Fermi National Accelerator Laboratory

FERMILAB-TM-2065

A Method to Calculate Limits in the Absence of a Reliable Background Subtraction

S. Geer

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

March 1999

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CH03000 with the United States Department of Energy

FERMILAB-TM-2065
APEX-Note 19
February 22, 1999

A method to calculate limits in the absence of a reliable background subtraction

S. Geer
Fermi National Accelerator Laboratory, Batavia, Illinois 60510

Abstract

We describe a method to calculate upper limits when there are background events, but no reliable background subtraction can be made. The method avoids using a final arbitrary cut to remove the remaining events.

1 Introduction

Searches for rare phenomena are common in particle physics. Typically, the search is for a new resonant state or a new decay mode that yields an enhancement in an appropriate kinematic distribution. The analysis proceeds by defining a set of requirements that would be satisfied by events originating from the signal process, but reject most, and hopefully all, of the events arising from background processes. Sometimes events survive the selection all the way to the "final" kinematic distribution, but do not populate the region expected for "signal". If no reliable background subtraction can be made, then upper limits on the rate at which the signal process occurs can be calculated after placing a somewhat arbitrary cut on the final distribution to remove some or all of the remaining background events. However, the dependence of the result on the exact position of the final cut is unsatisfactory.

In this note we discuss a method of setting an upper limit on the signal event rate that avoids using an arbitrary cut to remove the events remaining in the final distribution. The method is similar in spirit, but different in detail, to the one described in Ref. [1]. This note is organized as follows. In Section 2 the statistical method is described. In Section 3 the method is compared with a few alternative approaches. Finally, a summary is given in Section 4.

2 The Method

We begin by defining the statistical question we are trying to answer. Consider an experiment in which N events are observed after imposing a set of selection criteria. We can plot the number of surviving events as a function of the magnitude of an appropriate kinematic variable K, and compare the observed distribution with the distribution predicted for our signal–process. We wish to ask, for a given theoretical prediction :

"What is the probability P of observing a final distribution that is more signal–like than the distribution we have actually observed ?"

To make progress, we must define what we mean by "more signal–like". We will proceed by considering the simplest case, in which we have observed only one event ($N = 1$). After this we consider the $N = 2$ case, and then generalize the method to arbitrary N.

2.1 One observed event

Suppose we have observed one event passing our selection criteria. Let the value of the kinematic variable K for this event be K_1. We wish to test the viability of a theory that predicts that on average our experiment should have observed μ signal events with a known distribution $d\mu/dK$ populating the range from K_{min} to K_{max}. If the theory is correct, and if

