Confidence Limits and their errors

Rajendran Raja*

Fermi National Accelerator laboratory

Batavia, IL 60510

(Dated: March 6, 2002)

Abstract

Confidence limits are common place in physics analysis. Great care must be taken in their calculation and use especially in cases of limited statistics. We introduce the concept of statistical errors of confidence limits and argue that not only should limits be calculated but also their errors in order to represent the results of the analysis to the fullest. We show that comparison of two different limits from two different experiments becomes easier when their errors are also quoted. Use of errors of confidence limits will lead to abatement of the debate on which method is best suited to calculate confidence limits.

PACS numbers:

*Electronic address: raja@fnal.gov

I. INTRODUCTION

We introduce the concept of error of confidence limits by a simple Gaussian example. Consider a sample of n events, where $n = 10$, characterised by the variable x distributed as a unit Gaussian, with a mean value $\mu = 0$ and standard deviation $\sigma = 1$. Then the average value $\bar{x}$ of the n events will be distributed as a Gaussian of mean value zero and standard error $\sigma/\sqrt{(n)}$. The unbiased estimate of σ, the variance of the distribution is given by s where,

$$s^2 = \frac{1}{n-1}\Sigma_{i=1}^{i=n}(x_i^2 - \bar{x}^2) \tag{1}$$

Figure 1 shows the distribution $\bar{x}$ of our sample of 10 events for a large number of samples. The expected value $\bar{x}$ is zero and its standard deviation is 0.32 which is consistent with the theoretical value of $\sigma/\sqrt{(n)}$=0.316. Figure 2 shows a histogram of s deduced from a

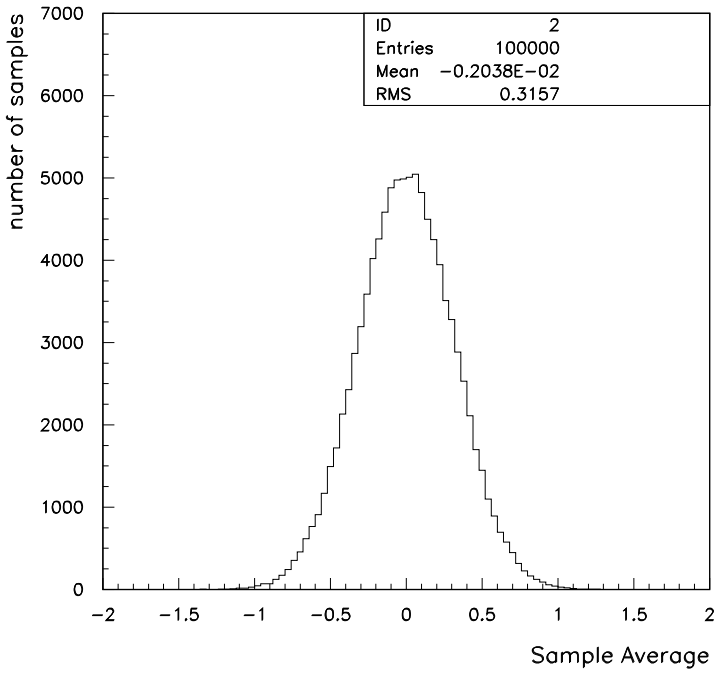

FIG. 1: The distribution of the sample average $\bar{x}$ over a large sample of events.

sample of 10 events for a large number of such samples. The average value of s is ≈ 1.0, showing that s is an unbiased estimator of σ. The important point to note is that s *also* has a variance and that its standard deviation is 0.23. This is as expected from theory where

the error on the standard deviation of a Gaussian sample [1] is $\approx \sigma/\sqrt{(}2n)$=0.223. Having

