PHYSTAT-Systematics: Setting the Scene

Louis Lyons

Imperial College London & Oxford

l.lyons@physics.ox.ac.uk

1st Nov 2021

PHYSTAT-Systematics: Setting the Scene

שְׁגִיאוֹת מִי יָבִין מִנִּסְתָּרוֹת נַקֵּנִי

Who can understand errors?
Save me from hidden effects.
Psalm 19 verse 13 ± 1

Louis Lyons

Imperial College London & Oxford

l.lyons@physics.ox.ac.uk

1st Nov 2021

What are 'Systematics'?

We are measuring a parameter; or setting a limit; or rejecting a hypothesis; etc.

What are statistical uncertainties ?

Limited accuracy of measurements,

Poisson fluctuations in counting experiments

Uncertainty usually reduces with more data, e.g. $1/\sqrt{n}$

Analyses with different data-sets uncorrelated

What are Systematics?

e.g. Uncertainty in calibration correction (but many other sources too)

Not measurable from spread of results

More data --> usually not reduced in magnitude

e.g. Measure my height to accuracy of $\pm$ 1mm. Repeat 10^6 times.

Height now known to 10^{-6}m?

Correlated between different analyses

Sometimes ambiguous whether statistical or systematic

More difficult that statistical uncertainty, & more open to personal judgement

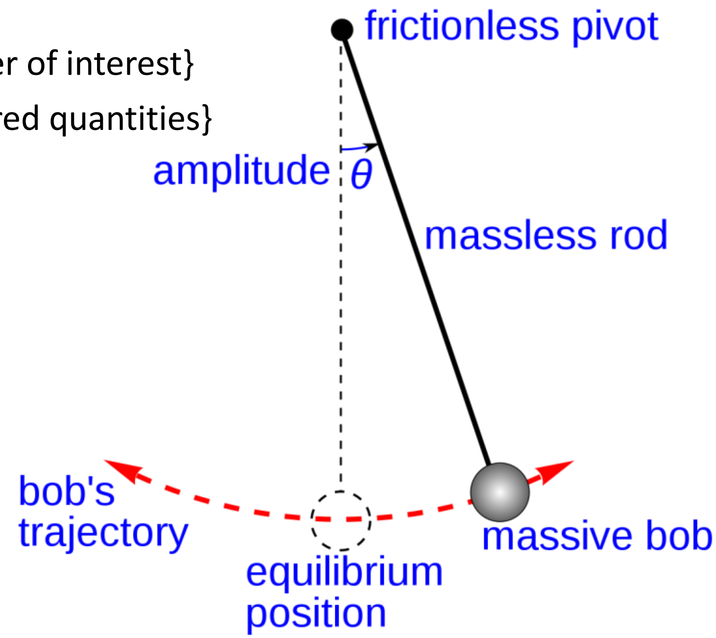
Universal in analyses. Big theme in Higgs 2021 meeting 2 weeks ago

~ Statisticians' nuisance parameters

Info/reading material on website https://indico.cern.ch/event/1051224/

Brief examples of pendulum experiment and from Particle Physics

Brief example of Systematics from school Physics: Pendulum experiment



Time swings of pendulum → g = local acceleration due to gravity {Parameter of interest}

$\tau = 2\pi\sqrt{L/g}$, τ = period of pendulum L = its length {Measured quantities}

Statistical uncertainty in measuring L and τ

Systematic uncertainties on L and τ, from uncertainties in calibrations.
Perform subsidiary measurements to estimate these.
BUT there are other systematics, not explicitly in formula e.g.
1: Derivation of formula requires swings of small angle θ, etc.
Soln: Either derive better approx for larger θ; or
Make measurements at a series of θ and extrapolate to θ = 0
2: Maybe we want g at sea level, and our lab is up a mountain:
Consult one or more geologists on how to correct.
See Sect 1.1 of 'Systematic Errors' for more details:
https://www.annualreviews.org/doi/abs/10.1146/annurev.nucl.57.090506.123052