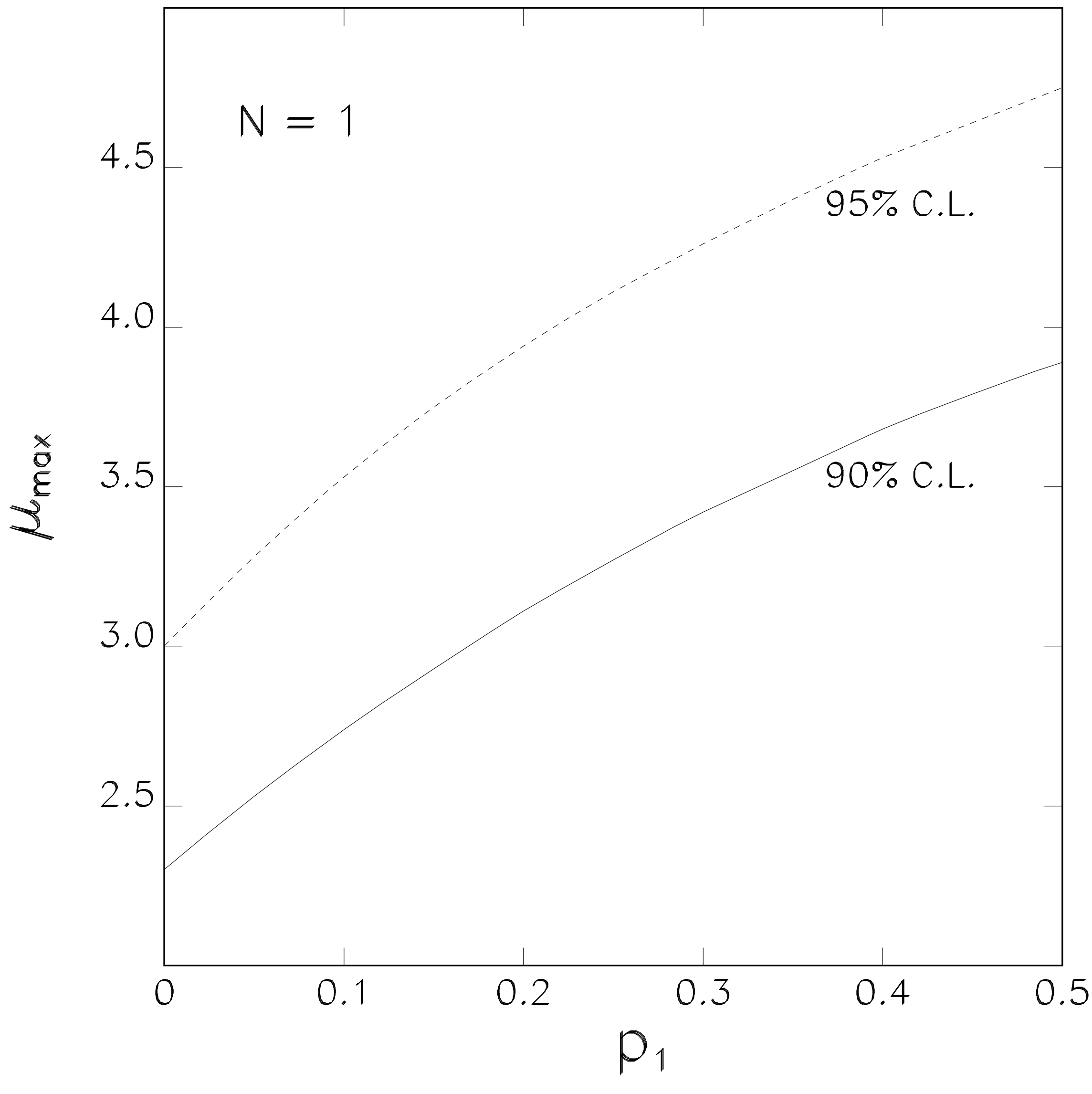

Figure 1: Upper limits μ_{max} on the number of predicted events that can be tolerated at 90% and 95% C.L. when one event is observed ($N = 1$) shown as a function of the single–event integrated probability (p_1).

f_1 is the fraction of signal events predicted to be less "signal–like" than the observed event, then :

$$P \ = 1 \ - \ e^{-\mu} \ - \ f_1 \, \mu e^{-\mu} \ , \tag{1}$$

where the second term is the statistical probability of observing no signal events (Poisson statistics), and the third term is the probability of observing one signal event that is less "signal–like" than the observed event. Note that in the limit $f_1 \to 0$ ($f_1 \to 1$) Eq. 1 reduces to the normal expression corresponding to the observation of $N = 0$ ($N = 1$) events.

To determine how "signal–like" our one observed event is, we introduce p_1 which is defined as the fraction of signal events that are predicted to have values of K in the tail of the distribution beyond K_1 :

$$p_1 \ \equiv \ min \ \left(\frac{1}{\mu} \int_{K_1}^{K_{max}} \frac{d\mu}{dK} \, dK \ , \ \frac{1}{\mu} \int_{K_{min}}^{K_1} \frac{d\mu}{dK} \, dK \right) \ . \tag{2}$$

In the following we will refer to p_1 as the single–event integrated probability. Note that p_1 lies within the range $0 \leq p_1 \leq 0.5$, and the predicted distribution of single–event integrated probabilities is uniform for the signal process. Hence, the fraction of signal events predicted to have single–event integrated probabilities less than p_1 is given by $2\,p_1$. We will use p_1 as our measure of how "signal–like" an event is. Therefore :

$$f_1 \ = \ 2\,p_1 \ , \tag{3}$$

and

$$P \ = 1 \ - \ e^{-\mu} \ - \ 2\,p_1 \, \mu e^{-\mu} \ . \tag{4}$$

If the theory we are testing is correct, then P is the fraction of an ensemble of identical experiments that will yield a more signal–like result than the observation of one event at $K = K_1$. The theory can therefore be excluded with a confidence level (C.L.) of P. To find the maximum value of μ that can be tolerated at a given C.L. we must solve Eq. 4 to find μ_{max}, the value of μ that yields P equal to the desired confidence level. The resulting 90% and 95% C.L. upper limits on μ are shown in Fig. 1 as a function of p_1. As expected, as $p_1 \to 0$ we find $\mu_{max} \to 2.3$ at 90% C.L. and $\mu_{max} \to 3.0$ at 95% C.L. These are just the Poisson upper limits [2] associated with the observation of zero events. Furthermore, if the observed event approaches the bisector of the kinematic distribution ($p_1 \to 0.5$), we find $\mu_{max} \to 3.9$ at 90% C.L. and $\mu_{max} \to 4.7$ at 95% C.L., which are just the Poisson upper limits associated with the observation of one event.

