Adding uncertainties

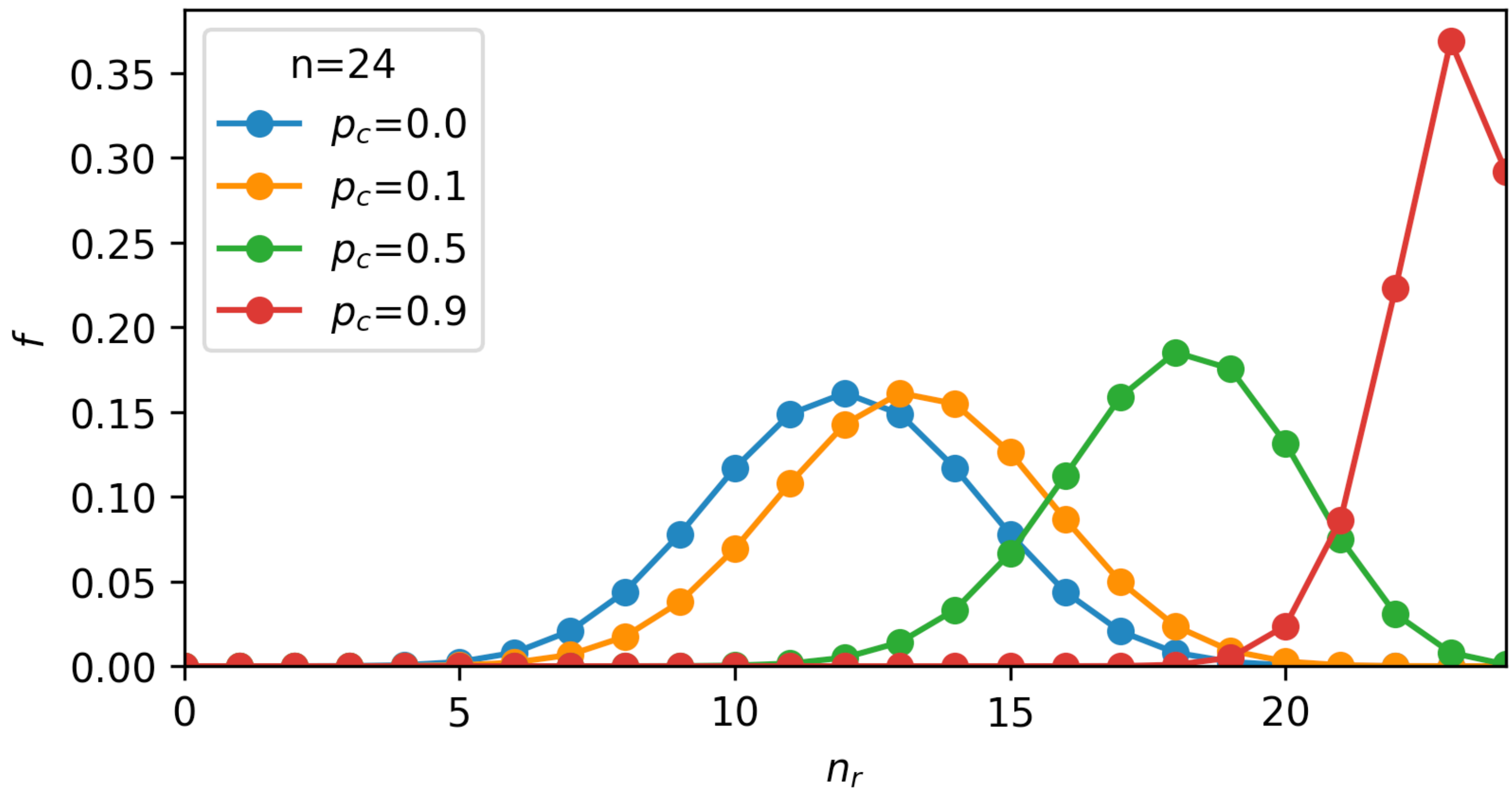
Topics: dealing with systematics

- Statistical model with auxiliary measurements
- Profiling and marginalizing nuisance parameters
- Tools and techniques for building models using simulation

Fermilab

Coming back to P(cheat)

- Visualizing $f(n_r; n, p_c) = \sum_{n_t=0}^{n} f_{Bi}(n_t; n, 1/2) f_{Bi}(n_r + n_t - n; n_t, p_c)$
- What if we observed $n_r = 6$?



