OPAL Technical Note TN619
August 12, 1999

A Simple Figure of Merit for Searches

Benno List
CERN–EP

12.8.1998

Abstract

In searches for new particles it is a common problem to determine the optimal value of cuts, for instance the final cut on some likelihood. Given the signal efficiency ϵ and the expected number of background events B as function of the cut, it is possible to determine the expected number of signal events that can be excluded. In this note, a simple figure of merit (FoM), given by $\mathrm{FoM} = 2.1\sqrt{B + 2.04}/\epsilon$, is proposed. This FoM is an estimate for the median of the number of signal events that can be excluded by the analysis if the likelihood ratio method is used, so that a smaller value of FoM corresponds to a better analysis.

http://www.cern.ch/opal/doc/technote/tn619/tn619.ps.gz

1 Introduction

To optimize a selection used in the search for a new particle one needs, in general, some figure of merit which tells whether a given selection is better or worse than another. This is especially important when one wants to optimize cut values, either in a cut–based selection, or at the end of a likelihood or neural network selection, if one wants to cut on the output of the likelihood or the artificial neural network.

In this note it is assumed that the search is not successful, i. e. no significant signal is observed, and therefore an upper limit is calculated. In any case one usually counts the number of events that survive the selection, then calculates an upper limit of observed signal events, which is translated into some upper limit of a cross section (or a coupling, a branching ratio or similar) by using the selection efficiency.

Generally, two numbers characterize the the performance of a selection: The selection efficiency for the signal events, and the expected number of background events that pass the selection. To decide which one of two selections is preferrable, one needs a single number that indicates the performance of a selection, a figure of merit.

The selection efficiency and expected number of background events passing the selection can be used to calculate the expected number of signal events that can be excluded at a certain (usually 95 %) confidence level. Since the number of excluded events is directly proportional to the number that is extracted from the analysis (cross section, coupling, branching ratio etc.), this is clearly the number that should be minimized, i. e. a good candidate for a figure of merit.

This calculation is, however, in general complicated, and a number of ways exist for the calculation of the limit.

Therefore in this note a simple approximation formula is proposed which can be used as figure of merit.

2 Definition of the Figure of Merit (FoM)

The proposed Figure of Merit (FoM) is defined as

$$\mathrm{FoM} = \frac{2.1\sqrt{B + 2.04}}{\epsilon}, \tag{1}$$

where B is the expected number of background events that pass the selection, and ϵ is the selection efficiency for background events.

The FoM is defined in such a way that it gives a reasonable approximation to the number of signal events that can be excluded at 95 % confidence level for half of all experiments performed under identical conditions, using the likelihood ratio method.

3 Details

3.1 The likelihood ratio method

The likelihood ratio method [Hel83] defines for a given number of observed events N_{obs} and a number of expected background events B the number of signal events S_{95} that can be excluded at a confidence level α as:

$$1 - \alpha = \frac{\sum\limits_{n=0}^{N_{\mathrm{obs}}} \mathcal{P}\left(n,\, S_\alpha + B\right)}{\sum\limits_{n=0}^{N_{\mathrm{obs}}} \mathcal{P}\left(n,\, B\right)} \tag{2}$$

$$= \frac{\exp\left(-[S_\alpha + B]\right) \sum\limits_{n=0}^{N_{\mathrm{obs}}} \frac{(S_\alpha+B)^{N_{\mathrm{obs}}}}{n!}}{\exp\left(-B\right) \sum\limits_{n=0}^{N_{\mathrm{obs}}} \frac{B^{N_{\mathrm{obs}}}}{n!}}, \tag{3}$$