2.2 Two observed events

Suppose we have observed two events passing our selection criteria ($N = 2$). Let the values of K for these events be K_1 and K_2. We define p_i to be the single–event integrated probability

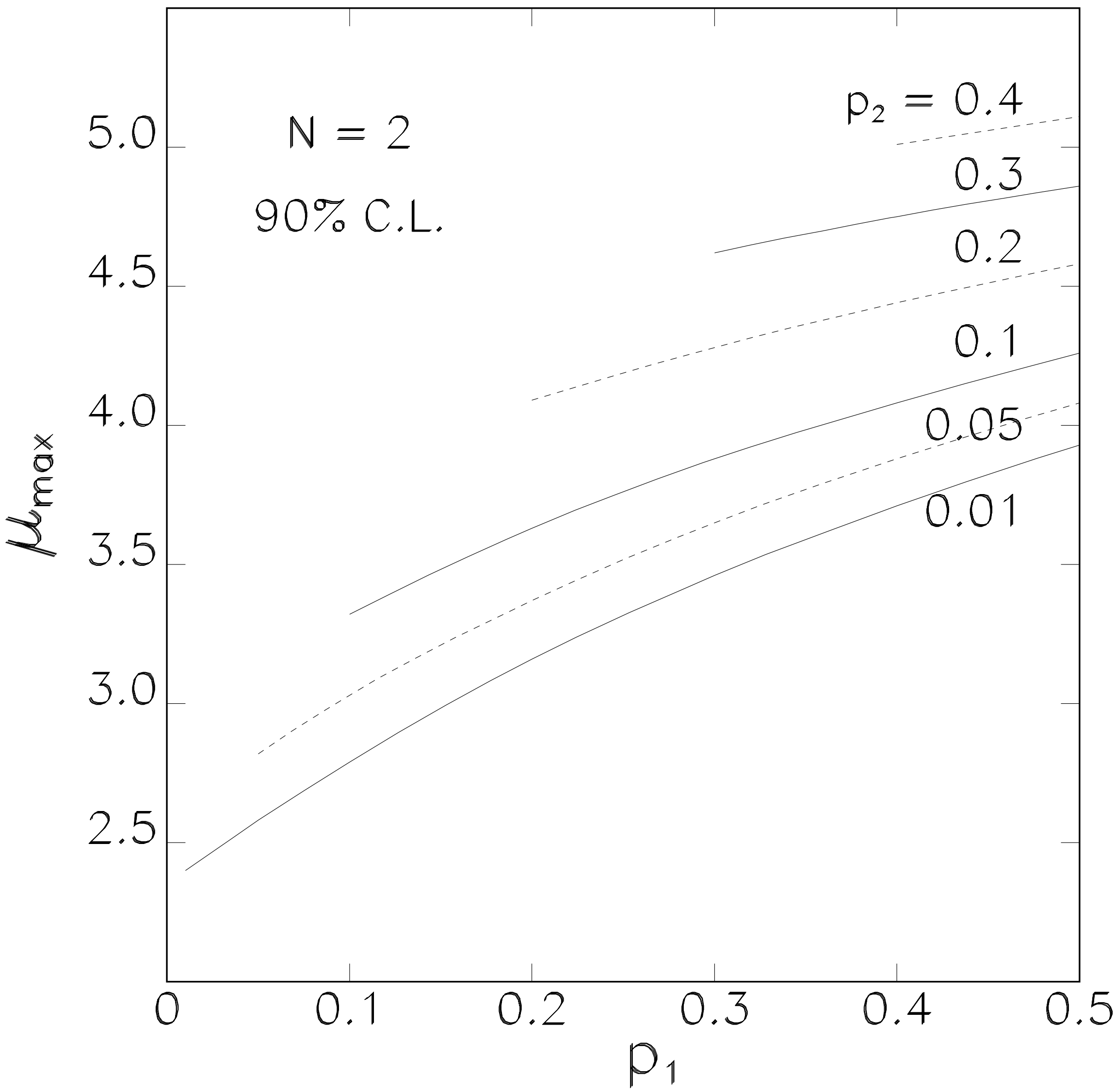

Figure 2: Upper limits μ_{max} on the number of predicted events that can be tolerated at 90% shown as a function of the single–event integrated probabilities (p_1 and p_2) for two observed events ($N = 2$).

