PHYSTAT-Systematics 2021
Statistician's (re)view

Sara Algeri

School of Statistics, University of Minnesota

BIRS workshop on
Systematics in Particle Physics Data Analyses

Banff, Alberta, Canada
April 24, 2023.

PhyStat - Syst. '21 vs BIRS - Syst. '23

PhyStat - Systematics, 2021

"A remote workshop devoted to the way systematic uncertainties are incorporated in data analyses in Particle Physics."

	Physicists	Statisticians
Contributions	12+	12+

BIRS - Systematics, 2023

Similar purpose but hybrid format ⇒ we expect interesting discussions to arise online and offline.

	Physicists	Statisticians
Contributions	At least 16	At least 11

Sources of errors

Let θ be a quantity (parameter) of interest, what is its true value?

Sources of uncertainties affecting our answers...

$$\underbrace{\widehat{\theta}}_{\text{Our estimate}} + \underbrace{\sigma}_{\text{Statistical uncertainty}} + \underbrace{\tau}_{\text{Systematic uncertainty}} + \underbrace{\epsilon}_{\text{Irreducible error}}$$

- σ: We know what we don't know and we know how to deal with it.

  $\Rightarrow$ 2+ centuries of statistical theory can typically help with that.

- τ: We know what we don't know but we don't always know how to deal with it.

  $\Rightarrow$ That is why we are here.

- ϵ: We don't know what we don't know.

  E.g., variables not included in our model, "hidden" systematics which we simply do not know are there.

A (maybe too) simplified statistical formulation

$$\underbrace{\widehat{\theta}}_{\text{Our estimate}} + \underbrace{\sigma}_{\text{Statistical uncertainty}} + \underbrace{\tau}_{\text{Systematic uncertainty}} + \underbrace{\epsilon}_{\text{Irreducible error}}$$

- σ is often thought of by statisticians as error that dissipate as $n \to \infty$.
  $\Rightarrow$ Essentially sources of variance.
- τ is often thought of by statisticians as error that dissipate as $n \to \infty$.
  $\Rightarrow$ Essentially sources of bias.

But is it really that simple?

- In practice, statistical and systematic errors may be correlated.
- Systematics may (even if not always) decrease with $n \to \infty$.
- Often systematics are the uncertainties associated with our corrections for the bias.

Some more warnings from **Tom Junk** (cf. `https://indi.to/bk2G8`)

Does this separation of statistical and systematic uncertainty meet the needs of the community?

- Can we tell from it if a result is "systematics limited"? There could be a component with large stat. uncertainty and small systematics that will start contributing but only with a much larger dataset.
- Does this distinction make sense for combinations, as some systematics have a statistical component

Main sources of systematic uncertainties

A non-exhaustive list

- Monte Carlo uncertainties
- Mismodeling
  ⇑**Received substantial attention at PhyStat**
- Uncertainties on parameters estimated in previous studies
  ⇑**Received a lot of attention at PhyStat**
- Instrument calibration
- Uncertainties associated with machine learning solutions

Mismodeling

Uncertainties arise from:

- Approximations
- Choice of the functional forms for the signal model
- Background mismodeling ⇐**Received lot of attention at PhyStat**

From a statistical perspective...

We have bias in our model that cannot be reduced by simply changing the value of the parameters (the functional form is incorrect).

Solutions proposed...

...by physicists

Mainly **parametric** solutions

- Discrete profiling method Dauncey et al. 2015 - arXiv:1408.6865
- Spurious signal/safeguard method Priel et al. 2016 - arXiv:1610.02643
- Yellin's maximum gap method Yellin 2002 - arXiv:physics/0203002

...by statisticians

Essentially **non-parametric/semi-parametric** methods e.g.,

- Smooth models e.g., Algeri 2020 - arXiv:1906.06615
- Optimal transport Manole et al. 2022 - arXiv:2208.02807

Question 1

- Paraphrasing **Larry Wasserman** (cf. https://indi.to/6rSZd) "Should we compare several methods? Does using the difference between methods as a measure of systematic bias make sense?"
- If we were to do so, how would we account for the uncertainty on the difference itself?

Systematics as nuisance parameters

Physicists often deal with systematics by introducing additional nuisance parameters in the model to the extent that the two terms are often used interchangeably.

