SETTING LIMITS AND MAKING DISCOVERIES IN CDF

John Conway
Rutgers University

Abstract
This paper presents the statistical methods used in setting limits and discovery significances in the search for new particles in the CDF experiment at the Fermilab Tevatron. For single-channel counting experiments the collaboration employs the classical Helene formula, with Bayesian integration over systematic uncertainties in the signal acceptance and background. For more complex cases such as spectral fits and combining channels, likelihood-based methods are used. In the discoveries of the top quark and $B_c$ meson, the significance was estimated from the probability of the null hypothesis, using toy Monte Carlo methods. Lastly, in the recent SUSY/Higgs Workshop, the Higgs Working Group used a method of combining channels and experiments based on the calculation of the joint likelihood for a particular experimental outcome, and averaging over all possible outcomes.

1. INTRODUCTION

In most new particle searches in high energy physics, one selects from a large number of recorded events those which bear characteristics of the new process while minimizing the retention of events from well-understood processes. This typically results in a small number of events passing the selection requirements, consistent with the expectation from a calculation of the expected background. At this stage one typically wishes to determine an upper limit on the number of signal events present in the sample, at some desired confidence level (usually 95%), employing a statistical method which allows one to take into account the systematic uncertainties in signal acceptance and expected background.

If, on the other hand, one observes an excess number of events passing the selection criteria, possibly consistent with the prediction of an as-yet-unobserved new particle, one would like to estimate the statistical significance of the observation in order to decide if a statistical fluctuation in the number of background events is more likely the cause of the excess.

This note discusses the method used by the CDF Collaboration to determine upper limits on Poisson processes in the presence of uncertainties (both statistical and systematic) simultaneously in the acceptance and background, and the methods for determining the statistical significance of an excess. The collaboration employs rather different methods for single-channel and multi-channel (spectral) searches, in the latter case using a likelihood-based approach which can also be used to estimate experimental sensitivity or expected limits.

2. SINGLE-CHANNEL LIMITS WITHOUT UNCERTAINTIES

Given $n_0$, the number of observed events, the probability $P$ for observing that number depends on $\mu$, the mean number of signal events expected, according to the Poisson distribution (assuming no background events are expected):

$$P(n_0 | \mu) = \frac{\mu^{n_0} e^{-\mu}}{n_0!}. \quad (1)$$

In new particle searches one wishes to determine the value of $\mu$. We define the upper limit $N$ on the number of expected events\(^1\) as that value of $\mu$ for which there is some probability $\epsilon$ to observe $n_0$ or

\(^1\)Note that $N$ is a real number, not an integer.
fewer events. The confidence level (CL) of the upper limit is then simply $1 - \epsilon$. One can calculate $\epsilon$ by summing over the Poisson probabilities:

$$\epsilon = \sum_{n=0}^{n_0} P(n | \mu) .$$

(2)

In practice, then, to calculate $N$ one varies $\mu$ until finding the value of $\epsilon$ corresponding to the desired CL; $N$ is the resulting value of $\mu$.

If one expects an average of $\mu_B$ background events among the $n_0$ observed, and if one knows $\mu_B$ precisely, then the method can be extended to calculating a Poisson upper limit $N$ on the number of signal events present in the observation. The value of $N$ represents that value of $\mu_S$, the mean number of signal events expected, for which the probability is $1 - \epsilon$ that in a random experiment one would observe more than $n_0$ events and have $n_B \leq n_0$, where $n_B$ is the number of background events present in the sample. This can be calculated as before by adjusting $N$ until the relation (known as the Helene formula) obtains.[1]

$$\epsilon = \frac{\sum_{n=0}^{n_0} P(n | \mu_B + N)}{\sum_{n=0}^{n_0} P(n | \mu_B)}$$

(3)

Note that if one obtains a value of $n_0$ significantly lower than $\mu_B$, the resulting limit is “better” in that it results in a lower value for $N$. This is viewed as a shortcoming by some authors,[2] though clearly on average the experiments with larger expected background will on average obtain “worse” (larger) limits on the signal.

The denominator on the right side of Equation 3 makes $\epsilon$ a conditional probability, and ensures that $N$ remains positive. This is clearly a desirable feature, and although the method has a frequentist interpretation, this feature is Bayesian in spirit in that the non-physical values are excluded.