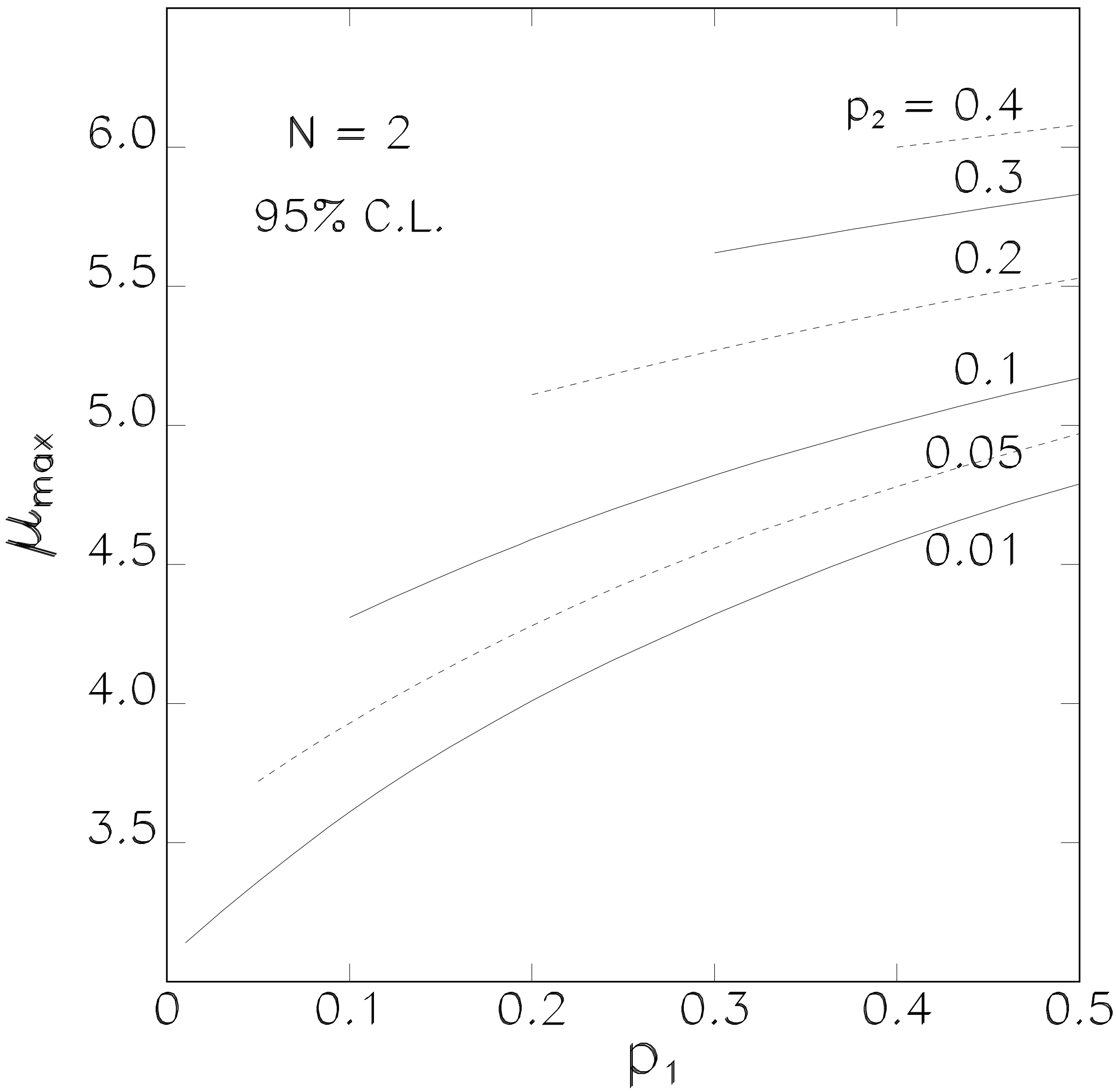

Figure 3: Upper limits μ_{max} on the number of predicted events that can be tolerated at 95% C.L. shown as a function of the single–event integrated probabilities (p_1 and p_2) for two observed events ($N = 2$).