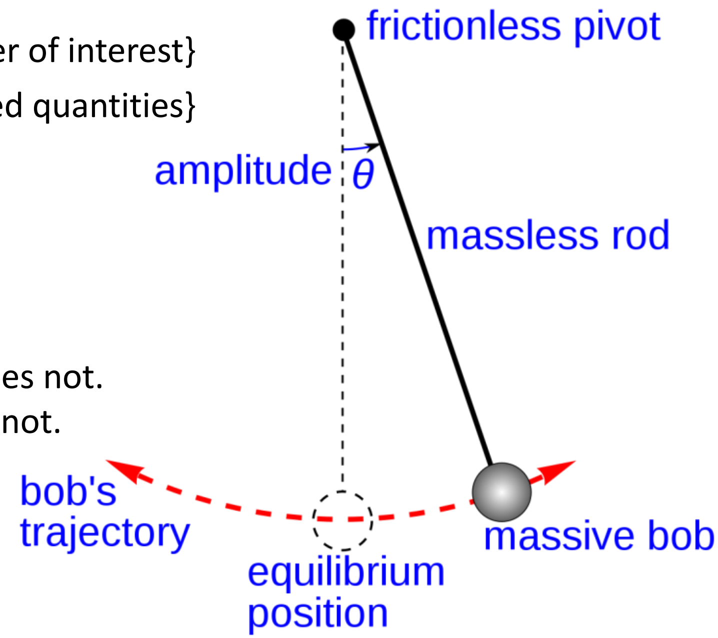
→

Pekka Sinervo's classification of systematics (PHYSTAT2003):

The Good, The Bad and The Ugly

Good = your own calibrations. Basically statistical
Bad = using other peoples results, poorly understood data or analysis technique, model assumptions
Ugly = different theoretical estimates, theory with limited number of terms; etc

Brief example of Systematics from school Physics: Pendulum experiment



Time swings of pendulum → g = local acceleration due to gravity {Parameter of interest}

$\tau = 2\pi\sqrt{L/g}$, τ = period of pendulum L = its length {Measured quantities}

Statistical uncertainty in measuring L and τ

Systematic uncertainties on L and τ, from uncertainties in calibrations.

Perform subsidiary measurements to estimate these.

N.B. Stat uncert decreases with more data in main measurement; syst does not.

For 2 different expts, stat uncerts are uncorrelated; syst uncerts are not.

e.g. Measure my height to accuracy of 1mm. Repeat 10^6 times.

Height now known to 10^{-6}m?

BUT there are other systematics, not explicitly in formula e.g.

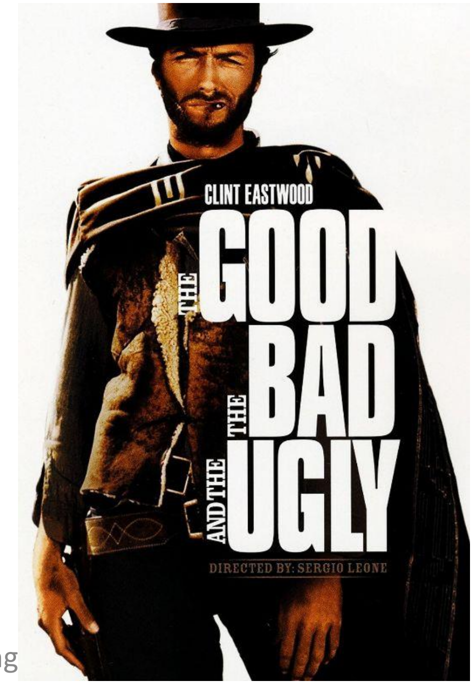
1: Derivation of formula requires swings of small angle θ, etc.

Soln: Either derive better approx for larger θ; or

Make measurements at a series of θ and extrapolate to $\theta = 0$

2: Maybe we want g at sea level, and our lab is up a mountain:

Consult one or more geologists on how to correct.