The Helene formula, like other similar methods, “overcovers”; if the new particle actually exists, the probability that the limit exceeds the true value of the expected signal is less than $\epsilon$, and depends on the true value. This is a result of the discreteness of the Poisson distribution.

3. SINGLE-CHANNEL UPPER LIMITS WITH UNCERTAINTIES

There is no generally accepted method in the high-energy physics community for the incorporation of systematic errors into upper limits on Poisson processes. CDF employs a method which is in essence a Bayesian-style integration over the uncertainties in the signal acceptance and expected background.

Suppose that one knows the value of $\mu_B$ to within an overall (statistical plus systematic) Gaussian uncertainty of $\sigma_B$, and the acceptance $A$ to within an overall uncertainty of $\sigma_A$. In this case the relative uncertainty on $\mu_S$ is $\sigma_A/A$. One can define the Poisson upper limit $N$ on $\mu_S$ as before: we seek that value of the true $\mu_S$ for which one would observe more than $n_0$ events and have $n_B \leq n_0$. In this case, however, one seeks the value of $N$ such that

$$\epsilon = \sum_{n=0}^{n_0} \frac{1}{\sqrt{2\pi} N \sigma_B} \int_0^\infty \int_0^\infty P(n | \mu_B + \mu_S) e^{-\frac{(\mu_B - \mu_S)^2}{2\sigma_B^2}} e^{-\frac{(N-\mu_S)^2}{2\sigma_N^2}} d\mu_S d\mu' \int_0^n P(n | \mu_B) e^{-\frac{(\mu_B - \mu_S)^2}{2\sigma_B^2}} d\mu'$$

(4)

where we take $\sigma_N = N \sigma_A / A$. In this way one assumes an a priori Gaussian distribution of the true values of $\mu_S$ and $\mu_B$ about the values obtained in subsidiary studies, with width given by the uncertainties obtained in those studies.

One can perform the integral in Equation 4 by various numerical techniques. The method employed in CDF uses a Monte Carlo integration, rather than performing the integral directly. For each test
value of \( N \) one generates a large ensemble of random pseudoexperiments, varying the expected number of signal and background about their nominal values according to a Gaussian distribution. In each experiment, the expected number of signal and background events are chosen from the Gaussians, and Poisson-distributed numbers of signal \( (n_S) \) and background \( (n_B) \) events are generated. For those trials where \( n_B \leq n_0 \), the fraction \( f \) in which \( n_B + n_S > n_0 \) is recorded. The confidence level for a given \( N \) is in fact equal to \( f \); one must then simply vary \( N \) until the desired CL \((1 - \epsilon)\) is obtained.

4. UPPER LIMITS WITH A BAYESIAN METHOD

One can also obtain upper limits on a Poisson process using a purely Bayesian approach, as discussed in the literature. [4] A Bayesian deems it sensible to treat the unknown expected number of signal events as a random variable, for which there is some “prior” probability density function (pdf) \( \mathcal{P}(\mu_S) \). Given the observation of \( n_0 \) events, one can then construct a “posterior” pdf \( \mathcal{P}(\mu_S|n_0) \) which depends on the likelihood \( \mathcal{L}(n_0|\mu_s) \) for observing \( n_0 \) events given \( \mu_S \) expected:

\[
\mathcal{P}(\mu_S|n_0) = \frac{\mathcal{L}(n_0|\mu_S)\mathcal{P}(\mu_S)}{\int_{0}^{\infty} \mathcal{L}(n_0|\mu_S)\mathcal{P}(\mu_S)d\mu_S} . \tag{5}
\]

One can set a Bayesian upper limit (or any other confidence interval) on the unknown parameter \( \mu_S \), then, simply from integration of \( \mathcal{P}(\mu_S|n_0) \).

