, e.g., the reconstructed mass, then becomes a separate search channel to be combined with all others, following the strategy of [8] and the neutrino-oscillation example of [9]. In this case, the expected signal in a bin of the reconstructed mass depends on the hypothesized true mass of the particle and also on the expected mass resolution. If the error on the reconstructed mass varies from event to event such that the true resolution is better for some events and worse for others, then the variables s, b, and d may be binned in both the reconstructed mass and its error to provide the best representation of the available information. By exchanging information in bins of the measured variables, different experimental collaborations may share all of their search result information in an unambiguous way without the need to treat the measured variables in any way during the combination.

For convenience, one may add the s_i's, the b_i's, and the d_i's of channels with similar s_i/b_i and retain the same optimal exclusion limit, just as the data from the same search channel may be combined additively for running periods with the same conditions. The same search with a new beam energy or other experimental difference should of course be given its own set of bins (which may be combined with others of the same s_i/b_i).

3. Confidence level calculation

The task of summing the terms of Eq. (4) can be formidable. For n channels, each with m possible outcomes, there are $\mathcal{O}(n^m)$ terms to compute. This sum is often carried out with a Monte Carlo [8,9], selecting representative outcomes of the experiment and comparing their test statistics with the test statistic computed with the data candidate event counts. Another alternative, described in this article, is to compute the probability distribution function (PDF) for the test statistic for a set of channels, and iteratively combine additional channels by convoluting with the PDFs of their test statistics.

The PDF of the test statistic for a single channel is a sum of delta functions at the accessible values of X_i. These may be represented as a list of possible outcomes

$$(X_i^j, p_i^j) \tag{7}$$

where X_i^j is the test statistic for the ith channel if it were to have j events, and p_i^j is the Poisson probability of selecting j events in the ith channel if the underlying average expected rate is $s_i + b_i$ when computing CL_{s+b}, or only b_i when computing CL_b. The list is formally infinitely long, but one may truncate it when the total probability sum of the outcomes in the list exceeds a fixed quantity, or one may select all j such that $X_i^j \leq X_{\mathrm{obs}}$.

For the case of two channels, one forms the probabilities and test statistics for the joint outcomes multiplicatively,

$$(X_i^j X_{i'}^{j'}, p_i^j p_{i'}^{j'}) \tag{8}$$

to form a representation of the PDF of the test statistic for the joint outcomes of two channels. One may then iteratively combine all channels together and use the list to compute the confidence level by adding the probabilities of outcomes with test statistics less than or equal to that observed. This reintroduces the computational difficulty of enumerating all possible experimental outcomes, and hence one needs to introduce an approximation to limit the complexity of the problem.

The approximation is to bin the PDF of the test statistic at each combination step. The cumulative PDF may be obtained from the listing of outcomes by sorting them by their test statistics and accumulating the probabilities. Then fine bins of the cumulative PDF may be filled with possible outcomes. A useful binning covers very small probabilities logarithmically in order to represent small CL's more exactly, and has a uniform binning for larger probabilities. The finer the bins, the more precise the computed CL will be; in the limit of infinitely fine bins, the problem reduces once again to adding the probabilities of all possible outcomes.

To guarantee a conservative CL for setting limits, one may, at each combination step, record as a possible experimental "outcome" the smallest test statistic within a bin coupled with the largest accumulated probability within the same bin. The list now consists of test statistics and the cumulative probability of observing that test statistic or less,

and the differential PDF of X may be recovered from it.

The process is then repeated iteratively for all channels to be combined. The running time on a computer is proportional to the number of channels, the number of bins kept in the PDF of X, and increases with the expected number of events in the channels. To improve the accuracy of the approximation, the search channels should be sorted in order of s_i/b_i, with the channels with the largest s_i/b_i combined last.

