MODIFIED FREQUENTIST ANALYSIS OF SEARCH RESULTS (THE CL_s METHOD)

A. L. Read
University of Oslo, Department of Physics, P.O. Box 1048, Blindern, 0316 Oslo 3, Norway

Abstract
The statistical analysis of direct Higgs searches at LEP is described. The likelihood ratio with respect to the background-only hypothesis (or a related test-statistic) is used to order experimental results. The ratio of the confidences in the signal+background to background hypotheses, so-called "CL_s", is used to set lower bounds on the Higgs boson mass. The excluded mass interval which results has an untraditional but useful interpretation which differs both from frequentist intervals which require coverage and from Bayesian credible intervals. Issues such as flip-flopping, experimental uncertainties, discovery significance and the transition to measurement are discussed.

1. INTRODUCTION

The interpretation of results of searches for new particles and phenomena near the sensitivity limit of an experiment is a common problem in particle physics. The loss of sensitivity may be due to a combination of small signal rates, the presence of background comparable to the expected signal, and the loss of discrimination between two models due to insufficient experimental resolution. The search for Higgs bosons at LEP is such an experiment. The LEP experiments have separately, and in collaboration through the LEP working group for Higgs boson searches, developed a nearly common strategy for carrying out and reporting the results of their direct searches.

For the time being no significant evidence of Higgs production at LEP has been observed and lower bounds on Higgs masses have been reported. In this report I hope to explain how the lower bound is derived with the so-called CL_s method, why this method is used, and how to interpret the result.

Since the SM Higgs search has the lowest number of free parameters (1) it will be used to illustrate the features of the CL_s method. The generalization to models with several parameters (e.g. the MSSM) is straightforward if more time-consuming in practice. The techniques described in this talk are in fact successfully used in general scans over the many-parameter space of the MSSM in searches for the h and A neutral bosons, in 2-parameter searches for charged Higgs bosons of a general 2-doublet Higgs model, and in combined searches for sparticles by the LEP working group for SUSY particle searches.

The individual LEP experiments use either the likelihood ratio, a close approximation of the likelihood ratio, or the integral of the likelihood function as their test-statistics in Higgs searches. Exhaustive studies [1] have shown that they have similar performances for exclusion. To simplify my presentation I only described the likelihood ratio and I will do the same here. I will also show how the CL_s method can be applied in other contexts with an example of a hypothetical search for new physics via deviations of a parameter which is measured with normal-distributed uncertainty.

2. GOALS

One of the goals of the Higgs working group is to combine the results of the searches for Higgs bosons carried out by the four LEP experiments in a framework in which the transitions between exclusion, observation, discovery and measurement are as small as possible. These are direct searches, so the influence of theoretical preferences is minimized as much as possible. The searches are designed, indeed tuned, to maximize the sensitivity of the searches to the models. A specific modification of a purely classical statistical analysis (the introduction of CL_s) is used to avoid excluding or discovering signals which the search is in fact not sensitive to. Experimental (systematic) errors are taken into account. At the time of this workshop the Higgs boson searches at LEP have been combined assuming that the systematic uncertainties are uncorrelated, but part of the focus of the current combination effort is precisely to take into account the most important correlations in the uncertainties.

The use of CL_s is a conscious decision not to insist on the frequentist concept of full coverage (to guarantee that the confidence interval doesn't include the true value of the parameter in a fixed fraction of experiments). The Higgs working group has also not insisted on an automatic procedure for the transition between one and two-sided confidence intervals. On the other hand, it will be shown that the non-frequentist confidence interval which results does not suffer seriously from the flip-flop effect that the unified approach [2] is designed to address.

It has not been an explicit goal of the Higgs working group to choose a frequentist(-like) analysis rather than a Bayesian analysis on philosophical grounds. Our attitude is rather practical, we want to do the best we can with the data we have, where the best we can means excluding the Higgs as strongly as possibly in its absence (in a mass region where a direct search can be sensitive) and confirming its existence as strongly as possible in its presence (again, in a mass region where a direct search can be sensitive).

The goal of a search is to either exclude as strongly as possible the existence of a signal in its absence or to confirm the existence of a true signal as strongly as possible while holding the probabilities of falsely excluding a true signal or falsely discovering a non-existent signal at or below specified levels.

3. SEARCH RULES

The analysis of search results can be formulated in terms of a hypothesis test. The null hypothesis is that the signal is absent and the alternate hypothesis is that it exists. An analysis of search results is simply a formal definition of the procedure which quantifies the degree to which the hypotheses are favored or excluded by an experimental observation.

The first step in defining an analysis of search results is to identify the observables in the experiment which comprise the search results. The simplest observable is the number of candidates satisfying a certain set of criteria. More advanced observables may be some feature of the candidates such as reconstructed invariant mass, b-quark tagging probability, or even composite properties such as the output of a multi-dimensional discriminant or artificial neural-network analysis. The next step is to define a *test-statistic* or function of the observables and the model parameters (particle mass, production rate, etc.) of the known background and hypothetical signal which ranks experiments from the least to most signal-like (most to least background-like). The last step is to define *rules* for exclusion and discovery i.e. specify ranges of values of the test-statistic in which observations will lead to one conclusion or the other. In practice one often wishes to specify the significance of the exclusion or discovery, and not simply give a true or false answer. In other words a *confidence level* for the exclusion will be quoted. A *confidence limit* for exclusion is defined as the value of a population parameter (such as a particle mass or a production rate) which is excluded at a specified confidence level. A confidence limit is a lower (upper) limit if the exclusion confidence is greater (less) than the specified confidence level for all values of the population parameter below (above) the confidence limit. *Note that confidence intervals obtained in this manner do not have the same interpretation as traditional frequentist confidence intervals nor as Bayesian credible intervals.*

For convenience the test-statistic Q is constructed to increase monotonically for increasingly signal-like (decreasingly background-like) experiments so that the confidence in the signal$+background$ hypothesis is given by the probability that the test-statistic is less than or equal to the value observed in the experiment, Q_{obs}:

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}), \tag{1}$$

where

$$P_{s+b}(Q \le Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_{s+b}}{dQ} dQ, \tag{2}$$

and where dP_{s+b}/dQ is the probability distribution function (p.d.f.) of the test-statistic for signal+background experiments. Small values of CL_{s+b} indicate poor compatibility with the signal+background hypothesis and favor the background hypothesis. Similarly, the confidence in the background hypothesis is given by the probability that the test-statistic is less than or equal to the value observed in the experiment, Q_{obs}:

$$CL_b = P_b(Q \le Q_{obs}), \tag{3}$$

where

$$P_b(Q \le Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_b}{dQ} dQ \tag{4}$$

and where dP_b/dQ is the p.d.f. of the test-statistic for background-only experiments. Values of CL_b very close to 1 indicate poor compatibility with the background hypothesis and favor the signal+background hypothesis.

3.1 Introducing CL_s

Taking into account the presence of background in the data may result in a value of the estimator of a model parameter which is "unphysical", e.g. observing less than the mean expected number of background events could be accommodated better if the signal cross-section was negative. It is important to make the distinction between the estimator, which may be expected to be "unphysical" with a probability of up to 50% for negligible or absent signals, from the parameter itself which may well be physically bounded. When an experimental result appears consistent with little or no signal together with a downward fluctuation of the background, the exclusion may be so strong that even zero signal is excluded at confidence levels higher than 95%. Although a perfectly valid result from a statistical point of view, it tends to say more about the probability of observing a similar or stronger exclusion in future experiments with the same expected signal and background than about the non-existence of the signal itself, and it is the latter which is of more interest to the physicist. Presumably a great deal of effort has already gone into verifying the correctness of the background model, so there is little point in obtaining a result which is more sensitive to fluctuations of the known background than to the hypothetical signal.

One of the reasons that there is no consensus on how to treat these situations is that the result is ambiguous. There is simply not enough information available to distinguish clearly between the signal and the signal+background hypotheses - we just don't know what the result means. This will be clearly illustrated when we look at distributions of the test-statistic and evaluate search potentials.