Pekka Sinervo's classification of systematics (PHYSTAT2003):

The Good, The Bad and The Ugly

Good = your own calibrations. Basically statistical

Bad = using other peoples results, poorly understood data or analysis technique, model assumptions

Ugly = different theoretical estimates, theory with limited number of terms; etc

Quoting result

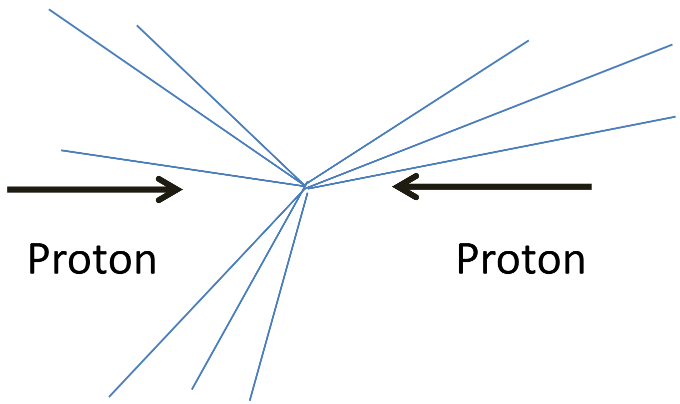
In Abstract and Conclusions: $\mu = x \pm \sigma_{stat} \pm \sigma_{syst}$
and perhaps $\sigma^2_{tot} = \sigma^2_{stat} + \sigma^2_{syst}$

In main part of paper: List all separate contributions of systematics to σ_{syst}

Why separate the stat and syst components?
1) Statistical uncertainties are more reliable than systematic ones.
2) Correlations with systematics in other measurements can be important.
3) In principle can improve result if some source of systematic is better determined.

Brief example from Particle Physics:

Quick aside on Particle Physics:
Experiment at Large Hadron Collider: protons collide, produce lots of particles



Detector for outgoing particles: Surrounds collision region

ATLAS is cylindrical, 44m long, 25m diameter

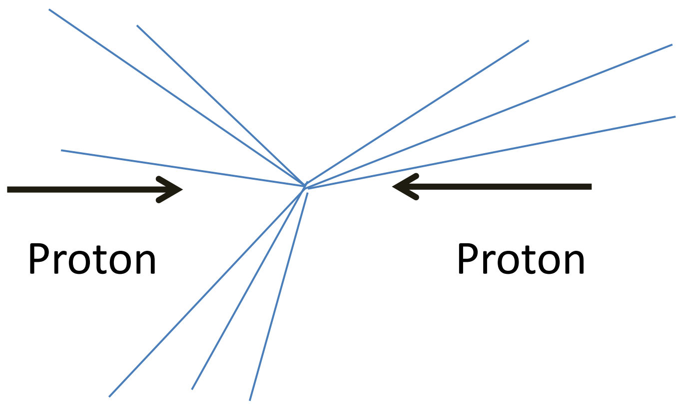
Too many interactions ('events') to record, so quickly decide which to keep ('Trigger')
Still too much background, so select sample with human designed cuts, and MVA/ML
Types of statistical procedures in analyses:

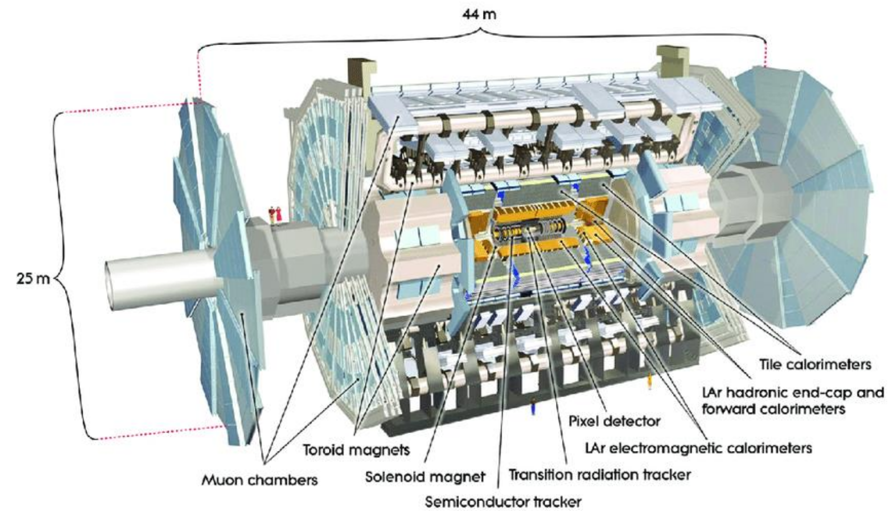
- Parameter determination e.g. Lifetime of specific particle, and uncertainty.
- Hypothesis testing e.g. search for New Physics →
  - Discovery (e.g. Higgs → Nobel Prize), or Upper Limit (SUperSYmmetry)