where $\mathcal{P}\left(n,\, \lambda\right) = \frac{\lambda^n}{n!} \exp\left(-\lambda\right)$ is the Poisson probability to observe n events when λ events are expected. That is, the likelihood ratio method excludes a number S_α of signal events that is so large that the probability to observe at most as many events as the number N_{obs} found in the experiment, given the

```
      FUNCTION PLIMIT (N, B, CL)
      INTEGER N
      REAL B, CL

      REAL*8 TERM, CLB, CLS, CLS1, F, F1, S, DS

      TERM = EXP (-B)
      CLB = TERM
      DO I = 1, N
        TERM = TERM * B/I
        CLB = CLB + TERM
      END DO
      S = 2.2*SQRT (MAX(1.*N,B)+1.86)
10    CONTINUE
        TERM = EXP (-(B+S))
        CLS = TERM
        CLS1 = 0.
        DO I = 1, N
          CLS1 = CLS1 + TERM
          TERM = TERM * (B+S)/I
          CLS = CLS + TERM
        END DO
        F = CLS - CL*CLB
        F1 = CLS1 - (S+B)*CLS
        DS = F/F1
        IF (DS .LT. S) THEN
          S = S - DS
        ELSE
          S = 0.5*S
        ENDIF
      IF (ABS (F) .GT. 0.0001*CLB) GO TO 10
      PLIMIT = S

      RETURN
      END
```

Figure 1: The function `PLIMIT` returns the limit given by the likelihood ratio mathod for a given number `N` of observed events and `B` expected background events at a confidence level `CL`$= 1 - \alpha$.

signal plus background hypothesis, is $1 - \alpha$ times more unlikely than the probability to observe at most as many events expecting only background events.

To calculate the actual number S_α of excluded events, given B and α, one has to solve the equation numerically, e. g. by the Newton method (Fig. 1 gives a simple Fortran function which does this).

This definition of an upper limit for Poisson processes in the presence was recommended by the Particle Data Group until 1996 [PDG96, p. 159]. In 1998, the PDG adopted [PDG98, p. 172] the unified approach of [Fel98].

3.2 Possible Figures of Merit

Starting from the likelihood ratio formula, several possibilities arise to define a number that characterises the number of signal events that can be excluded:

1. The number of events S_{95} that can be excluded in an experiment at 95 % confidence level if the number of observed events N_{obs} is exactly the number of expected background events B.

2. Taking an ensemble of experiments with the same number of expected background events B, the mean number of excluded events $\langle S_{95} \rangle$ is calculated.

   In this calculation it is assumed that the number of observed events is distributed according to a POISSON distribution with a parameter B, i. e. that the background only hypothesis is true.

3. Taking an ensemble of experiments with the same number of expected background events B, it is determined, how many events can be excluded in at least half of the experiments (i. e. the median of the distribution of excluded event numbers).

   If the median $S_{\text{med},95}$ lies between two numbers, it is interpolated linearly.

Possibility (1) has the disadvantage that it works only for an integral number of expected background events.

Possibility (2) und (3) represent the mean and the median of the distribution of exclusion limits obtained for a large number of identical events, and it is somewhat a matter of taste which number one prefers.

The proposed FoM approximates (3), i. e. the median, among other reasons because it is easier to approximate. Also it is felt that the median is a more robust estimator of the expectation one has concerning the limit than the mean, because the distribion of the actual limits obtained is highly asymmetric for small numbers of expected background events.

3.3 The Form of the FoM

It is well known that for a large parameter λ the POISSON distribution becomes a GAUSSian, with mean and variance λ.

Therefore, for sufficiently large numbers of expected background events, one expects

$$S_{95} \propto \sqrt{\lambda}, \tag{4}$$

with the proportionality constant around 2.

For small numbers of expected background events, more and more experiments will have zero observed events, in which case the limit becomes simply

$$S_\alpha = \ln \frac{1}{1-\alpha}, \tag{5}$$

i. e. $S_{95} = \ln(20) = 2.996$.

This means that, no matter how low the expected background is, one can never exclude less than three (or, more precisely, 2.996) signal events at 95 %

confidence level[1].

Therefore, the following ansatz was chosen:

$$S_{95} \approx a\sqrt{B+b}, \tag{6}$$

with two constants a and b which are chosen as to yield an acceptable approximation to the true number of excluded events. Additionally, b is subjected to the constraint that

$$S_{95}(B=0) = 2.996, \tag{7}$$

so that one has

$$b = (2.996/a)^2. \tag{8}$$

3.4 Calculation of Expected Limits

Using a special program, the numbers defined in section 3.2 were evaluated for different amounts of expected background, and compared to the following approximations:

$$\begin{aligned} S_{95} &\approx 2.3\sqrt{B+1.70} && (9)\\ \langle S_{95}\rangle &\approx 2.4\sqrt{B+1.56} && (10)\\ S_{\mathrm{med},95} &\approx 2.1\sqrt{B+2.04} && (11) \end{aligned}$$

To compute the mean and median of the distribution of the limits, the limit was calculated for each number of observed events, and weighted with the Posisson probability that this number of events is actually observed, assuming that the expected number of background events is correct. If the median lies between two (necessarily integer) numbers of events, it is interpolated. For instance, if for 33.3 % of all events the exclusion limit was 4.4 events or better, and for 67.7 % of events 5.0 events or worse, the median would be defined as 4.8 events.