One possible technique for dealing with this situation is to normalize the confidence level observed for the signal+background hypothesis, CL_{s+b}, to the confidence level observed for the background-only hypothesis, CL_b. This is a generalization of the modified classical calculation of confidence limits for single channel counting experiments presented in [3]. This also makes it possible to obtain sensible exclusion limits on the signal even when the observed rate is so low that the background hypothesis is called into question. Of course, the experimentalist should be aware that a low background confidence may also indicate underestimated or forgotten systematic errors. It may be said that this *modified frequentist* or CL_s procedure (as it will be called here) is performed in order to obtain conservative limits on the signal hypothesis. That this procedure is conservative is undeniable, but I prefer to add that it gives an approximation to the confidence in the signal hypothesis, CL_s, one might have obtained if the experiment had been performed in the complete absence of background, or in other words, if it had been possible to discard with absolute certainty the selected events due to background processes.

The modified frequentist re-normalization described above is simply

$$CL_s \equiv CL_{s+b}/CL_b. \tag{5}$$

Although CL_s is not, strictly speaking, a confidence (it is a ratio of confidences), the signal hypothesis will be considered excluded at the confidence level CL when

$$1 - CL_s \leq \mathrm{C}L. \tag{6}$$

The consequence of CL_s not being a true confidence is that the hypothetical false exclusion rate is generally less than the nominal rate of $1-CL$. The difference between CL_s and the actual false exclusion rate will in fact increase as the p.d.f.'s of the signal+background and background hypotheses become more and more similar. Thus the use of CL_s increases the "coverage" of the analysis, i.e. the range of model parameters for which an exclusion result is possible is reduced, but it also avoids the undesirable property of CL_{s+b} that of two experiments with the same (small) expected signal rate but different backgrounds, the experiment with the larger background may have a better expected performance.

3.2 Other definitions of CL_s

Three of the four LEP experiments use the above definition of CL_s, while ALEPH [4] uses

$$CL_s = CL_{s+b} + (1 - CL_b) \times e^{-s}.$$

There is some skepticism on the part of the other LEP experiments to adopt this alternate definition. One of the objections is that the appearance of the global parameter s, the total expected signal rate, opens the way for absurd optimizations. Adding a new channel with a moderate signal rate and a completely overwhelming background to an existing search will give an improvement to the search sensitivity out of proportion to the signal-to-noise ratio in the additional channel (a microscopic S/N should indicate that the new channel contains practically no information about the signal). Another objection, which is more of a philosophical nature, is that this definition of CL_s can not be applied to searches which consist of looking for small deviations of parameters measured with normal-distributed errors.

4. THE LIKELIHOOD RATIO TEST-STATISTIC

The likelihood ratio, $Q(\vec{X})$, is the ratio of the probability densities for a given experimental result $\vec{X}$ for two alternate hypotheses. In searches for new particles an appropriate likelihood ratio is $Q = \mathcal{L}(\vec{X}, s + b)/\mathcal{L}(\vec{X}, b)$, that is the ratio of probability density for the signal+background hypothesis to the signal-free or background hypothesis.

The likelihood ratio for an experiment with independent channels is simply a product of the likelihood ratios of the individual channels, so that the combination of additional histogram bins, independent search channels, experiments or center-of-mass energies is straightforward and unambiguous.

The likelihood ratio can be thought of as a generalization of the change in χ^2 for a fit to a distribution including signal plus background relative to a fit to a pure background distribution. In the high-statistics limit the distributions of $-2\log Q$ are in fact expected to converge to $\Delta\chi^2$ distributions.

The likelihood ratio Q for experiments with N_{chan} independent search channels and measurements of a discriminating variable x (for multidimensional discriminants replace x with $\vec{x}$) for each candidate, and where the absolute signal and background rates are known, can be written as

$$Q = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i}b_i^{n_i}}{n_i!}} \frac{\prod_{j=1}^{n_i} \frac{s_i S_i(x_{ij}) + b_i B_i(x_{ij})}{s_i+b_i}}{\prod_{j=1}^{n_i} B_i(x_{ij})} \tag{7}$$

which can be simplified to

$$Q = e^{-s_{tot}} \prod_{i=1}^{N_{chan}} \prod_{j=1}^{n_i} \left(1 + \frac{s_i S_i(x_{ij})}{b_i B_i(x_{ij})}\right), \tag{8}$$

where n_i is the number of observed candidates in each channel, x_{ij} is the value of the discriminating variable measured for each of the candidates, s_i and b_i are the integrated signal and background rates per channel, s_{tot} is the total signal rate for all channels, and $S_i(x)$ and $B_i(x)$ are the probability distribution functions of the discriminating variable for the signal and background of channel i respectively.

If the p.d.f.'s of the discriminating variable are identical for the signal and background, if none is measured or if the distributions are expressed as binned histograms, the likelihood ratio simplifies further to

$$Q = e^{-s_{tot}} \prod_{i=1}^{N_{chan}} \left(1 + \frac{s_i}{b_i}\right)^{n_i}. \tag{9}$$

Note that in the *complete* absence of background ($b = 0$) and the observation of one or more candidates, an alternate null-hypothesis must be chosen, such as that the signal is the one that maximizes the likelihood function $\mathcal{L}(s)$. In such a situation the existence of the signal is undeniable and the setting of confidence limits is firmly in the realm of measurement.

A simple derivation shows that the likelihood ratio method is effectively based on counting weighted events. Since $Q > 0$ and $P(Q \leq Q_{obs}) = P(\ln(Q) \leq \ln(Q_{obs}))$ we can write

$$\ln(Q) = -s_{tot} + \sum_{k=1}^{n} n_k w_k \tag{10}$$

where n is the total number of events observed in all channels and the weight for each candidate k is given by

$$w_k = \ln\left(1 + \frac{s_k}{b_k}\frac{S_k(m_k)}{B_k(m_k)}\right), \tag{11}$$

where the k index also assigns the candidate to the search channel in which it was observed. Since the constant s_{tot} appears on both sides of the expression $\ln(Q) \leq \ln(Q_{obs})$, the method consists basically of comparing the observed number of weighted events with the distributions expected for the signal+background and the background hypotheses.

4.1 Single channel counting experiment

For a counting experiment with a single channel all the candidate events have the same weight, $\ln(1 + s/b)$, so that Eqn. (5) takes the form

$$\mathrm{CL}_s = \frac{P(X \leq X_{obs})}{P(X_b \leq X_{obs})} = \frac{P(n \leq n_{obs})}{P(n_b \leq n_{obs})}, \tag{12}$$

where n_b and n come from the Poisson distributions of the number of events for the background and signal+background hypotheses respectively, and n_{obs} is the number of candidates observed in the experiment. Thus the modified frequentist signal exclusion confidence becomes

$$\mathrm{C}L = 1 - \frac{\sum_{n=0}^{n_{obs}} \frac{e^{-(b+s)}(b+s)^n}{n!}}{\sum_{n=0}^{n_{obs}} \frac{e^{-b}b^n}{n!}}. \tag{13}$$

An *identical* result is obtained by computing the Bayesian credible interval (with uniform prior probability density for the signal s')

$$\mathrm{C}L = \frac{\int_s^\infty \mathcal{L}(s', b)ds'}{\int_0^\infty \mathcal{L}(s', b)ds'}. \tag{14}$$

Without going into detail, let's just say that this interesting coincidence is responsible for a lot of confusion.

4.2 Search potential optimization

The illustration of the search optimization follows closely the simple presentation of the well-known Neyman-Pearson theorem in [5]. The same conditions that make the maximum likelihood estimator the most efficient statistic for parameter estimation in many situations exist also for hypothesis testing. The basic principle is illustrated in Fig. 1.

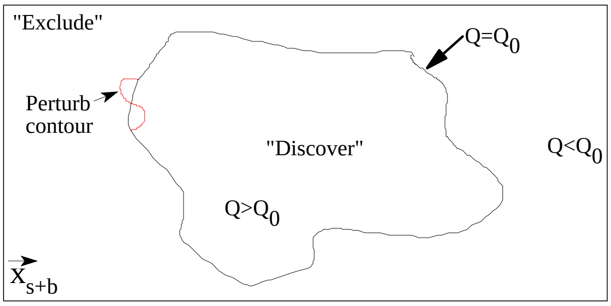

Fig. 1: Illustration of Neyman-Pearson theorem applied to searches.