Coming back to P(cheat)

- Re-introduce p_t: $f_c(n_r; n, p_t, p_c) = \sum_{n_t=0}^{n} f_{Bi}(n_t; n, p_t) f_{Bi}(n_r + n_t - n; n_t, p_c)$
- We could ask everyone to flip the same fair coin
  - If tails (T), raise your hand (✋); if heads, don't raise
  - The sampling distribution for $n_{r'}$ raised hands is $f_{Bi}(n_{r'}, n, p_t)$
- This is a *control region* for p_t, with *auxiliary measurement* $n_{r'}$

Fermilab

Full statistical model

- Split likelihood parameters into *parameters of interest* (POIs) θ and nuisance parameters ν, and define *auxiliary measurements* y that target the latter
  - Then the joint pdf factorizes $P(x, y; \theta, \nu) = P(x; \theta, \nu)P(y; \nu)$
  - N.B. *y* are *global observables* in RooFit

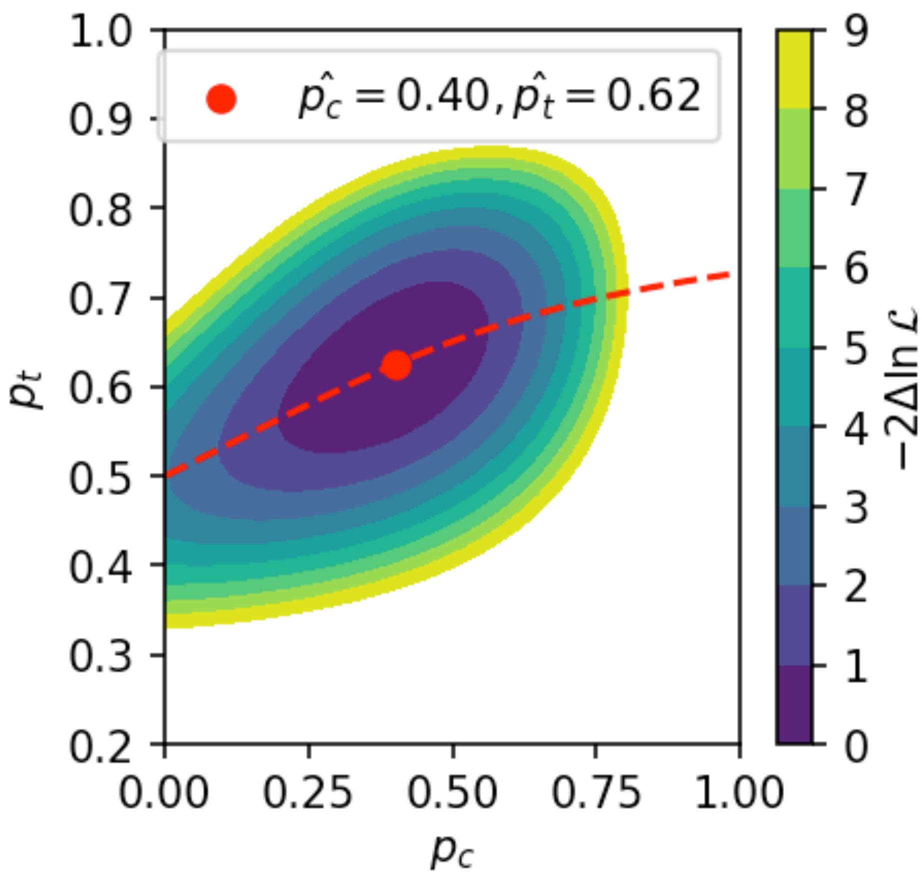
- To frequentists, $P(x, y; \theta, \nu)$ can be used as a likelihood in (θ, ν)
  - *Profile* away ν-dependence: $\mathscr{L}_p(\theta; x, y) = \max_\nu \mathscr{L}(\theta, \nu)$
  - With $\mathscr{L}_p$ we can do all of what was shown before (in approximation)

