The signal estimator limit setting method

Shan JIN [a*]

[a]Department of Physics, University of Wisconsin–Madison, Madison, WI 53706, USA

In this paper, a new method of background subtraction is presented which uses the concept of a signal estimator to construct a confidence level which is always conservative and which is never better than e^{-s}. The new method yields stronger exclusions than the Bayesian method with a flat prior distribution.

1. MOTIVATION

We wish to determine if a given signal hypothesis s is excluded. If we could know the observed test-statistic based on events truly from signal only, which we refer to as the signal estimator $(\epsilon_s)_{obs}$, the confidence level would be rigorously defined as

$$CL \equiv 1 - c \equiv 1 - \mathcal{P}(\epsilon_s \leq (\epsilon_s)_{obs}) \tag{1}$$

where $\mathcal{P}(\epsilon_s \leq (\epsilon_s)_{obs})$ is the probability that an experiment with signal expectation s yields a value of the signal estimator less than or equal to $(\epsilon_s)_{obs}$.

Unfortunately, we cannot directly know $(\epsilon_s)_{obs}$ from an experiment as it is not possible to unambiguously determine if an event comes from signal or background. We can only directly know a test-statistic value based on the total observation

$$\epsilon_{obs} = (\epsilon_{s+b})_{obs}. \tag{2}$$

Although it is not possible to know $(\epsilon_s)_{obs}$ directly, it is still possible to produce an estimate of it, with which we can calculate Eq. 1. This is most straightforward for test-statistics of the form

$$\epsilon_{s+b} = \epsilon_s \oplus \epsilon_b \tag{3}$$

where '$\oplus$' represents a sum or product. For example, in simple event counting,

$$\epsilon = n \tag{4}$$

$$n_{s+b} = n_s + n_b. \tag{5}$$

In this case, we can use a Monte Carlo simulation of the background expectation to remove the background contribution in the observed test-statistic $(\epsilon_{s+b})_{obs}$, i.e., to estimate $(\epsilon_s)_{obs}$,

*Corresponding address: CERN/EP Division, CH-1211, Geneva 23, Switzerland. Tel: (41 22) 767 7331; fax: (41 22) 782 8370; email: jin@wisconsin.cern.ch

and to calculate Eq. 1. In each Monte Carlo experiment, the estimate of $(\epsilon_s)_{obs}$ is defined as

$$(\epsilon_s)_{obs} = \begin{cases} \epsilon_{obs} \ominus \epsilon_b & if \epsilon_{obs} \ominus \epsilon_b \geq (\epsilon_s)_{min} \\ (\epsilon_s)_{min} & if \epsilon_{obs} \ominus \epsilon_b \leq (\epsilon_s)_{min} \end{cases} \tag{6}$$

where '$\ominus$' represents difference or division, and $(\epsilon_s)_{min}$ is the minimum possible value of the signal estimator, which corresponds to the physical boundary (zero signal events).

The confidence level can be computed with Monte Carlo methods in the following way for an observed test-statistic ϵ_{obs}. First, generate a set of Monte Carlo experiments with test-statistic values distributed as for experiments with the expected background but no signal to determine a distribution of possible signal estimator values for the observation according to Eq. 6. Next, using a sample of Monte Carlo with test-statistics distributed as for experiments with signal only, and for each possible signal estimator value, calculate

$$c(\epsilon_{obs}, \epsilon_b) = \mathcal{P}(\epsilon_s \leq max[\epsilon_{obs} \ominus \epsilon_b, (\epsilon_s)_{min}]). \tag{7}$$

The value of $c(\epsilon_{obs}, \epsilon_b)$ averaged over all of the signal estimator values determined with background Monte Carlo forms an estimate of $\mathcal{P}(\epsilon_s \leq (\epsilon_s)_{obs})$, or

$$c \equiv \mathcal{P}(\epsilon_s \leq (\epsilon_s)_{obs}) \approx \overline{c(\epsilon_{obs}, \epsilon_b)}. \tag{8}$$

2. CONFIDENCE LEVEL FOR SIMPLE EVENT COUNTING

The Monte Carlo procedure described above is very slow, and without generalization, it can only be used for the class of test-statistics which satisfy Eq. 3. The method can be generalized into a much simpler mathematical format which can be used for any kind of test-statistic. The generalization can best be illustrated with an example. In the case of simple event counting, the boundary condition for the signal estimator can be understood intuitively. For an observation of n_{obs} events, the confidence level is computed by allowing the background to vary freely, and according to Eq. 6, the signal estimator will be

$$(n_s)_{obs} = \begin{cases} n_{obs} - n_b & \text{if } n_{obs} - n_b \geq 0 \\ 0 & \text{if } n_{obs} - n_b \leq 0. \end{cases} \tag{9}$$

Using Eq. 8, one can easily compute the confidence coefficient to be

$$\begin{aligned} c = {} & [\mathcal{P}(n_b = 0) \times \mathcal{P}(n_s \leq n_{obs}) \\ & + \mathcal{P}(n_b = 1) \times \mathcal{P}(n_s \leq n_{obs} - 1) + \ldots \\ & + \mathcal{P}(n_b = m) \times \mathcal{P}(n_s \leq n_{obs} - m) + \ldots \\ & + \mathcal{P}(n_b = n_{obs}) \times \mathcal{P}(n_s \leq 0)] \\ & + \mathcal{P}(n_b \geq n_{obs}) \times \mathcal{P}(n_s \leq 0) \\ & = \mathcal{P}(n_{s+b} \leq n_{obs}) + [1 - \mathcal{P}(n_b \leq n_{obs})] \times e^{-s}. \end{aligned} \tag{10}$$

This probability reduces to $e^{-(s+b)} + (1 - e^{-b})e^{-s} = e^{-s}$ when one observes no candidates, so it is independent of expected background level, and because the confidence level is always strictly greater than $\mathcal{P}(n_{s+b} \leq n_{obs})$, it is conservative.