Brief example from Particle Physics:

Quick aside on Particle Physics:

Experiment at Large Hadron Collider: protons collide, produce lots of particles





Detector for outgoing particles: Surrounds collision region

ATLAS is cylindrical, 44m long, 25m diameter

Too many interactions ('events') to record, so quickly decide which to keep ('Trigger')

Still too much background, so select sample with human designed cuts, and MVA/ML

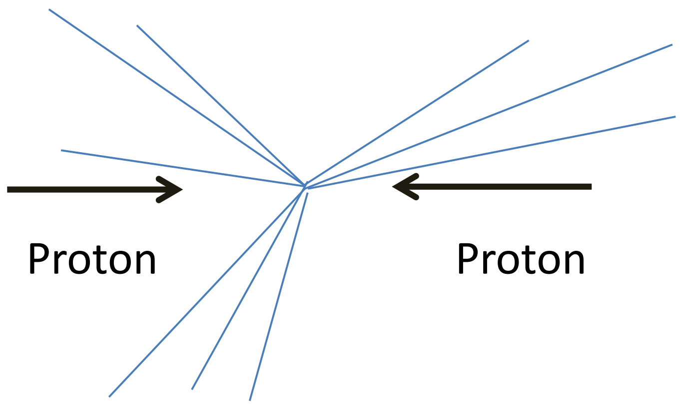
Types of statistical procedures in analyses:

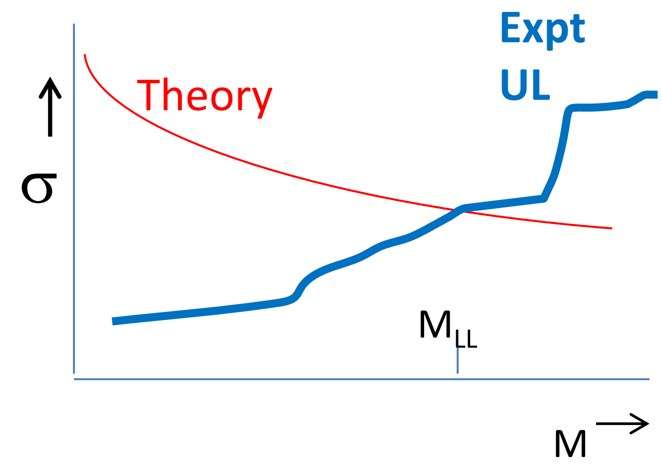
- Parameter determination e.g. Lifetime of specific particle, and uncertainty.
- Hypothesis testing e.g. search for New Physics →
  - Discovery (e.g. Higgs → Nobel Prize), or Upper Limit (SUperSYmmetry)

Brief example from Particle Physics:

Quick aside on Particle Physics:

Experiment at Large Hadron Collider: protons collide, produce lots of particles





Detector for outgoing particles: Surrounds collision region

ATLAS is cylindrical, 44m long, 25m diameter

Too many interactions ('events') to record, so quickly decide which to keep ('Trigger')

Still too much background, so select sample with human designed cuts, and MVA/ML

Types of statistical procedures in analyses:

- Parameter determination e.g. Lifetime of specific particle, and uncertainty.
- Hypothesis testing e.g. search for New Physics →
  - Discovery (e.g. Higgs → Nobel Prize), or Upper Limit (SUperSYmmetry)

