Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nuclear Instruments and Methods in Physics Research A 534 (2004) 15–23
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

PHYSTAT2003

Louis Lyons[a,b,*]

[a]Department of Physics, University of Oxford, UK
[b]KEK, High Energy Research Organisation, Tsukuba, Ibaraki, Japan

Abstract

PHYSTAT2003 was a Conference about statistical techniques as used in particle physics, astrophysics and cosmology. A flavour is given of the presentations at the Conference, and of the issues that were discussed there.

1. History

PHYSTAT2003 was a Conference with the theme of "Statistical Problems in Particle Physics, Astrophysics and Cosmology". It was held at the SLAC Laboratory in September 2003, and attracted 120 participants. It followed on from three previous meetings at CERN, Fermilab and Durham, but differed from them in two important respects. While the first three meetings had been attended almost entirely by particle physicists, the SLAC Conference was designed to appeal also to astrophysicists and cosmologists. Secondly, there was a very strong presence of statisticians, who had a significant impact on the meeting. They contributed to the planning of the Conference, gave invited and contributed talks, and also interacted with physicists through animated discussions both during and between the Conference sessions.

*Department of Physics, University of Oxford, UK.
E-mail address: llyons@oxpc07.fnal.gov (L. Lyons).

0168-9002/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.nima.2004.07.051

2. Structure of Conference

Because the participants came from different backgrounds, there were two introductory talks by Eric Feigelson and Roger Barlow, on astrophysics and on particle physics, respectively. These were to inform non-experts in these fields of the range of statistical issues of interest for those subjects. These were followed by the keynote address on "Bayesians, Frequentists and Physicists" delivered by Bradley Efron, the President of the American Statistical Association.

The invited talks were more or less equally divided between statisticians and physicists. Two of the invited talks were devoted to the statistical issues confronting the large SLAC experiments-BaBar [2] in particle physics and GLAST [3] in astrophysics. The number of contributed talks necessitated two parallel sessions.

The Conference ended with a Panel Discussion on questions submitted by conference attendees, and with John Rice's concluding address.

Table 1
Contrasting views concerning $\mu_L < \mu < \mu_U$

	What it is	Frequentist view	Bayesian view
μ	Physical param, fixed value	Unknown constant	Treated as random variable
μ_L, μ_U	Determined by expt	Random when expt repeated	Fixed constants
$\mu_L < \mu < \mu_U$	Result of analysis	How often range contains μ_{true}	Degree of belief that μ is in fixed range

Further details, as well as copies of the transparencies and write-ups of most of the talks, can be found at the Conference web-site and in the Proceedings [1].

3. Bayes versus frequentism

An almost constant theme running through the Conference was the Bayesian versus frequentist discussion. The statisticians at the Conference covered a wide range of statistical philosophies, but as even the Bayesian Persi Diaconis put it, "When you service your car, you don't have to decide whether to use just a spanner or a screwdriver".

A large part of statistics is devoted to analysing data with the aim of extracting the value of one or more parameters. Given the strong difference in approaches, one might well wonder how it is possible to spend a lifetime analysing data, while remaining blissfully unaware of the controversy. The answer is that, for well-measured quantities with small errors and where there is no restriction on the possible values of the parameter μ, the Bayesian and the frequentist approaches can give the same answer e.g. $\mu_L < \mu < \mu_U$ at 68% probability, where the μ_L values are identical for the two methods, and similarly for μ_U. The result thus does not depend on the method. However, the interpretations of the equation are very different (see Table 1).

The Bayesian approach starts from Bayes' theorem

$$P(A|B) = P(B|A) * \frac{P(A)}{P(B)} \tag{1}$$

where $P(A|B)$ is the probability for A, given the fact that B has happened, while $P(A)$ is the unconditional probability for A. This formula is also accepted by frequentists, for the case where the probabilities are acceptable probabilities in the frequentist sense.[1]

The controversy begins when Bayesians replace A by 'parameter value μ and B by 'data x'. They then write

$$P(\mu|x) \sim P(x|\mu) * P(\mu) \tag{2}$$

and where the P's may now be probability densities. The three factors in the above equation are known as the posterior probability density distribution, the likelihood function and the Bayesian prior, respectively. Bayesians see this as a way of using the result of the experiment, embodied in the likelihood, to update one's prior knowledge about the parameter to give the posterior density. From this, parameter ranges or upper limits can be extracted.