In order to compare the performances of this method with the Bayesian method[1], the confidence levels for a simple experiment are analyzed in Fig. 1. For this example, the analysis is assumed to expect three events from a possible signal, and three events from Standard Model background processes. For both methods, when zero events are observed, the confidence level reduces to e^{-s} while for observations of more events, the signal estimator method yields a lower confidence coefficient, and thus a better exclusion confidence level. For large numbers of events, $\mathcal{P}(n_b \leq n_{obs})$ approaches one, meaning that both methods approach the classical confidence level and give very similar results.

3. GENERALIZED FORMULA

This method can then be generalized to any kinds test-statistics, to include discriminating variables. The natural generalization takes the form

$$c = \mathcal{P}(\epsilon_{s+b} \leq \epsilon_{obs}) + [1 - \mathcal{P}(\epsilon_b \leq \epsilon_{obs})] \times e^{-s}. \tag{11}$$

For an observation of zero events, the generalized method continues to give a confidence level e^{-s}, and the confidence level computed with this method is always conservative, with c strictly greater than $\mathcal{P}(\epsilon_{s+b} \leq \epsilon_{obs})$.

4. CONCLUSION

The Signal Estimator method proposed in this paper is a conservative limit setting method. For an observation of zero events, this method gives a confidence level e^{-s}, which is independent of number of expected background events. The Signal Estimator method yields better limits than the Bayesian method, which should result in an increase in search senstitivity.

REFERENCES

1. O. Helene, Nucl. Instr. and Meth. **212** (1983) 319.

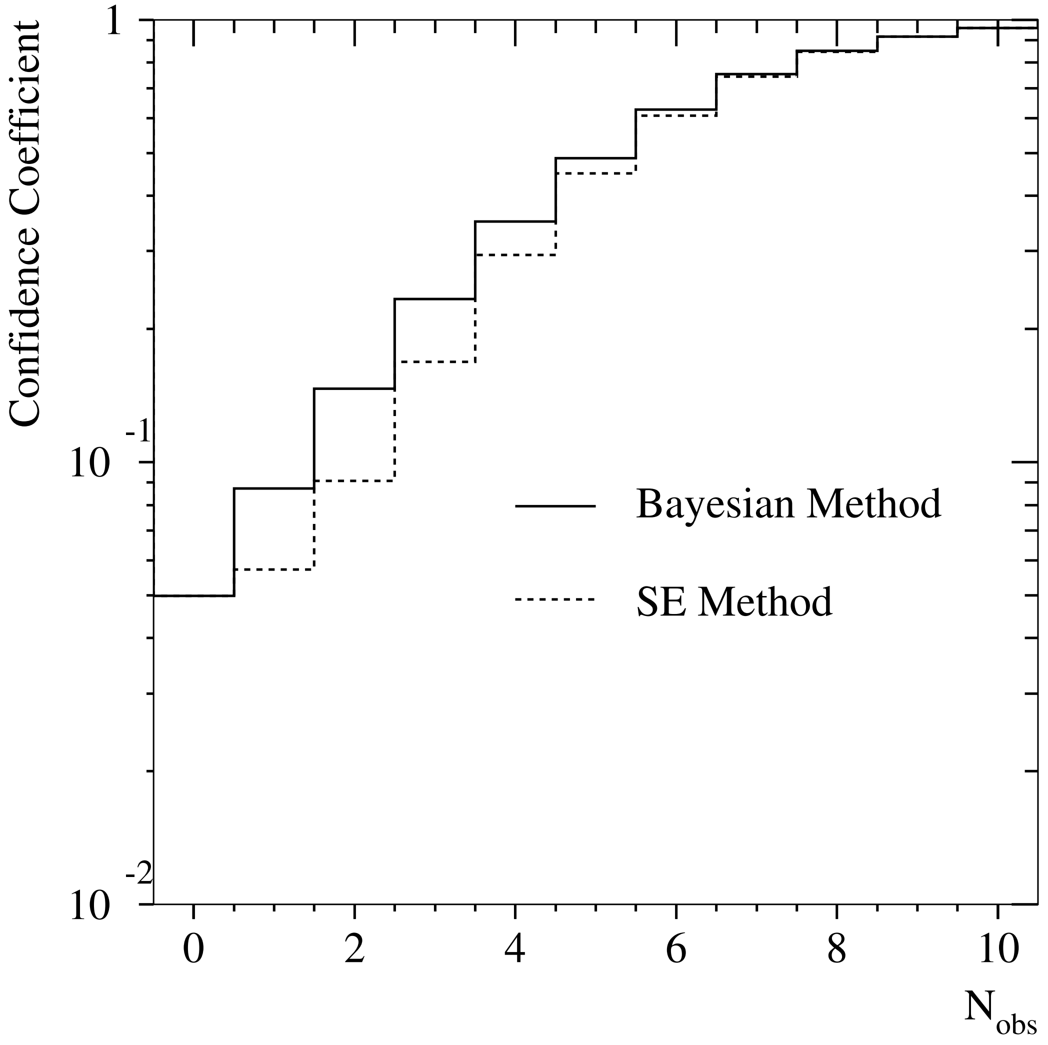

Figure 1. A comparison of Signal Estimator method performance to the Bayesian method performance. For an experiment with three signal and three background events expected, the confidence levels are shown for different numbers of observed events. The Signal Estimator method gives either an equal or better confidence level for all possible observations.