for event i, and order the event labels so that $p_1 > p_2$. Equation 1 now becomes :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; g_1 \,\mu e^{-\mu} \;-\; f_2 \,\frac{\mu^2 e^{-\mu}}{2} \;, \tag{5}$$

where g_1 and f_2 will define what we mean by "more or less signal–like" than the observation of two events at $K = K_1$ and $K = K_2$. There is some freedom in defining what we mean by "signal–like". However, we will require that as $p_2 \to 0$ we recover our one–event result, and hence :

$$\text{As } \; p_2 \to 0 \;, \quad g_1 \to 2\,p_1 \text{ and } f_2 \to 0 \;. \tag{6}$$

We also require that as both events approach the bisector of the distribution ($p_2 \to 0.5$) we recover the standard expression for a limit based on the observation of two candidate events, and hence :

$$\text{As } \; p_2 \to 0.5 \;, \quad g_1 \to 1 \text{ and } f_2 \to 1 \;. \tag{7}$$

The simplest way to satisfy Eqs. 6 and 7 is to set :

$$g_1 = f_1 = 2\,p_1 \;, \tag{8}$$

and

$$f_2 = 2\,p_2 \;. \tag{9}$$

Equations 8 and 9 define what we mean by "signal–like". Explicitly, we consider a result to be more signal–like than the observation of two events at $K = K_1$ and $K = K_2$ if (i) more than two events are observed, or (ii) two events are observed, both having single–event integrated probabilities greater than p_2, or (iii) one event is observed with single–event integrated probability greater than p_1. Although there are certainly other ways of defining how "signal–like" a particular observation is, our choice seems at least to be reasonable, and has the added virtue of simplicity. Using our definition of "more signal–like" we obtain :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; f_1 \,\mu e^{-\mu} \;-\; f_2 \,\frac{\mu^2 e^{-\mu}}{2} \;, \tag{10}$$

or more explicitly :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; 2\,p_1 \,\mu e^{-\mu} \;-\; 2\,p_2 \,\frac{\mu^2 e^{-\mu}}{2} \;. \tag{11}$$

To find the maximum value of μ that can be tolerated at a given C.L. we must solve Eq. 11 to find μ_{max}, the value of μ that yields P equal to the desired confidence level. The resulting 90% and 95% C.L. upper limits on μ are shown respectively in Figs. 2 and 3 as a function of p_1 and p_2. Note that the curves corresponding to $p_2 = 0.01$ are very similar to the single–event curves in Fig. 1. Furthermore, as p_1 , $p_2 \to 0.5$ we find $\mu_{max} \to 5.3$ at 90% C.L., and $\mu_{max} \to 6.3$ at 95% C.L., which are just the normal Poisson upper limits based on the observation of two events.

2.3 N observed events

We now generalize our method of computing P to apply to an arbitrary number of observed events N. We will label the events $i = 1, 2, \ldots N$ and order the labels so that $p_1 > p_2 > \ldots. > p_N$, where p_i is the single–event integrated probability for event i. The expression for P becomes :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; f_1\, \mu e^{-\mu} \;-\; f_2\, \frac{\mu^2 e^{-\mu}}{2!} \ldots \;-\; f_N\, \frac{\mu^N e^{-\mu}}{N!} \;, \tag{12}$$

with

$$f_i \;=\; 2\, p_i \;. \tag{13}$$

Equations 12 and 13 provide our definition of "signal–like". We consider a result to be more signal–like than the observation of N events at $K = K_1, K_2, \ldots K_N$ if (i) more than N events are observed, or (ii) i events are observed $(i \le N)$ all of which have single–event integrated probabilities greater than p_i. With this definition :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; 2\, p_1\, \mu e^{-\mu} \;-\; 2\, p_2\, \frac{\mu^2 e^{-\mu}}{2!} \ldots \;-\; 2\, p_N\, \frac{\mu^N e^{-\mu}}{N!} \;. \tag{14}$$

To find the maximum value of μ that can be tolerated at a given C.L. we must solve Eq. 14 to find μ_{max}, the value of μ that yields P equal to the desired confidence level.

3 Comparison with alternative methods

In the following we consider several alternative methods of setting an upper limit when N events have been observed and no reliable background subtraction is possible.

