View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CERN Document Server

ATLAS and CMS Statistics Wish-List

Eilam Gross
Weizmann Institute, Rehovot, Israel

Abstract
A wish-list of statistics related issues, which were regarded by ATLAS and CMS as requiring a deeper understanding and perhaps the response of a professional statistician, is given.

1 Introduction

The first Phystat meeting was a workshop at CERN on Confidence Limits followed by a similar workshop at Fermilab. Fred James who organized the meeting with Louis Lyons presented then his personal wish list titled: "What I would like to see" (see Figure 1). Fred wishes that physicists learn the vocabulary of statistics, all searches use Feldman and Cousins unified method [1] to derive confidence intervals and Bayesian methods are used only in policy decisions. When accepting upon myself to collect a wish list from Atlas and CMS my only experience was with LEP statistics [2] and the so called CL_s method for deriving limits which was used at LEP. In the two months of preparation of this lecture I had a steep learning curve during which I partially fulfilled the first item in Fred's wish list but found myself in mild debate with his other two points. However when enquiring around I discovered that most physicists are mainly concerned with old fashioned systematics issues and the majority have only a vague idea of the meaning of the term "Nuisance parameters" and the meaning and difference between the Feldman and Cousins vs the profile likelihood methods etc. It became clear to me that my mission in this lecture is not only to communicate to the statisticians our unsolved difficulties but also to make sure that when Atlas and CMS publish a combined limit or discovery significance not only the few statisticians amongst the Physicists (which I propose to call "Phystatisticians") will understand but also the majority of the HEP experimentalists. Therefore I decided to expand the contents of my talk to include also a pedestrian guide to LHC statistics. This guide which also provides the separate title to these proceedings [3]also provides the reader with the statistics background required to understand the wish list. The statisticians and phystatisticians are exempted from reading it.

2 A Wish List of ATLAS and CMS

The wish-list is made of statistics related issues which were regarded by ATLAS and CMS as requiring a deeper understanding and perhaps a professional statistician response.

2.1 Modeling of the underlying process

The raison d'etre of our meeting here are 10^9 Protons that will collide with 10^9 protons per second in the 27 km long LHC tunnel. The Proton is made of partons which are Quarks and Gluons. The underlying process is a collision between two partons. To understand the process we need to know the Parton Distribution Function $f_i(x;Q^2,\alpha_s(Q^2))$ of parton i in a Proton. x is the fraction of momentum carried by the parton and Q^2 is the energy scale squared. To obtain the Parton Distribution Functions a global fit analysis is performed[4], mainly the CTEQ and MRST, [5]. However some strange phenomena occur with these fits. For example, for most of the individual experiments from which the data is taken the χ^2/dof is around 1. However, in the global fit, the CTEQ group set $\Delta\chi^2 \sim 100$ in order to get reasonable errors [6]. Stump [7] argues: "What we have are estimates on the uncertainties, not the true ones. The increase of χ^2 if the estimators are biased or wrong might be bigger than 1. We find that alternate pdfs that would be unacceptable differ in χ^2 by an amount of order 100". This is a very vague statement. Robert Thorne, the "S" in MRST concludes with a wish: "It would be nice to have a more