The values obtained depend, of course, on the choice of the prior \( \mathcal{P}(\mu_S) \). In considering the results of a particular experiment, usually one uses an “uninformed” prior pdf; that is, one wants to give no \textit{a priori} bias to certain values of \( \mu_S \). This usually results, then, in choosing \( \mathcal{P}(\mu_S) \) to be uniform for all physical values of \( \mu_S \): \( \mathcal{P}(\mu_S) = \text{const.} \) for \( \mu_S \geq 0 \).\(^2\)

Extension to the case where one expects \( \mu_B \) background is straightforward:

\[
\mathcal{P}(\mu_S|n_0, \mu_B) = \frac{\mathcal{L}(n_0|\mu_S + \mu_B)\mathcal{P}(\mu_S)}{\int_{0}^{\infty} \mathcal{L}(n_0|\mu_S + \mu_B)\mathcal{P}(\mu_S)d\mu_S} . \tag{6}
\]

For uniform prior \( \mathcal{P}(\mu_S) \) this reduces to

\[
\mathcal{P}(\mu_S|n_0, \mu_B) = \frac{\mathcal{L}(n_0|\mu_S + \mu_B)}{\int_{0}^{\infty} \mathcal{L}(n_0|\mu_S + \mu_B)d\mu_S} . \tag{7}
\]

Remarkably, as Cousins points out [4], the upper limits obtained with this expression match exactly those obtained with Equation 3. Note also that the denominator of Equation 6 can simply be regarded as a normalization constant whose value depends on \( n_0 \) and \( \mu_B \). Thus we see that

\[
\mathcal{P}(\mu_S|n_0, \mu_B) \propto \mathcal{L}(n_0|\mu_S + \mu_B) . \tag{8}
\]

To incorporate uncertainties on the signal and background one treats the expected background and signal as unknown parameters with uniform prior pdf, with Gaussian likelihood about the estimates from subsidiary studies, just as in the frequentist case. One thus obtains

\[
\mathcal{P}(\mu_S|n_0, \mu_B) \propto \frac{1}{2\pi\sigma_B\sigma_S} \int_{0}^{\infty} \int_{0}^{\infty} \mathcal{L}(n_0|\mu'_S + \mu'_B)e^{-\frac{(\mu_B - \mu'_B)^2}{2\sigma_B^2}} e^{-\frac{(\mu_S - \mu'_S)^2}{2\sigma_S^2}} d\mu'_B d\mu'_S , \tag{9}
\]

where \( \sigma_S = (\sigma_A/A)\mu_S \) comes from the relative uncertainty on the acceptance.

\(^2\)Note that such a pdf is formally non-normalizable.
To calculate upper limits, one can simply calculate the right hand side of Equation 9 for an appropriate range of $\mu_S$, and then define the upper limit on $\mu_S$ as that value for which

$$\epsilon = \frac{\int_{\mu_S}^{\infty} P(\mu_S|n_0, \mu_B) d\mu_S}{\int_0^{\infty} P(\mu_S|n_0, \mu_B) d\mu_S}$$

(10) obtains for some desired confidence level $1 - \epsilon$.

In general the upper limits obtained using this method exceed those obtained with the frequentist version in Equation 4; that is the Bayes intervals “overcover” the frequentist (or more properly speaking, frequentist/Bayesian) ones. This is regarded as a shortcoming by some authors, and as laudably “conservative” by others. The difference lies, however, in the different meaning of the two statistics.

5. DISCOVERY SIGNIFICANCE: TWO EXAMPLES

In searching for new particles the possibility exists that the result will be an excess of observed events in the selected sample. The standard in the community is to quote a significance for the excess in terms of the number of Gaussian sigma the result deviates from the null hypothesis. For Poisson processes with small numbers of events this is almost always based on the probability that the background alone can account for the observed number of events. Given $n_0$ observed events, with $B \pm \sigma_B$ expected background, one typically wishes to calculate the probability of observing $n_0$ or more, taking into account the uncertainties present. Then one relates this probability to the number of Gaussian standard deviations to quote a significance.

If the uncertainty in the expected number of background events is zero or negligible, then the calculation of the probability $P_{null}$ of the null hypothesis is a straightforward sum over Poisson probabilities:

$$P_{null} = \sum_{n=n_0}^{\infty} \frac{B^n e^{-B}}{n!} .$$

(11) To relate this probability to a Gaussian deviation (in units of sigma), one simply finds that value of $x$ for which

$$P_{null} = \sqrt{\frac{2}{\pi}} \int_x^{\infty} e^{-x^2/2} dx!$$

(12) obtains. Note that the normalization constant corresponds to finding that fraction of the integral over the positive half of the Gaussian lying beyond $x$. This effectively means that one is calculating the probability that, for a positive fluctuation, one would get $x$ or larger in a Gaussian-distributed quantity. Such a convention is necessary to ensure consistency with the confidence intervals determines from the Helene equation (3) and the Bayesian equation (6).