Once all channels have been combined, the test statistic is computed for the candidate events observed in the experiment and CL_{s+b}, CL_b and CL_s may be computed using Eqs. (3), (5) and (6). Furthermore, the PDFs of X in the signal + background and background hypotheses allow computation of the expected confidence levels $\langle CL_{s+b}\rangle$, $\langle CL_b\rangle$, and $\langle CL_s\rangle$, assuming the presence only of background. These are indications of how well an experiment would do on average in excluding a signal if the signal truly is not present, and are the important figures of merit when optimizing an analysis for exclusion.

When computing $\langle CL_b\rangle$, the outcomes are already ordered by their test-statistic and only the probabilities are needed:

$$\langle CL_b\rangle = \sum_{i=1}^{N_{\text{blist}}} \left[p_i^b \sum_{j=1}^{i} p_j^b \right] \tag{9}$$

where N_{blist} is the number of entries in the table of the PDF of X for the background-only hypothesis, and p_j^b is the jth probability in the list, where the test statistic X increases with increasing j. For total expected backgrounds of more than about 3.0 events in channels with non-negligible sensitivity to the signal, $\langle CL_b\rangle \approx 0.5$.

The values of $\langle CL_{s+b}\rangle$ and $\langle CL_s\rangle$ can be computed similarly, although the PDF of X is needed in the $s+b$ hypothesis as well as the background-only hypothesis.

$$\langle CL_{s+b}\rangle = \sum_{i=1}^{N_{\text{blist}}} \left[p_i^b \sum_{X_j^{s+b} \le X_i^b} p_j^{s+b} \right] \tag{10}$$

and

$$\langle CL_s\rangle = \sum_{i=1}^{N_{\text{blist}}} \left[p_i^b \frac{\sum_{X_j^{s+b} \le X_i^b} p_j^{s+b}}{\sum_{j=1}^{i} p_i^b} \right] \tag{11}$$

where p_j^{s+b} is the jth entry in the PDF table of X for the $s+b$ hypothesis, and X_j^{s+b} is its corresponding value of X.

The difference between this method and that described by Cousins and Feldman [9] is the choice of test statistic (referred to as the "ordering principle" in Ref. [9]). The likelihood ratio of Eq. (2) has the advantages that it is the most powerful test statistic for distinguishing the $s+b$ hypothesis from the background-only hypothesis, and also because it does not depend on the range of possible models of new physics considered when testing a particular signal hypothesis. With the test statistic of [9,10], a signal hypothesis can be excluded because other signal hypotheses fit the data better. The use of the test statistic of [9,10] does not allow the exclusion of the entire model space under study – one must be careful to include the null hypothesis of no new particle production in the space of models to be tested. In addition, there may be more than one new physics signal present in the data. The method of [9] is ideal for the case in which the possible model space is fully known, and it is known that exactly one of the points in model space corresponds to the truth.

For purposes of discovery, $1 - CL_b$ indicates the probability that the background could have fluctuated to produce a distribution of candidates at least as signal-like as those observed in the data. This probability depends on the signal hypothesis because channels with small s_i/b_i do not contribute as much to the computation of CL_b as those with large s_i/b_i. In the case that a particle of unknown mass is sought, analyses which reconstruct the mass provide discrimination among competing signal hypotheses when a clear signal is present, rather than the presence of an excess of candidates. Nonetheless, the probability in the upper tail of the X distribution in the $s+b$ hypothesis may be used to exclude a signal hypothesis because it does not predict enough signal to explain the candidates in the data.

4. Systematic uncertainty on signal and background

The effect on the confidence levels from systematic uncertainties in the signal estimations $\{s_i\}$ and

background estimations $\{b_i\}$ can be accommodated by a generalization of the method of Cousins and Highland [11]. This approach was originally created for one-channel searches with systematic uncertainty on the signal estimation only. A very similar approach for handling background uncertainty is described by Giunti [12]. The generalization of this technique to the case of many channels with errors on both signal and background is summarized here.