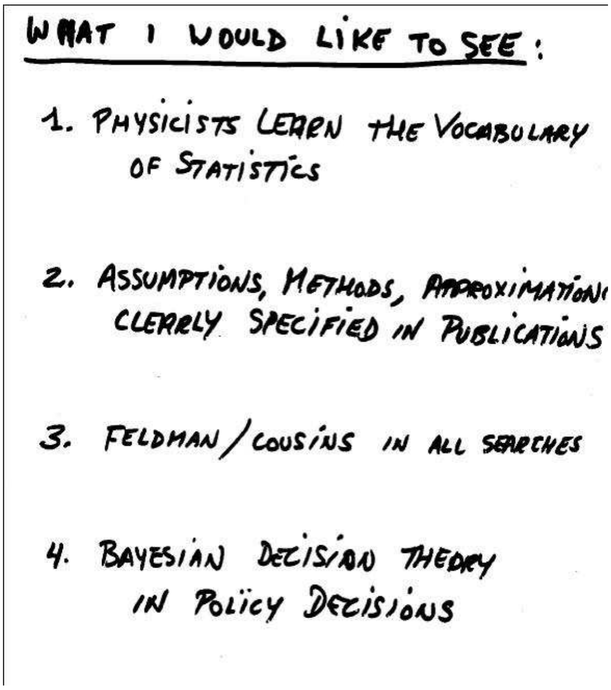

Fig. 1: Fred James personal wish-list presented in the first Phystat meeting, year 2000.

systematic way of accounting for this, e.g. a modified definition of goodness of fit to account for non-Gaussian nature of errors, a quantitative way of accounting for theoretical errors, etc..." [8]. Since Throne is elaborating the issue in this conference [6] I decided not to dwell anymore with it, though it is in the heart of proton-proton collisions and must be sorted out in order to reduce the systematics involved with all processes involved.

2.2 Why 5σ?

The null hypothesis is usually taken to be the background only hypothesis. The alternate hypothesis is the signal+background hypothesis. When analyzing results in HEP (High Energy Physics) it became a habit either to reject the signal hypothesis at the 95% Confidence Level or announce an observation at the 3σ level or a discovery at the 5σ level. Statistically speaking a 5σ discovery corresponds to a fluctuation of a Gaussian distributed background expectation at the level of $5.4 \cdot 10^{-7}$ (here we adopt the 2-sided interpretation). At that level one is probing tails of distributions. In order to make such statements one needs to understand the data and the detector response to that level. Nobody really knows where this habit of 5σ was born. A back of the envelope calculation reveals a possible explanation which has to do with the "look elsewhere effect". Suppose when searching for a new phenomena (Higgs boson...) one is combining 100 search channels each with a discriminating variable distributed within 100 resolution bins. The false discovery rate of 5σ will be $10^4 \cdot 2.7 \cdot 10^{-7} = 0.27\%$. This degrades the discovery sensitivity to 3σ. More examples and insights into this problem can be found in [9, 10, 11].

So is there a way to clarify how many σs are needed for discovery? Is there a problem in seeking for an effect at at tail of a pdf? How can we take a fluctuation into account?

2.3 Look Elsewhere Effect in Time

An ongoing discussion in the LHC collaborations is the need and possibility to perform a blind analysis. Even if from the scientific integrity point of view the pros are clear, with each collaboration having over 2000 physicists it is hard to believe such a habit can be adopted. Moreover, it will take years till the detectors are understood, and understanding the detector necessitates looking at the data as it comes. However, that raises another issue, the issue of sequential analysis. Should a statistical stopping rule be established? Is it possible that by adopting a stopping rule we would achieve a discovery with less luminosity than needed with a blind analysis? Understanding stopping rules is a complicated issue. It is probably more relevant in medical experiments where one would like to minimize the damage that might be referred to as patients trying new drugs. No doubt the HEP community must adopt very strict rules for looking at the data and publishing results to minimize human bias. The HEP Physicist might be bothered by another related issue, which might be refereed to as a "look elsewhere in time" effect: How much the p-value is increased as a result of the fact that we have already looked at the data a few times before and got no satisfactory significance?

2.4 Estimating Systematics

There are two issues related to systematics. Classifying and estimating them and implementing them in the analysis. Implementation of the systematics in the statistical hypothesis tests is discussed at length in [3]. Knowing the type of error one is dealing with is very important and make its estimation clearer. Sinervo [12] classified the sytematics into three types: Class I: Statistics-like uncertainties that are reduced with increasing statistics. Example: Calibration constants for a detector whose precision of (auxiliary) measurement is statistics limited. Class II: Systematic uncertainties that arise from one's limited knowledge of some data features and cannot be constrained by auxiliary measurements. One has to make some assumptions. Example: Background uncertainties due to fakes, isolation criteria in QCD events, shape uncertainties. These uncertainties do not normally scale down with increasing statistics. Class III: The "Bayesian" kind. The theoretically motivated ones, uncertainties in the model, Parton Distribution Functions, Hadronization Models. *The most accurate way to communicate the systematic error is to separate one type from another and quote them separately*. Some bad habits should cease. For example adding all sorts of systematics in quadrature and quote only the final result.