Great, but how do we deal with them?

- Hybrid Bayesian/Frequentist (e.g., Cousins and Highland, 1992 doi:10.1016/0168-9002(92)90794-5)
  - Frequentist for main measurement, Bayes for nuisance parameters.
- Marginalizing vs Profiling
  - Different experiments use different approaches
  - E.g., **Christopher Bronner**'s PHYSTAT talk (cf. https://indi.to/8JvPB) for marginalizing vs profiling in neutrino experiments.
- Pragmatic vs fully Bayesian (e.g., Xu et al., 2014 doi:10.1088/0004-637X/794/2/97)
  - Should we use the data in our current experiments to update the systematics? When possible, yes. If model is too complex "pragmatic Bayesian" can help substantially.

Question 2

Is it at all possible to reach a consensus on what to do when? And if not, is there any hope in comparing/combining results of studies adopting different approaches?

Combining the results of different studies

A substantial challenge

One needs to account for the correlation between systematic uncertainties across different analyses (**Sasha Glazov** `https://indi.to/QvSZ7`)

A big help in combining results from different experiments

Publishing likelihoods (**Kyle Cranmer** `https://indi.to/fXDSp`)

Question 3

Can a statistician effectively access them?

What is the ultimate goal here?

It is NOT "just an estimation problem"...

Let's keep in mind that more that estimating accurately our nuisance parameters, we want to make sure we incorporate their effect when

- discriminating signals from background
- testing of hypotheses/goodness of fit.

Therefore it is particularly important to assess how (and which) nuisance parameters/systematic effects impact on the result of our analysis.

The need of "regularity" to trust our asymptotics

As emphasized by **Alessandra Brazzale** (cf. `https://indi.to/M2gwT`)

Searching for new phenomena with profile likelihood ratio tests

Sara Algeri[1], Jelle Aalbers[2,3], Knut Dundas Morå[2,3,4] and Jan Conrad

Box 2 | Necessary conditions for Wilks' theorem

Asymptotic
Sufficient data are collected.

Interior
Only values of the parameters of interest μ and nuisance parameters θ that are not on the boundaries of their parameter space are admitted.

Identifiable
Different values of the parameters specify distinct models.

Nested
The null hypothesis H_0 is a limiting case of the general case hypothesis H_1, for example, with some parameter constrained to a subrange of the entire parameter space.

Correct
The true model is specified either under H_0 or under H_1.

Unfortunately, failure of these conditions is extremely common even in simple setups
E.g.,

$$(1-\eta)\underbrace{f(y,\phi)}_{\text{background}} + \underbrace{\eta}_{\text{signal relative intensity}}\underbrace{g(y,\overbrace{\theta}^{\text{signal location}})}_{\text{bump}}, \qquad 0 \leq \eta \leq 1$$

Question 4

Are these regularity conditions effectively checked in practice? How can this be done when dealing with complicated likelihoods?

How to keep the conversation up?
(...and ideally involve more statisticians along the way)

Paraphrasing **Richard Lockhart** (cf. `https://indi.to/WgH2F`)

> "Statisticians need to see abstraction at the level of mathematics to be confident that they are given a valuable contribution."

Some considerations (based on personal experience):

- Often, the statistical issues arising in particle physics translates into fundamental problems in statistics. Which means that the statistical theory to be developed/studied is already pretty complicated on its own.
- When feasible, formulating the problem using simple toy models (and which can be generalized to more realistic scenarios), can be of great help.
- When feasible, providing "realistic" synthetic data (e.g., data challenges) can also be of great help.

Question 5

Realistically, is this enough to "bridge" the two communities? What else can be done? (While keeping in mind that our students/postdocs will still need to satisfy certain criteria to be competitive on the job market)

To summarize...

Some possible points of discussion simulated by Phystat-Systematics...

Q1- In the context of background mismodelling, can the difference between methods be used to acquire some notion/measure systematic bias?

Q2- When dealing with nuisance parameters is it at all possible to reach a consensus on what to do when? (e.g., marginalizing or profiling)

Q3- Can a statistician effectively access published likelihoods?

Q4- How to check the validity of regularity conditions needed by classical statistics when dealing with complex models?

Q5- What do we need to robustly bridge the statistics and physics communities?

Thank you all for your time and for accepting our invitation!