When forming the list of the probabilities and test statistics of possible outcomes for a channel, each entry in the list is affected by the systematic uncertainties on the signal and background estimations for that channel. This effect is computed by averaging over possible values of the signal and background given by their systematic uncertainty probability distributions. For purposes of implementation, these probability distributions are assumed to be Gaussian, with the lower tail cut off at zero, so that negative s or b are not allowed.

When computing the PDF of X for the $s+b$ case, the probability to observe j events in channel i with estimated signal $s_i \pm \sigma_{s_i}$ and estimated background $b_i \pm \sigma_{b_i}$, is

$$p_i^j = \frac{\int_0^\infty ds' \int_0^\infty db' \dfrac{e^{-((s'-s_i)^2/2\sigma_{s_i}^2 + (b'-b_i)^2/2\sigma_{b_i}^2)}}{2\pi\sigma_{s_i}\sigma_{b_i}} \dfrac{e^{-(s'+b')}(s'+b')^j}{j!}}{\int_0^\infty ds' \int_0^\infty db' \dfrac{e^{-((s'-s_i)^2/2\sigma_{s_i}^2 + (b'-b_i)^2/2\sigma_{b_i}^2)}}{2\pi\sigma_{s_i}\sigma_{b_i}}} \tag{12}$$

which is used in each entry in the list of Eq. (7). While the denominator is a product of error functions, the numerator may be computed numerically. When computing the PDF of X for the background-only case, the averages are only done over the background variation.

To extend this to the multichannel case, additionally the test statistic must be averaged over the systematic variations because it, too, depends on s_i and b_i:

$$X_i^j \to \frac{\int_0^\infty ds' \int_0^\infty db' \dfrac{e^{-((s'-s_i)^2/2\sigma_{s_i}^2 + (b'-b_i)^2/2\sigma_{b_i}^2)}}{2\pi\sigma_{s_i}\sigma_{b_i}} X_i^j}{\int_0^\infty ds' \int_0^\infty db' \dfrac{e^{-((s'-s_i)^2/2\sigma_{s_i}^2 + (b'-b_i)^2/2\sigma_{b_i}^2)}}{2\pi\sigma_{s_i}\sigma_{b_i}}} \tag{13}$$

This average is also computed numerically. It is computed both when the sum over all possible experimental outcomes is performed and when the test statistic is computed for the data candidates, ensuring that the data outcome is identical with one of the possible outcomes in the PDF tables. This is important for confidence levels computed with a single channel, when all outcomes are listed in the PDF table.

5. Numerical examples

The above algorithm has been tested in a variety of ways. A program implementing the algorithm is available.[1]

- If a single channel has 3.0 expected signal events, no expected background events, and no observed candidates, then $CL_s = 4.9787\%$ as expected from an exact computation. $CL_b = 1.0$ in this case. For experiments with few possible outcomes, this technique yields exact CLs.
- If this single channel is broken up into arbitrarily many pieces (say, a few hundred), equally dividing up the three expected signal events, each with no background or candidates, the limit is the same as that for the single channel.
- If a channel with no expected signal, but some expected background (and corresponding data candidates) is added to the combination, then CL_s is not changed significantly, while CL_{s+b} and CL_b reflect the relationship between the expected background and the observed candidate count.

[1] For general use, a program implementing the algorithm described in this article is available at http://home.cern.ch/ ~ thomasj/searchlimits/ecl.html.

- A more realistic example requiring the binning of search results and combination of those bins has been explored by simulating a typical search for the Higgs boson (or any new particle) in high-energy particle collisions, where the mass of each observed candidate may be reconstructed from measured quantities. The mock experiment has an expected background of 4 events, uniformly distributed from 0 to 100 GeV/c^2 in the reconstructed mass. The resolution of the reconstructed mass of signal events, were a signal to exist, decreases linearly from 10.5 GeV/c^2 at $m_H = 10$ Gev/c^2 to 3.3 GeV/c^2 at $m_H = 80$ GeV/c^2, where m_H is the mass of the Higgs boson (or other new particle). In a real search, the signal resolutions and background levels are typically obtained from Monte Carlo simulations. Three candidates were introduced with measured masses of 34, 35, and 55 GeV/c^2.