Imagine that the box represents the uniformly distributed set of all possible experimental outcomes for the signal+background hypothesis. The likelihood ratio Q defines a set of contours with a constant ratio of signal+background density to background density. Suppose we choose one contour given by $Q = Q_0$ and separate all possible experiments into "discovery" (conclusion that signal hypothesis is true) and "exclusion" (conclusion that signal doesn't exist) classes. In practice we usually choose much more stringent criteria for discovery than for exclusion (we set up at least two contours) and we accept that experimental results may not always be conclusive. In order to maximize the probability for correctly confirming (excluding) the signal, the region with $Q > Q_0$ ($Q \leq Q_0$) is defined as the "discovery" region ("exclusion" region). To show that no further optimization is possible, the likelihood ratio contour is perturbed in such a way that that the fraction of signal+background experiments is constant. But since the background density outside the contour is larger than inside, the probability that a background experiment will lead to a false confirmation of the signal has increased. Similarly, if we imagine that the perturbation is done in such a way as to hold constant the fraction of background experiments, the probability of falsely excluding the signal is increased since the fraction of signal+background experiments in the exclusion region has increased. Since, for a fixed exclusion rate for background experiments or for a fixed "discovery" rate for signal+background experiments, any perturbation of the contour given by the likelihood ratio increases the false exclusion rate and the false discovery rate, the likelihood ratio is shown to be the optimal test-statistic for searches.

The use of any other test-statistic (ordering principle) represents a perturbation of the optimal contours defined by the likelihood ratio and thus yield less sensitive hypothesis tests for searches. All test-statistics, including frequentist confidence intervals with or without exact coverage or Bayesian credible intervals with reasonably unbiased or finely tuned prior probability distributions have an expected distribution for background experiments and another for signal+background experiments. The distribution for signal-only experiments (the one physicists would like to draw conclusions from) simply doesn't exist experimentally in the large majority of searches and using other test-statistics won't make this particular problem go away.

4.3 Terminology

It is probably worth defining the language used in the Higgs working group in terms of traditional statistical terminology (in italics). *Accepting the null-hypothesis* ($\mathcal{H}_0$: there is no signal or it is too small to be seen) is what we call exclusion. The *power* of this test, that is the probability to correctly exclude an absent signal is what we call the exclusion potential. The probability to falsely exclude a true signal, that is to commit a *type II error*, is what we call the false exclusion rate. In cases where there is complete separation of the distributions of the likelihood ratio for the signal+background and the background hypotheses, the false exclusion rate will be specified by one minus the confidence level of the exclusion (for discrete probabilities there are unavoidable deviations from this ideal behavior). The *power* of the discovery test is the probability to correctly confirm the signal+background hypothesis; this is what we refer to as the discovery potential of the experiment. The probability to falsely discover an absent signal, that is to commit a *type I error*, is what we call the false discovery rate. In ideal cases where the likelihood ratio distributions for the background and signal+background hypotheses are completely separated, the *significance level* (here, for once, we use the same language) of the discovery is equal to the false discovery rate.

4.4 Consequences of CL_s

The use of CL_s as the figure of merit for signal exclusion in general causes the false exclusion rate to be lower than the ideal rate give by the specified value of the exclusion confidence level. Similarly, the use of $(1 - CL_b)/(1 - CL_{s+b})$ instead of $1 - CL_b$ for discovery [6] causes the false discovery rate in general to be lower than the stated significance level. An example (taken from one of the Higgs searches at LEP) of the reduced false exclusion rate for exclusion at the 95% confidence level is shown in Fig. 2. The dashed line for the false exclusion rate, if CL_{s+b} would be interpreted as the confidence in the signal hypothesis, actually continues to the right until the expected signal event rate (sum over search channels of cross-section times luminosity times branching fraction times detection efficiency) falls identically to zero. The vertical dotted line at 92.5 GeV shows where the ideal background-free CL_s would drop to zero suddenly when the probability to observe zero candidates is larger than 5% (expected rate of signal events 3, as seen in middle plot of the figure) and exclusion at the 95% confidence level is no longer possible. The bottom plot in the figure shows the potential for exclusion (reminder: the fraction of background experiments leading to exclusion at 95% CL or higher) versus Higgs mass for CL_s (the solid curve) and the potential if CL_{s+b} were interpreted as the confidence in the signal hypothesis. In the region to the right of 92.5 GeV, indicated by the vertical dotted line, the expected signal rate is less than 3, and the exclusion potential for a background-free search would be identically zero.

To summarize the previous paragraph, the exclusion potential for CL_{s+b} flattens out at 5% even when the expected signal rate is microscopic. The false exclusion rate for CL_{s+b} is also 5% for microscopic signal rates, which is in fact entirely correct from the purely frequentist viewpoint since we know a mistake is being made at the rate of 5% when a signal is excluded for which the experiment has no sensitivity (we might as well throw dice). This is the main motivation for adopting the CL_s method.

Figure 3 shows an example of the likelihood ratio (Q) distributions (in fact minus twice the log-likelihood ratio) for experiments with varying degrees of sensitivity. The confidences CL_b and CL_{s+b} are the integrals of the normalized distributions from right to left. The example is taken from the Higgs search at LEP. For light Higgs the cross-section is large and the distributions for signal+background and background are well separated. In this case the most probable results are either strong exclusion ($CL_b \sim 0.5$, tiny CL_{s+b}) or a strong confirmation of the signal ($CL_b \to 1$, $CL_{s+b} \sim 0.5$). As the hypothetical mass of the Higgs increases, the cross-section falls, the overlap of the likelihood ratio distributions increases and the most probable results for the two hypotheses move closer to each other. One is no longer able to conclude that one of the hypotheses is much more strongly supported than the other - the result tends to be ambiguous. The CL_s method can be seen as a way of taking this ambiguity into account in the extraction of a single result characterizing the possible presence of a signal.

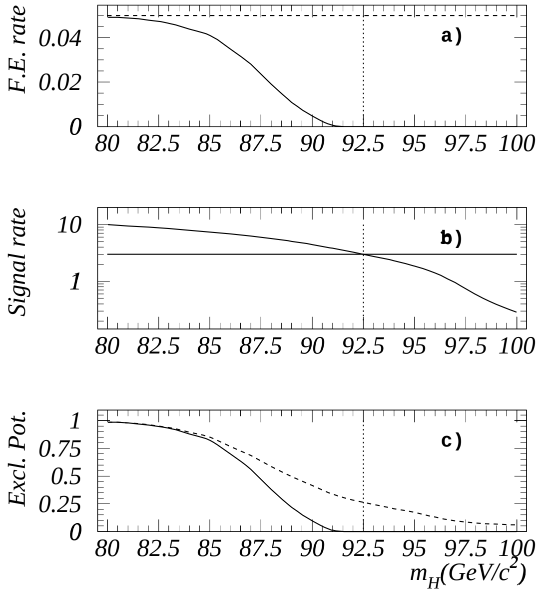

Fig. 2: False exclusion rate (a), expected signal rate (b) and exclusion potential (c) for exclusion at the 95% confidence level versus Higgs mass for a typical Higgs search at LEP. The solid curves in (a) and (c) are for CL_s and the dashed line and curve for CL_{s+b}. The vertical dotted lines show where the expected signal rate in (b) falls below 3.