When there is uncertainty in the background, and when there is more than one channel, calculating $P_{null}$ becomes complicated. Typically in CDF a toy Monte Carlo is used to actually perform the calculation; two examples of actual new particle discoveries illustrate this, those of the $B_c$ meson and the top quark.

In the case of the search for the $B_c$ meson, one sought events where a $J/\psi \rightarrow \mu^+\mu^-$ decay from a secondary vertex was accompanied by an additional lepton ($e$ or $\mu$) from the same vertex, coming from the semileptonic decay of the $b$ quark. The backgrounds were estimated from the sidebands of the $J/\psi$ peak. Table 1 shows the results, the expected background, and the probability that the background alone could give the observed number of events or more in the electron and muon channels.

250
One might at this stage be tempted to simply quote the product of the two probabilities, or add the observed and expected numbers of events together and calculate a probability that way. But the collaboration first determined the number of signal events present in the sample by minimizing a complicated likelihood function which took into account systematic uncertainties and correlations in the expectation. This yielded a value of 20.4 signal events. To estimate the probability of the null hypothesis a toy Monte Carlo was used to generate over 350,000 pseudoexperiments in which the number of observed events was generated according to Poisson distributions of expected background events, putting in fluctuations and correlations as estimated in the experiment. For each pseudoexperiment the same likelihood fit was performed, and the fraction of such fits which yielded more than 20.4 events was determined from a fit to the shape of the distribution of number of signal events. This fraction, $6 \times 10^{-7}$, then, corresponded to a 4.8$\sigma$ significance. However, this fraction included the results of those pseudoexperiments in which the fitted signal contribution was zero (negative values were not allowed). Thus, strictly speaking, the prescription of considering only positive fluctuations was not adhered to in this case; had it been, the resulting statistical significance would have been close to 4.2$\sigma$.

The case of the top quark discovery was more complicated in that there were three overlapping search channels involved, the so-called SVX, SLT, and DIL searches. In the SVX channel, events with a high-$p_T$ lepton ($\ell$ or $\mu$) plus three or more jets were accepted, and at least one of the jets was required to have been tagged as a $b$ jet with a reconstructed secondary vertex. In the SLT analysis, the same sample was selected, and one jet had to have been tagged as a $b$ by the presence of a low-$p_T$ lepton. In the DIL (dilepton) channel, events with two leptons, large missing $E_T$, and two or more jets were selected. Table 2 shows the observed number of events, the expected background, and the probability or that channel that the background alone could give rise to the observed number of events or more.

The acceptance for the SVX and SLT channels clearly overlap to a great extent; they are based on the same kinematic selection and only differ by the $b$-tagging algorithm. To take this into account, the probabilities in the table are calculated by considering the only that set of pseudoexperiments that give the same number of lepton plus jets events as were observed in the actual data sample before $b$ tagging. The overlap in acceptance for the different tagging methods, as well as other uncertainties in the expected background, are modelled in each pseudoexperiment by appropriate Gaussian smearing of the parameters.

To determine the overall significance, the three resulting probabilities are multiplied together, yielding $3.6 \times 10^{-9}$. The probability of the null hypothesis is then taken to be the probability that the product of three random numbers, uniformly distributed in the range $[0,1]$, is less than this value. This probability, in fact, can be calculated from a straightforward equation:

$$P(r_1r_2...r_n < \epsilon) = \epsilon \sum_{i=0}^{n-1} \frac{1}{i!} (\ln \epsilon)^i$$

This yields $10^{-6}$, which was claimed to be equivalent to a 4.8$\sigma$ Gaussian significance. However, this value would have been 4.9$\sigma$ had only positive fluctuations been considered, as discussed above.

<table>
<thead>
<tr>
<th>$J/\psi + e$</th>
<th>$J/\psi + \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed</td>
<td>19</td>
</tr>
<tr>
<td>expected</td>
<td>5.0 ± 1.1</td>
</tr>
<tr>
<td>probability</td>
<td>0.00002</td>
</tr>
</tbody>
</table>

Table 1: Results from the CDF search for the $B_s$ meson.
Table 2: Results from the CDF search for the top quark.