Example from Particle Physics:
Production rate of Higgs compared with Prediction of Standard Model

Higgs detected via its decay . e.g H → $\gamma\gamma$ or H → ZZ → 4l, l = e or μ

$n = L*A*\sigma$ n = number of signal events;
L = time integrated 'luminosity'
A = acceptance *efficiency
σ = 'cross-section'= param of interest. (Dartboard analogy) Cf with SM prediction

STATISTICAL UNCERTAINTIES:
n = t-b, t = total events , b = number of background events;
or from fit to mass spectrum
Luminosity from machine experts
Acceptance & efficiency estimated within analysis

SYSTEMATIC UNCERTAINTIES:
All above quantities. Many related to experimental features, e.g.:
- Trigger efficiency
- Accuracy of calorimeter energy determination
- Misalignment of detector
- Identification of different types of particles e.g. γ, e, μ, π, K, ν, …….
- Unfolding (~ Deconvolution), incuding response matrix (see talk by **Mikael Kuusela**)

But also:
Monte Carlo statistics, and theory uncertainties e.g. background shape
Mis-modelling (see talk by **Adinda de Wit**)
Machine Learning Uncertainties (see talk by **Kyle Cranmer**)
Parton uncertainties (See future PHYSTAT Seminar by **Mandy Cooper-Sarkar**)
Other theory uncertainties. e.g. Missing higher order corrections (see talk by **Frank Tackmann**)

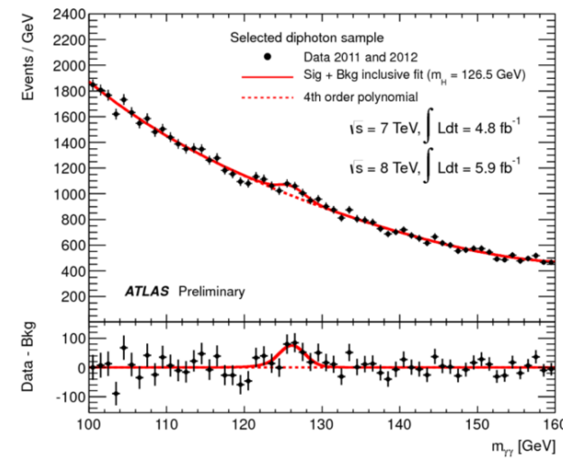
More details in talks by **Lukas Heinrich** and **Sasha Glazov**

Example from Particle Physics: Production rate of Higgs compared with Prediction of Standard Model

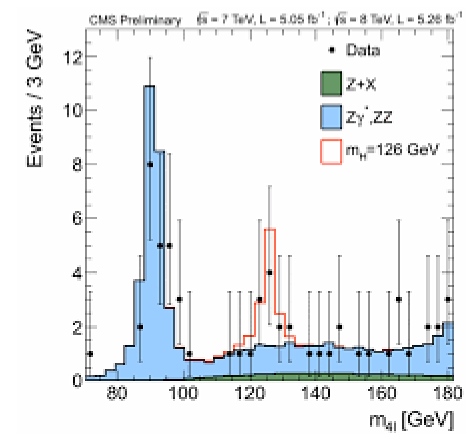
Higgs detected via its decay . e.g H → γγ or H → ZZ → 4l, l = e or μ

n = L*A*σ n = number of signal events;
L = time integrated 'luminosity'
A = acceptance *efficiency
σ = 'cross-section'= param of interest. Cf with SM prediction

STATISTICAL UNCERTAINTIES:
n = t-b, t = total events , b = number of background events;
or from fit to mass spectrum
Luminosity from LHC machine experts
Acceptance & efficiency estimated within analysis



SYSTEMATIC UNCERTAINTIES:
All above quantities. Many related to experimental features, e.g.:
- Trigger efficiency
- Accuracy of calorimeter energy determination
- Misalignment of detector
- Identification of different types of particles e.g. γ, e, μ, π, K, ν, …….
- Unfolding (~ Deconvolution) (see talk by **Mikael Kuusela**)