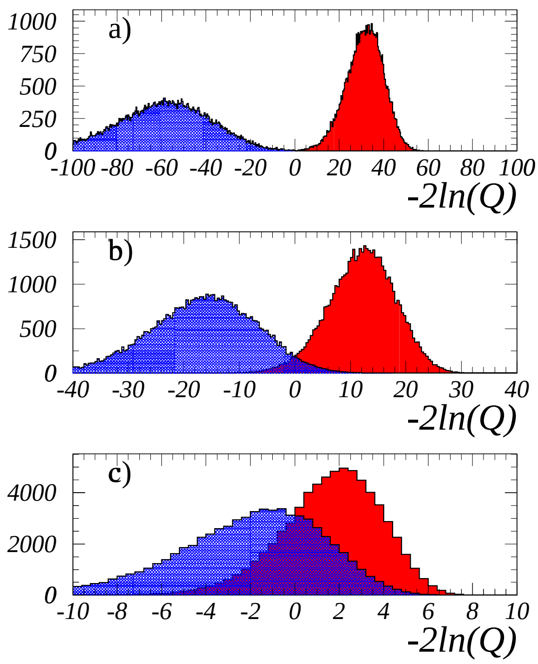

Fig. 3: Examples of distributions of minus twice the log-likelihood ratio ($-2ln(Q)$) for the signal+background (light shaded histograms on the left) and background (dark shaded histograms on the right) hypotheses from the Higgs search at LEP: a) for a light Higgs with large cross-section, b) for a moderate Higgs with moderate cross-section, c) for a heavy Higgs with small cross-section. The vertical scales are arbitrary.

Another way of interpreting CL_s is that it serves as an approximation of the confidence one might obtain if the background events could be removed from the sample of selected events. Obviously, if this were possible the experimentalist would have done it already! But this *is* possible in a Monte Carlo study and an example of such a study is shown in Fig. 4(a)-(d). In (a) and (b) the confidence distributions are uniform since the distributions are formed for the hypotheses being tested, except for the small peak in (a) which is due to the probability e^{-b} of observing zero background candidates. The distribution of confidences in (c) is obtained with signal-only experiments; this is possible with gedanken experiments but *not* in the real experiment. The peak at the left of (c) is due to the probability e^{-s} of observing zero signal candidates. The additional structure in (c) is caused by the use of histograms to describe the signal and background discriminant distributions instead of continuous functions and is only a technical distraction here. In (d) one sees that the peak at $CL = e^{-s}$ is reduced with respect to (a); this is because for signal+background experiments the probability to observe zero candidates is $e^{-(s+b)}$. In addition there is a tail from the peak on top of the uniform distribution; this is due to the experiments in the overlap region of the signal+background and background distributions of the test-statistic. These features of (d) lead to overcoverage but (c) is experimentally inaccessible.

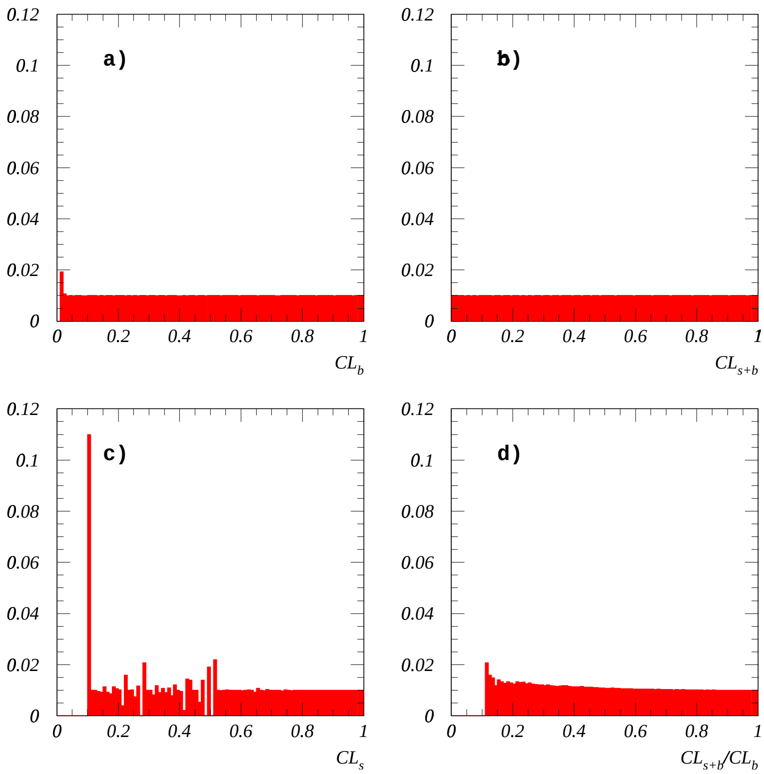

Fig. 4: Example of distributions of confidences from a Higgs search at LEP: a) distribution of CL_b expected for background experiments, b) distribution of CL_{s+b} expected for signal+background experiments, c) distribution of CL_s expected for *signal-only* experiments, and d) distribution of modified CL_s expected for signal+*background* experiments.

5. 'LOOK-ELSEWHERE' EFFECT

When establishing the significance of a possible signal from a model with a free parameter, e.g. the mass of the Higgs boson predicted by the Standard Model, our attention is naturally drawn to the point where the likelihood function is maximized (corresponding to the minimum of $-2ln(Q)$). However it is not quite sufficient to test if the expected rate of false discovery conservatively given by $(1 - CL_b)/(1 -$

CL_{s+b}) at the most likely point is small enough to meet our discovery criteria, since the background can fluctuate anywhere and not just at the point we focus on because of the data at hand. This is the equivalent of the "many-histograms" problem in data analysis. If you look at enough histograms, sooner or later you *must* find large deviations from the standard result, even if there is no new phenomenon at work. The "look-elsewhere" effect, as it is called in the Higgs working at LEP, may be estimated roughly by the ratio of the search region (e.g. the range in Higgs mass) to the experimental mass resolution. The dilution of the significance level was estimated to be a factor ~ 4 for the combined search at LEP with $\sqrt{s} \leq 189$ GeV [14]. This is less dramatic than it sounds, since it would imply e.g. reducing the significance of a 5σ observation to 4.5σ. In addition, if the search sensitivity is far from uniform over the mass range under consideration (such is usually the case for the Higgs search at LEP), the background will tend to give signal-like fluctuations mostly in the region of reduced sensitivity, thus leading to smaller dilution factors than the rough estimate.

6. NORMAL DISTRIBUTION - AN ILLUSTRATION

A study of a search for deviations of a parameter measured with a normally distributed uncertainty is a useful illustration of the properties of the CL_s method. The background will be a normal distribution with mean 0 and standard deviation 1. The hypothetical search will be for a signal that gives a small, positive deviation from 0 and for simplicity it is assumed that the standard deviation remains constant independent of the true value of the signal. Fig. 5 shows the distribution of the observable X for the background ($X_{model} = 0$) and for a hypothetical signal ($X_{model} = 1$). The log-likelihood ratio distributions for the background and signal+background hypothesis will also be normal distributions separated by one standard deviation, so it is sufficient to use X as the test-statistic.

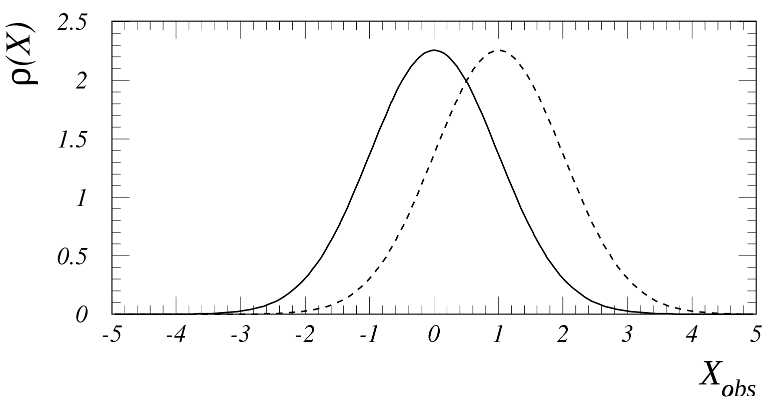

Fig. 5: Example of the unnormalized probability distributions of an observable X for background-only (solid curve) and signal+background (dashed curve) hypotheses (example of Sec. 6.).

