Discussion Questions for Day 2

We like to combine data! And produce simple results based on joint interpretation of many data samples (or just bins). Data reduction is a name of what we do.

Doing so increases our discovery sensitivity, so it maximizes the value of our experiment.

But there is tension in our results.

Good, Bad, and Ugly systematic errors.

Supposed to be covered by Good and Bad systematic uncertainties.

Cross-Checks in Data Subsamples

Do not improve discovery sensitivity

Give confidence that the model(s) are predicting the data adequately.

See an excess of events in a specific energy range? Look in all the corners of the detector!

See something weird in the data – does it persist in other kinematic bins, run periods, etc.

Detector triggers almost always have this built in to them.

Coincidence triggers require consistency to reduce backgrounds. (Andrey Sheshukov's talk)
  Digression (not really) Trigger Bias --> you will only see events you seek.

Combination analyses are more like OR triggers instead of AND triggers.

Data Subsamples

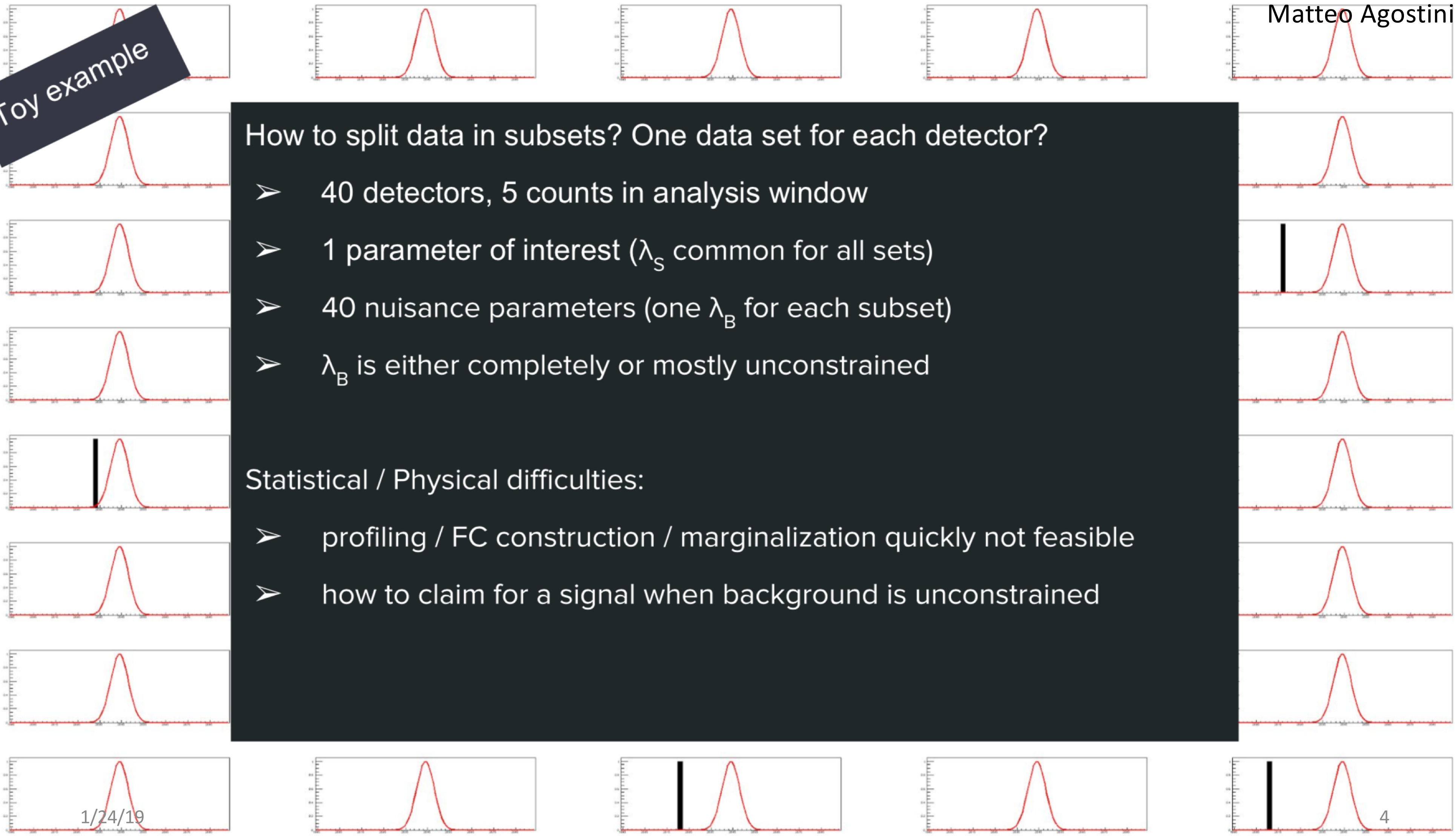
Smaller samples have larger statistical uncertainties and may have less power to test the effect seen in the aggregate.