But also:
Monte Carlo statistics, and theory uncertainties e.g. background shape
Mis-modelling (see talk by **Adinda de Wit**)
Machine Learning Uncertainties (see talk by **Kyle Cranmer**)
Parton uncertainties (See future PHYSTAT Seminar by **Mandy Cooper-Sarkar**)
Other theory uncertainties. e.g. Missing higher order corrections (see talk by **Frank Tackmann**)
More details in talks by **Lukas Heinrich** and **Sasha Glazov**

Systematics $\nu \rightarrow$ Uncertainty on parameter μ

Simple uncertainty propagation

OK if uncertainties are small

χ^2

$S = \sum\sum(d_i-m_i)C_{ij}(d_j-m_j)$, where d = data, m = model and

C = covariance matrix, including systematic correlations

Equivalent to constraint terms in likelihood

Likelihood

Include nuisance param as constraint, based on subsidiary measurement

e.g. For counting expt, with uncertain background

$$L(\mu,\nu) = \text{Poisson}(d|\mu,\nu)*\text{Gaussian}(\nu_{meas}|\nu,\sigma_\nu)$$

Use profile likelihood

Bayes:

Posterior of subsidiary expt = Prior for main measurement.

Marginalise posterior over nuisance params

Frequentist Neyman construction (including Feldman-Cousins likelihood ordering)

For more than 2 or 3 parameters, use $L_{profile}$

Mixed Frequentist/Bayes

Frequentist for main measurement, Bayes for nuisance params

e.g Cousins & Highland NIM A320 (1992) 331

Systematics $\nu \rightarrow$ Uncertainty on parameter μ , continued

OPAT = One Parameter At a Time

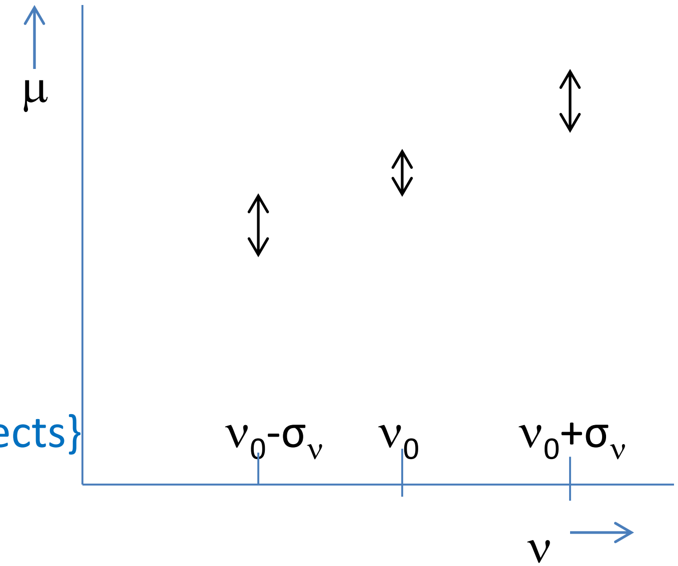
Change in μ for $\pm 1\sigma$ change in ν;

or could do MC of 1-D Gaussian (or whatever) for ν

If MC, new MC for each value of ν, or reweight?

Change in μ has uncertainty u from limited amount of MC, i.e. $\Delta\mu \pm u$

If Δ^+ differs from $-\Delta^-$, need quadratic term? But then other second order terms



All Params at Once

MC for multi-D Gaussians for all ν

For methods that determine effect of independent sources of systematics, total syst=$\sqrt{\Sigma}$ {squares of independent effects}

*****But using full likelihood, as product of likelihoods from main & subsid expts, is better ******

PHYSTAT-SYSTEMATICS

Emphasis on **systematics**
Emphasis on **incorporating systematics into analyses**
'How to estimate' = another interesting topic