Since we restrict the search to positive signals, the confidences are integrals over these distributions from $-\infty$ to X_{obs} as shown in Fig. 6. Recall that the upper bound on X_{model} is found when $CL_s(X_{obs}) = 0.05$. The solid curve in the figure for CL_b is independent of the signal model and indicates the compatibility of the observation with the background hypothesis. Values of CL_b close to 1 indicate a signal-like (non background-like) result. The dashed curve shows CL_{s+b} for $X_{model} = 1$. For other values of X_{model}, CL_{s+b} is obtained by sliding the dashed curve to the left (but not to the left of CL_b) or to the right. A family of CL_s curves (dotted curves) for $X_{model} = 0.1 - 4.0$ are also shown.

Several features are apparent in the figure.

- CL_s approaches CL_{s+b} for $X_{model} >\sim 3$ for any value of X_{obs}.
- CL_s approaches CL_{s+b} for $X_{obs} >\sim 2$ even for small values of X_{model}.
- For increasingly large, negative values of X_{obs} the upper bound on X_{model} given by CL_s approaches zero *slowly* but never reaches it.

It is also apparent from the figure that lower bounds at the 95% confidence level, defined by the value of X_{model} that solves $CL_s(X_{obs}) = 0.95$, also exist. These lower bounds make sense when the evidence for a signal is strong but as long as the observation is consistent with background they don't contain much information. In fact, it is not hard to show that when evidence for a signal is strong, that the confidence intervals found by more traditional techniques are recovered from the CL_s method. For example, the 84% confidence level upper and lower bounds correspond exactly to the traditional frequentist 68% confidence interval and the 68% Bayesian credible interval (with uniform prior from zero to ∞). This is accomplished with little flip-flopping on the part of the physicist. The upper bound is computed with a procedure which is entirely independent of the observed result, be it very compatible with background or an outstanding discovery of a signal. The only flip-flopping is the subjective decision whether or not to quote the lower bound. A confidence interval which doesn't contain zero can be misunderstood if the signal is poorly established and, for example, the LEP Higgs searches will probably not quote upper bounds on the Higgs mass until at least "possible observation" criteria have been met.

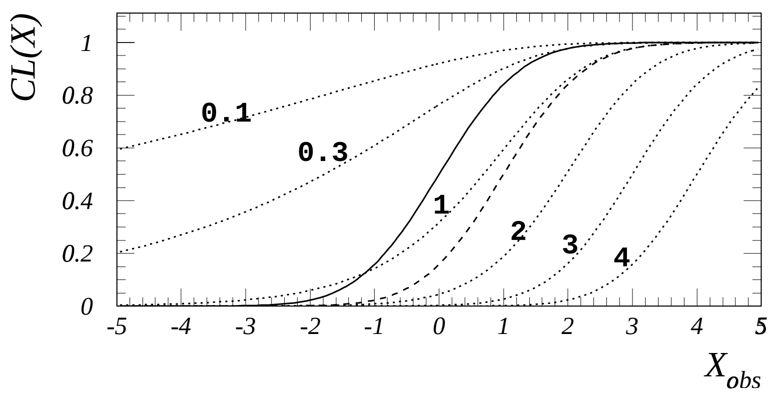

Fig. 6: Confidences versus observed value of X for the background-only hypothesis (CL_b, solid curve), the signal+background hypothesis for $X_{model} = 1$ (CL_{s+b}, dashed curve) and various CL_s curves (dotted) for X_{model} ranging from 0.1 to 4 (example of Sec. 6.).

6.1 Normal distribution - exclusion

If an experiment is entirely without sensitivity to a model, it should be forbidden to exclude it, and if the sensitivity is poor it should be extremely difficult to exclude it. In the present example the use of the purely frequentist CL_{s+b}, which gives optimal sensitivity for exclusion, has the frequentist property of being wrong a fixed fraction of the time, also for microscopic values of X_{model} where the experiment is clearly not sensitive to the model being tested. Fig 7 shows the false exclusion rate versus signal model for both frequentist and CL_s methods. The CL_s curve has a slow turn-on from zero for no signal towards the specified false exclusion rate (5%) as the background and signal+background distributions separate.

The exclusion potentials for the purely frequentist and CL_s methods are compared in Fig. 8. The exclusion potentials converge as they both approach 100%, the region where the distributions of X are well-separated ($\sim 3\sigma$ separation). For $X_{model} <\sim 2$ the overlap of the distributions of X is large, the sensitivity of the experiment is obviously poor, and the exclusion potential of the CL_s method is naturally suppressed.

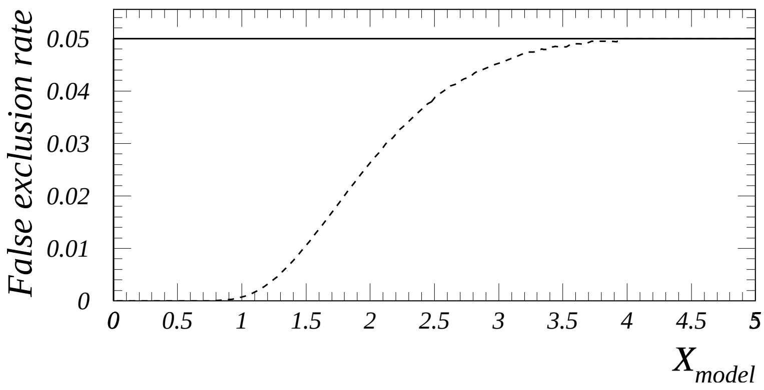

Fig. 7: Probability of falsely excluding the signal versus the signal model parameter X_{model} when using CL_{s+b} (solid line) and CL_s (dashed curve) to set exclusion limits at the 95% confidence level (example of Sec. 6.).

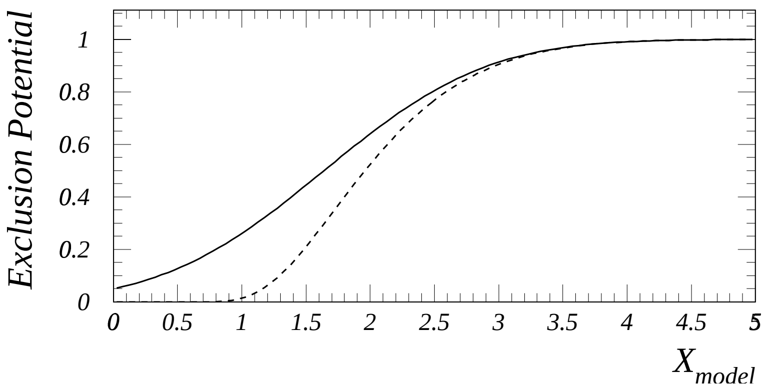

Fig. 8: The probability of excluding the *false* signal hypothesis versus the signal model parameter X_{model} when using CL_{s+b} (solid curve) and CL_s to set exclusion limits at the 95% confidence level (example of Sec. 6.).

6.2 Normal distribution - discovery

If an experiment is entirely without sensitivity to a model, it should be forbidden to discover it whether or not it might observe large background fluctuations. Fig. 9 shows how the discovery potential is affected by the generalization of CL_s for the determine of the signal significance. The plots contain the same information on log and linear scales. One sees that using $1-CL_b < 5.7\times10^{-7}$ as the discovery criterion allows experiments with no sensitivity to the signal to make discoveries (admittedly with a small rate, but even so this is not reasonable) whereas this is very strongly suppressed by using $(1-CL_b)/(1-CL_{s+b})$ instead. This suppression has mostly disappeared by the time the background and signal+background distributions of X are separated at about the 5σ level. The false discovery rate for $1 - CL_b$ is the stated value of $1 - CL_b$ whereas it is effectively zero for $(1 - CL_b)/(1 - CL_{s+b})$ for less than 2σ separation of the background and signal+background models and converges towards the stated value for $\sim 5\sigma$ separation.