To explore the limits one may set on Higgs production, the space of possible values of m_H was explored from 10 to 70 GeV/c^2, and the total expected signal count was studied between 2 and 6.5 events. For each pair of m_H and the signal count, histograms of the expected signal and background were formed in fine bins from 0 to 100 GeV/c^2. The candidates were also histogrammed using the same binning as the signal and background. Each bin of these histograms was considered a separate search channel, and the confidence level CL_s was formed.

The 95% CL upper limits ($CL_s < 0.05$) on the signal $s = \sum_{i=1}^{n} s_i$ are shown in Fig. 1 for two choices of the test statistic X_i: the likelihood ratio of Eq. (2), and the test statistic $X_i = d_i s_i / b_i$. This latter test statistic is the event count weighted by the signal/background ratio, and it is combined additively from channel to channel.

The two test statistics perform differently under these circumstances, and the method described in this article can be used to evaluate the effects of changing the test statistic. The expected confidence levels $\langle CL_{s+b} \rangle$ and $\langle CL_s \rangle$ provide discrimination of which test statistic is the best choice.

- The probability coverage of the techinique was explored by testing to see how often a true signal would be excluded at the 95% CL. The same mock experiment as described above was used,

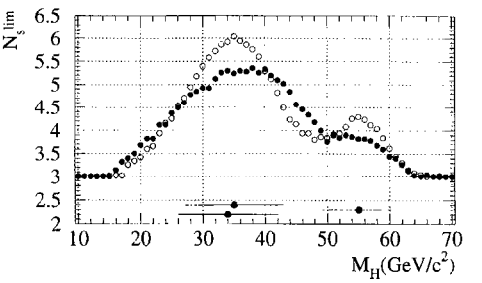

Fig. 1. The 95% CL upper bound on the number of events as a function of a hypothetical Higgs mass, using two test statistics, the likelihood ratio (filled circles) and events weighted by s_i/b_i (empty circles). Candidates are shown with their respective mass resolutions at the bottom of the figure. The total background is four events expected to be uniformly distributed from zero to 100 GeV/c^2.

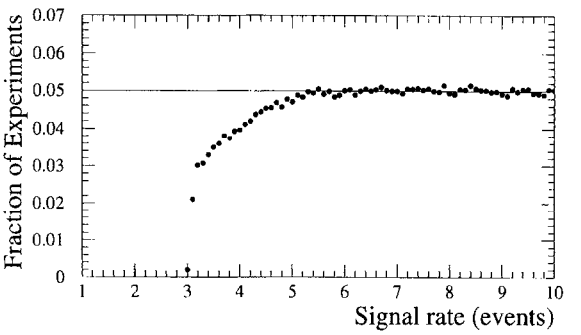

Fig. 2. The false exclusion rate for the mock Higgs search experiment in the presence of a real signal at $m_H = 77$ GeV /c^2, for 95% CL computation. The error bars are hidden within the plot symbols. If a pure frequentist approach were taken (using CL_{s+b}), then the false exclusion probability would be flat at 5%.

but the candidates were distributed according to a signal + background expectation with signal levels varying from 3 events to 10 events, with a true mass of 77 GeV/c^2. Many experiments were simulated with different populations of candidates according to the hypothesis, and the probability of excluding a true signal, hypothesized to have the same strength as was used to simulate the experiments, at 95% CL is shown in Fig. 2. The exclusion fraction is smaller than 5% for low expected signal rates, a consequence of the use of the Bayesian $CL_s = CL_{s+b}/CL_b$, where some of the exclusion power is lost by dividing by CL_b. Alternatively, one may use