Frequentists object to this on two grounds. First, probabilities are being used in contexts that they consider inappropriate. Secondly one needs to specify a prior. This may be relatively straightforward if there is already a well-measured estimate of the parameter μ, in which case a Gaussian prior centred on the measured value might be appropriate. But in other situations, very little may be known about μ. One would then perhaps like an 'uninformative prior', for which a constant might seem to be a suitable choice, since it does not

[1] Frequentists define probability in terms of the results of a repeated large number of essentially identical trials. Thus frequentist probability simply is not meaningful for single events (e.g. Was it raining in Oxford yesterday? Will the first astronaut to set foot on Mars return alive to Earth?), or for whether a particular physical constant lies in a specified range. In contrast, Bayesian probability is a measure of degree of belief. It can vary from person to person, depending on their knowledge, but is quantified by the concept of a "fair bet".

favour one value as compared with any other. However, it is not clear whether one is uniformly uncertain about μ, μ^2, $\ln\mu$, etc. and these are all different. A criticism of this then, is that priors are supposed to parametrise prior knowledge, not prior ignorance.

Frequentists seek a method which avoids what they consider to be subjective choices based on personal beliefs, and which does not require a prior, or indeed any statement about probabilities of parameter values. They instead use the Neyman construction, which considers only $P(x|\mu)$ i.e. the probability for different experimental results, once the parameter value is specified. For each value of the parameter μ, a likely range of data values x is defined (say, at the 68% or 90% level). This is repeated for every possible physical value of the parameter. This results in a band in the (μ, x) plane of likely measured values for any μ. Then the actual value x_0 of the measurement is used to read off the range of μ values for which x_0 was a likely result (see Fig. 1).

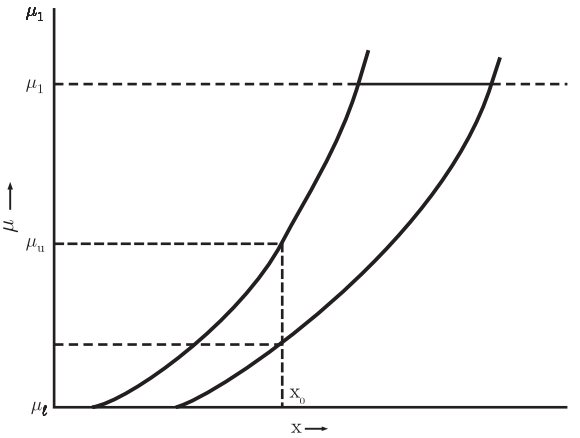

Fig. 1. Construction of the Neyman confidence band. μ is the parameter to be determined, and x is a measurement that depends on μ. For any μ (such as μ_1 in the diagram), we find an interval in x such that the probability that x lies in this range is 90%. This is repeated for all values of μ, to build up the region between the solid curves. Then a given measurement x_0 defines a range μ_l to μ_u such that for an value of μ in this range, x_0 was a likely measurement. For a parameter which cannot be negative (e.g. the temperature at the centre of the sun), the diagram does not exist for μ below zero.

A criticism of the method is that it is possible to obtain a result x for which there are *no* values of the parameter for which this result is likely. The statistician who recommended this method would not be particularly perturbed by this, since the confidence region is only at the 90% level, and this particular result is presumably one of the expected 10% for which the (empty) range does not contain the true value of the parameter. It is nevertheless disconcerting for the experimenter who has spent years of his life designing, building, testing and running the experiment, as well as analysing the data, to end up with a range which is known to be wrong.

Another criticism of the method is that the choice of the 90% (or whatever level) range in x for a given μ is not unique. There are many ways in which this could be specified, and each can give a different confidence interval for μ.[2]

This ambiguity in the choice of range has been exploited, for example, in the approach of Feldman and Cousins [4]. They define an 'ordering rule' for different x values, which is based on likelihood ratios. The range is then uniquely defined, and has the effect of making empty or zero-length intervals for the parameter much less likely. It also provides a unified way of obtaining either upper limits, or two sided-intervals; which it turns out to be is decided by the data, and not by the physicist. Another positive feature is that the ordering rule extends naturally to data in two or more dimensions. These situations can be problematic for other procedures. The Feldman–Cousins approach is gaining popularity in a range of usages in particle physics, especially in situations involving constraints on the physical parameter (e.g. the mass squared of a particle cannot be negative, the sine of an angle must be between -1 and $+1$, etc.).