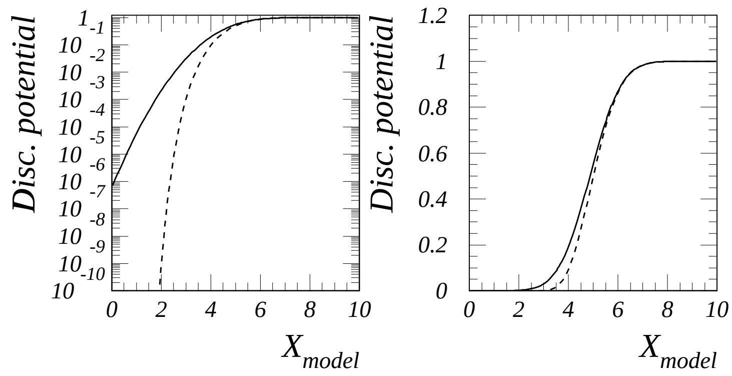

Fig. 9: The probability of claiming a discovery with a significance corresponding to 5σ versus the signal model parameter X_{model} when using $1 - CL_b$ (solid curves) and $(1 - CL_b)/(1 - CL_{s+b})$ (dashed curves) as the discovery criteria. The two plots show the same information on logarithmic and linear scales (example of Sec. 6.).

7. BACKGROUND SUBTRACTION

Background in the search results is accounted for in several ways. First, it appears in the likelihood ratio (or other test-statistic). Second, even if doesn't appear in the test-statistic (which would thus be non-optimal), the confidences are computed by comparison of the value of the test-statistic observed in the experiment with the distributions of the test-statistic expected for the background and background+signal hypotheses.

It is often tempting to shift the background estimate in order to obtain conservative results (conscious overcoverage). Increasing the background estimate leaves less room for the hypothetical signal thus leading to conservative discovery significances. However, this also leads to overly aggressive exclusion results (undercoverage) if, when the experiment is carried out, the observed result is reasonably compatible with the background expectations. If the background estimate is decreased then exclusion becomes conservative and discovery overly aggressive.

If the expected background rate is set to zero for the gedanken experiments in the computation of the distributions of the test-statistic, then all selected data events are considered to originate from the signal, exclusion results are maximally conservative and no conclusions whatsoever can be drawn about observation or discovery (since $CL_b = 1$ by construction). The advantage of this extreme procedure is that it tolerates unknown systematic uncertainties in the background estimates [7]. A disadvantage, in addtion to the extremely conservative exclusion and the complete absence of discovery potential, is that in such a case there is a mismatch between the hypotheses being tested and the hypotheses used to generate the distributions of the test-statistic. Thus the likelihood ratio is no longer guaranteed to be optimal and methods with tunable parameters, for example [8] and [9], will perform better.

A final comment on background subtraction is that taking the ratio $CL_{s+b}/CL_b = CL_s$ to define an approximate signal-only hypothesis test may appear to be a background-subtraction procedure, but if the background is properly accounted for in the computations of CL_{s+b} and CL_b, they are already "background-subtracted" quantities and no further background subtraction is possible.

8. SEARCH OPTIMIZATION

In section 3. it was shown that given a certain amount of information about a search, the choice of the the likelihood ratio with respect to the background-only hypothesis as the test-statistic will maximize the sensitivity of the search for both exclusion and discovery. However, the choice of what information to put in the likelihood ratio is a critical aspect of optimization. Analysis that assigns events to two classes, the rejected class (none of these participate in the confidence computation) or the accepted class (all of these participate with equal weight in the confidence computation) is the basis of the simple counting experiment. Adjustment of the cut(s) that define the two classes will affect the discovery and exclusion potentials. At this point minimizing the average (or median) value of CL_s expected for background experiments by adjustment of the cuts will maximize the exclusion potential [10]. It should be kept in mind that although the exclusion and discovery potentials are maximized for a specific information content with the use of the likelihood ratio, this is *not* the same as saying that the information content that globally maximizes the exclusion potential also globally maximizes the discovery potential.

If the search has several well-defined final states (e.g. $HZ \rightarrow 4jets, HZ \rightarrow 2jets + l^+l^-$, etc.) with different signal to noise ratios (S/N) , the search sensitivity is improved by splitting the search into separate channels so that events selected in a channel with lower S/N are weighted less than those selected in a channel with a good S/N. Since the likelihood ratio accounts for the variations between channels of S/N in an optimal way, the addition of a channel always improves the search sensitivity, even if the background rate is large and/or S/N is poor (but uncertainties on the background can dampen or even reverse the improvement, so there is an optimal amount of background to allow [11]).

If in addition to topology, the measured values of some feature(s) of the event are different for the signal and background, this can be introduced into the likelihood ratio with additional improvements in the sensitivity. Of particular importance is the identification of observables directly related to a parameter in the signal model being tested (e.g. the reconstructed mass of the Higgs candidate in the Higgs search), but also roughly model-independent observables are quite useful (e.g. b-tagging for $H \rightarrow b\bar{b}$ is only mildly m_H-dependent due to reconstruction effects).

One danger of optimization is that the event selection gets sub-divided enough that statistical fluctuations in the detector simulation of either the background or the signal produce spurious peaks of large S/N, resulting in an artificial improvement of the search sensitivity. One way to detect the onset of this over-training is to split the detector simulation into sub-samples. If the search sensitivity is better for both of the sub-samples than for the combined sample, this is clear evidence of over-training in the sub-samples. If the full-sample gives results compatible with the mean of the sub-samples, then the full sample is most likely not suffering from over-training.

The performance of a statistical analysis of search results should not improve by the increase of the background with no additional efficiency for the signal (and it should not improve significantly if the added signal efficiency comes at the cost of an overwhelming background). The CL_s method is relatively immune to this kind of false optimization - this is a strong point in its favor.

9. UNCERTAINTIES

Very seldom are all the ingredients of a search without experimental (systematic) uncertainty. Background rates and signal detection efficiencies, even the theoretical input may be uncertain (e.g. missing higher order corrections). Since a confidence interval is already an expression of uncertainty, one doesn't want to quote an uncertainty on the confidence limits, but rather modify the confidence limits to allow for the experimental uncertainty. A simple procedure is to shift all the relevant parameters (backgrounds, efficiencies, etc) coherently by one standard deviation of each of the individual parameters in the direction which weakens the confidence limit. Cousins and Highland have shown in [12] that such a procedure is far too pessimistic.

What is done in the Higgs searches at LEP is to use traditional Bayesian techniques to infer a likelihood distribution for the parameters in question and then treat them as probability distributions in the generation of gedanken experiments (with MC, FFT or whatever technique). For each gedanken experiment a new set of "smeared" efficiencies and background rates are generated and from these, background and signal events are generated to form the input to the likelihood ratio for this gedanken experiment. This procedure is the generalization of the Monte Carlo computation which is compared to the analytic approximations derived in [12] for simple counting experiments in the presence of background and reproduces those results. This should not be a surprise since the CL_s method applied to the likelihood ratio for single-channel counting experiments is equivalent to the Bayesian credible limits computed with a flat prior distribution used in [12].

The consequences of this "smearing" procedure is that the likelihood ratio distributions get widened and that especially the background tail under the signal+background distribution and the signal+ background tail under the background distribution are enhanced. In other words the overlap of the distributions has increased. This reduces both the exclusion and discovery potential of the search and tends to weaken both discovery-like and exclusion-like observations. As was described above, the effect on moderate exclusion (95% CL) tends to be minor, but for extreme exclusion and especially for discovery the effect can be relatively dramatic (dramatic meaning that a discovery significance can *easily* be reduced by a sigma when the experimental uncertainty is accounted for). The experience with the LEP Higgs searches confirms the conclusion in [12] that even moderately large uncertainties (say $\sim 20\%$) have little effect on exclusion limits (lowering expected and observed lower bounds on m_H at LEP by typically a few hundred MeV).

Important correlations between the parameters that describe the signal and background rates and distributions should certainly be taken into account. This is one of the current activities of the Higgs working group. The effect of correlations is expected to be most noticeable in case of discovery-like results.

10. TECHNICAL CHALLENGES

The brute-force method of computing the distributions of the likelihood ratio expected for the background-only and signal plus background hypotheses is to use a Monte Carlo computer program. Of course, this is unnecessary in certain situations, for example high-statistics searches where all uncertainties, statistical and experimental, can be described by analytic distribution functions; and relatively simple counting experiments in the absence of uncertainties where direct sums of Poisson probabilities can describe the results.