3.1 Using an aggressive cut

Suppose we use the technique of placing an aggressive cut on our final kinematic variable by requiring $p_i > p_1$, i.e. we place a cut right up against the "most signal–like" event so that we are left with no candidate events. If we then compute an upper limit based on the observation of zero events we will obtain at 90% C.L. :

$$\mu_{max} \;=\; \frac{2.3}{1 \;-\; p_1} \;, \tag{15}$$

where the factor of $1/(1 - p_1)$ takes account of the reduction in the selection efficiency for signal events associated with the cut $p_i > p_1$. The upper limits on the number of predicted events μ_{max} obtained using the method proposed in Section 2 divided by the corresponding limits obtained using the aggressive cut technique are shown in Fig. 4 for the $N = 1$ case as a function of the single–event integrated probability. When p_1 is small the aggressive cut

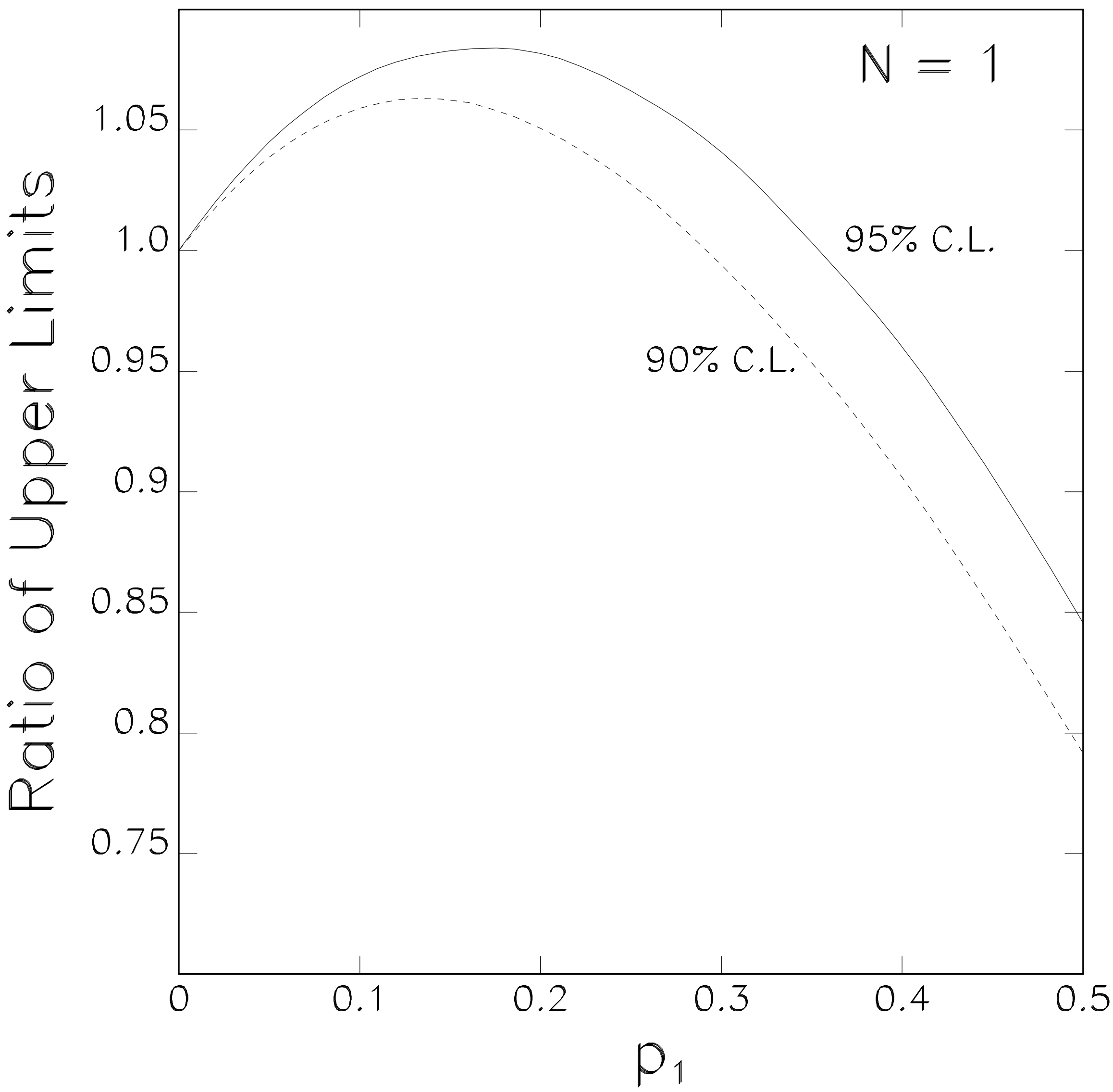

Figure 4: The upper limits on the number of predicted events obtained using the method proposed in the text divided by the corresponding limits obtained using the aggressive cut technique. The ratio of limits is shown as a function of the single–event integrated probability for one observed event.