6. LIMITS FROM SPECTRA AND COMBINING CHANNELS

Quite often, particularly in recent years, one uses fits to the spectra of kinematic variables in order to maximize the sensitivity to new particles. Such fits can be made in variables such as the new particle mass, other kinematic quantities which distinguish signal and background, or even the output value of a neural network trained to distinguish signal and background.

The Helene formula applies to only single-channel counting experiments, and thus cannot be used in this case. The natural practice in the case of fitting spectra is to perform a $\chi^2$ or maximum-likelihood fit. For the likelihood, the Poisson probability of observing the number of events $n_i$ in each bin, given the expected background $B_i$ and signal $S_i$ is multiplied together:

$$\mathcal{L} = \prod_{i=1}^{n_{\text{bins}}} \frac{\mu_i^n e^{-\mu_i}}{n_i!}$$

where $\mu_i = B_i + S_i$. The likelihood can be maximized (or, more usually, $-\ln \mathcal{L}$ is minimized) with respect to the normalization of the signal, or more generally calculated as a function of the signal normalization. This can be expressed as a variable $f$ which multiplies the signal prediction, such that we have $\mu_i = B_i + fS_i$. Though it is not often made explicit, if one assumes a flat prior pdf in $f$, then the posterior pdf in $f$ is, via Bayes’ Theorem, proportional to the likelihood:

$$\mathcal{P}(f|n_i, S_i, B_i) \propto \mathcal{L}(n_i|B_i, fS_i)$$

One can then, by plotting the likelihood as a function of $f$, set confidence intervals on $f$, the signal normalization. For example, to set a 95% CL limit on the signal, one finds that value of $f$ beyond which 5% of the total integral of the likelihood lies. If this value is less than $f = 1$, then one can conclude that the theoretical prediction is excluded at least the 95% level. Stated more precisely, one can conclude that, if there is equal a priori probability that the signal could have any normalization from zero to infinity, then it is less than 5% probable that the true value is more than the theoretical value.

Such a technique has been applied in numerous searches in CDF, including the search for fourth-generation $b'$ quarks decaying to $bZ$ [5], the search for the Standard Model neutral Higgs [6], and other searches. In fact, in these cases, the likelihood is written in such a way as to take into account uncertainties in the signal and background, and correlations in these uncertainties, by integrating over them in the same way as described above for single channel counting experiments. Also, in these cases, there is more than one channel involved. This is handled by simply multiplying the likelihoods for the different channels.

This illustrates powerfully the flexibility inherent in likelihood-based methods: combining channels and taking into account uncertainties is a trivial extension of the definition of the likelihood. The main difficulty lies in actually calculating the likelihood in cases where the correlations are complicated. This can be made tractable by Monte Carlo integration over these uncertainties.
7. ESTIMATING EXPERIMENTAL SENSITIVITY

Often in new particle searches one wants to know the sensitivity of a particular analysis, to know how strong a limit can be set with a certain amount of integrated luminosity, or conversely how much integrated luminosity is needed to set a limit or, more optimistically, discover the new particle. This information can be used to optimize analyses, or to estimate the discovery reach of a new machine or detector.

Most often one finds a simple approach is used, in which the ratio of the signal to the square root of the background, \( S/\sqrt{B} \) is used as the main indicator of experimental sensitivity. One can then estimate the integrated luminosity necessary for, say, a 5\( \sigma \) discovery by finding when \( S/\sqrt{B} = 5 \). A 95\% CL limit would correspond to \( S/\sqrt{B} = 1.96 \), using the one-sided formulation discussed above. This procedure gives a reasonable estimate of the required integrated luminosity only when uncertainties are negligible, and the statistics are well in the Gaussian range. It is possible to consider combining single channel counting experiments this way, by adding the values of \( S/\sqrt{B} \) in quadrature, but doing this procedure for spectral fits is not possible.

The most straightforward way to estimate experimental sensitivities is to use the likelihood as a function of the signal cross section (or cross section multiplier). This immediately allows for the possibility of incorporating systematic errors and correlations, combining channels, and using spectral fits, just as in the methods outlined in the previous sections.

The main new element in estimating experimental sensitivities is including the fact that there are many possible future experimental outcomes: how does one average over or otherwise take into account the relative probability for all the possible outcomes?