The major drawback of the Monte Carlo method is that it is computationally intensive, but modern computers are relatively inexpensive and powerful so this is not nearly as strong an objection as it was only a few years ago. In addition, since the likelihood ratio is the ratio of local probability densities, efficient computation of discovery significances (and extremely strong exclusion) is possible by generating weighted Monte Carlo experiments. In the tiny tail of the background distribution which one integrates to find the discovery significance via $(1 - CL_b)/(1 - CL_{s+b})$ ones generates signal+background experiments and weights them by the inverse likelihood ratio. This is highly efficient since it is in this region that the density of signal+background experiments is large. In the tiny tail of the signal plus background distribution which one integrates to find the exclusion confidence via CL_{s+b}/CL_b ones generates background experiments and weights them by the likelihood ratio. This is highly efficient since it is in this region that the density of background experiments is large. With only a few thousand Monte Carlo experiments discovery significances for the Higgs search at LEP in the 5σ region can be computed with a relative statistical precision of a few per cent (ignoring, of course, the significance reduction of the "look-elsewhere" effect).

One of the current challenges in the LEP working group for Higgs boson searches is to handle ever-increasing numbers of selected events which slow the Monte Carlo and other numerical integrations down considerably. Since the likelihood ratio distribution for N selected events is the convolution (N times) of the likelihood ratio distribution for 1 selected event, the likelihood ratio distributions can quickly be computed with a fast Fourier transform (FFT) [13]. This technique, developed and used by one the LEP experiments, promises to make revolutionary reductions in the computing power needed to obtain the likelihood ratio distributions. Recently it has also been shown in the working group that analytic approximations work quite well in this situation even if the large-statistics limit of normal distributions hasn't been reached yet.

11. HIGGS SEARCH AT LEP WITH $\sqrt{s} < 189$ GeV

In this section the preliminary results of the search in data taken at LEP with $\sqrt{s} < 189$ GeV for the neutral Higgs boson predicted by the Standard Model are described [14]. The results are obtained with the combination of the results of the four experiments. At that time the working group was using $1 - CL_b$ as the significance indicator and so I refrain from using $(1 - CL_b)/(1 - CL_{s+b})$ here.

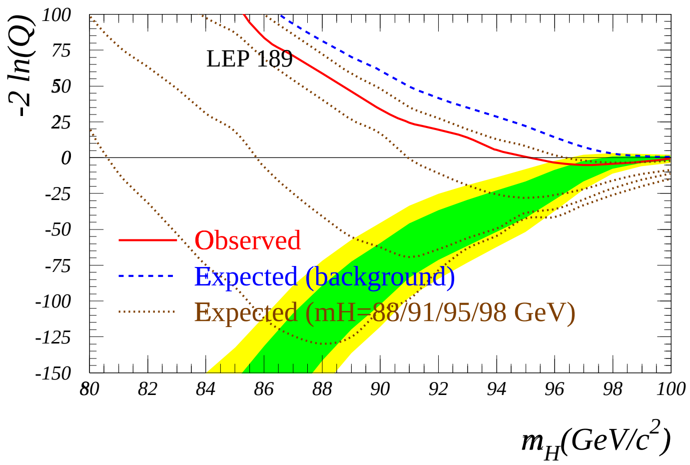

Fig. 10: The negative log-likelihood ratio versus m_H. The shaded bands show the 68 and 91% probability bands for the signal at the "true" mass. The expected signal curves (dotted) show the median response away from the "true" mass for four different Higgs masses.

The test-statistic $-2\log Q$ versus m_H computed for the observed results should have a minimum near the true Higgs mass and the more negative the value at the minimum the more significant the result. There is indeed a minimum with a negative value near 97 GeV in the data, shown in Fig. 10, indicating a slight preference for the signal hypothesis (results from data taken at higher values of $\sqrt{s}$ shown at the same conference showed that the slight preference for signal was in fact due to a fluctuation).

The significance of the result is given by $1 - CL_b$, which is plotted as a function of m_H in Fig. 11. Values of $1 - CL_b$ below 6×10^{-7}, corresponding to a five standard deviations fluctuation of the background, are considered to be in the discovery region. However, it is not enough just to read off the value of $1 - CL_b$ at the minimum of $-2\log Q$ since this only gives the probability that the background fluctuated at precisely *that* mass and in principle it could have fluctuated anywhere in the mass region not already strongly excluded by previous searches and up to the limit of sensitivity. A rough estimate based on Monte Carlo studies shows that $1 - CL_b$ must be multiplied by about a factor of four, corresponding roughly to the width of the mass search region divided by the typical mass resolution. This gives an effective $1 - CL_b$ of about 5%, in other words a significance corresponding to a bit less than two standard deviations.

Regardless of the interpretation of the result at 97 GeV, a 95% confidence level lower limit on the Higgs mass may be set by identifying the mass region where $CL_s < 0.05$, as shown in Fig. 12. The average limit expected in the absence of signal is 97.2 GeV and the limit observed by LEP is 95.2 GeV.

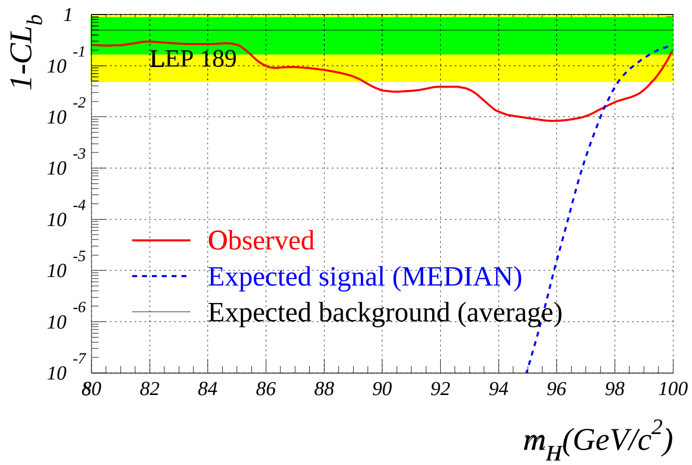

Fig. 11: The confidence level $1 - CL_b$ as a function of the Higgs mass. The straight horizontal line at 50% and the shaded bands represent the mean result and the symmetric 68 and 91% probability bands expected in the absence of a signal. The solid curve is the observed result and the dashed curve shows the median result expected for a signal when tested at the "true" mass.

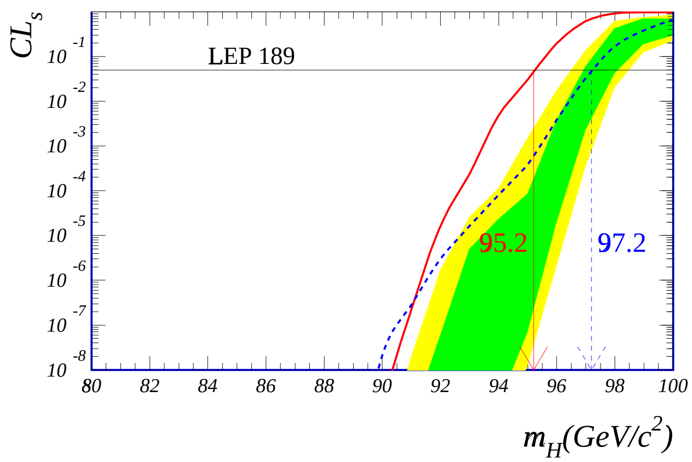

Fig. 12: The confidence level for the signal hypothesis CL_s versus Higgs boson mass. The solid curve is the observed result, the dashed curve the mean expected in the absence of a signal. The shaded areas represent the symmetric 68 and 91% probability bands of CL_s in the absence of a signal. The intersections of the curves with the horizontal line at $CL_s = 0.05$ give the mass limits at the 95% confidence level.

