May 3, 2009

Statistics for Nuclear and Particle Physicists: an Update

Louis Lyons
Particle Physics, Keble Rd, Oxford OX1 3RH, UK
and
Blackett Lab, Imperial College, Kensington, London SW7 2BW
e-mail: l.lyons@physics.ox.ac.uk

Abstract

Over the years since this book was published in 1986, I have discovered that various statements are incorrect, could have been emphasised differently, or should have been included but were not. This attempts to remedy some of these.

1 Introduction

This update refers to my book 'Statistics for Nuclear and Particle Physicists', published by Cambridge University Press in 1986. Below are listed statements that were incorrect, those where the emphasis could have been better, omissions and typos. In general I attempt to give references to useful information, rather than to provide a full discussion here. Most of these updates are covered in refs. [1] and [2].

If you are aware of any other possible improvements, please e-mail me.

2 Errata

2.1 Coverage issues

Page 90: sentence around eqn (4.26). This refers to the likelihood method of determining a range for a parameter by finding the values where the logarithm of the likelihood function decreases by 0.5 units from its maximum value. The sentence in the text states that, if the measurement is repeated again and again, this procedure results in a series of ranges of which 68% will (in the absence of biases in the measurement) contain the true value of the parameter. In general this statement is **untrue.**

For a given procedure for determining a range for a parameter, the fraction of times this results in a range that does include the true value is called the **coverage** C. It is important to be aware of the fact that coverage is a property of the statistical procedure being used, and does not apply to the particular measurement you are making. Ideally we would like C to equal the nominal expected confidence level for the intervals (68% for the $\Delta \log L = 0.5$ rule quoted above), and to be independent of the values of the parameter. Joel Heinrich[3] has investigated the problem of estimating ranges for the parameter μ of a Poisson distribution from the number of observed counts n; the methods used include likelihood, χ^2, central frequentist intervals, Feldman-Cousins,.... Because in this situation the observations are discrete, the coverage has discontinuities as a function of μ. Plots of C for each method show that C varies by large amounts at small μ, and is as low as $\sim 30\%$ for $\mu \sim 0.5$ for the $\Delta \log L = 0.5$ method. For frequentist methods, C is guaranteed not to fall below the nominal value for any value of μ, but the discontinuities then result in overcoverage for most or all values of μ.

The conclusion is that coverage is not guaranteed by the likelihood method, and undercoverage is a possibility. If this is regarded as an important issue, the coverage properties of the method should be investigated for the particular measurement being made.

2.2 Unbinned maximum likelihood as goodness of fit?

Comment (xii) in Section 4.4.3 of the book states that the value of L_{max}, the observed maximum of the (unbinned) likelihood function, can be used to assess whether the assumed functional form with the best fit values of the parameters provides a satisfactory fit to the data. This is in general **incorrect**.

For the case of a likelihood determination of a lifetime, using a probability density $(1/\tau)\exp(-t/\tau)$, Heinrich[4] has shown that L_{max} depends on the observed decay times only through their average, but not at all on their distribution. Thus L_{max} is incapable of distinguishing between n decays which are approximately exponentially distributed with average time $\bar{t}$, and a sample of n decays all of which decay at the same time $\bar{t}$. This means that at least in this case L_{max} provides no information concerning the goodness of fit of the data to the assumed distribution.

Other examples show similar behaviour. There are even cases where a larger value of L_{max} indicates a worse fit.

In contrast, a maximum likelihood approach to a histogram of data (rather than to a series of individual observations) can yield goodness of fit information if a likelihood ratio is used[5].

3 Clarifications

3.1 Punzi effect for PID

On page 101, likelihood functions are given for determining how many pions and kaons there are in a sample of events for which particle identification information is available. The first formula uses a normalised fit to determine the fractions of each type of particle, while the second uses the 'extended maximum likelihood' approach to estimate the actual numbers of pions and of kaons (see next paragraph). These formulae are suitable for particles of a fixed momentum, but if they have a momentum spectrum, and the spectra for pions and kaons differ, the formulae as they stand will give a biassed result, because of the Punzi effect[6]. This arises because for most particle identification techniques, the lower momentum particles are easier to separate than those at higher momentum, and so a fit to a complete spectrum of momenta is likely to give estimated fractions that are based on the actual fractions at low momenta. The way to modify the formulae to provide unbiassed estimates for a complete spectrum of momenta is discussed by Catastini and Punzi[7].