Dividing data into subsamples and combining them back is a test of consistency of the combination procedures. But assumes that all subsamples have the same properties (e.g. time periods) Built into our ability to make histograms of long runs.

Relaxing assumptions of correlated modeling in each subsample hits sensitivity: smaller control samples and no ability to combine those.

Making a detector out of multiple pieces with different technologies invites more nuisance parameters. Extrapolatability of knowledge from one to the others.

Toy example



How to split data in subsets? One data set for each detector?

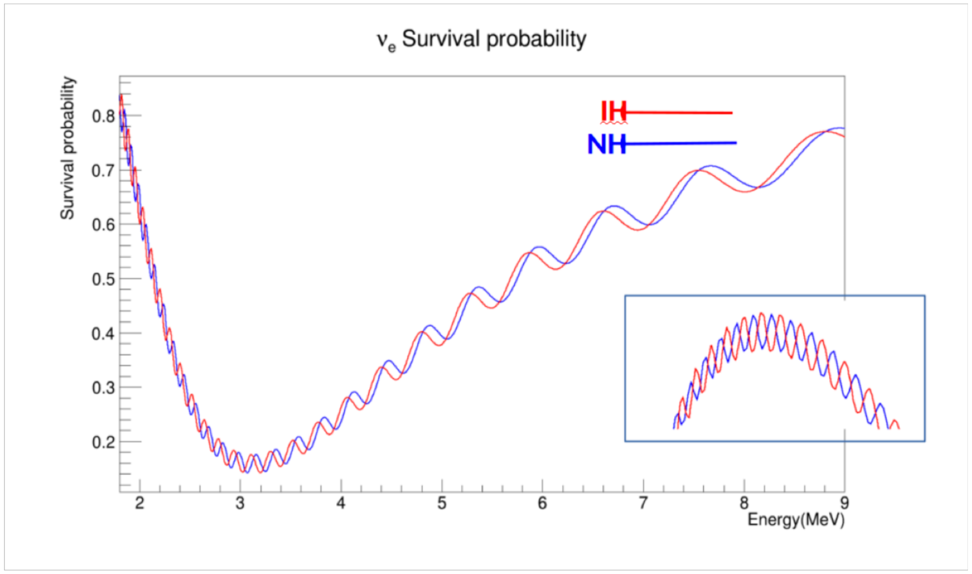
- 40 detectors, 5 counts in analysis window
- 1 parameter of interest (λ_S common for all sets)
- 40 nuisance parameters (one λ_B for each subset)
- λ_B is either completely or mostly unconstrained

Statistical / Physical difficulties:

- profiling / FC construction / marginalization quickly not feasible
- how to claim for a signal when background is unconstrained

Reducing Data to a Smaller Number of Quantities to Improve Sensitivity

or reduce sensitivity to systematics



F. Sawy

F^{NH}= sum of data in Ereco bins where NH is predicted to have more events

F^{IH} = sum of data in Ereco bins where IH is predicted to have more events

Distills NH/IH separation information into two numbers
May reduce sensitivity to some systematics, but need external E scale, resolution, and delta msquared

But the entire spectrum contains more information
-- test of the energy resolution
The F's aren't expected a priori to be the same

