

D.R. Cox at PhyStat

Bob Cousins
Univ. of California, Los Angeles

PhyStat seminar
March 23, 2022

In 2004, I was searching the statistics literature while preparing for a talk at 2005 Oxford PhyStat meeting,

“Treatment of nuisance parameters in high energy physics, and possible justifications and improvements in the statistics literature”

Among the papers that I found interesting and relevant were:

D.R. Cox, N. Reid,

“A Note on the Difference Between Profile and Modified Profile Likelihood” Biometrika 79, 408 (1992)

D.R. Cox, N. Reid,

“Parameter Orthogonality and Approximate Conditional Inference”, Jour. Roy. Stat. Soc. Series B 49, 1 (1987)

Meanwhile, I received some emails,

From : Louis Lyons <l.lyons1@physics.ox.ac.uk>

Subject : A few more items of interest to Particle Physicist

To : cousins@physics.ucla.edu

Hi Bob

I met with **Sir David Cox** recently. He is an eminent statistician, and said he would be interested in discussing with me 'interesting problems'. This sounds as if it could be useful.[...].

Let me know if you have suggestions. [...]

Louis

From : Louis Lyons <l.lyons1@physics.ox.ac.uk>

Subject : Nuisance parameters and Frequentist approaches

To : cousins@physics.ucla.edu

Here is the previous message I sent to Sir David Cox
Louis

-----Original Message-----

From: Louis Lyons

Sent: 23 June 2004 18:05

To: 'David Cox'

Subject: Nuisance parameters and Frequentist approaches

Dear Sir David,

I said I would send you an outline of some of the statistical issues on which we would appreciate expert guidance. Here is one that I sent to the Statisticians who had been at the PHYSTAT2003 Conference.

More questions to follow

Louis

It was some months before I realized that
“D.R. Cox” and “Sir David” were one and the same!

Slow to catch on, eventually I also realized that
statistician **Nancy Reid**, who was invited to the
conference, was D.R. Cox’s co-author, **N. Reid**.

At that point I realized that I was planning to talk about
their papers with them in the audience...

Fortunately, it all worked out OK, and I had the pleasure of sitting between Nancy and David at the conference banquet, as they reminisced and discussed various matters.

And we all learned how kind and gentle David was with high energy physicists.

PhyStat 2005 in Oxford

FREQUENTIST AND BAYESIAN STATISTICS: A CRITIQUE (KEYNOTE ADDRESS)

D.R. COX

Nuffield College, Oxford OX1 1NF, UK

E-mail: david.cox@nuf.ox.ac.uk

The broad distinctions between the frequentist and Bayesian approaches to statistical inference are outlined and some brief historical background given. The advantages and disadvantages of the frequentist discussion are sketched and then two contrasting Bayesian views given. The difficulties with the notion of a flat or uninformative prior distribution are discussed.

https://www-worldscientific-com.ezproxy.cern.ch/doi/suppl/10.1142/p446/suppl_file/p446_chap01.pdf

Very nice 6-page introduction, which I highly recommend. Interesting that a lot of it is devoted to Bayesian methods, with repeated warning about flat priors, especially in high dimensions: “representing ignorance as such by a flat prior is treacherous”.

Describes “major advantages” of frequentist approach, while noting,

“A key problem in principle in frequentist formulations is that of ensuring that the long-run used in calibration is relevant to the analysis of the specific data being analysed.”

No other allusion to conditioning (!).

**PhyStat 2007 at CERN: Major talk,
“The Last Fifty Years of Statistical Research and their
Implications for Particle Physics”**

Model construction absolutely critical

In testing many hypotheses (labeled by j), there are two distinct problems (quite different answers):

- 1) some small but almost certainly nonzero number of the hypotheses are false; it is required to assess which those are**
- 2) it is quite possible that all null hypotheses are true: how strong is the evidence against this on the basis of $m = \min(p_j)$**

Advocates plot of $-\log(p_j)$.

Slides: <https://indico.cern.ch/event/17772/contributions/1444577/attachments/232667/325582/cernmod.pdf>

Video: <http://cds.cern.ch/record/1565849>

Proceedings: <https://cds.cern.ch/record/1021125> , p. 3,

For large number of tests, emphasizes “sensitivity to the assumptions involved in calculating the individual p_j to a degree that will often be quite unreasonable.”