The wording explaining the difference between the two types of likelihood fit could have been better. For example:
(ii) What is our estimate of the fraction of pions and the corresponding uncertainty based on samples like ours, given it contained n_{obs} tracks? This is answered by the maximum likelihood approach.
(iii) What is our estimate of the number of pions and the corresponding uncertainty for samples like ours, but in which the total number of tracks varies in a Poissonian manner? Here the extended maximum likelihood approach is needed.
Thus with 100 tracks of which 96 are identified as pions, our estimate of the pion fraction in (ii) is 0.96 ± 0.02, while in (iii), the estimated number of pions is 96 ± 10.

In the last paragraph on page 101, the remark about the covariance being zero in the absence of ambiguities applies to the extended maximum likelihood case (as stated). For the ordinary likelihood determination, where n_{obs} is regarded as fixed, the estimates of the pion and kaon fractions are completely anti-correlated (even with no ambiguities in identification) because the fractions must add up to unity.

3.2 Number of degrees of freedom

On the middle of page 112, point (b) states that the number of degrees of freedom ν in assessment of a χ^2 value in a goodness of fit to some data is given by $b-p$, where b is the number of data points included in the comparison, and p is the number of free parameters in the fit. This may be true **asymptotically**, but not for smaller amounts of data.

For example, in a situation where we have two neutrino flavours that are mixing, the survival probability P for a neutrino of type i is given by

$$P = 1 - A\sin^2(k\Delta m^2 L/E), \tag{1}$$

where A and Δm^2 are the two parameters of interest, L and E are respectively the distance the neutrino has travelled and its energy, and k is a known constant. We appear to have 2 free parameters. However, as explained by Feldman and Cousins[8], when Δm^2 is such that the angle $k\Delta m^2 L/E$ is small, the sine of this angle is approximately equal to the angle itself, and so $P \approx 1 - A(k\Delta m^2 L/E)^2$. Then the parameters appear in the fitting function only in the combination $A * (\Delta m^2)^2$. Thus in this region of small Δm^2, there is essentially only one free parameter, while there are effectively two free parameters for larger Δm^2. However asymptotically, we would have enough data to distinguish between $\sin(k\Delta m^2 L/E)$ and $k\Delta m^2 L/E$, and so we would then have two parameters everywhere.

The assymptotic requirement discussed here is in addition to the necessity for a histogram to have enough data so that the number of entries in each bin is approximately Gaussian distributed.

3.3 'Goodness of Fit' and 'Hypothesis Testing'

It is becoming customary in Particle Physics to use the term 'Goodness of Fit' for testing the consistency between data and a single hypothesis H_0, while 'Hypothesis Testing' implies seeing which of two hypotheses H_0 and H_1 (typically the Standard Model and some version of New Physics respectively) is favoured by the data, perhaps also considering other input such our prior beliefs about the models.

Although there is discussion in my book of Goodness of Fit by the χ^2 method, there is almost no mention of methods for Hypothesis Testing. This is discussed, for example, in ref. [9]. Also there are methods other than χ^2 for assessing Goodness of Fit.

3.4 Kinematic Fitting

This is discussed in Section 5.2 in the book.

Steffen Lauritzen[10] has pointed out that the usual weighted sum of squared deviations S is not expected to follow a χ^2 distribution if the kinematic constraints (usually momentum and energy) are non-linear in the variables which have Gaussian experimental uncertainties. If the magnitude of S is used to select wanted events, the loss of signal resulting from the requirement $S < S_{max}$ needs to be investigated, rather than simply using χ^2 tables.

4 Additions

If a new edition of the book were to be produced, here are some extra topics that should be included.

4.1 Bayes versus frequentism

Given that there are these two fundamentally different approaches to statistical analyses, there should have been a discussion of these two methods, emphasising their different philosophies (including their definitions of probability) and practicalities. For a discussion, see for example ref. [11].

4.2 Signal/background separation

Almost every experiment in High Energy Physics uses a multivariate technique for separating signal from unwanted background. A wide variety of methods is available. Information and details of software are available for the packages TMVA[12] and StatPatternRecognition[13].