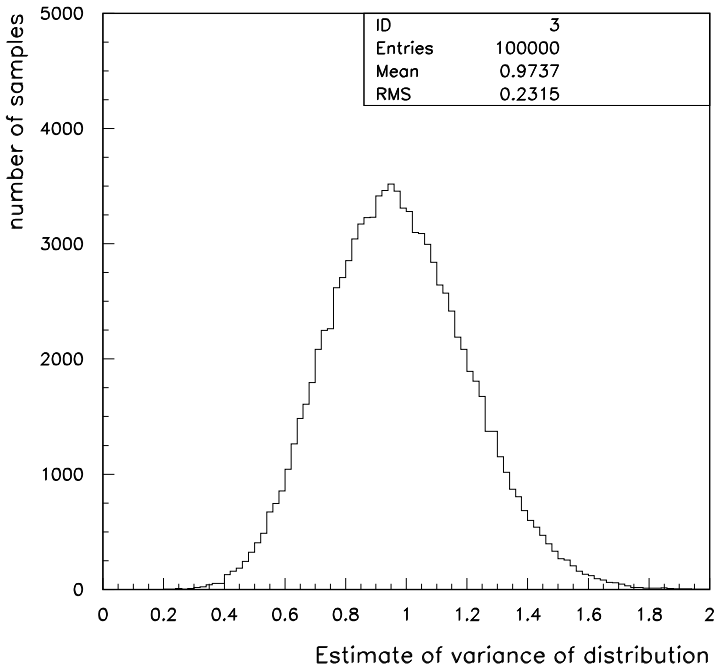

FIG. 2: Unbiased estimate s of the standard deviation of the σ of the Gaussian distribution deduced from a sample of $n = 10$ events. The average value of s is ≈ 1.0 and *its* standard deviation is 0.23.

got the value of $\bar{x}$ and s for our sample, one can proceed to work out confidence limits for our observation. The two sided 68% CL limits for our observation of $\bar{x}$ will be given by the standard error $\sigma(\bar{x})$ of $\bar{x}$ and we would write the observation of $\bar{x}$ from our sample as

$$\bar{x} \pm \sigma(\bar{x}) = \bar{x} \pm s/\sqrt{(}n) = -0.188 \pm 0.408 \tag{2}$$

where the numbers correspond to our sample of 10 events. Note that the standard error $\sigma(\bar{x}) = 0.408$ derived from our sample of 10 events is quite different from the theoretical value of 0.32, but this is merely due to statistical fluctuation.

One can also work out the two sided 90% CL limits for our observation of $\bar{x}$ which would correspond to $\pm 1.64\ \sigma(\bar{x})$ and quote the 90% CL limits as -0.188 ± 0.669, which is the value observed for our sample of 10 events.

Figure 3 shows the distribution of the 90% CL two sided errors on the sample average, over a large number of samples. The mean value of the distribution is 0.505 which is close to

the theoretical value of 1.64 $\sigma(\bar{x})$=0.519. Note that the standard deviation of the 90% CL errors in Figure 3 is 0.12. We can also calculate the standard deviation of the 90% CL limit from our sample as 1.64 $\sigma(\bar{x})/\sqrt{(2n)}$ and this is plotted in figure 4. The mean value of the standard deviation of the 90% CL figure 4 is 0.113, in line with the theoretical value of 0.116. When the mean value is of interest, we quote the mean value and the standard error

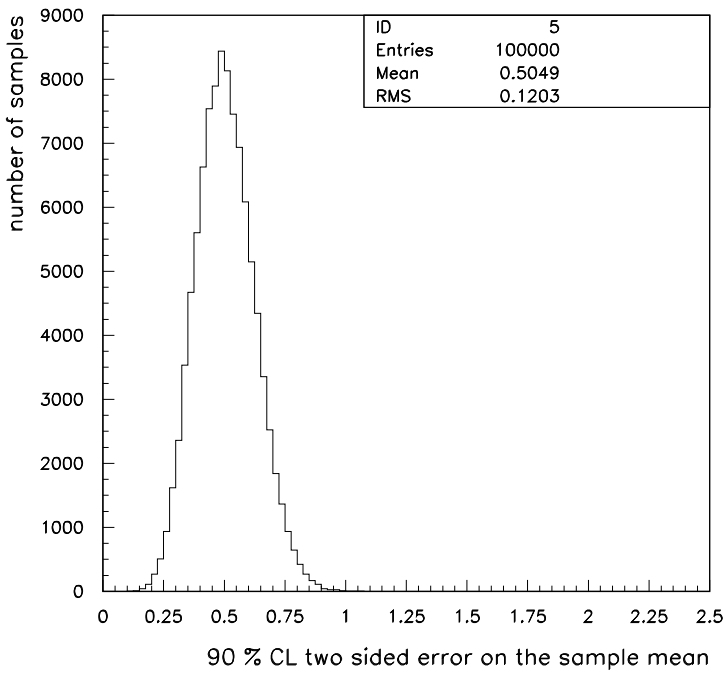

FIG. 3: The distribution of the calculated two sided 90% CL errors of the mean value of the sample.