The results are collected in table 1. It can be seen that the apprximations presented here describe the various expectations to within 10 percent, in fact more precisely for many values of the expected background.

[1]Actually, for the likelihood ratio method the exclusion limit for zero observed events is *independent* of the expected background. This appears to be sensible, because obviously, if one has observed zero events one knows that one has observed exactly zero background events and zero signal events, so no matter how many background events were expected, the conclusion regarding the signal events remains valid.

If one observes, e. g., one event, the situation is different: The event may be a signal or a background event—assuming that this distinction makes any sense, that is if interferences can be neglected. It is reasonable to assume that the probability for the observed event to come from the background is higher if the expected background is, say, 3 events, compared to a situation where only 0.1 background events are expected, and that therefore the limit is more stringent if the expected number of background events is higbher.

Contrary to this expectation, in the unified approach [Fel98] of Feldman and Cousins, which was adopted by the PDG in 1998 [PDG98, p. 172], the limit set depends on the number of expected background events even in the case of zero observed events, and can in fact become considerably smaller than three. E. g., for two expected background events the upper limit is 1.78.

However, the expected (mean) exclusion limit given by the unified method deviates hardly from the expected limit obtained for the likelihood ratio method.

B	S_{95}	Appr.	Ratio	$\langle S_{95} \rangle$	Appr.	Ratio	Unif	$S_{\mathrm{med},95}$	Appr.	Ratio
0	3.00	3.00	1.001	3.00	3.00	1.000	3.09	3.00	3.00	1.001
1	4.11	3.78	0.919	4.18	3.84	0.918	4.05	3.40	3.66	1.078
2	4.82	4.42	0.918	4.95	4.53	0.914	4.76	4.17	4.22	1.013
3	5.39	4.99	0.925	5.57	5.12	0.921	5.36	4.77	4.71	0.988
4	5.89	5.49	0.933	6.09	5.66	0.929	5.86	5.28	5.16	0.977
5	6.32	5.95	0.941	6.56	6.15	0.937	6.32	5.74	5.57	0.971
10	8.08	7.87	0.974	8.43	8.16	0.968	8.13	7.53	7.29	0.967
15	9.44	9.40	0.995	9.87	9.77	0.989	9.54	8.92	8.67	0.972
20	10.60	10.71	1.011	11.09	11.14	1.004		10.09	9.86	0.977
25	11.62	11.88	1.023	12.17	12.37	1.016		11.12	10.92	0.982
30	12.55	12.95	1.032	13.15	13.48	1.025		12.05	11.89	0.987
35	13.40	13.93	1.040	14.05	14.51	1.033		12.91	12.78	0.990

Table 1: Results for limits and the corresponding approximations. The column labeled "Appr." contains the approximations as defined by equations (9), (10), and (11), and the column "Ratio" gives the ratio of the approximation to the true values. The column "Unif." contains the expected confidence limit using the "unified approach" of [Fel98].

3.5 Introduction of the FoM Itself

After an approximation has been obtained for the expected number of events that can be excluded in a counting experiment, given a certain number of expected background events, it is straightforward to define the figure of merit: The FoM is defined as the approximation to the median of the number of *signal* events which can be excluded at 95 % confidence level, therefore we have

$$\mathrm{FoM} \approx \frac{S_{\mathrm{med},95}}{\epsilon} \tag{12}$$

$$\mathrm{FoM} = \frac{2.1\sqrt{B+2.04}}{\epsilon}. \tag{13}$$

It should be noted the the FoM does not depend on the (a priori unknown) signal cross section, but it does depend on the absolute number of expected background events. Therefore the optimal working point for a search depends on the luminosity used for the analysis, and in principle a cut which is optimal for the complete dataset need not be optimal for a small portion of the dataset. This is a situation commonly encountered when a seach which was optimized for a given dataset (say, the 189 GeV data corresponding to 180 pb^{-1}) is applied to the first data of the next year, which ususally corresponds to much less luminosity, say 20 pb^{-1}.