In the Tevatron Run 2 SUSY/Higgs Workshop [7], the Higgs Working Group adopted a statistical procedure based on the joint likelihood for all the various search channels. To take into account all possible future outcomes, the procedure generated large numbers of pseudoexperiments, and for each pseudoexperiment the same procedure which would be applied in a real experiment was applied to that particular outcome. In the case of no signal actually present, for example, the outcome would have only background events present, with Poisson fluctuations around the expected mean background. Then, the integral of the likelihood as a function of Higgs cross section was determined, and the 95\% point compared with the theoretical value. To determine the integrated luminosity threshold, then, the integrated luminosity was increased or decreased until in 50\% of the pseudoexperiments one could obtain a 95\% CL limit. (This follows the convention set by the LEP-II Working Group.

In the case of determining discovery thresholds, again many pseudoexperiments were generated, this time with signal present at the appropriate rate, given the theoretical cross section. To determine whether the particular outcome represented a 5\( \sigma \) discovery, for example, the ratio of the maximum likelihood to the likelihood at zero cross section was used. If this ratio was greater than the equivalent ratio for a Gaussian at 5\( \sigma \), then the outcome was deemed a 5\( \sigma \) discovery. As in the case of setting a limit, if in 50\% of the pseudoexperiments this was the case, then the threshold was said to be met. Figure 1 illustrates graphically the technique, as applied in both cases.

One could also imagine using more standard confidence interval definitions, such as the 68\% central interval, to determine whether the pseudoexperiment represented a 5\( \sigma \) discovery. In the limit of Gaussian statistics, the methods should be equivalent. But in the case of a likelihood which is asymmetric about the maximum, there is no set convention for setting such confidence intervals anyway. The bottom line was that the likelihood ratio was much easier to calculate numerically, and with the integral over systematic errors, compute time was very limited.
Fig. 1: Illustration of likelihood versus cross section multiplier for two cases in new particle searches, above where there is no signal, and below with a small signal present.

8. SUMMARY AND CONCLUSIONS

The techniques in CDF for setting limits and discovery significances in new particle searches have evolved, beginning early on with the Helene formula, extending the formula to include uncertainties on backgrounds and acceptance. In recent years the collaboration has shifted to likelihood-based methods, which allows the use of fits to spectra, and allows combining channels and the results from different experiments.

For discovery significances, typically CDF has used toy Monte Carlo techniques to estimate the probability of the null hypothesis, the probability that, in the case of no signal, the background alone could produce the observed number of events or more. But clearly this question as well can be addressed, in future analyses, using the same likelihood methods by which we would otherwise set limits, estimate experimental sensitivity and estimate integrated luminosity discovery thresholds.

A clear conclusion is thus that basing the estimates of limits, significances, and sensitivities on the likelihood offers the greatest hope of meeting the needs for incorporating uncertainties, fitting to spectra, and combining channels. Yet it leaves open many questions: Should the field abandon the frequentist view and adopt a purely Bayesian viewpoint? If so, what about the issue of the choice of prior pdf? If a frequentist approach is the goal, should the field adopt the Feldman-Cousins unified approach of likelihood ratio ordering or choose another statistic, such as in the LEP-II $C_{L_s}$ method? [8] Hopefully the field can overcome the present surfeit of methods and adopt a simply understood, explainable, and meaningful method for making these statistical estimates.

References


[8] See A. Read, these Proceedings.

Bob Cousins

What is the interpretation of your result? That is, you do all this and you say you’ve excluded at 95% confidence. So what does that number 95% mean?

John Conway

In which case? In the modified frequentist approach or in the likelihood ...

Cousins

The way you’re saying these things, the way of the future in CDF.

Conway

We know that in the limit of Gaussian statistics, if you apply that method, and in the case of no uncertainties, it converges to the same meaning as the frequentist case, and we also know that it doesn’t have that meaning as soon as you have systematic uncertainties, or start combining channels. It’s just a convention, I guess I would say.

Cousins

The flat prior has this nice property that in general the limits you get are conservative, by frequentist standards. So my question is, are you really being Bayesian, or are you using the mathematical machinery of Bayesian statistics to get an answer that you consider acceptable because it’s conservative on frequentist grounds?