Table 1
Reproduction of Table 1 of Ref. [11], together with the computation of the same quantity using the method of this article. Listed are the 90% CL upper limits on the signal for a single counting measurement with no background, no uncertainty on the background, and n candidates. The relative uncertainty on the signal is $\sigma_r = \sigma_s/s$. The Monte Carlo column (MC) is also from Ref. [11]. The missing entry in the column for Eq. (17a) has a square root of a negative argument, indicating that the expansion used to derive the formula has reached its limit of validity

n	σ_r	MC	Eq. (17a)	This work
0	0.00	2.30	2.30	2.30
	0.10	2.33	2.33	2.33
	0.20	2.42	2.41	2.42
	0.30	2.60	2.58	2.61
1	0.00	3.89	3.89	3.89
	0.10	3.94	3.95	3.95
	0.20	4.13	4.14	4.14
	0.30	4.51	4.57	4.53
2	0.00	5.32	5.32	5.32
	0.10	5.41	5.41	5.42
	0.20	5.71	5.72	5.71
	0.30	6.30	6.78	6.32
3	0.00	6.68	6.68	6.68
	0.10	6.80	6.81	6.81
	0.20	7.21	7.27	7.22
	0.30	8.05	—	8.05

CL_{s+b} exclusively, which would give the proper limit. In the latter case, the sensitivity $\langle CL_{s+b} \rangle$ should be quoted with experimental results as well to cover the case of much fewer candidate events than the background expectation, giving a more stringent limit than would be warranted by the sensitivity of the experiment.

- For combining the search results from four LEP experiments for the MSSM Higgs, nearly 100 separate search analyses from different energies, performed by different collaborations, have been combined using this technique. For a model point with m_h and m_A near the exclusion limit for the combined data from 1997 and before, this method computes $CL_s = 5.380\%$, while an exact computation yields $CL_s = 5.332\%$, both corresponding to an exclusion not quite at the 95% level. For this test, the bin width for the PDF of X was 0.03% above probabilities of 1%, and 20 bins per decade below 1%.
- To test the correctness of the strategy for handling systematic uncertainty in the signal, the results of Table 1 in Ref. [9] have been reproduced. In all cases, the Monte Carlo confidence levels of Ref. [9] were reproduced at least as well as by Eq. (17*a*) in the same paper. This equation is

$$U_n = U_{n0}[1 + \{1 - (1 - \sigma_r^2 E_n^2)^{1/2}\}/E_n] \qquad (14)$$

where U_n is the upper limit, including the effects of systematic uncertainty, on the signal at a desired CL if n candidate events were observed in the data, U_{n0} is the upper limit on the signal at the same CL without the effects of systematic uncertainty, σ_r is the relative uncertainty on the signal (e.g., from uncertainty on the efficiency or luminosity), and $E_n \equiv U_{n0} - n$. The results of this test are shown in Table 1.

6. Limitations

Because the binning of the PDF of the test statistic X has a finite resolution, experimental outcomes with very small probabilities of occurring are not represented correctly. When using the conservative choice of filling the bins described above, these outcomes are overrepresented in the final outcome. For the purposes of discovery, however, this approach is not conservative. When computing the CL for a potential discovery, one must compute the sum of probabilities of fluctuations of the background giving results that look at least as much like the signal as the observed candidates, or more. Conversely, one may add up all the probabilities for outcomes less signal-like than observed and subtract it from unity. This involves precise accounting of many outcomes with small probabilities, and the approximation presented here will not suffice. The most useful case for this technique is in forming CL limits near the traditional 90%, 95%, and 99% levels.

Another limitation is that correlations between the systematic uncertainties of different search channels are not incorporated. If the results of a search are binned in a discriminant variable, the signal estimations in neighboring bins may share common uncertainties, as may the background estimations. Similarly, if several experimental

collaborations perform similar searches using similar models for the signal and background, then their results will share common systematic uncertainties. A Monte Carlo computation of the confidence levels is needed when the effects of correlated errors are expected to be large. The effect can be estimated by replacing blocks of correlated parameters s_i and b_i with biased values and recomputing the confidence levels.