on the mean value as in equation 2. This enables us to gauge the fluctuations in the mean value from sample to sample. When the confidence limit is of interest, we propose that we quote the confidence limit along with *its* standard error. This would enable us to gauge the significance and stability of the confidence level. In our example we would write this as

$$\bar{x} - 1.64\sigma(\bar{x}) \pm \sigma_{90} < \mu < \bar{x} + 1.64\sigma(\bar{x}) \pm \sigma_{90} \; at \; 90\% \; CL \tag{3}$$

where μ is the expectation value of $\bar{x}$ and the standard error σ_{90} on the 90% CL limit would be given by

$$\sigma_{90} \approx \sigma(\bar{x})\sqrt{(1 + (1.64)^2/(2n)} \tag{4}$$

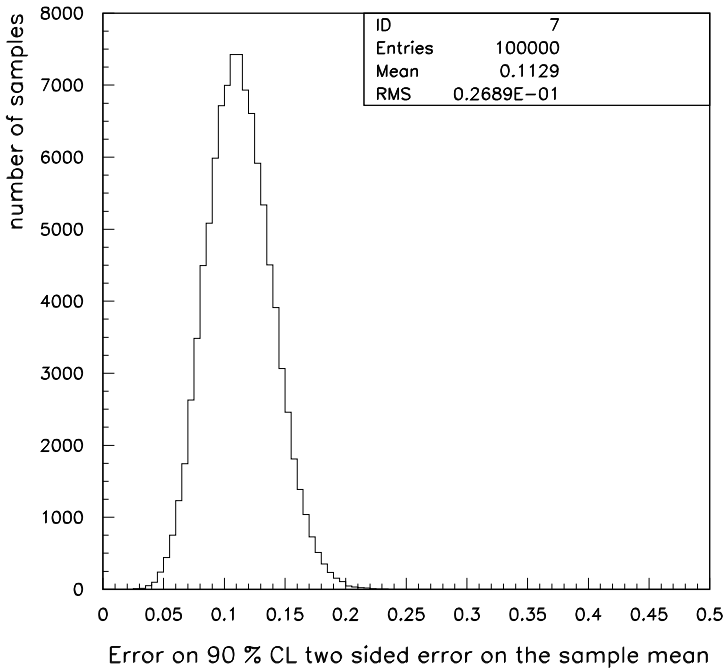

FIG. 4: The distribution of the calculated error on the two sided 90% CL error of the mean value of the sample.

In our sample of 10 events, this would lead to

$$-0.857 \pm 0.434 < \mu < 0.481 \pm 0.434 \; at \; 90\% \; CL \tag{5}$$

Note that the error on the lower and upper 90% CL limits are correlated by the error on $\bar{x}$ which they have in common. Half the difference between the lower and uper 90 % CL limits is $1.64\sigma(\bar{x})$ and its error is $1.64\sigma(\bar{x})/\sqrt{(}2n)$. These two errors added in quadrature yield the formula in equation 4. The error in the 90% CL limit indicates to the reader the stability of the limit and the statistical significance of the result.

Very often, we are not interested in the mean value of our observations but are more interested in the confidence limits, due to the low statistics of the observation. We may only be interested in an upper (one sided) bound. So we would quote a 95% CL upper bound on μ as

$$\mu < 0.481 \pm 0.434 \; at \; 95\% \; CL \tag{6}$$

A second sample of 10 events from the same distribution may yield a result

$$\mu < 0.354 \pm 0.335 \; at \; 95\% \; CL \tag{7}$$

but we do not fall into the trap of declaring the second result a better limit than the first, because both the limits are the same within errors. If we did not quote the errors on the limits, we would be tempted to declare the second limit superior to the first.

Similarly, as analyses proceed in discovery searches, events can go in and out of samples, as cuts are refined and more data is accumulated. Appearance of a single event in a sample can change the confidence limit drastically, as was the case in the search for the top quark. These changes can be understood as fluctuations of the confidence limit within errors, if we were to quote not only the confidence limit but also its error.

II. AN ILLUSTRATIVE EXAMPLE

We can illustrate the need for confidence limits errors using the following example. In 1995, the DØ collaboration published limits on the top quark mass and cross section [2]. Figure 5 shows [2] the 95% CL upper limit on top quark production as a function of top quark mass using 13.5 pb^{-1} of data. The confidence limit curve is used to derive a lower limit of 128 GeV/c^2 for the top quark mass at 95% CL. In the same paper, another figure, reproduced here as (Figure 6) shows the top quark production cross section as a function of the top quark mass. This curve has a 1 σ error band around it. But the top quark production cross section may be thought of as the 50% CL upper/lower bound on the cross section. Surely, if the 50% CL limit has an error band around it, the 95% CL limit should also have its own error band. In what follows, we show how to calculate errors in confidence levels in general and use the method to calculate the error in the 95% CL curve shown in Figure 5.