Conway

I think that the main opinion of my colleagues is that they don’t tend to think about deep philosophical issues about probability and they’ll adopt a standard method just to go along with the flow. That’s most of my colleagues. Now the people who think about this, and I would count myself among them, still tend to regard this as use of the mathematical machinery from a practical point of view, and if the community at large were to adopt that, we would all know what each other means by a 95% confidence level limit. Personally I don’t, and I’ve changed over the last two years my own opinion, I was pretty strongly frequentist two years ago, but I realized that from a practical point of view, if we want to be able to combine channels and take into account correlations and uncertainties, we have got to use a method that is straightforward and understood by people.

Cousins

For example a flat prior for Poisson, if you use it for 90% lower limits you will always under-cover rather than always over-cover. So my guess is you’re evaluating it from a frequentist point of view. If someone showed you that 70% out of 90% confidence limits were, on average, going to be wrong by your technique, that you would switch to a different prior.

Conway

I think we’d be happy to adopt a prior which was standard in the community.
Bill Murray

Halfway through your talk you advocated the use of the change in likelihood interpreted as chi-squared as a discovery indicator rather than doing Monte Carlo experiments to establish the significance. But for the establishment of the exclusion, you’re not prepared to make the same extraction, you’d rather use the Bayesian integration. Why is that?

Conway

The question becomes: how do you take the distribution of likelihood versus this cross-section multiplier, and determine a significance. Suppose you had the case that’s shown right here, perhaps on a log scale. There are choices, as we have seen at this meeting, of the conventions for defining the confidence interval around a maximum, and I was just sort of throwing this out as a proposed convention that’s easy to understand, and in fact it’s the same convention as the likelihood ratio method that we heard about yesterday in one talk, I forget which talk that was.

Murray

I’m not sure which talk you’re referring to there, but it just seems very odd to use two different conventions, one for discovery and one for exclusion, when you could use the same convention for both. It seems unsatisfying.

Conway

It’s the same as these various likelihood ratio methods, is it not?

Murray

Well, for example, the Higgs group at LEP would use a frequentist fraction of times for both occasions, not ....

Michael Woodroofe

It seems to me that if you can write down what the intervals are, then you can compute the frequentist probability of coverage doing Monte Carlo. It may take a while, but you can do it.

Conway

And I would note that in the case of the Higgs Working Group, it already took quite a while to do these countless pseudo experiments, and making it much more complicated would be computationally intractible at this stage.

Harrison Prosper

It’s also true, if we were to use the suggestion of Bob Woodroofe, rather than using Gaussians which are really a pain, if we used Gamma distributions, one could actually do much of this analytically, and reduce the amount of computation.

Stephane Keller

Why do you call it the Bayesian integration of the systematic uncertainty? Is there a different way to calculate the probability of the data given in the theory? I thought it was the same in the frequentist
approach as the Bayesian approach. Can you explain this? Why do you call it the Bayesian integration of the systematic uncertainty when you calculate the likelihoods?

**J. Conway**

We’re just treating the true signal and true background as unknown random variables, much as in the previous talk to this one, and integrating them out.

**Stephane Keller**

Does that mean it’s different in the frequentist approach, you would do a different calculation of the probability of the data given in the table? I’m confused.

**Conway**

I don’t have an answer to that.

**Fred James**

Maybe I can answer that. The idea is that in the frequentist method you have to cover for any possible value of the unknown parameters, including the nuisance parameters like systematic effects, whereas in the Bayesian method you integrate over them. In the frequentist method, this integral doesn’t make sense, because parameters (including nuisance parameters) don’t have distributions, they just have a true value, which is unknown. Even in the Bayesian method, the parameters don’t have a distribution, only our beliefs have a distribution, and Bayesians are willing to integrate over beliefs, but not Frequentists.

**Giovanni Punzi**

Just a question. When you combine channels in this way, is this taking into account the fact that the actual systematic deviations belong to the same detector?

**Conway**

You can!

**Punzi**

Are you not, by combining channels, kind of simulating having the two different detectors extracted at random for the two channels, while they actually should be the same? Isn’t there some over-counting on this?

**Conway**

If you’re referring to the specific case of the result of the Higgs working group, we didn’t take into account correlations between channels. We regarded the systematic uncertainties as uncorrelated, which I think was in the conservative direction. Certainly in the future, once we do know the correlations, they can be taken into account in a joint likelihood in a more or less straightforward way in the Monte Carlo integration.