Statistical Issues II

Run II Advanced Analysis Workshop

Harrison B. Prosper

Florida State University

Outline

- Introduction
- Questions
- Summary

Introduction

- High energy physics is a field characterized by frequent, often bold, intellectual innovation in experimental technique and theory. Ideas are constantly challenged because none is held to be sacrosanct.

- Yet our statistical practice, based largely on the ideas of
  - Karl Pearson — χ^2
  - Jerzy Neyman — confidence intervals, likelihood ratio
  - Ronald Fisher — likelihood theory, null hypothesis tests

  has remained fossilized for seventy years. Why is this?

- It could be that statistical analysis in our field is already optimal. If so, by definition, it cannot be improved.

- But presumably we wouldn't be at this workshop if we believed that!

…Introduction

- We shall encounter increasingly difficult analyses. But it may be that the methods of the 1930s will remain adequate. On the other hand, it is also possible that they will not and that we could benefit from a critical re-assessment of our statistical ideas.
- Some re-assessment and renovation is needed if only to illuminate and clarify current practice.
- A goal of the Run II Advanced Analysis Workshop is to assist this renaissance through vigorous discussion of, and hard thinking about, statistical issues.
- A great deal has happened in statistics during the past seventy years. It would be astonishing if none of it is useful to us.

Questions: Uncertainty

- How should uncertainties, statistical and systematic, be handled? Are Frequentist notions adequate? What about Bayes?
  - What are the mathematical rules?
  - Are they unique? If not, does it matter?
  - Do we care if some are ad-hoc?
- What precisely is a statistical uncertainty? What precisely is a systematic uncertainty? Can these concepts be made precise, or is some imprecision inevitable? Do systematic errors have probability distributions? Is so, what do we want them to mean?
- Can the the two kinds of uncertainty always be disentangled?
- When the two kinds of uncertainty are added in quadrature what do we want the resulting uncertainty to mean? ($m_t = 174 \pm 5$ GeV)

Questions: …Uncertainty

- Under what circumstances is it reasonable to use Gaussians to model uncertainty? When does it not make sense? For example, is it okay to use a Gaussian for an efficiency, which necessarily assigns a non-zero probability density at zero efficiency?
- Is there a way to report the results of experiments in a way that would allow other parties to combine the results easily?
- How should we combine the results of two or more experiments when our knowledge of their mutual correlation is poor?
- How should we handle systematic uncertainties that are merely "assigned?", that is, guessed!

Questions: Model Comparison

- Does a "goodness of fit" test really do what we want?
  - On the face of it no: it does not tell us how good a fit is; rather it tells us the degree to which a fit is bad!
  - It tests how badly the null hypothesis (a curve, perhaps, or a mean background count) fits the data. It uses a measure of badness called a p-value; if the p-value is small enough, assuming the null hypothesis to be true, then the latter is assumed not to be and is therefore rejected!
- Can we construct a "goodness of fit" test that does what we want? Can Bayes help? On the face of it not easily, because of the need to specify a sufficiently large space of alternative hypotheses. But how big must this space be to be adequate for our purposes?
- How should we compare models? Is there a way to compare, for example, the models: background vs (unspecified) signal + background, when we go bump-hunting? Or our we stuck with p-values?

Questions: Probabilistic Reasoning

- Should we strive for conceptual clarity and consistency? Or should we reserve the right to remain "exuberantly irrational?"
- Since, in practice, we use Bayesian ideas, why not do so correctly?
- Is it practical to do so? (How do we handle goodness of fit?)
- What about prior probabilities? How should we choose them?
- Can we arrive at a default set that our community can use for the classes of problems we encounter?
- What about priors on multi-dimensional spaces?
- How should we sample such parameter spaces?
- How can we perform efficient multivariate marginalization?

Questions: Odds and Ends

- Given two analyses, how do we decide which is the better?

- Is a lack of bias (bias(x_{true}) = <x> - x_{true}) a sensible criterion?
- What about min[root mean-square]? Or using consistency and rate of convergence to x_{true} as criteria?
- We insist on combining results by a weighted average. But is that the only way to do this?
- How much are we hurt, in the sense of inefficient use of data, by inflating the variance in order to achieve bias(x_{true}) = 0?

- To understand our background models and object identification we must of necessity subject our data, the vast bulk of which is background, to excruciatingly detailed investigation. So what precisely is the import of the commandment:
  - “thou shallt not open thy data box until thy cuts be carved in stone?”

Summary

- Our analyses are becoming increasingly difficult, yet the statistical ideas we have been using have hardly changed in seventy years.
- Since the 1930s, however, a huge amount of theoretical and applied work has been done in statistics and many ideas have changed.
- It is probable that some small fraction of this work could be useful to us, both to illuminate what we are already doing and to suggest better ways to do it.
- But to learn enough of the useful bits will require real intellectual effort on our part.
- Our hope is that this workshop will provide a stimulating forum for our collective education.