III. A GENERAL ALGORITHM TO CALCULATE ERRORS IN CONFIDENCE LIMITS

Most experiments have elaborate algorithms to calculate confidence limits for their results. Such algorithms will include detailed calculations and parametrizations of efficiencies

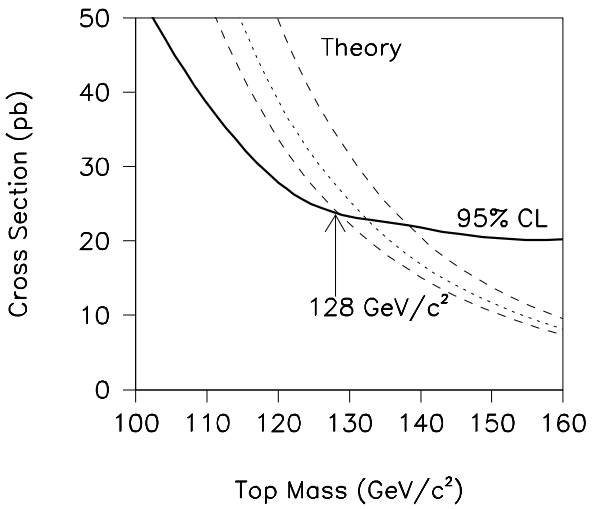

FIG. 5: The 95% confidence level [2] on $\sigma_{t\bar{t}}$ as a function of top quark mass. Also shown are central (dotted line) and low (dashed line) theoretical cross section curves [3].

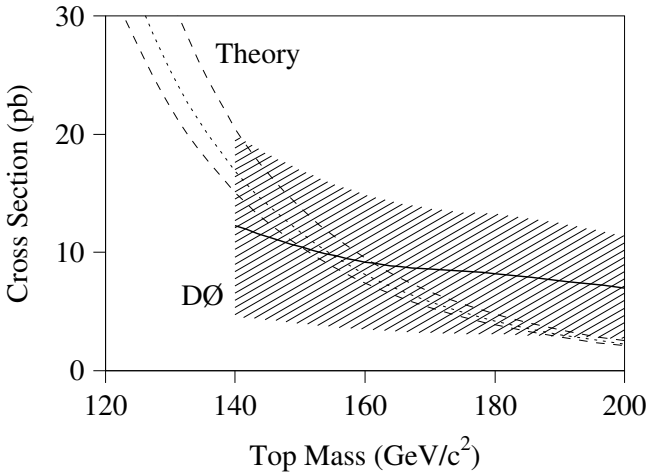

FIG. 6: Measured $t\bar{t}$ production cross section (solid line, shaded band = one standard deviation error) as a function of top mass [2]. Also shown are central (dotted line), high and low (dashed lines) theoretical cross section curves [3].

and acceptances. In addition, they will have several other input parameters such as the number of events observed, total integrated luminosity and the error on the luminosity. Let us denote the input parameters as $a_i,\ i = 1, n$. The output of such a program will be the

confidence limits $C_\alpha,\ \alpha = 1, k$. Figure 7 illustrates this general case. Then, for small changes

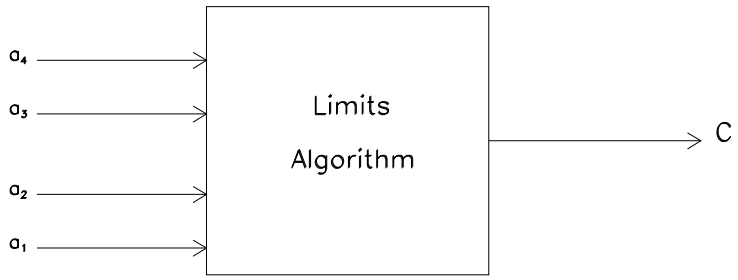

FIG. 7: Schematic "black box" representation of a general confidence limit calculating algorithm, that has input parameters $a_1, a_2..a_4$ and outputs a confidence level C in a single variable.

in the input parameters, the following equations hold.