Especially in situations where the experimental measurement results only in a limit on the physical parameter, the result can be very sensitive to the method used to extract it. It is thus crucial to explain carefully in any report of the experiment exactly how the result was obtained. Otherwise, one experiment may appear to be more sensitive

[2]This ambiguity is in some way mirrored by the choices available in the Bayesian approach for extracting a range in μ from its posterior probability distribution.

Table 2
Comparison of Bayes and frequentist philosophies

Method	Bayes	Frequentist
Basis of method	Bayes' theorem	Neyman construction
Probability	Degree of belief	Limit of frequency ratios
Prob of params?	Yes	Unwilling to consider
Need for prior?	Yes	No
What is used?	$\mathscr{L}(\mu\|x)$	Prob$(x\|\mu)$
Which data?	Just the data you have	All possible data in ensemble
Need for ensemble?	No	Yes
Obeys likelihood principle?	Yes	No
Choice for defining interval	Yes	Yes (except Feldman–Cousins)
Final statement	Posterior prob dist	Values of param for which data is likely
$\mu_L<\mu<\mu_U$	See Table 1	See Table 1
Treatment of nuisance params	Marginalise (easy)	Enlarge dimensionality (hard)
Unphysical values of params	Excluded by prior	May give empty range
Concept of coverage	Not seen as relevant	Incorporated in Neyman construction
Is coverage achieved?	When averaged over parameter	For all values of parameter

than another simply because of the different analysis techniques used.

A summary of the differences between the Bayesian and frequentist approaches is given in Table 2. In an invited talk at the Conference, Fred James [5] explained how teaching the two techniques helps illustrate their complementary advantages.

4. Multivariate analysis

Another major theme at the Conference was the use of multivariate techniques for separating wanted signal from unwanted noise (confusingly called 'background' by particle physicists, but 'foreground' by astrophysicists looking at the signal from the cosmic microwave 'background'). Techniques available range from cuts and Fisher discriminants to neural networks and support vector machines. Many of the contributed talks either addressed this issue directly, or simply made use of it as a necessary part of their analysis.

The invited talks on this topic were given by Harrison Prosper [6], and by Jerry Friedman [7], who emphasised the advantages of using decision trees, especially in their boosted form. In his welcome address at the opening of the Conference, SLAC Director Jonathan Dorfan made reference to work he had done with Jerry Friedman when, years earlier when both were at the Lawrence Radiation Laboratory in Berkeley, they had used decision trees. They do not seem to be used very much in particle physics at present, and certainly seem to be worthy of being looked at more seriously.

5. Blind analyses

In complicated analyses, there is the danger that we might (perhaps inadvertently) make choices in the analysis of the data, in order to influence the result in a particular direction. This bias could be introduced via the selection of cuts, decisions whether to include or exclude particular data sets, correction procedures, choices between different analyses, etc. In order to avoid this possibility, 'blind analyses' can be employed, as was discussed in a talk by Aron Roodman [8]. The idea is that the final result is kept unknown from the physicist until the analysis has been completed. Techniques include:

- Cuts and corrections are chosen by looking at real data, except that the signal region is excluded. For the final analysis, the 'signal box' is finally opened.

- Cuts and corrections are chosen by looking at Monte Carlo simulations, rather than using the data.
- Cuts and corrections are chosen by looking at part of the data. The analysis details are then frozen, for the actual analysis on the remainder of the data.
- Parameter fitting is performed by comparing the data with Monte Carlo simulation, which is generated for parameter values which are kept hidden. The Monte Carlo is then weighted, to correspond to different parameter values, and these weighted Monte Carlo distributions are compared with the data. At the time of the fit, the physicist knows only the *change* in parameter value required. This is converted to an absolute value only at the end of the analysis.

An advantage of the last of the above techniques is that the actual data to be used for evaluating the final result can be examined at all stages of the blind procedure. For other methods, there is a potential problem that when the data are finally looked at, some really anomalous feature might be apparent e.g. a large number of events is observed every time the super-express train pulls out of the local railway station. The generally agreed advice is that, even for supposedly blind analyses, further procedures may be applied after looking at the data, in situations where everyone would agree that it would be stupid not to do so.