12. CONCLUSION

A modified frequentist analysis of search results used in searches for Higgs bosons at LEP, the so-called CL_s method, has been presented. It offers a general, practical (robust, if you like) solution to the problem of dealing with confidence limits for small signals in the presence of backgrounds. The definition of the confidence interval obtained is useful but somewhat untraditional. It neither adheres to the frequentist principle of coverage (it overcovers *by design* as the experimental sensitivity to the hypothetical signal vanishes) nor does it indicate the bounds of a Bayesian subjective probability distribution. Instead it indicates the boundary (or boundaries if it is reasonable to quote both) of a region where one would not have expected to observe equally or less signal-like results than the actual observation in case the signal hypothesis were true (at or below a specified rate). Let me try to make an important point about the previous sentence as clearly and simply as possible (even my friends claim I got it wrong all the three times I tried to explain this in my presentation): The lower bounds on the Higgs mass that are quoted for the direct Higgs searches at LEP say *absolutely nothing* about the *probability of the Higgs mass* being higher or lower than some value. To make such a statement the direct search results must be first folded with a prior probability distribution for the Higgs mass [15].

The Higgs searches at LEP use the likelihood ratio with respect to the background-only hypothesis (which could be called more generally the insensitivity limit or bound) or closely related test-statistics to order the results of their searches. Simple application of the Neyman-Pearson theorem shows that this is the optimal way of distinguishing between the signal/no-signal hypotheses - which is the first objective of a search.

The CL_s method, together with the use of the likelihood ratio with respect to the insensitivity limit is general enough to be applicable to different types of searches (counting experiments, parameter measurements, multichannel searches with measurements of multidimensional discriminants such as the Higgs searches at LEP). There exists a complement of CL_s for discovery significance which strongly reduces the chances of making a discovery with an experiment which is in fact insensitive to the signal in question, at the cost of a small reduction in the discovery potential for truly sensitive experiments.

Experimental uncertainties of all types can be accounted for by “smearing” the gedanken experiments of the confidence computations. The experience of the Higgs searches at LEP is that except in extreme situations their inclusion does not lead to unintuitive results.

The issue of flip-flopping (deciding whether to quote one or two-sided confidence intervals based on the data) is mostly avoided by the CL_s method. For example, the procedure used by the Higgs groups to find the lower bound of the Higgs mass is independent of how compatible the data are with either the background or signal+background hypotheses. Two-sided intervals are not very meaningful when there is no significant evidence of a signal, but as the significance increases, the interval defined by CL_s will approach those of traditional measurement techniques (even if the interpretations differ). This and the use of the likelihood ratio as the test-statistic give a clear point of contact with those techniques. Thus there can be a rather smooth transition from exclusion, to observation, to discovery and finally to measurement (assuming of course that the signal is there somewhere and that we are clever enough to build an experiment to find it, otherwise the story ends with exclusion).

Acknowledgements

On behalf of the LEP experiments and the LEP working group for Higgs boson searches, I would like to thank F. James and L. Lyons for their initiative to organize this workshop and for their invitation to present our work. The local organizers, F. James and Y. Perrin, have done a superb job. This work has been supported in part by the University of Oslo and the Norwegian Research Council.

References

[1] LEP Working Group for Higgs Boson Searches (P. Bock et al.), CERN preprint CERN-EP/98-046 (1998).

[2] G. Feldman and R. Cousins, Phys Rev **D57** (1998) 3873-3889.

[3] G. Zech, Nucl. Instr. and Meth. **A277** (1988) 608.

[4] S. Jin, “The Signal Estimator Limit Setting Method”, these proceedings.

[5] S. L. Meyer, *Data Analysis for Scientists and Engineers*, John Wiley and Sons, 1975, ISBN 0-471-59995-6.

[6] Suggested recently by W. Murray, LEP working group for Higgs boson searches.

[7] J.-F. Grivaz and F. Le Diberder, Nucl. Instr. and Meth. **A333** (1993) 320.

[8] P. Janot and F. Le Diberder, Nucl. Instr. and Meth. **A411** (1998) 449.

[9] P. Bock, Heidelberg University preprint HD-PY-96/05 (1996).

[10] A.L. Read, DELPHI collaboration note 97-158 PHYS 737 (1997).

[11] A. Favara, M. Pieri, L3 internal note 2066 (1997).

[12] R.D. Cousins and V.L. Highland, Nucl. Instr. and Meth. **A320** (1992) 331.

[13] H. Hu, J. Nielsen, "Analytic Confidence Level Calculations using the Likelihood Ratio and Fourier Transform", these proceedings.

[14] A. Read, proceedings of International Europhysics Conference on High Energy Physics, July 14-21, 1999, Tampere, Finland.

[15] R. Cousins, "Difficulties with Frequentist and Bayesian approaches to limit setting", these proceedings. G. D'Agostini, "Confidence limits: what is the problem? Is there the solution?", these proceedings.

Discussion after talk of Alex Read. Chairman: Roger Barlow.

Masahiro Kuze

You mentioned about the experimental uncertainty. Sometimes it's not easy to assume Gaussian errors with well-known sigma. Some quantities are best described by a flat distribution with strict bounds. Are there any discussions in the Higgs group ... ?

A. Read

In Delphi, one year we had one channel where there was a convolution of a Gaussian with the uniform distribution over some region. Since we do it by Monte Carlo, such things are easy to put in.

M. Kuze

As you said, it makes a small change in the limits, but when there is really a positive signal and you want to get the significance, this can change by an order of magnitude.

A. Read

I haven't given you any examples. You can imagine that errors in the background (if I show this on a log scale you might see it better) can easily give fluctuations out here to the left in the region where the distribution is very sparse, and in fact those fluctuations will be much bigger than the fluctuations that you get at the signal under the background when you're doing exclusion. So what we've seen in the Higgs group is just what you said earlier: Unless your errors are really big, they don't make such a big impact on exclusion results, but they can easily take one or two sigma off your discovery significance.

M. Kuze

Yes, in this case this problem can be rather subjective. It depends on each physicist and can be a difficult problem.

Glen Cowan

I want to understand better this business of dividing CL_{s+b} by CL_b. I understand that if you do that it makes your interval more conservative. So my first question is can you quantify by how much it becomes more conservative, can you state that if you want to give a 95% confidence level upper limit what is your coverage probability, say as a function of the hypothesized Higgs mass. Could you quantify that in the following way? Suppose you were just to use the number of candidate events as the basis of your test statistic. You don't measure any invariant masses or anything like that, but you just use the number of candidate events, and if you were to make the plot which has appeared on various transparencies today, of the limit as a function of the expected background, and you get a family of curves for a number of candidates, how would that family of curves look like in your method ? You can ignore the second part of the question if that wasn't clear, but I'll get back to it. The main thing was that I do not understand the theoretical justification for dividing by CL_b.

A. Read

It's related to conditional probability, and the idea is that you make an observation, and according to CL_b the compatibility of the background confidence, the result is unlikely. But it's even more unlikely that it can be accounted for by signal plus background and somehow the ratio of these two is telling you something about this probability. As I say, it's an approximation, it's not stringent.

M. Woodroofe

If I could elaborate a little bit on what was just said. If there are only counts, then that's absolutely right that you have the conditional probability given background less than or equal to N. I tried to justify that in my talk; I don't know if I succeeded. If there are X's present it becomes a little more complicated, and I can't see my way through the calculations but it's not at all clear to me that it's still conditional probability in the case that there are X's present.

Bob Cousins

Virgil Highland's criticism of Günter Zech's original paper was that the conditional probability in the denominator is actually not the probability conditioned on what you measure. I think that this speaker made it clear that he was conditioning on a number known in the Monte Carlo. He was not conditioning on a number that the experimenter can know, so Virgil's criticism was to say that the conditional probability should be calculated using Bayes Theorem. So the first thing I did when I got your paper was to check that you calculated the conditional probability the same way Virgil did, which you did, and Günter Zech's reply (if I can speak for him) was basically to say we're going to condition on this number which is in our Monte Carlo. Let's call it a convention but it gives reasonable results I think. So there is this technical detail of the conditioning on something that's only in the Monte Carlo.

A. Read

I stress again it's only an approximation.

Günter Zech

Sorry but I must make a remark. In this paper I never claimed that there is coverage, so this was just a frequentist interpretation of a Bayesian formula. I think this interpretation is correct as it was. It does not fulfill coverage, and it has the property that it is equal to the standard frequentist method as long as you have no background expectation, and it fulfills the likelihood principle in other cases, and this defines it fully.