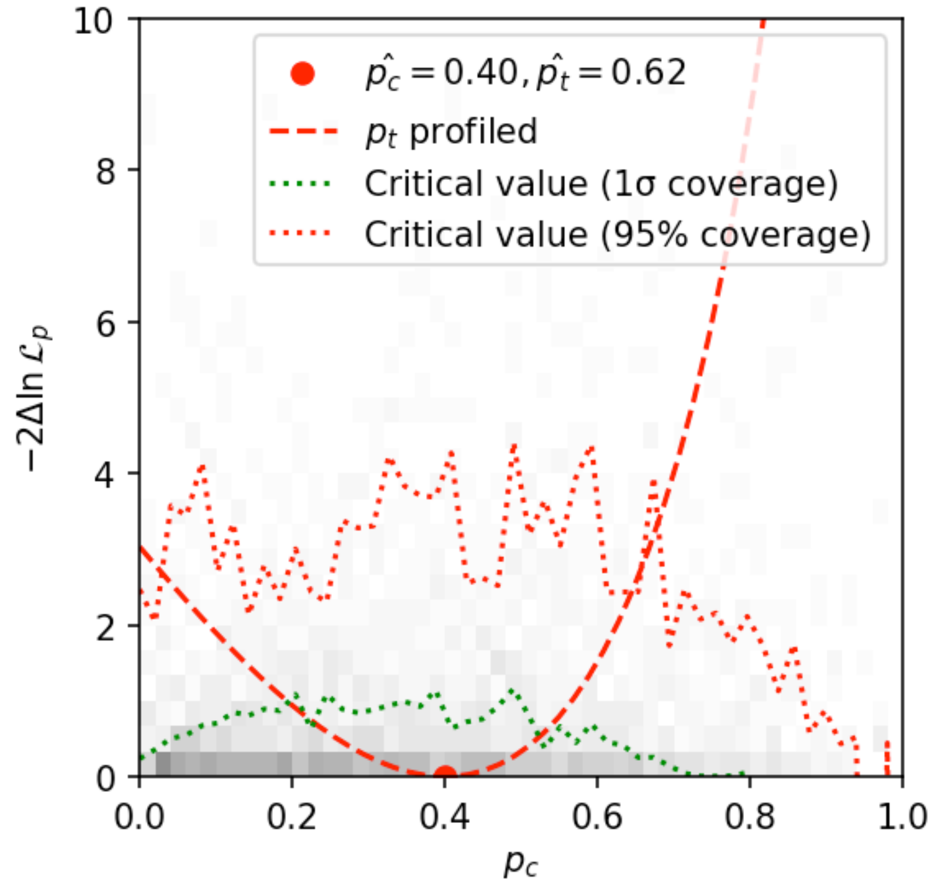
- Bayesians can insert a *ur-prior* $\pi(\nu)$ and use Bayes' theorem to get $P(\nu \,|\, y)$
  - Then *marginalize* out ν: $P(x \,|\, \theta) = \int P(x \,|\, \theta, \nu)P(\nu \,|\, y)\, d\nu = \int P(x \,|\, \theta, \nu)\frac{P(y \,|\, \nu)\pi(\nu)}{P(y)}d\nu$
  - Proceed as before with $P(x \,|\, \theta)$

- Renewed interest in publishing full statistical models: arxiv:2109.04981
  - Enables recasting in either language

Profiling P(tails)