$$\delta C_\alpha = \frac{\delta C_\alpha}{\delta a_i}\delta a_i \tag{8}$$

$$< \delta C_\alpha \delta C_\beta >= \frac{\delta C_\alpha}{\delta a_i}\frac{\delta C_\beta}{\delta a_j} < \delta a_i \delta a_j > \tag{9}$$

where the repeated indices i, j are meant to be summed over and the symbols $<>$ indicates the average over the enclosed quantities. The quantity on the left hand side of the equation is the error matrix in the confidence limits C_α, denoted E_{CC}. The above equation can be re-written in matrix form as

$$E_{CC} = \tilde{T} E_{aa} T \tag{10}$$

where E_{aa} is the error matrix of the input parameters $a_i,\ i = 1, n$ and T is the transfer matrix, such that $T_{\alpha,i} = \frac{\delta C_\alpha}{\delta a_i}$. T can be determined numerically by varying the input parameters to the limits algorithm. The error matrix E_{aa} should be known to the experimenter, yielding the required error matrix E_{CC}.

A. An Example

Let us consider the calculation of C, the 95% CL upper limit to the top quark cross section as published in reference [2]. The output of the limits algorithm is C. The input

parameters can be taken as three, namely a_1, the total number of top quark events observed, a_2, the luminosity$\times$efficiency$\times$branching ratio of the channels under consideration, summed over the channels and a_3, the error in the luminosity. We have used a single parameter a_2 summed over the channels to simplify the calculation. In principle, all channels may be varied independently, but since they are uncorrelated, and the dominant error is due to the common luminosity factor, the above simplification will result. We use this example for illustrative purposes to show how such a calculation may proceed.

The error matrix of the parameters E_{aa} is a 3×3 diagonal matrix, since the parameters are uncorrelated. The variance of a_1 is the number of events observed, the variance of a_2 is calculated using the error in luminosity, and the variance of a_3 is calculated assuming that there is a 50% uncertainty in the *error* in the luminosity. The transfer matrix T is calculated by numerical differentiation.

Figure 8 shows the contribution to σ_C, the error in the 95% CL upper limit to the cross section, due to the three parameters a_1, a_2 and a_3 as a function of the top quark mass. The overall error σ_C, obtained by adding the component errors in quadrature, is also shown as a function of the top quark mass. It can be seen that the contribution due to uncertainties in a_1, is negligible. So we are not sensitive to errors in our guess of 50% uncertainty to the error in the luminosity. The overall error is dominated by the fluctuation in the total number of events. This example thus graphically illustrates why confidence limits fluctuate up and down as events fall in and out of the selected sample as the analysis proceeds and more data is accumulated. The 95% CL upper limit to the cross section is merely fluctuating within its error as all statistical quantities do. When we are interested in a confidence limit, it thus behooves us to compute not only that limit but also its error. We may superimpose these errors on Figure 5 yielding Figure 9. The 95% CL lower limit to the top quark mass can then be quoted as 128^{+14}_{-18} GeV/c^2, the error bars indicating the range of fluctuation for the mass limit. This implies that if one were to repeat the DØ experiment numerous times with an integrated luminosity of 13.5 pb^{-1} fluctuating within its errors, one would expect to get a top quark lower mass limit that fluctuates within the errors quoted.