Dimensional reduction may make computations quicker

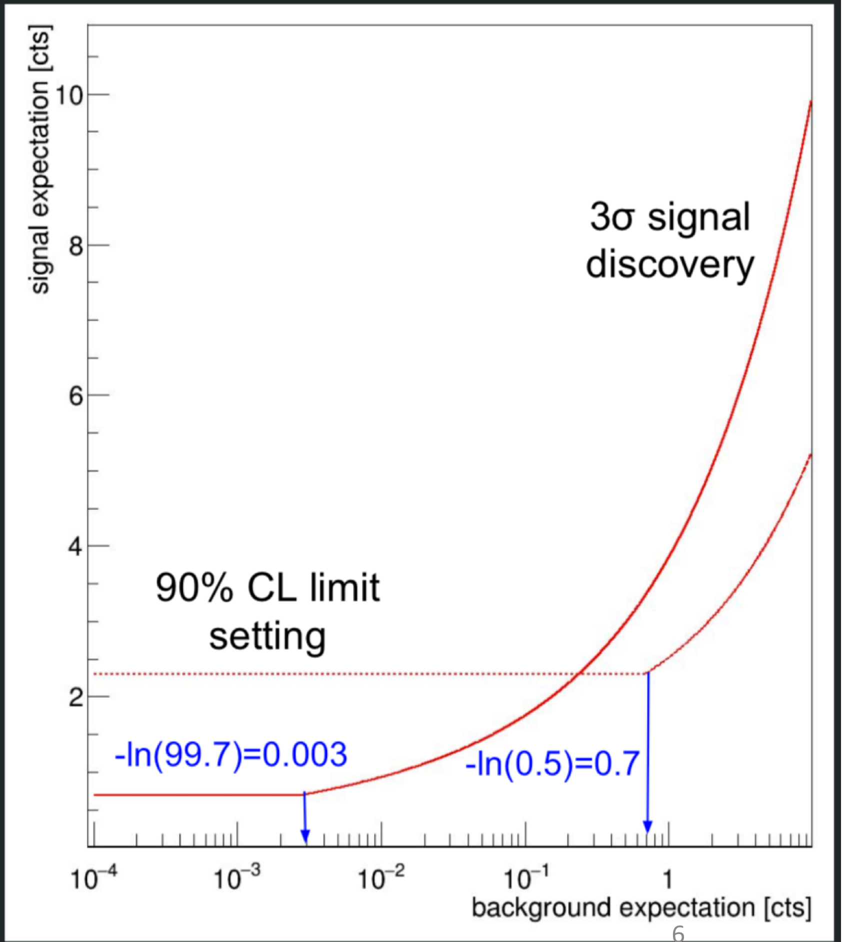
Counting Experiment Sensitivity

Simple calculation for signal + known background:

- ➢ based on distributions of expected frequency of observations
- ➢ Poisson CDF approximated with gamma functions

Background free regime:

- ➢ **limit setting sensitivity:** flattens for a background expectation of .7 cts (median of H_0 is at 0)
- ➢ **signal discovery sensitivity:** flattens for a background expectation of 0.003 (99.7% quantile of H_0 is at 0)
- ➢ 1 count is a discovery for low enough background



[**M.A.**, G Benato and J A Detwiler, PRD 96, 053001 (2017)]

Modified Likelihoods

Anthony Davison: I'm no expert on these, but do have some figures of merit.

We care about discovery sensitivity, and expected upper limits on new effects.
Also are figures of merit for choosing detector design, reconstruction, analysis cuts,
and how to package our results for downstream consumption. But downstream consumers may have other things in mind. Use your data for things you had no intention of testing.

Methods must have coverage (frequentist) or credibility (Bayesian)

Neutrino example: Both NH and IH possibilities fit into mass sum < 120 meV.

But with the delta msquareds used (75 meV^2 and 2524 meV^2), can exclude IH if mass sum < 100 meV. Why is there any sensitivity at 120 meV? Some shape assumed for the likelihood of the inputs? More important than the technique.

Look-Elsewhere Effect Comments

Generally, no LEE for limits.

BUT: Taking the union of excluded parameter spaces is not well defined (breaks coverage. Points would have multiple chances to be excluded).

We overlay exclusion contours all the time.

LEE for p-values for sure. Bayes Factor? Question for Jim Berger

LEE depends on how many independent testable models. More dimensionality of parameters of interest not present on the null, the larger the LEE
Example: LEP MSSM Higgs search – multiple small excesses. Theorists found models to predict them all simultaneously, did not compute LEE.

LEE when most of the model space has been excluded already?

Not all dimensions are created equally. (delta msquared, sin^2 theta)

Look-Elsewhere Effect Comments

Gross and Vitells: Application of Ledbetter's method to bump-hunting in HEP
LEE depends on significance of effect! More models become distinguishable if the effect size is stronger.

Phil Litchfield – can we generalize Gross and Vitells to non-bump-hunting cases.
E.g. neutrino oscillation spectra?

Comment from the audience: Feldman-Cousins already has LEE built in to it.

Comments on FC: Results make sense when the model space is complete – contains the true values.

- It is impossible to exclude the entire space under study, even if it is all wrong.
- You may be able to "discover" a model point by ruling out all the others, even if it is not true.
- E.g. Karagiorigi: Do we live in a 3+1 world, or 3+2? 3+0?

Unfolding:

M. Kuusela, S. Schmitt, P. Rodrigues

When is it appropriate?

- Who are the consumers of unfolded data and what are they going to do with it?
- Some unfolding techniques may have good coverage and other properties, but bias the measurement of a parameter of interest. E.g. width of a peak
- Do we need to provide guidance of what an unfolded spectrum is good for and not?
- Is unfolding Asimov data representative?

- What do we publish, and provide electronically?
- Original data and smearing matrix, for example.
- Giving consumers too many options – can get multiple results.