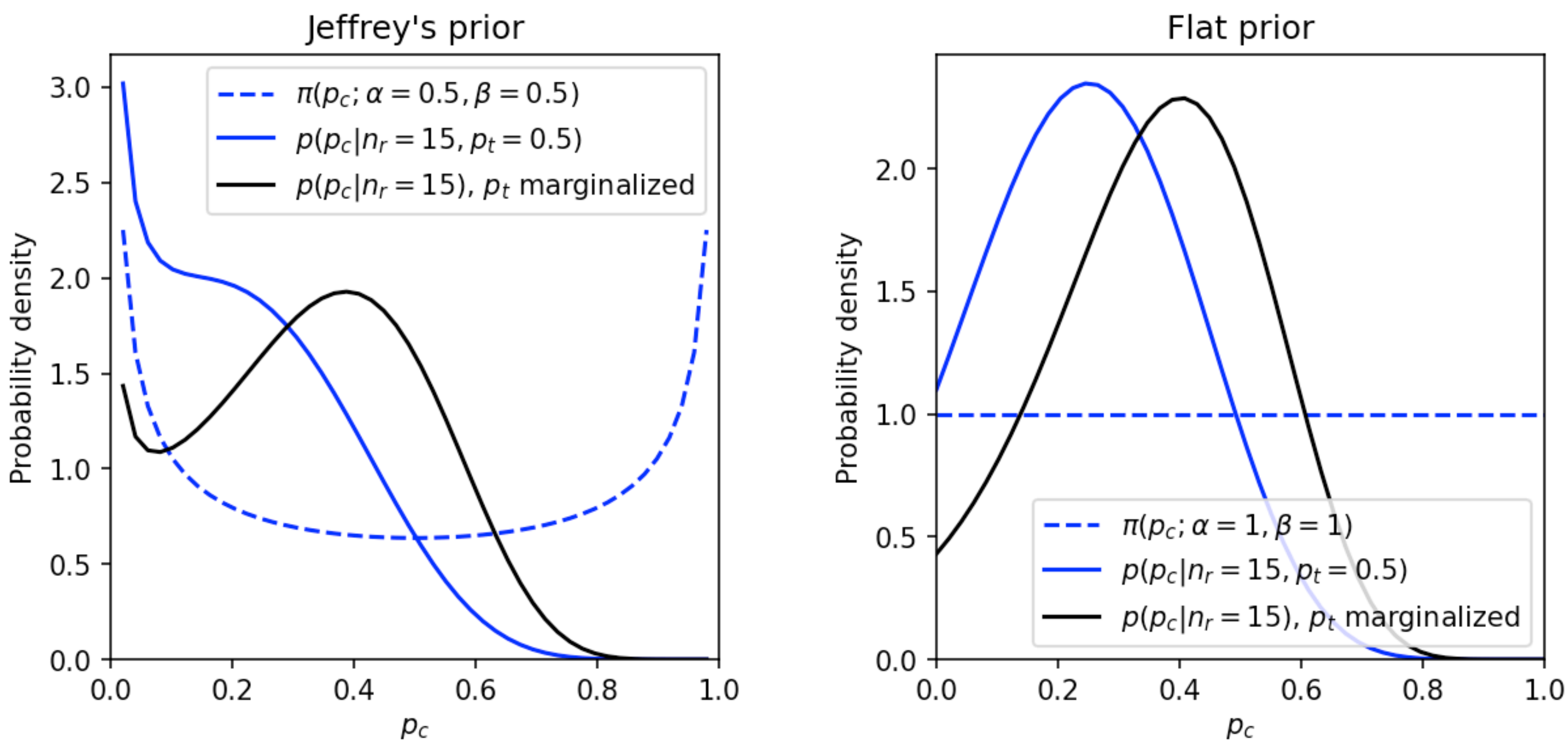
- If we observe $n_r = 15, n_{r'} = 15$, we get the following result:
- Building toys for given p_c now requires p_t assumption to sample $(n_r, n_{r'})$
  - A-priori vs. a-posteriori: our initial parameter guess (0.5) vs. best-fit (0.62)





Marginalizing P(tails)

- If we observe $n_r = 15$, $n_{r'} = 15$, we get the following result:
  - Using flat prior $f_{Beta}(p_t; 1,1)$, get a posterior $f(p_t \mid n_{r'}) = f_{Beta}(p_t; 1 + n_{r'}, 1 + n - n_{r'})$
  - Marginalize $f(n_r \mid p_c) = \int f(n_r \mid p_c, p_t) f(p_t \mid n_{r'}) dp_t$ and proceed as before



Fermilab

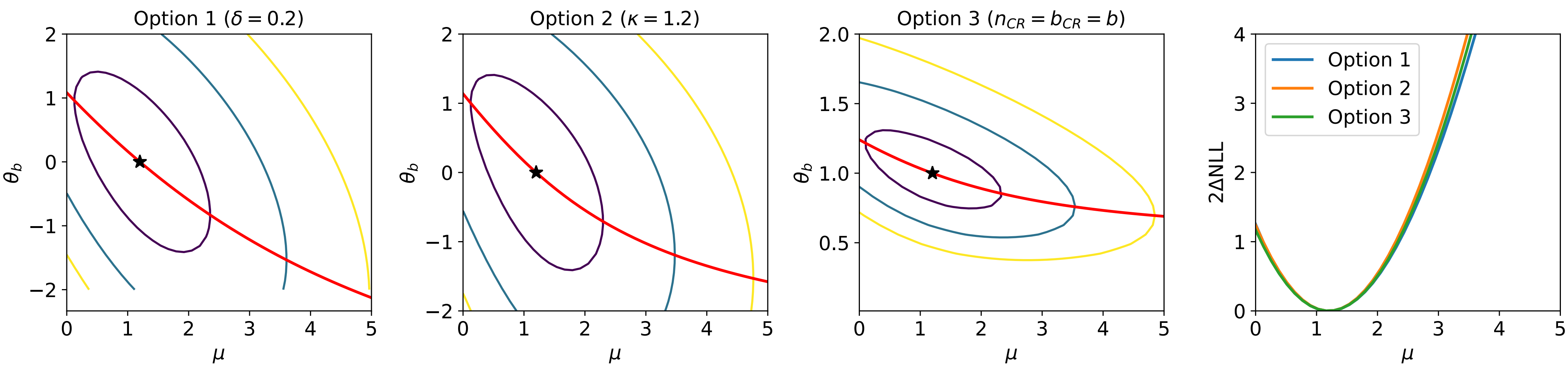
An uncertain background

- Adding some background uncertainty to our counting model
  - $f_P(n; \mu s + b)$ for some fixed s, b, and a varying signal strength μ