Another unfortunate habit in estimating systematics is when Physicists do not differentiate between cross-checks and identifying the sources of the systematic uncertainty. For example we shift cuts around and measure the effect on the observable. Very often the observed variation is dominated simply by the statistical uncertainty in the measurement [13].

2.5 Reference Priors in Demand of a Code

Analytical derivation of reference priors might be technically complicated. Bernardo [14] proposes an algorithm (pseudo code) to obtain a numerical approximation to the reference prior in the simple case of a one parameter model. The pseudo code should work for any number of parameters (of interest and nuisance) provided you make non informative priors for ALL! If the code could be extended to multiple parameters, including some with informative priors, it would be more useful for the HEP community. Another complication is that the order of the parameters matter. This should be further investigated and clarified. The wish is to have a generalized routine (REAL CODE) to numerically calculate reference priors for parameters $\{\theta\}$ given the Likelihood $L(\{\theta\})$ as an input.

2.6 Subsequent Inference

Often the background distribution is fitted with a polynomial, $\sum_{i=0}^{n} a_i x^i$ with the degree n determined with a stepwise test. However, the fitted coefficients a_i were obtained as if we know a priori the degree of the polynomial. How does one take the prior test into account? Perhaps the degree was wrong to start with?

Table 1: Hypothesis Test Methods. The columns indicate if the method obeys the Likelihood Principle (LP), if it has a coverage and if it uses priors.

		LP	Coverage	Priors	Comments
1	F&C	No	Yes	No	Pioneering in HEP
2	F&H&C^2	No	No	Yes	
3	Profile Likelihood (PL)	Yes	Asymptotic	No	`Best Value for Money`
4	PL F&C Construction	No	Satisfactory	No	
5	PL Full Construction	No	Yes	No	Cumbersome
6	Bayesian	Yes	No	Yes	Choose priors with care
7	CL$_s$ with C&H	Yes	Partial	Yes	For upper limits only

2.7 Multivariate Analysis

The number of Physicists objecting to MultiVariate (MV) analyses (like ANN, Decision Trees) is getting smaller as the average year of birth of the active physicists go up. Evaluating the systematics with MV analyzes is very unclear. Many physicists have the habit of changing the input parameter by what they believe is a standard deviation, do it one at a time or randomly with all of them together. There must be a better way to do it. Can the community come up with good figures of merit for the robustness optimization of a MV analysis (and not only for the significance)?

Note that the articles by Linnemann ('A pitfall in estimating systematic errors') and by Reid ('Some aspects of experiment design') in these proceedings also deal with this issue of problems with changing one variable at a time.

2.8 Telling Between Multi Hypotheses

Is the scalar particle we have just discovered a Standard Model one, a CP-odd SUSY one or a CP-even SUSY one [15]? Here are three hypotheses regarding the nature of the Higgs Boson. The Neyman Pearson lemma tell us the best test statistic to tell between two simple hypotheses. In case of more than one equivalent alternate hypotheses, what is the best test statistics to use besides testing them one against the other? Is there anyway to do it without a Bayesian assumption that all hypotheses have an a priori degree of belief?