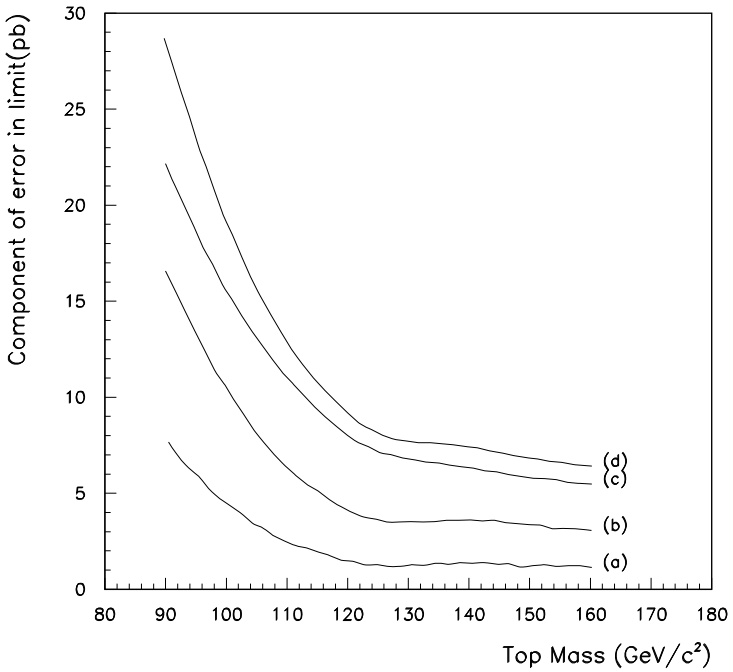

FIG. 8: The components of σ_C, the error in the 95% CL top quark cross section upper limit, due to uncertainties in (a) error in luminosity (b) Luminosity×efficiency×branching ratio (c) The overall number of events observed as a function of top quark mass. (d) shows the overall error σ_C.

IV. COMBINING LIMITS

Combining limits from two different experiments is difficult at best. We remark here that in simple Gaussian cases, quoting the limit and its error provides us with enough information to make a combined result, as may be seen by examining equations 3 and 4. Using the value of the limit and its error, we may deduce $\bar{x}$ and $\sigma(\bar{x})$, if the number of events n in the sample (the dependence on n in equation 4 is weak and may be neglected for large enough n) is known. Having the mean and its variance in each case, we can combine the results to get a new limit. One can also ask if the two limits are consistent with each other, if the errors on the limits are quoted.

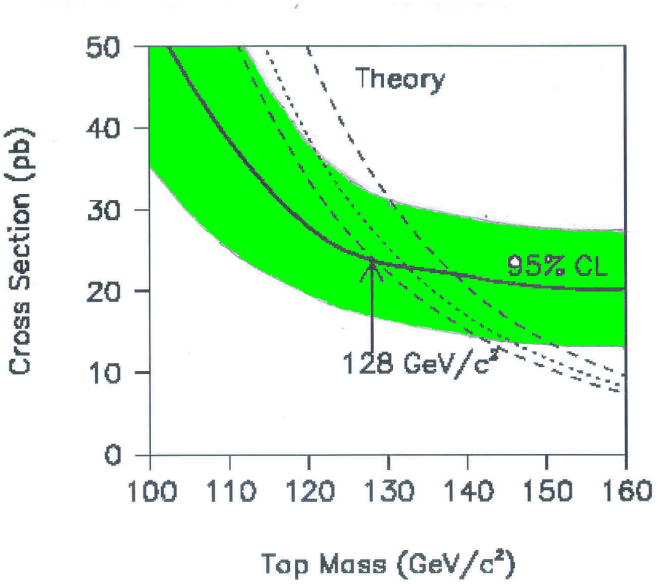

FIG. 9: The DØ 95% CL upper limit to the top quark cross section [2] with its accompanying error band, as calculated by the method in the text.

V. COMPARING LIMITS FROM TWO DIFFERENT ALGORITHMS

When two different algorithms are used on the same data, two different limits will result that are correlated. The correlations will be due to the common input into the two algorithms. We can think of the "black box" in Fig. 7 as consisting of two different algorithms producing as output C_1 and C_2, the two confidence levels in question, using the same common input $a_i,\ i = 1, n$. We can then use equation 10 to work out E_{CC}, the error matrix of the two confidence level algorithms and use this matrix to decide whether the two confidence levels are significantly different from each other.

VI. CONCLUSIONS

We have motivated the concept of statistical error for a confidence limit. In cases of limited statistics, our estimates of the confidence limits can fluctuate significantly. Comparing confidence limits is only meaningful if these errors are quoted. Different methods exist (e.g Bayesian, Frequentist) for calculating these limits. The differences between limits computed in the same experiment using different methods will lose their significance if the limits are

accompanied by their error. Often in analyses with limited statistics, the appearance of a new event can make significant differences to the limit calculation. An error analysis of the limit will show that the limit is exhibiting statistical fluctuation as it is entitled to. We propose that experimenters publish confidence limits to their data accompanied by the error on the limits.

VII. ACKNOWLEDGEMENTS

The author wishes to thank Louis Lyons and Tom Trippe for helpful comments.

---

[1] C.E. Weatherburn, "A first course in Mathematical Statistics", Cambridge University press, Page 137.

[2] "Top quark search with the DØ 1992-1993 data sample", Phys.Rev.D52:4877(1995)

[3] E.Laenen, J.Smith, W. van Neerven, Phys. Lett. 321B, 254 (1994); Nucl. Phys. B368, 543(1992).