- Option 1:
  - $f(n, \nu_{b0}; \mu, \nu_b) = f_P(n; \mu s + (1 + \delta\nu_b)b) f_N(\nu_{b0}; \nu_b, 1)$
  - (for some δ close to 0) Not great: b can go negative
- Option 2:
  - $f(n, \nu_{b0}; \mu, \nu_b) = f_P(n; \mu s + b\kappa^{\nu_b}) f_N(\nu_{b0}; \nu_b, 1)$
  - (for some κ close to 1) Better: log-normal
- Option 3:
  - $f(n, n_{cr}; \mu, \nu_b) = f_P(n; \mu s + \nu_b b) f_P(n_{cr}; \nu_b b_{cr})$
  - Best, if such a background-pure control region can be constructed
- …and many more
  - In all cases we now have a new observable (ν_{b0}, n_{cr}), a new nuisance parameter ν_b, and several new constants (δ, κ, b_{cr}) to compute (e.g. from simulation)

Profiling example

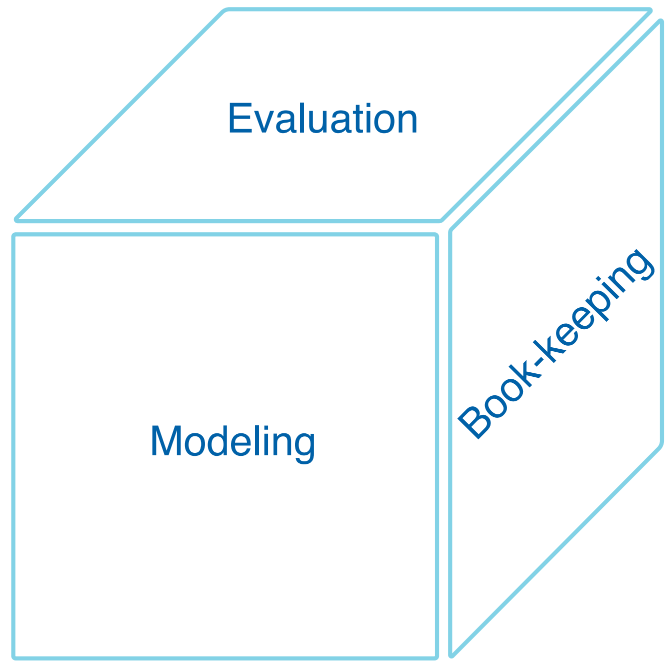
- Profile likelihood for the single-bin background uncertainty example
  - Option 1: Normal-distributed auxiliary constraint
  - Option 2: Log-normal auxiliary constraint
  - Option 3: Poisson-distributed auxiliary control region
  - s=10, b=25, x=37



Typical tasks in building a model

- Enumerate effects to get dimension of ν
  - Don't forget anything! Unknown unknowns?
- Find good auxiliary measurements y
- Choose a parameterization $f_\phi(y;\nu)$
  - e.g. the options 1-3 from before
- Evaluate the constants ϕ
  - In practice: interpolate between shifted or weighted MC

- Iterate
  - Compromise: fidelity/computability/practicality
    - Prune low-impact effects
  - Initial model might not fit observed y well