A related approach in searches for new effects is to use a general program like QUAERO or SLEUTH [9], which operate in a fairly user-independent mode. Otherwise it is not obvious how to conduct such searches for interesting new physics in a blind manner, and then this complicates the assessment of the significance of any possible exciting discovery.

6. Significance

In general, discovery of a new phenomenon in particle physics is set at the 5σ level. This is equivalent to a probability of a fluctuation of the background of a few 10^{-5} times. Why is this set at such a low level, rather than at around, say, 1%?

One answer is the 'look elsewhere' effect. The probability is usually calculated as that for a fluctuation in a specified region of phase space. (e.g. the bin of a particular histogram, produced by a pre-defined set of cuts.) But there are typically several bins in a histogram, and many possible histograms in an experiment, each of which can be produced by variable selection cuts. Then the chance of a fluctuation somewhere is very much larger than that in a single specified location.

Another reason is that, given our data, we are interested in the (Bayesian) relative probabilities of the Standard Model and some alternative involving New Physics. These interesting probabilities are extracted from the more readily obtainable (frequentist) probability of observing our data, given either of these model, by Bayes' theorem. But we also require prior probabilities for the standard model or for the new physics being true. We might believe that the latter was low. In a not-too-different scenario, we could imagine an experiment to test special relativity. If we found that our data was such that Prob (data|S.R.) was 10^{-3}, this does not imply that the probability that Special Relativity is true is only 10^{-3}, and that we should immediately publish an article that Einstein was wrong. Thus the 5σ criterion is chosen in part to compensate for the intrinsic unlikeliness of any specific new effect, especially when compared with a well-established model.

For a histogram bin with an estimated background B and an excess of events S, the best approach for evaluating the significance is to calculate the p-value. This is the probability that Poissonian fluctuations of the background could have given an effect at least as large as that observed. There has been much discussion as to whether $S/\sqrt{B}$, $S/\sqrt{B+S}$, $2(\sqrt{B+S}-\sqrt{B})$, etc. are suitable for approximating the effective number of standard deviations of the observed effect. For large numbers of events, and $B \gg S$, these become similar, but this is not so in other circumstances. It seems as if $S/\sqrt{B}$ should be most relevant for how often the background can fluctuate up to the observed number of events. This would be so if the fluctuations were Gaussian, but for small numbers of events, the difference

between the Poisson and the Gaussian distributions can make $S/\sqrt{B}$ misleading.

Problems arise in incorporating systematics (e.g. that arising from an uncertainty in B) in the procedure. Punzi [10] has also pointed out that it can be misleading to use any of the above approximations to optimise the design of a future experiment. This is because they could result in an 'optimisation' with, for example, an expected background and signal of 0.0001 and 0.5 events, respectively. For these values, there is, in fact a distinct possibility that the experiment will observe no events at all. Punzi's criterion involves having the best chance of actually making a discovery (or exclusion).

An alternative to simply counting events in a signal region is to compare fits to the data with background only, and with background plus signal. This does not even require the data to be binned. The difficulty comes in interpreting the log likelihood ratio λ, because -2λ will not behave like a χ^2 variable in non-asymptotic situations. This requires the significance attributed to λ to be calculated by Monte Carlo. In the lucky case of a very significant signal, it may be difficult to generate enough Monte Carlo to determine this value.

7. Systematics

An ongoing issue is how to incorporate systematic effects in parameter estimation, especially in situations involving upper limits. A typical situation is where the rate or cross-section r is determined from an observed number of events by

$$n = Ar + b \tag{3}$$

where b is the expected background and A is a factor which allows for the beam intensities, experimental acceptance and inefficiencies, etc. Both A and b will usually have uncertainties, from being estimated from other measurements, limited accuracy of Monte Carlo simulations, theoretical uncertainties, etc. They are known as nuisance parameters. Several methods exist for including the effects of these uncertainties:

- *Move the systematic parameters*. This involves recalculating the best estimate r as A and then b are each individually changed from their best estimate by their own error; assuming the errors on A and b are uncorrelated, the shifts in r are combined in quadrature. This may be appropriate for Gaussian situations, but is not used for upper limits.
- *Bayesian approach*. The systematics are assigned Bayesian priors (typically Gaussians). Then the posterior probability for the physics parameter r simply requires the nuisance parameters to be integrated out of their joint probability density. Problems are encountered, in that the posterior diverges for the combination of a uniform prior for r and a Gaussian for A (even if it is truncated to remove negative values).
- *Frequentist approach*. The dimensionality of the problem is increased to include nuisance parameters (as well as the physics parameter), and their respective measurements. The usual Neyman construction is used to determine a confidence region in (r, A, b)-space, which is then projected onto the r-axis to determine the accepted values of r. This approach consumes much computer time, and a maximum of three parameters have been used to date. The projection phase of the procedure usually results in overcoverage. Another feature, which is after regarded as paradoxical, is that the upper limit on r for a small uncertainty in A can be below than when A is known exactly. This is related to the fact that the observable n is discrete.
- *Approximate frequentist approaches*. Because the exact frequentist approach is in many cases not practical with more than a small number of parameters, approximate methods have been investigated. Nancy Reid [11] spoke about a modified likelihood method, while at the earlier Fermilab Workshop both Feldman [12] and Rolke [13] described the profile method, which uses the best value of the nuisance parameters at each value of the physics parameter(s).
- *Mixed Bayes–frequentist*. Again because of the difficulties with the pure frequentist method, Cousins and Highland [14] discuss a method of incorporating the nuisance parameters in a Bayesian way, in what is otherwise a frequentist

method for extracting the physics parameter of interest. This results in debates about the suitability of a mixed Bayes and frequentist approach in a single analysis. Cousins claims that, although the method does not guarantee this, in practical cases it has frequentist coverage.

Of the two sources of systematics mentioned above, b is the less problematic in that it occurs in an additive fashion.

A related problem of incorporating systematic effects in calculations of significance was dealt with by Demortier [15]. Cranmer's contribution [16] was about another related issue—allowing for systematic effects in hypothesis testing. This will be very important in experiments looking for the Higgs boson at hadron colliders, where there are large uncertainties in the expected backgrounds.

This problem is certainly one of the big statistical issues in particle physics analyses, it is very common, and there is currently no obviously preferred method. It was reviewed at the Conference by Pekka Sinervo [17].

8. Goodness of fit

The well-loved method of determining goodness of fit of our data compared with some hypothesis is the χ^2 approach. This requires data to be put into binned histograms, and there to be enough expected events in each bin for the actual Poisson distribution to approximate to a Gaussian. While this may be all right for plentiful data in one dimension (i.e. one physical variable), it may well become problematic with sparse data in several dimensions. For d-dimensions with 10 bins in each variable and a minimum of 10 events per bin, at least 10^{d+1} events are required. In these sparse situations, parameters are often optimised using an unbinned maximum likelihood approach. The question then arises as to whether $\mathscr{L}_{\max}$, the maximum of the likelihood, provides a measure of the agreement between the data and the expected functional form. After all, the parameters are determined by maximising the likelihood, so surely the larger its value, the better the fit? It turns out, however, that even if this is obvious, it is not necessarily true.

An instructive example [18] is provided by a fit of some times t_i to an exponential

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \mu\, \mathrm{e}^{-\mu t}, \quad t \geqslant 0. \tag{4}$$

Very little algebra is required to show that $\mathscr{L}_{\max}$ is a function only of $\bar{t}$, the average of the observed times. That means that *any* distribution of non-negative times with the same $\bar{t}$ is guaranteed to give the identical value of $\mathscr{L}_{\max}$. The value of $\mathscr{L}_{\max}$ is thus completely incapable of distinguishing an exponential distribution of times from a delta function at the time $\bar{t}$.

Another simple example is provided by the angular distribution

$$\frac{\mathrm{d}n}{\mathrm{d}\cos\theta} = (1 + a * \cos^2\theta)/N \tag{5}$$

where N is a normalisation constant $2 + 2a/3$. This angular distribution is symmetric about $\cos\theta = 0$. However, the likelihood function involves only $\cos^2\theta$, and hence is completely insensitive to the sign of the $\cos\theta$ values in the data. Thus, a sample with all events at negative $\cos\theta$ would give the same value for $\mathscr{L}_{\max}$ as one in which the events were more plausibly distributed between the hemispheres.

This is an example of a more general principle, that events can be moved around in observable space between regions of equal probability density, with the likelihood being unaffected, even though the distribution becomes very unlikely.