technique results in a more restrictive upper limit than the method we are proposing. Note however that the ratio of the limits never exceeds 1.1. As p_1 becomes large the aggressive cut technique yields a less restrictive upper limit than the method we are proposing. This makes sense since, if we had observed one event exactly at the bisector of the predicted distribution, and calculated an upper limit based on the observation of one event, we would indeed obtain a more restrictive limit than if we reduced our selection efficiency by a factor of two and calculated a limit based on the observation of zero events.

3.2 Background populating only one side of the distribution

Consider an experiment in which the background processes are expected to populate only one side of the observed kinematic distribution. Can we make use of this knowledge to obtain a more restrictive upper limit μ_{max} ?

To explore this possibility, consider an experiment in which one event is observed ($N = 1$) having $K = K_1$, corresponding to a single–event integrated probability p_1. If K_0 is the bisector of the predicted signal distribution, we can define two regions of the kinematic space: region A defined as $K_{min} \leq K < K_0$, and region B defined as $K_0 \leq K \leq K_{max}$. The theory we are testing predicts that on average we should observe $\mu/2$ signal events in region A, and $\mu/2$ signal events in region B. Let the background preferred side of the distribution be region B, and let the observed event also be in region B. We want to know if we can we modify our definition of "more signal–like" to exploit the knowledge that the observed event is in region B rather than A. Therefore, we will consider a result to be "more signal–like" if any of the following are true : (i) One or more events have been observed anywhere in region A; (ii) No events in have been observed in region A, but one event has been observed in region B with single-event integrated probability exceeding that of the an event at $K = K_1$; (iii) No events have been observed in region A, but more than one event has been observed in region B. The probabilities associated with each of these "more signal–like" cases can be added to yield :

$$P \;=\; \left(1 - e^{-\mu/2}\right) \;+\; \left((1 - f_B)\,\frac{\mu}{2}e^{-\mu}\right) \;+\; \left(e^{-\mu/2} - e^{-\mu} - \frac{\mu}{2}e^{-\mu}\right)\;, \tag{16}$$

where the brackets delineate the contributions associated with (i), (ii), and (iii) above, and $f_B = 2\;p_B$ with p_B defined as the single event integrated probability for an event observed at $K = K_1$ in an experiment sensitive only to region B. Note that $p_B = 2\;p_1$; if 10% of the tail of the predicted signal distribution within the region (A + B) lies beyond K_1, then 20% of the part of the signal distribution within region B will lie beyond K_1. Hence, $f_B = 4\;p_1$, and we find that :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; 2\,p_1\,\mu e^{-\mu}\;. \tag{17}$$

This is just Eq. 4. The knowledge that the observed event lies in the half of the distribution which might be populated by background processes has made no change to our result. It is easy to see why. No matter how we divide the distribution, we will always end up with two regions in integrated probability space which we consider to be associated with "less signal–like" observations, each having area p_1. If we are to obtain a more restrictive limit we will need more information about the background than just the region within which it lies.

3.3 Using the distance from the peak of the distribution

It has been suggested [1] that the distance of an event from the peak of the predicted distribution for the signal process may provide a good measure of how signal–like the event is. However, in the method proposed in Section 2 this distance has no special significance. Indeed, instead of plotting the K–distribution for signal events we could have chosen, without loss of information, to plot the distribution of some function $f(K)$, and discovered that events near the peak of the K–distribution are no longer near the peak of the $f(K)$–distribution. As an example let us assume we measure an interaction energy E for each event, and that we wish to test a theory that predicts $dN/dE \sim 1/E$. If we plot the predicted distribution of interaction energies for the signal, the distribution will peak at $1/E_{min}$, and events near the peak will have energies near E_{min}. However, if instead we decide to plot the distribution of the inverse of the interaction energies $dN/d(1/E)$ we will obtain a peak at $(1/E)_{min}$, and events near the peak will have energies near E_{max}. Hence, we choose not to use the distance from the predicted peak of the distribution as a measure of how signal–like any given event is.