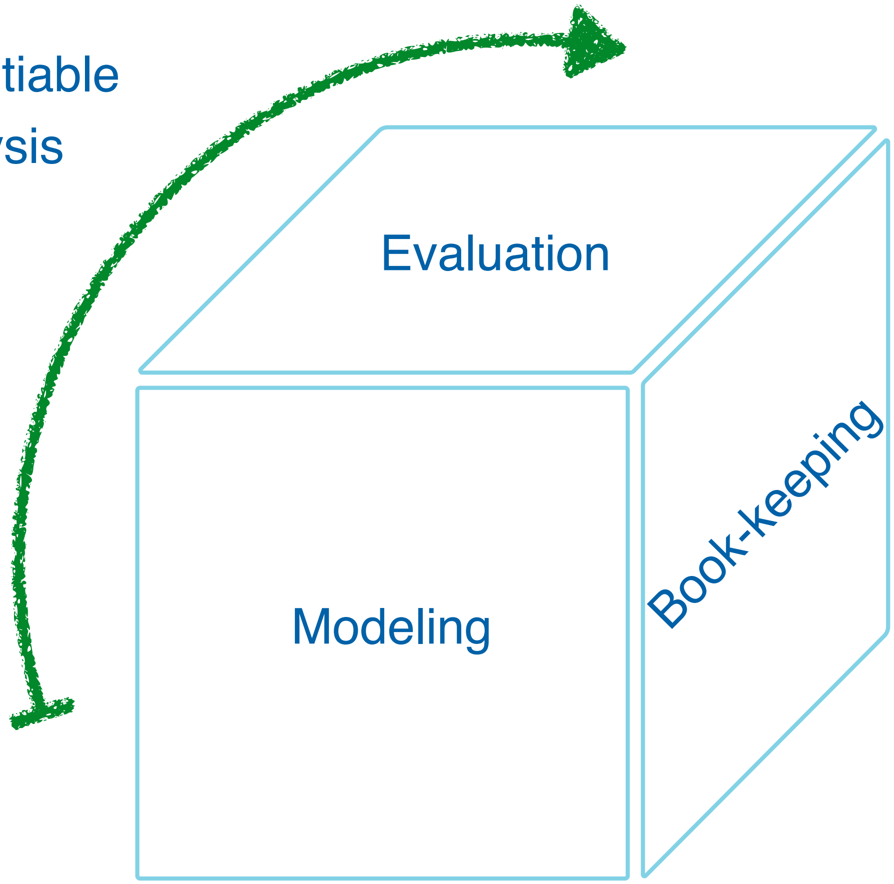
Relevant topics in systematics

Analysis task graph

Reduced data formats

On-the-fly evaluation

Differentiable analysis



Systematic-aware optimization

Metadata systems

Modeling techniques

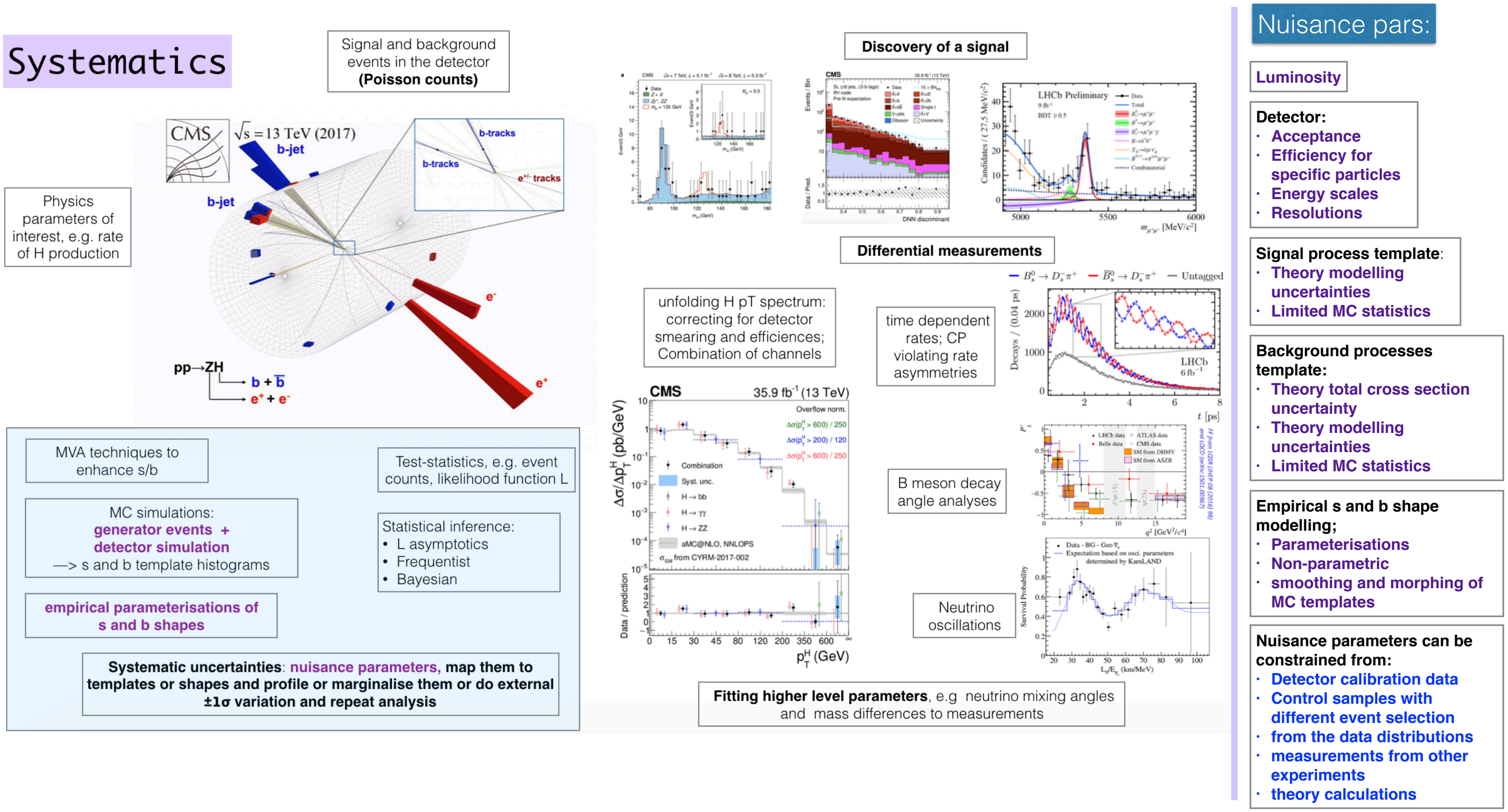
Experimental prescriptions

Publishing statistical models

Template storage

Modeling techniques

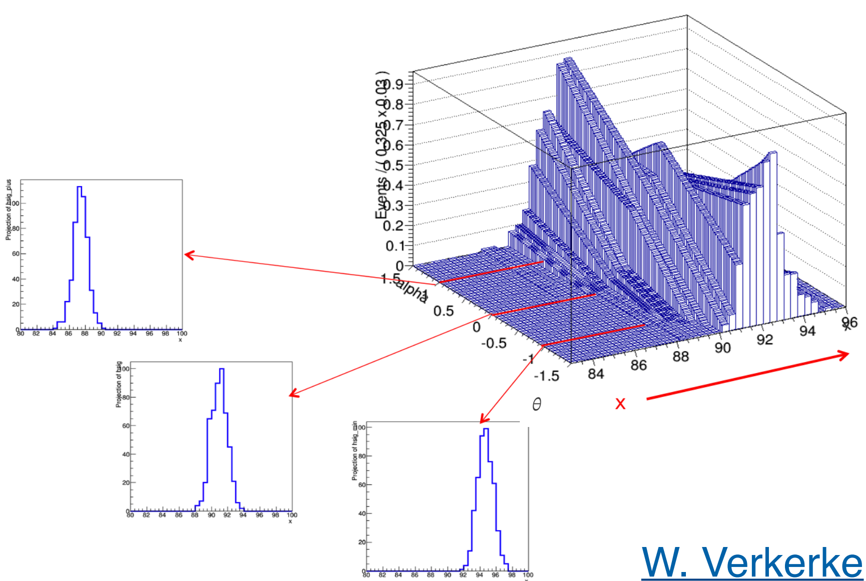
- Can be a whole workshop
  - Was: [PHYSTAT-Systematics 2021]
  - Excellent presentations covering a wide range of techniques