The conclusion is that $\mathscr{L}_{\max}$ is at best a very weak way of examining goodness of fit, and that it is strongly recommended to use other techniques. One way of overcoming this difficulty for a binned distribution is to use a likelihood *ratio*; this is applicable when the number of events per bin is below the level where we can comfortably use χ^2. It involves calculating the product of the Poisson probabilities for observing n_i events in a bin, when the expected number is μ_i. This is then compared that with the largest value this probability could have if μ_i were allowed to vary in each bin i.e. $\mu_i = n_i$. Then the ratio λ is such that $-2\ln\lambda$

behaves like χ^2, at least asymptotically (see, for example, Ref. [19]).

Even lower statistics require an unbinned method. Kolmogorov–Smirnov is possible, but this does not extend easily to more that one observation. That is because the observations have to be ordered according to their numerical values. There is no obvious way to do this for more than one dimension. As always with multivariate distributions, it is possible to look at one-dimensional projections, but this ignores any possible correlations.

Several talks at the Conference were devoted to this topic. Jerry Friedman [20] spoke about it at the Panel Discussion, while Pia and Ribon [21] described the general facility which the Genoa group is making available, which incorporates a whole variety of methods.

Zech [22] described the performance of various goodness of fit techniques. An interesting one which performs well is the 'energy test'. It compares two distributions (e.g. data and Monte Carlo) by calculating the electrostatic energy of the combined data set, assuming positive charge is assigned to each data point, and negative to the Monte Carlo. Rather than using $1/r$ for the energy, Zech chooses $1/(r+\varepsilon)$ or $-\ln(r+\varepsilon)$, where ε is a cut-off parameter to avoid problems when two points coincide. The performance of the method is insensitive to the exact choice of value for ε. Another issue is the choice of metric. For similar distributions, the energy should be very negative, while for very separate distributions it would tend to zero. The expected distribution for the energy for samples from the same parent distribution is found by Monte Carlo.

9. Where are we now?

As a result of these meetings, there has hopefully been an increased understanding of statistical techniques. This includes the realisation that:

- *Prob* (*data*|*theory*) is not the same as *Prob* (*theory*|*data*).
- Integrating the likelihood with respect to parameters does not have any theoretical basis.
- It is important to explain in detail how your upper limits are determined.
- χ^2 is not the only technique for measuring goodness of fit.
- Neural networks and support vector machines are not the last word in multivariate separation techniques.

The talks and presence of statisticians greatly enhanced the appreciation of physicists for some of the more recent statistical techniques on the market. Several of the statisticians also very kindly offered their services in getting involved with some of the more interesting analyses.

We hope that this form of collaboration will strengthen, and that the links that were started at PHYSTAT2003 in SLAC will be even more productive by the time of PHYSTAT05.

References

[1] L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[2] F. Porter, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 186, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[3] S. Digel, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 176, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[4] G. Feldman, R. Cousins, Phys. Rev. 57 (1998) 3873.

[5] F. James, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 272, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[6] H. Prosper, in: M. Whalley, L. Lyons (Eds.), Proceedings of Institute of Particle Physics Phenomenology's Conference on Advanced Statistical Techniques in Particle Physics, Durham, 2002.

[7] J. Friedman, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHY-

STAT2003, SLAC-R-703, eConf C030908, 2003, p. 196, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[8] A. Roodman, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 116, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[9] B. Knuteson, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 8; and these Proceedings.

[10] G. Punzi, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 79, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[11] N. Reid, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 265, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[12] G. Feldman, Multiple measurements and parameters in the unified approach, Workshop on Confidence Limits, Fermilab, 2000.

[13] W. Rolke, Confidence intervals and upper bounds for small signals in the presence of noise, Workshop on Confidence Limits, Fermilab, 2000.

[14] R. Cousins, V. Highland, Nucl. Instr. and Meth. A 320 (1992) 331.

[15] L. Demortier, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 256, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[16] K. Cranmer, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 261, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[17] P. Sinervo, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 122, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[18] J. Heinrich, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 52, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[19] S. Baker, R. Cousins, Nucl. Instr. and Meth. 221 (1984) 437.

[20] J. Friedman, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 311, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[21] M.G. Pia, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 106.
A. Ribon, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 110, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.

[22] G. Zech, B. Aslan, in: L. Lyons, R. Mount, R. Reitmeyer (Eds.), Proceedings of Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology: PHYSTAT2003, SLAC-R-703, eConf C030908, 2003, p. 97, Web-site is <http://www-conf.slac.stanford.edu/phystat2003/>.