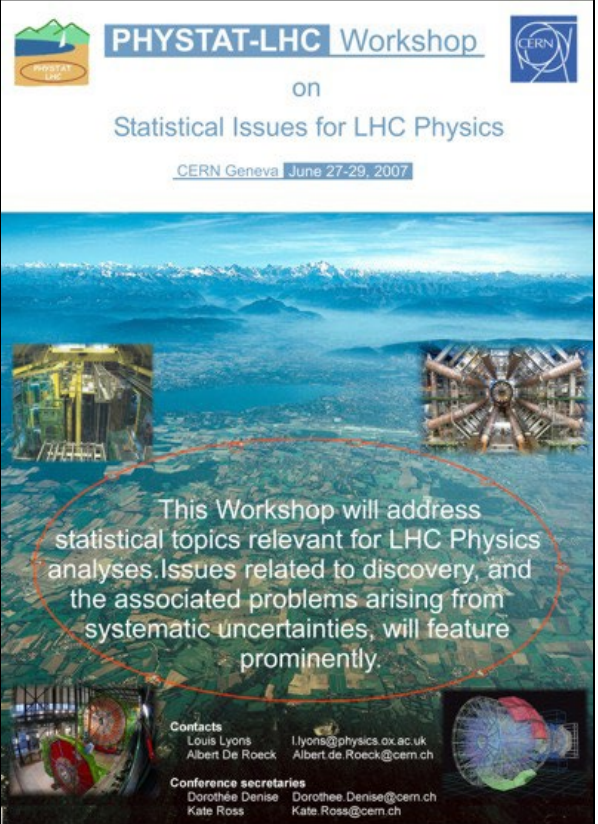
The killer question: What if there are a lot of parameters (relative to the amount of information)? Can cause problems for both frequentist (Max Likelihood) and Bayesian methods.

Big takeaway for me was his “Five faces of Bayesian statistics” – I have shown next two slides many times.


Sir David Cox at PhyStat-LHC 2007

Five faces of Bayesian statistics

- empirical Bayes; number of similar parameters with a frequency distribution
- neutral (reference) priors: Laplace, Jeffreys, Jaynes, Berger and Bernardo
- information-inserting priors (evidence-based)
- personalistic priors
- technical device for generating frequentist inference



The poster for the PHYSTAT-LHC Workshop features a blue and white color scheme. At the top left is the PHYSTAT-LHC logo, and at the top right is the CERN logo. The main title is 'PHYSTAT-LHC Workshop' in a blue box, followed by 'on Statistical Issues for LHC Physics' and the date 'CERN Geneva June 27-29, 2007'. The background is a landscape with mountains and a city. Two large particle detector structures are shown on either side of the central text. A red circle highlights the central text. At the bottom, there are two columns of contact information and a small image of a particle detector.

PHYSTAT-LHC Workshop 

on
Statistical Issues for LHC Physics

CERN Geneva June 27-29, 2007

This Workshop will address statistical topics relevant for LHC Physics analyses. Issues related to discovery, and the associated problems arising from systematic uncertainties, will feature prominently.

Contacts
Louis Lyons l.lyons@physics.ox.ac.uk
Albert De Roeck Albert.de.Roeck@cern.ch