  - A one-slide overview was [produced] (more on types than techniques)



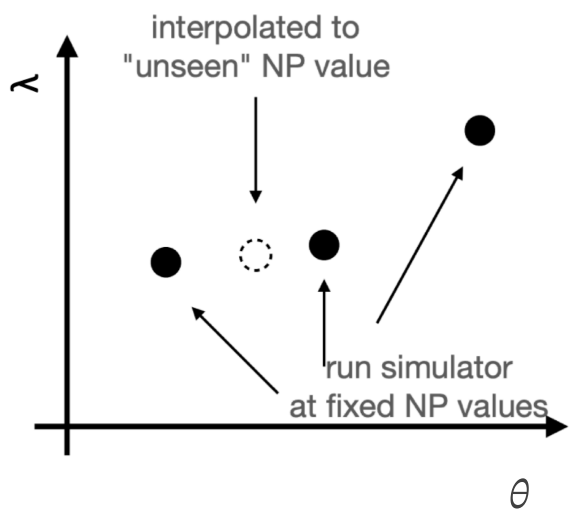
Modeling techniques

- **Rich** set of interpolation/extrapolation techniques at end-stage
  - Morphing: vertical, horizontal, moment; splines; gaussian process; asymmetric shift interpolation; additive/multiplicative effects; MC stat uncertainty, [BB-lite](); …
  - i.e. what is done in [RooFit]()/[pyhf]()/[zfit]()/[iMinuit]()/[combine]()/etc.
    - What features do each of these tools offer? Nobody has it all!

