

QUESTIONS FOR PANEL DISCUSSION

The following questions were submitted by PHYSTAT2005 participants for the Panel Discussion session.

1. **GENERAL** : Do the statisticians on the Panel consider there are any techniques we should be using, but currently we are not?
2. **PARAMETER INTERVALS** : What are the properties that confidence intervals should respect? e.g. Frequentist coverage? Short but not too short? No empty intervals? Invariance with respect to reparametrisation? Robustness? Consistent approach to incorporating nuisance parameters?
3. **FREQUENTIST PARAMETER DETERMINATION** : Particle Physicists like to use frequentist approaches. Is it practically possible to use a Neyman construction in several dimensions i.e. of data and/or parameters? And if only one of the parameters is a physics parameter and the rest are nuisance parameters, so that it is necessary to project the multi-parameter confidence region in order to obtain that for the single physics parameter, what ordering rule will give optimal behaviour for the resulting one-dimensional intervals?
4. **P-VALUES WITH NUISANCE PARAMETERS** : An experiment is looking for some rare or perhaps non-existent process. It involves simply counting events. There is an uninteresting background b which also contributes to the counting rate. The number of observed events N is expected to be Poisson distributed with mean b , which has been measured in a subsidiary experiment as $b_0 \pm \sigma_b$. (This can be thought of as being determined as c/r , where c is the number of events in a situation which is sensitive only to the background, and r is a scale factor which typically could be 5, and is accurately specified with zero uncertainty. Larger r results in a smaller error σ_b). We want to calculate the p-value for the null hypothesis (only background), for observing at least N events. To be specific, we could take $N = 9$ and $b = 3.1 \pm 0.4$. What is the recommended statistical technique? It is desirable that it could easily be extended to a larger number of nuisance parameters.
5. **PROFILE LIKELIHOOD AND BAYES PRIORS** : If I understand correctly, eliminating a nuisance parameter by using the profile likelihood is equivalent to using a delta function prior for the nuisance parameter in the Bayesian philosophy. More refined likelihood methods can probably also be interpreted in the Bayesian way. If this is correct, why doesn't one use directly Bayesian methods?
6. **BAYESIAN TREATMENT OF SYSTEMATIC UNCERTAINTIES** : In a Bayesian approach to parameter estimation, it is straightforward to include nuisance parameters in a Monte Carlo Markov Chain, then marginalize over them in order to recover high probability regions for the parameters of interest which include the effect of our imperfect knowledge of the nuisance parameters. I am uncertain about the proper way to treat systematical errors in this context: an example could be the uncertainty associated with numerical inaccuracies of the code used, or the error induced by the fact that some second-order physical processes have been neglected in the code. This results in an uncertainty associated with the output of the code itself (which I'd classify as "systematical"), rather than a statistical uncertainty, associated with the data used. A common way to deal with this is to add the statistical and the estimated systematical errors in quadrature (or linearly if one wants to be conservative), then use this new artificial error on the data at hand. I would like to know whether there are more satisfactory ways of dealing with systematical errors of the kind described above, and in particular methods which recognise the different nature of the statistical and systematical errors.