The technique described in this article also requires that the value of the test statistic is defined for each single-bin counting search channel, and that these test statistics may be combined to form a joint test statistic.[2] More complicated test statistics which cannot be separated into contributions from independent channels cannot be used with this technique. A Monte Carlo approach is suggested in order to use such test statistics. The likelihood ratio test statistic of Eq. (2), because it combines multiplicatively, is well suited for this technique.

Special care has to be taken in the case that candidate events can have more than one interpretation. A single event may appear in more than one bin of an analysis or may appear in two separate analyses due to ambiguities in reconstruction or interpretation. The most rigorous treatment of such cases is to construct search bins which contain mutually exclusive subsets of the search results. For example, one may wish combine three counting channels, A, B, and C, and candidate events may be classified as passing the requirements of A, B, or C separately, while some may pass the requirements of both A and B, or both A and C, etc. In this case, one would construct seven exclusive classification bins, A, B, C, AB, AC, BC, and ABC, and proceed as before. In general, if a combination has a total of n bins, then there are $2^n - 1$ possible classifications of each event if multiple interpretations are allowed. The nature of the analyses will necessarily reduce the size of this possible overlap problem, and only cases in which significant overlap is expected for signal or background events need to be considered.

7. Summary

An efficient technique for computing confidence levels for exclusion of small signals when combining a large number of counting experiments has been presented. The results of sophisticated channels with reconstructed discriminating variables are binned and the separate bins are treated as independent search channels for combination. A variety of test statistics may be used to evaluate their effects on the confidence levels. The approximate confidence levels obtained are very close to the values of computationally intensive direct summations of probabilities of all final outcomes, or to those obtained by Monte Carlo simulations, and the accuracy of the approximation is adjustable. The confidence levels are either exact or more conservative than the true values from explicit summation. Average expected confidence levels may easily be calculated from the results, and the probability distributions of the test statistic may be used to construct confidence belts using the techniques described in Ref. [7]. Uncorrelated systematic uncertainties in the signal and background models are incorporated in a natural manner. Monte Carlo alternatives are suggested when the effects of correlated systematic uncertainties are expected to be large and in the case of potential discoveries. This technique is useful for efficiently scanning many possible models for production of signals with different signatures and combining the results of searches sensitive to these different signatures.

[2] The combination rule for the test statistic needs to be associative in order for the iterative combination of one search channel to a list of combined results of other search channels to be well defined. The combination rule also needs to be commutative so that the order in which the combination is performed does not affect the outcome.

References

[1] A.G. Frodesen, O. Skeggestad, H. Tøfte, Probability and Statistics in Particle Physics, Universitetsforlaget, 1979, ISBN 82-00-01906-3.

[2] W.T. Eadie, D. Drijard, F.E. James, M. Roos, B. Sadoulet, Statistical Methods in Experimental Physics, North-Holland, 1971, ISBN 0 7204 02395.

[3] Maurice G. Kendall, Alan Stuart, The Advanced Theory of Statistics, Vol. 2, 2nd Edition, Charles Griffen & Company Limited, London, 1967, SBN 85264 011 0 (Chapter 24).
[4] B. Escoubes, S. De Unamuno, O. Helene, Nucl. Instr. and Meth. A257 (1987) 346.
[5] O. Helene, Nucl. Instr. and Meth. A228 (1984) 120.
[6] O. Helene, Nucl. Instr. and Meth. A212 (1983) 319.
[7] Particle Data Group, R.M. Barnett et al., Phys. Rev. D54 (1996) 1 and earlier editions.
[8] A. Favara, M. Pieri, Confidence Level Estimation and Analysis Optimization, hep-ex/9706016, 1997.
[9] G.J. Feldman, R.D. Cousins, Phys. Rev. D57 (1988) 3873.
[10] S.S. Wilks, Mathematical Statistics, Princeton University Press, Princeton, NJ, 1943.
[11] R.D. Cousins, V.L. Highland, Nucl. Instr. and Meth. A 320 (1992) 331.
[12] C. Giunti, Treatment of the Background Error in the Statistical Analysis of Poisson Processes, hep-ex/9901015, 1998.