4.3 Blind analyses

Although they are not applicable to all search or measurement experiments, or to early analyses with a new detector, blind analyses are becoming more popular. They avoid unconscious bias on the part of the physicist, but this has to be weight against the longer times taken in defining the analysis procedure. A review is by Klein and Roodman[14].

4.4 Searches

Given the large number of experimental papers reporting on unsuccessful searches for various forms of new particles or new interactions, an industry has grown up for calculating upper limits on effects that have not been seen. The first two workshops of the PHYSTAT series[21, 27] were devoted to just this topic (see also ref. [15]).

Hopefully, future experiments will be successful in searches for new physics beyond the Standard Model. Statistical issues related to discovery claims were discussed in the 2007 PHYSTAT meeting[26].

4.5 $\Delta\chi^2 = \chi^2$?

In comparing data with two hypotheses, the one which gives a much larger χ^2 than the other may be rejected. For example, we could be comparing a mass histogram with the Standard Model which may predict a smooth mass distribution, or with a model which includes a new particle, when we would also expect a narrow bump in the distribution.

The question is how the difference in χ^2 values should be assessed. When the model with the fewer free parameters is true, and when certain conditions are satisfied, Wilks' Theorem[16] says that this difference should be distributed as χ^2 with the number of degrees of freedom equal to the difference in the number of free parameters in the two models. For the case of the search for a new particle with an unknown mass, the required conditions are not satisfied, and so the significance of any suspected effect cannot be assessed from χ^2 tables. This is discussed in refs. [17] and [18].

4.6 Sensitivity of experiment

In comparing different experiments, or different analyses of the same experiment, it is useful to define which is more sensitive for the determination. This could be, for example, the expected number of standard deviations for the significance claim of the discovery of a new particle, whose expected mass, production rate, etc, are known; or the upper limit expected on the production rate of a hypothesised particle if it really is not produced; or the length of the shortest two-sided interval for a measurement of a given parameter; etc. The sensitivity of a measurement is independent of the actual data, and so is not affected by its particular statistical fluctuations. Thus it is possible for one experiment to have a larger upper limit than another, even though its sensitivity for exclusion is better.

Punzi[19] has argued that, in an experiment searching for a new particle where possible outcomes are discovery, exclusion or no decision, the sensitivity should be defined as the integrated luminosity required for there to be no ambiguous 'no decision' region.

4.7 Systematics

Systematic errors were discussed briefly in Section 1.3 of the book. A fuller review, where other references can be found, is by Heinrich and Lyons[20].

4.8 Bibiliography

Some new literature relevant to the statistical analysis of data in Particle Physics and related subjects has appeared since the book was published. There have been several meetings in the PHYSTAT series; the Proceedings of all of these have appeared[21] - [26], except for the one at Fermilab[27]. The book by Eadie et al[28] has been substantially updated by Fred James[29], and several books have been written by Particle Physicists[30] - [32]. Interesting relevant issues are also discussed at the Statistics Committees that have been established by several of the large experiments[33] - [35].

5 Typos

- Page 15, 4 lines below eqn 1.18 (variance of sum of r_i for Central Limit Theorem):
  'and variance $\Sigma\sigma_i^2/n$' should be 'and variance $\Sigma\sigma_i^2/n^2$'.

- Page 142, in comment (i):
  'δy_i' should be replaced by 'δx_i'.

6 Acknowlegements

I wish to thank the many people who have contributed to my improved knowledge of Statistics over the years. These have included my experimental High Energy Physics colleagues with whom I have discussed data analysis issues; members of the Statistics Committees of the CDF and CMS experiments; and

the Statisticians with whom I have interacted, mainly through the PHYSTAT series of Workshops. Among the Particle Physicists, I would particularly like to mention Bob Cousins, Luc Demortier, Joel Heinrich, Tom Junk and Giovanni Punzi; and the Statisticians include David Cox, Brad Efron, Jerry Friedman and Steffen Lauritzen. I also wish to acknowledge the e-mails I have received for numerous readers of this book. To all of you, I am most grateful.