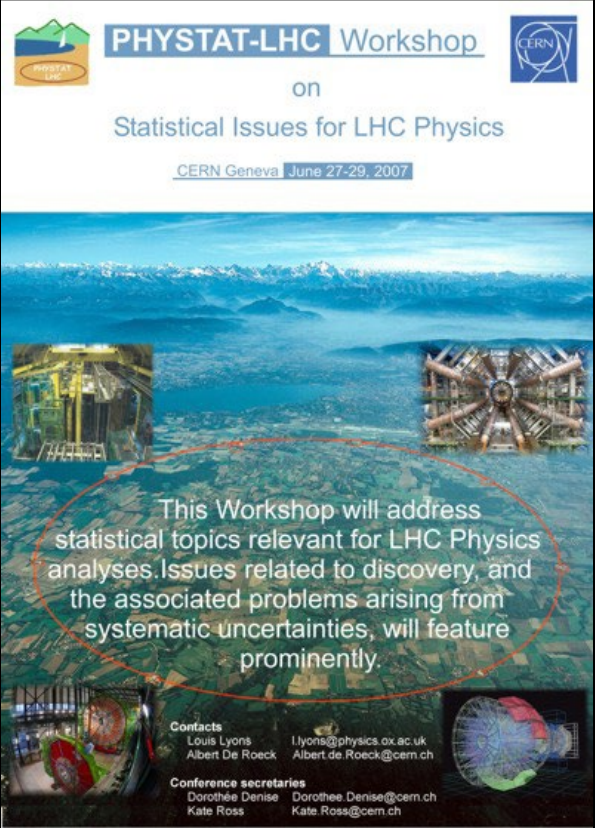
Conference secretaries
Dorothee Denise Dorothee.Denise@cern.ch
Kate Ross Kate.Ross@cern.ch

Further information and registration at <http://cern.ch/physstat-lhc>


From my Summary Talk: his list, as augmented in HEP

Six faces of Bayesian statistics

- empirical Bayes; number of similar parameters with a frequency distribution
- neutral (reference) priors: Laplace, Jeffreys, Jaynes, Berger and Bernardo
- information-inserting priors (evidence-based)
- personalistic priors
- technical device for generating frequentist inference
- Priors flat in arbitrary variables.



The poster for the PHYSTAT-LHC Workshop features a blue and white color scheme. At the top left is the PHYSTAT-LHC logo, and at the top right is the CERN logo. The main title is 'PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics', with the location and date 'CERN Geneva June 27-29, 2007' below it. The background is a landscape with mountains and a city. Two large particle detector structures are shown on either side of the central text. A red circle highlights the central text. At the bottom, there are contact and conference secretary information, and a small image of a particle detector.

PHYSTAT-LHC Workshop 

on
Statistical Issues for LHC Physics

CERN Geneva June 27-29, 2007

This Workshop will address statistical topics relevant for LHC Physics analyses. Issues related to discovery, and the associated problems arising from systematic uncertainties, will feature prominently.

Contacts
Louis Lyons l.lyons@physics.ox.ac.uk
Albert De Roeck Albert.de.Roeck@cern.ch

Conference secretaries
Dorothee Denise Dorothee.Denise@cern.ch
Kate Ross Kate.Ross@cern.ch

Further information and registration at <http://cern.ch/phystat-lhc>

Also at PhyStat 2007 at CERN

The Wish-lists: Some Comments

D.R. Cox and N. Reid

Nuffield College, Oxford and University of Toronto, Canada

Abstract

We provide brief comments on some common threads arising from the ‘wish-lists’ set out in some of the other papers in this volume. The discussion is necessarily incomplete: in particular we have dealt only with points for which a reasonably compact answer seems possible.

1 Introduction

The wish-lists are wide-ranging and raise issues of varying difficulties, ranging up to the seemingly impossible. The following comments concern just some of the topics raised.

D.R. Cox and N. Reid, “The Wish-lists: Some Comments”, p. 119, <https://cds.cern.ch/record/1021125>

PhyStat 2011 at CERN: Another major talk, with more technical aspects

As in 2007, with many tests, “crucial distinction”:

- 1) there may be no signal present or just one.**
- 2) there are likely to be a limited but nonzero number of signals present and the challenge is to find as many as possible of them with few false alarms.**

Again advocates ordered plots of $-\log(p)$; relates max value to extreme value theory

Discusses monograph of Efron (2010), including a fully nonparametric empirical Bayes procedure.

Issues of sequential testing as more data accumulated.

Slides: <https://indico.cern.ch/event/107747/contributions/32679/attachments/24373/35062/cerncox.pdf>

Video: <https://indico.cern.ch/event/107747/contributions/32679/attachments/24372/35061/go>

Proceedings: <https://indico.cern.ch/event/107747/contributions/32679/attachments/24374/35063/cox.pdf>

Conclusion:

My overall impression from examining these talks, including the wish-list comments with Nancy Reid, is that we did not absorb as much as we might have, and it is worth our while to study them again.