4 Other Methods to Calculate Limits

In the last years much more sophisticated methods than the likelihood ratio method have been introduced in the calculation of limits, *e. g.* the fractional event counting method [Boc97] of Peter Bock. This method weights events depending to the value of an additional variable, *e. g.* the mass of a candidate event, according to the signal over background ratio for the particular value of the variable (which means that events with a candidate mass close to the value where the mass distribution is expected to peak get a higher weight than far away from the peak).

If such a method is employed, the FoM is not a good estimate of the expected exclusion power anymore, because often (if signal and background have markedly different distributions of the test variable) the exclusion limit is more stringent than the limit obtained from simple event counting. However, the FoM may still be used as an indicator of a good cut value.

A typical example is the Higgs search [OPA99]; basically all analyses calculate a likelihood and perform a cut on it. Then the candidate mass is used as test variable in the limit calculation. If we now assume that the shape of the background and signal mass distributions does not depend on the likelihood cut (or only weakly), then we can assume that the exclusion power of a particular analysis is still better if the FoM is lower.

However, the question how to optimize an analysis which is used in a combination with other analyses is not easily answered. Consider the case where of an analysis with relatively little background. Since a single analysis can never exclude less than three signal events, the FoM improves very little if the expected background is less tan around one event, and therefore "prefers" a cut which leads to higher efficiencies, at the price of a lower purity. This may not be the optimal choice if several such channels are combined. In such a situation it may be necessary to optimize each analysis by caculating the expected limit for the combination of all channels and optimizing this limit by changing the cut value of the channel under consideration. However, for channels which have a much larger expected rate than others (*e. g.* the four–jet channel in the Higgs analysis) than all others the optimal cut is little influenced by the other channels and the FoM is a good estimator of the optimal cut value.

5 Using the FoM with the PC Package

The PC (Projections and Correlations) package [Kar97] is a very useful tool to interactively optiize likelihood selections. Among other things it allows to automatically scan a set of likelihoods which emerge if one likelihood variable at a time is dropped from the set of likelihood variables. The only drawback of the PC package in its curent (PC103) incarnation is that it judges the relative merit of different likelihood selections by the product of efficiency times purity, which is directly related to the expected statistical error of a cross section measurement in the presence of background. This quantity is however not the quantity to optimize in the case of a search, but the proposed figure of merit or something similar should be used. The PC package can relatively easily be adapted to the use of any figure of merit, and a private verion exists which incorporates the FoM given in this note. This version may be obtained from the author.

References

[Boc97] PETER BOCK (1997): Determination of exclusion limits for particle production using different decay channels with different efficiencies, mass resolutions and backgrounds. — *Updated version of Heidelberg preprint HD–PY/96–05, unpublished.*

[Fel98] GARY J. FELDMAN & R. D. COUSINS (1998): Unified approach to the classical statistical analysis of small signals. — *Phys. Rev.* ***D57****, 3873–3889.*

[Hel83] O. HELENE (1983): Upper limit of peak area. — *Nucl. Instr. Meth.* ***212****, 319.*

[Kar97] DEAR KARLEN & ERIC TORRENCE (1997): PC102 manual. — *http://www.cern.ch/opal/manuals/pc/pro/pc.ps. Unpublished program manual.*

[OPA99] OPAL Collaboration, G. ABBIENDI et al. (1999): Search for neutral Higgs bosons in e^+e^- collisions at $\sqrt{s} \approx 189\,\mathrm{GeV}$. — *CERN preprint CERN–EP/99–096. Electronic preprint* ***hep–ex/9908002.***

[PDG96] Paticle Data Group, R. M. BARNETT et al. (1996): Review of particle physics. — *Phys. Rev.* ***D54****, 1–720.*

[PDG98] Paticle Data Group, C. CASO et al. (1998): Review of particle physics. — *Eur. Phys. J.* ***C3****, 1–794.*