3.4 Using the distance from the bisector of the distribution

Unlike the distance of an event from the peak of the predicted distribution, the distance in integrated probability space from the bisector of the distribution $(0.5-p_i)$ does have a special significance in the method proposed in Section 2. Indeed, the bisector of the distribution is the most signal–like location for an event. Furthermore, the distance $(0.5-p_i)$ is independent of which function $f(K)$ we choose to plot, provided we do not lose information in going from K to $f(K)$. We might therefore think that a quantity like:

$$r_N \equiv \sqrt{\sum_{i=1}^{N}(0.5-p_i)^2} \tag{18}$$

would provide a good measure of how signal–like N observed events are. Note that r_N can be interpreted as the radius of an N–dimensional hypersphere in probability density space, with the center of the hypersphere coincident with the center of a uniformly populated N–dimensional unit hypercube. The fraction F_N of the hypercube that is outside of the hypersphere is equal to the probability that N randomly selected events from the signal process will be "less signal–like" than the N observed events. Equation 12 is then replaced

Table 1: Volumes of N-dimensional hyperspheres of radius r, for $N = 2$ – 7.

N	V_N	N	V_N
2	$\pi\, r^2$	5	$\frac{8}{15}\pi^2\, r^5$
3	$\frac{4}{3}\pi\, r^3$	6	$\frac{1}{6}\pi^3\, r^6$
4	$\frac{1}{2}\pi^2\, r^4$	7	$\frac{16}{105}\pi^3\, r^7$

by :

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; F_1\, \mu e^{-\mu} \;-\; F_2\, \frac{\mu^2 e^{-\mu}}{2!} \;...\; -\; F_N\, \frac{\mu^N e^{-\mu}}{N!}\;. \qquad (19)$$

Note that as $p_N \to 0.5$, $r_N \to 0$ and $F_N \to 1$, satisfying Eq. 7. Furthermore, as $p_N \to 0$, $F_N \to 0$, and the expression for P with N observed events reduces to the corresponding expression for $(N-1)$ observed events.

The F_N can be easily computed by a Monte Carlo technique. If $r_N < 0.5$, then $F_N = (1 - V_N)$, where V_N is the volume of an N–dimensional hypersphere of radius r_N, and we have:

$$P \;=\; 1 \;-\; e^{-\mu} \;-\; 2p_1\, \mu e^{-\mu} \;-\; (1 - V_2)\, \frac{\mu^2 e^{-\mu}}{2!} \;...\; -\; (1 - V_N)\, \frac{\mu^N e^{-\mu}}{N!}\;, \qquad (20)$$

where simple analytical expressions can be calculated for the V_N. As an example, the V_N are summarized in Table 1 for $N = 2$, 3, 4, 5, 6, and 7.

We conclude that Eqs. 18 and 19 provide a viable alternative to the method described in Section 2. However, this alternative method is more complicated than the Section 2 method, and the associated definition of "signal–like" does not seem to be better (or worse) than the corresponding Section 2 definition. We therefore advocate using the simpler method described in Section 2.

4 Summary

We propose a simple method of computing upper limits when N candidate events have been observed but no reliable background subtraction can be made. The method avoids using a final arbitrary cut to remove events, but requires a definition of how "signal–like" each observed event is. As a measure of how "signal–like" an event is, the method uses the single–event integrated probability p_i, defined by Eq. 2. After ordering the event labels so that $p_1 > p_2 > \ ... > p_N$, a result is considered to be "more signal–like" than the observation of a particular set of N events (set A) if (i) more than N events are observed, or (ii) i events are observed ($i \leq N$) all of which have single–event integrated probabilities greater than the single–event integrated probability for the ith event of set A. With this definition, if the theory we are testing predicts on average the observation of μ events, then Eq. 14 can be used to find μ_{max}, the maximum value of μ that can be tolerated at a C.L. of P. Although our definition of what is more or less "signal–like" is not unique, it has the virtue of being simple, and leads to a method that provides sensible asymptotic limits as the p_i approach their extreme values (0 and 0.5).

References

[1] R.D. Cousins, "A method which eliminates the discreteness in Poisson confidence limits and lessens the effect of moving cuts specifically to eliminate candidate events", Nucl. Instr. and Methods in Phys. Research **A337** (1994) 557.

[2] A useful tabulation of Poisson upper and lower one-sided limits is given in: N. Gehrels, "Confidence limits for small numbers of events in astrophysical data", **AJ 303**, 336 (1986).