- 7. MAXIMUM ENTROPY PRIORS :** Setting the prior correctly by making use of all available information is a central problem of Bayesian model selection, where the result is strongly dependent on the prior scale and does not disappear with better data (as it is the case for parameter estimation). I am getting interested in ways of setting the prior by maximum entropy arguments. I would like to hear the opinion of the Panel regarding this method, and in particular whether this way of determining priors is now well accepted in the Bayesian community. I would be interested in comments about the applicability and limitations of maximum entropy priors, if possible with examples illustrating situations where those kind of argument has been proven successful (or has failed for a clear reason).
- 8. PARAMETER DETERMINATION/HYPOTHESIS TESTING :** To what extent are hypothesis testing and parameter determination equivalent? Are there simple examples to illustrate when they are equivalent and when they are not?
- 9. MULTI-DIMENSIONAL CLASSIFICATION WITH VERY MANY VARIABLES :**
Analyzing HEP data, physicists more and more often deploy multivariate classification methods. We have seen a bunch of HEP publications where analysts separate signal and background by training a neural net on 10 or more input variables. Byron Roe and his associates in their recent work on PID at MiniBoone used 100 input variables for classification by boosted decision trees; this seems to set a record on the dimensionality used in HEP analysis. At the same time, there is a number of conservative physicists who refuse to adopt such multivariate methods with many input dimensions. They argue that it is very hard to assess how well Monte Carlo models data in so many dimensions, especially if one needs to take various systematic effects into account. Is there a generic prescription that relates the maximal reasonable dimensionality to the size of available Monte Carlo and data samples, in the context of a specific classification method? Can professional statisticians recommend good literature on variable selection, perhaps for two different problems: a) statistics- and b) systematics-dominated analysis? Should we attempt multidimensional analysis with dozens of input variables only when systematic effects do not matter much - for example, in rare signal searches - and stick to more robust and simple-minded techniques when systematics are important?
- 10. MULTI-DIMENSIONAL CLASSIFICATION WITH VERY MANY VARIABLES :** I have a data set where each entry is described by 18 variables. I want to reduce the dimensionality from 18 to, say, 4. Are there well-understood techniques for doing this, so that some measure of information loss is minimised?
- 11. NUMBER OF VARIABLES IN MULTIVARIATE PROBLEMS :** I assume we have 11 variables per event. The reason this is of interest is the observation that the D0 experiment obtained its most precise measurement of the top quark mass using the so-called matrix element method, a method that CDF is working on also. In the matrix element method one writes an explicit formula for the N-dimensional differential density, based on one's knowledge of the matrix element squared and the mapping from partons to observed objects. The premise is that, if one knew this N-dimensional density, one need look no further in terms of the search for new variables - one would just use the density directly. So the question is this: Given $p(x)$ where x is N-dimensional and given $q(y)$ where y is M-dimensional and $y = f(x)$ (and perhaps $M > N!$), can the use of $q(y)$ yield better signal/background discrimination than the use of $p(x)$? We spend a lot of time constructing $y = f(x)$, by hand! Is this necessary, if we have $p(x)$?
- 12. KOLMOGOROV-SMIRNOV :** Is there a good method of using the Kolmogorov-Smirnov goodness of fit test with multi-dimensional data?

13. BLIND ANALYSES : What do you do when you unblind your analysis and find something in there that wasn't predicted by either the background or signal estimates? This is usually the case where there's no data off source region to estimate the background and the background estimate is purely simulation based. You predict the background, open the box and find a big excess but then see that the excess lies in an event-observable region that doesn't look like either signal or background predictions. You guess it's an unsimulated background, but then what? I would have thought that maybe you could possibly cut it away and recalculate. But then I learned from Kath Rawlins that LIGO saw a similar thing in their analysis and she had showed that coverage got all screwed up if one removes events post-unblinding, even if one could tell for sure that the extra events were a background from some source you hadn't considered and that, had you known about this class in the first place, you would have designed cuts to never allow them into the final analysis.

14. PULLS : If we have a complicated parameter-estimation technique, we may want to use Monte Carlo simulation to check whether the procedure is behaving sensibly. One way to do this is to look at the distribution of

$$\text{pull} = (p_f - p_t)/\sigma_f^2$$

where p_f is the fitted parameter, p_t is its true value, and σ_f is the estimated error in p_f . The distribution is calculated for repeated simulations. Asymptotically and if all is well, we expect the distribution of pull to be Gaussian centered on zero with unit width. However there are simple non-asymptotic examples where this is not so. How do I know for small sample simulations whether deviations from standard Gaussianity is a cause for worry or not?

15. ASYMMETRIC ERRORS : Are there recognised methods for dealing with asymmetric errors? These often arise when estimating parameters in low statistics experiments? For example, we may measure a lifetime as $1.6_{-0.3}^{+0.6}$ picoseconds. Then we might be interested in comparing it with another measurement (e.g. taking the ratio of it with another lifetime); incorporating another contribution to the error, possibly also with asymmetric errors; or combining this result with another to obtain a weighted average.

16. ASYMPTOTICS : Can there be or has there been progress to improve asymptotic techniques' convergence in the tails? In particular the modified and adjusted profile likelihood techniques attempt to improve convergence of the first and second moments, but for a 5σ test we are more interested in describing the tails.

17. IMPROVED LIKELIHOOD TECHNIQUES : Under what circumstances do improved profile likelihood methods help? We have a range of N from 10-10,000 events, we have a range of likelihoods from Gaussian to highly-non Gaussian, and we are interested in a range significance levels from $2-5\sigma$. Under the circumstances that they do help, how much do we stand to gain?

18. JAMES-STEIN ESTIMATOR : There exist other shrinkage estimators that could be helpful for high-dimensional problems like Supersymmetry. The estimators are biased, which will alarm most physicists and provide an obstacle for the methods to be accepted; however, these estimators can significantly improve the mean-squared error. Do you have any words of wisdom regarding these estimators: when are they a good idea when are they a bad idea?