Talks (ample time for discussion):
Introductions: LL, David van Dyk
Specific physics areas: LHC, Neutrinos, Dark Matter, Flavour Physics
Specific topics: Mis-modelling, Machine Learning, Theory, Unfolding, Bootstrap
Short talks by Statisticians: second half of Wednesday
Responders after some of talks
Concluding talks: Physicist: Nick Wardle; Statistician: Sara Algeri .
N.B. Wednesday Nov 10th

Not using Mattermost, etc; send emails to speakers.
Comments /Feedback (during or after Workshop): Send to organisers:
olaf.behnke@cern.ch lockhart@sfu.ca l.lyons@physics.ox.ac.uk nckw@cern.ch

Topics for discussion*

* A version of this list, with comments to make it more accessible to Statisticians, starts at the bottom of page 2 of https://www.desy.de/~obehnke/PHYSTAT_SYS/StatisticianGloss.pdf

a) Are systematics for event selection (efficiency and contamination), parameter determination, discovery claims and upper limits dealt with differently?

b) Some systematics have a statistical origin. Is it unimportant what they are called, so long as their correlations (with other possible measurements) are dealt with correctly?

c) Are there situations where profiling a likelihood is better than marginalising a posterior probability?

d) How do we deal with asymmetric uncertainties? e.g. How to combine several measurements with asymmetric uncertainties; calculating Goodness of Fit; etc.

e) Do uncertainties in theory present special problems? And similarly when different theories make discrepant predictions.

f) Are there issues in situations with thousands of nuisance parameters? Can we run into problems as pointed out in the Neyman-Scott paper (Econometrica, 1948)?

g) The various issues that arise when Monte Carlo simulation is compared with the data, e.g. M.C./data discrepancies; weighted M.C. events; reweighting M.C. or generating new M.C. as parameter(s) of M.C. are changed; limited M.C. data.

h) If the effect of changing a nuisance parameter on the parameter of interest is x+-s (where s is an uncertainty on the estimate x, perhaps arising from limited number of events in the M.C.), what do we take as the contribution of this nuisance parameter to the total systematics?

i) Relative merits of Bayesian and Frequentist (likelihood) ways of incorporating nuisance parameters.

j) Validity of combining Bayes for nuisance parameters with frequentist methods for parameter of interest.

k) Problems in the shift method for estimating the effect of nuisance parameters. This includes OPAT = changing One Parameter At a Time. How does it compare with changing all parameters simultaneously? Do possible non-linearities affect the choice? How does it compare with the full likelihood approach?

l) A paper by Dauncey et al. (J Instrumentation 10 (2015) 04015) deals with discrete nuisance parameters (e.g. the functional form chosen for fitting a spectrum) in analogy with the way continuous nuisance parameters are profiled. Is this a known statistical procedure?

Statisticians

Statisticians have always been essential part of PHYSTAT meetings and Seminars
Very welcome at this remote meeting. Encouraged to interact.
Very sorry missing live interactions, in sessions, coffee breaks, mealtimes, etc.

Talks: Introduction (David van Dyk)
Bootstrap (Brad Efron)
Unfolding (Mikael Kuusela)
Summary (Sara Algeri)

Most of Responders: Alessandra Brazzale, Anthony Davison, Chad Schafer, Larry Wasserman (+ Tom Junk = Physicist)

Short talks:
- Jim Berger
- Richard Lockhart
- Tudor Manole
- Xiao-Li Meng
- Larry Wasserman

Hopefully in future, chance to have live meetings.

Future

Summary talks: Sara Algeri & Nick Wardle. **November 10th**

Interaction among different Particle Physics areas
e.g. LHC, ν, DM, Flavour

Feedback from Physicists and especially Statisticians

PHYSTAT Seminars and possibility of short (~90') meetings? e.g.
Estimating systematics;
More Statisticians;
Systematics in Cosmology;
Systematics in Combinations;
Parton Distribution Functions;
Interesting specific analyses;
Machine Learning;
etc.

Enjoy the meeting!

Olaf Behnke
Richard Lockhart
Louis Lyons
Nick Wardle
(emails on webpage)