References

[1] L. Lyons, Annals of Applied Statistics **2** (2008) 887

[2] L. Lyons, 'Statistical issues in Particle Physics', to appear in Handbook of Physics (2009).

[3] J. Heinrich, "Coverage of error bars for Poisson data", `http://www-cdf.fnal.gov/publications/cdf6438_coverage.pdf`

[4] J. Heinrich, 'Pitfalls of Goodness of Fit from Likelihood', ref. [23], page 52.

[5] S. Baker and R. Cousins, Nucl. Inst. and Meth. **A221** (1984) 437.

[6] G. Punzi, "Comments on likelihood fits with variable resolution", ref. [23], page 235.

[7] P. Catastini and G. Punzi, "Bias-free estimation of multicomponent maximum likelihood fits with component-dependent templates", ref. [24], page 60.

[8] G. J. Feldman and R. D. Cousins, Phys. Rev. **D57** (1998) 3873.

[9] L. Lyons, 'Methods of comparing two hypotheses' (2009) `http://www-cdf.fnal.gov/physics/statistics/statistics_recommendations.html`

[10] S. Lauritzen, private communication.

[11] L. Lyons, "Bayes or Frequentism?", `http://www-cdf.fnal.gov/physics/statistics/statistics_recommendations.html`

[12] A. Hocker et al. 'TMVA, Toolkit for Multi-Variate data Analysis with Root', ref. [26], page 184.

[13] I. Narsky, 'StatPatternRecognition in analysis of HEP and Astrophysics data', ref. [26], page 188.

[14] J. R. Klein and A. Roodman, Annual Review of Nuclear and Particle Physics **55** (2005) 141.

[15] J. Heinrich, 'Review of the Banff challenge on upper limits', ref. [26], page 125.

[16] S. S. Wilks, 'The large-sample distribution of the likelihood ratio for testing composite hypotheses', Annals of Math. Stat. **9** (1938) 60.

[17] R. Protassov et al., 'Statistics: Handle with care. Detecting multiple model components with the likelihood ratio test', Astrophysics Journal **571** (2002) 545.

[18] L. Demortier, 'p-values and nuisance parameters' ref. [26], page 23.

[19] G. Punzi, 'Sensitivity of searches for new signals and its optimisation', ref. [23], page 235.

[20] J. Heinrich and L. Lyons, Annual Reviews of Nuclear and Particle Science, **57** (2007) 145.

[21] Workshop on Confidence Limits, CERN Yellow Report 2000-05.

[22] Advanced Statistical Techniques in Particle Physics, Durham IPPP/02/39.

[23] Proceedings of PHYSTAT2003, eConf C030908, SLAC-R-703.

[24] "PHYSTAT05: Statistical Problems in Particle Physics, Astrophysics and Cosmology", Imperial College Press (2006) `http://www.physics.ox.ac.uk/phystat05/` .

[25] BIRS Workshop on "Statistical inference problems in High Energy Physics and Astronomy", Banff 2006 `http://www.birs.ca/birspages.php?task=displayevent&event_id=06w5054` .

[26] PHYSTAT-LHC Workshop on "Statistical Issues for LHC Physics" (2007), `http://phystat-lhc.web.cern.ch/phystat-lhc/2008-001.pdf`

[27] FNAL Confidence Limits Workshop (2000) `http://conferences.fnal.gov/CLW/`

[28] W. T. Eadie, D. Drijard, F. E. James, M. Roos and B. Sadoulet, "Statistical Methods in Experimental Physics", American Elsevier, (1971).

[29] F. James, "Statistical Methods in Experimental Physics", World Scientific Publishing Co (2007).

[30] R. Barlow, "Statistics: a Guide to the Use of Statistical Methods in the Physical Sciences", Wiley (1989).

[31] G. Cowan, "Statistical Data Analysis", Oxford University Press (1998).

[32] B. Roe, "Probability and Statistics in Experimental Physics", Springer Verlag (1991).

[33] BaBar Statistics Working Group `http://www.slac.stanford.edu/BFROOT/www/Statistics/`

[34] CDF Statistics Committee http://www-cdf.fnal.gov/physics/statistics/statistics_home.html

[35] CMS Statistics Committee https://twiki.cern.ch/twiki/bin/view/CMS/StatisticsCommittee