2.9 Hypothesis Test

Testing a preferred hypothesis includes the estimation and incorporation of systematic errors. The result is then interpreted in terms of exclusion, measurement or discovery. Hypothesis testing is a science by itself. The LEP collaboration has chosen the CL_s method integrating out the systematics using the C&H method. This was one possibility out of many. In the years since LEP we have grown up to understand many more methods. The frequently used ones in HEP are introduced in [3] and compared in Table 1. Since each method has its pros and cons the ATLAS and CMS combined statistics forum has expressed its wish that the analyses be interpreted in a few of them. The frequentist based Profile Likelihood and a Bayesian method are highly recommended. CL_s will certainly be practised for exclusion (the power of a habit...). It is also recommended to try one of the Neyman construction methods. If all methods agree a trust in the result will be established, however, if one method gives completely different inference from others, this should be further investigated.

3 Conclusion and a Personal Wish

Some of the statistical issues raised in this paper are quite picky. The level of the discussions on hypothesis tests is quite advanced, which indicates that the HEP community's understanding of statistics has matured since the days of LEP (thanks in large measure to the Phystat meetings). It is therefore my personal wish that when the real data analysis phase arrives (one hopes soon) every physicist will make the effort to become a Phystatistician to some degree, so he or she understands what is a p-value, what is Profile Likelihood, a Confidence Limit, a Confidence Interval etc.

4 Acknowledgements

This work would have never been possible without the patience and amazing support of Bob Cousins, Kyle Cranmer and Glen Cowan. They helped me to become as close as possible to a Phystatistician in two months.

I would also like to thank Alex Read, Bill Quayle, Luc Demortier and my student Ofer Vitells for always being there to answer my questions and requests. Last but not least, applause to Louis Lyons for keeping the Phystat meetings going. These meetings are the ultimate source of statistics for the High Energy Physicists. Long may these meetings live! Toda Louis. I am obliged to the Benoziyo center for High Energy Physics for their support of this work. This work was also supported by the Israeli Science Foundation(ISF), by the Minerva Gesellschaft and by the Federal Ministry of Education, Science, Research and Technology (BMBF) within the framework of the German-Israeli Project Cooperation in Future-Oriented Topics(DIP).

References

[1] Gary J. Feldman and Robert D. Cousins. A unified approach to the classical statistical analysis of small signals. Phys. Rev., D57:3873-3889, 1998.

[2] Eilam Gross and Amit Klier, Higgs Statistics for Pedestrians, hep-ex/0211058.

[3] Eilam Gross, LHC Statistics for pedestrians, these proceedings.

[4] A nice introduction can be found in John Pumplin's talk: Light Front Physics and Parton Distributions (Talk at LC2006 Minneapolis May 17, 2006); `http://www.pa.msu.edu/people/pumplin/talks/lc.pdf`

[5] A CTEQ and MRST Parton Distribution Functions generator can be found in `http://durpdg.dur.ac.uk/hepdata/pdf3.html`

[6] Robert Throne in these proceedings.

[7] Daniel Stump, Uncertainties of Parton Distribution Functions, Phystat2003.

[8] Private Communication with Robert Throne.

[9] See talk by Kyle Cranmer in these proceedings.

[10] Gary Feldman, Concluding Remarks: Phystat 2005

[11] 95% CL limits and 5σ discoveries in CMS TDR Appendix A. See also talk by Alexey Drozdetskiy, this conference.

[12] P. Sinervo, Definition and treatment of systematic uncertainties in high energy physics and astrophysics PhyStat2003.

[13] See an excellent discussion on estimating sysyematics in the 2000 SLUO Lectures on Statistics and Numerical Methods in HEP, lecture number 5, Roger Barlow, `http://www-group.slac.stanford.edu/sluo/Lectures/stat_lecture_files/sluolec5.pdf`

[14] Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. J. American Statistical Association 84, 200-207; Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, D. V. Lindley and A. F. M. Smith, eds). Oxford: Oxford University Press, 61-77 (with discussion). See also Luc Demortier, Bayesian Reference Analysis for Particle Physics, Phystat05.

[15] D. Rainwater, D. Zeppenfeld and K. Hagiwara, Phys . Rev. D 59, 014037 (1999); T. Plehn, D. Rainwater and D. Zeppenfeld, Phys. Lett. B 454, 297 (1999); Phys. Rev. D 61, 093005 (2000).