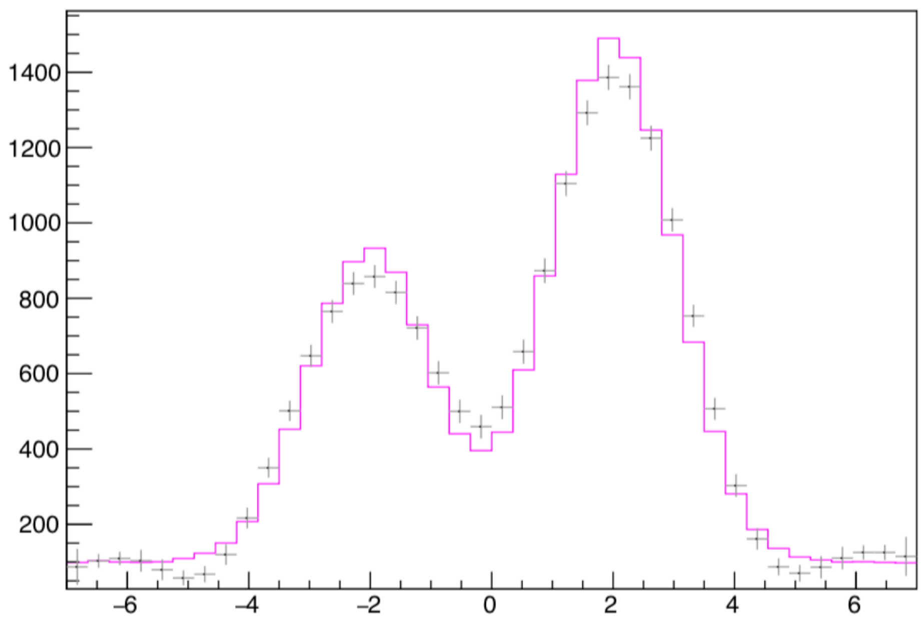


Figure: L-curve, $\tau = \sqrt{\delta}$=0.01186

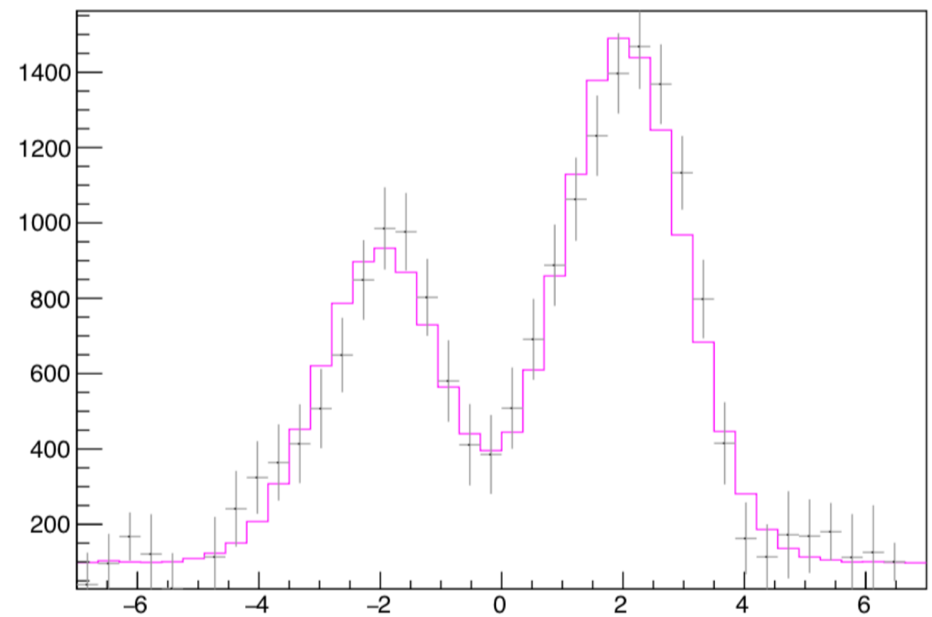


Figure: Undersmoothing, $\tau = \sqrt{\delta}$=0.00177

Some outstanding questions

P. Rodrigues

- Can we unfold and quantify bias s.t. it's small enough to not matter?
- How can we adapt unfolding techniques from the literature to work with the multi-universe/multisim method used by Minerva and MicroBooNE?
- Evaluating unfolding bias by comparing to a model "warped like the data": what's the range of validity?
- How do we assign systematic uncertainties to "the unfolding technique" without double-counting?
- How do we deal with PPP? Or, what's the best way to preserve the features of our detailed systematic error estimates in a way that's digestible to users (theorists, other experiments)?