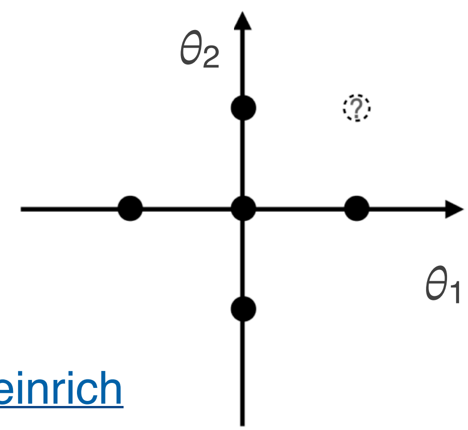
Visualization of bin-by-bin linear interpolation of distribution



W. Verkerke



Combining effects



L. Heinrich

Modeling techniques

- Simpler taxonomy of techniques to get inputs to fitting tools?
  - This is the dominant analysis-stage computation expense (process billions of events)
- Posit three basic techniques for binned fits
  - Just need functions w(x,θ) and Δ(x,θ)

$$\int_{\text{bin}} P(x)\mathrm{d}x \approx \frac{\sigma}{N} \sum_{x_i \sim P(x)}^{N} 1(x_i \in \text{bin})$$

(nominal)

$$\int_{\text{bin}} P(x|\theta=\theta_1)\mathrm{d}x \approx \frac{\sigma}{N} \sum_{x_i \sim P(x|\theta=\theta_1)}^{N} 1(x_i \in \text{bin})$$

(alternative sample, e.g. 2-point)

$$\int_{\text{bin}} P(x|\theta=\theta_1)\mathrm{d}x \approx \frac{\sigma}{N} \sum_{x_i \sim P(x)}^{N} w(x_i, \theta=\theta_1) 1(x_i \in \text{bin})$$

(reweight, e.g. efficiency)

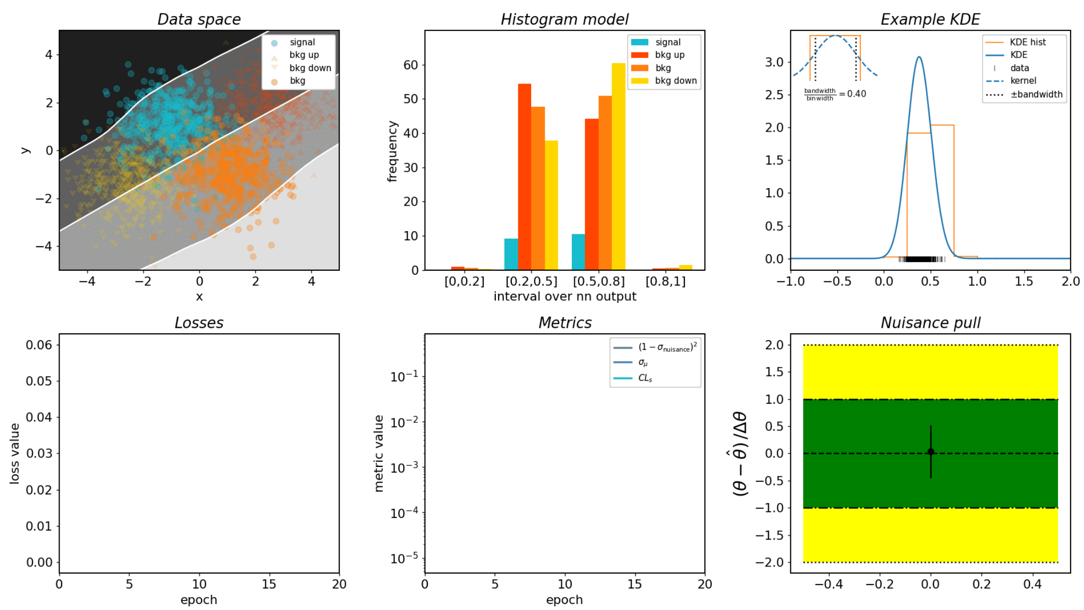
$$\int_{\text{bin}} P(x|\theta=\theta_1)\mathrm{d}x \approx \frac{\sigma}{N} \sum_{x_i \sim P(x)}^{N} 1(x_i + \Delta(x_i, \theta=\theta_1) \in \text{bin})$$

(shift, e.g. energy scale)

Systematic-aware optimization

- Analysis design and optimization often involves ML these days
- Learn salient features, ignore features affected by nuisance params
- Dozens of proposals, see HEPML LivingReview sections:
  - Decorrelation methods allow for construction of control regions
  - Inference-aware: maximize sensitivity or exclusion power of POI in full likelihood model
    - Can be deployed in more “traditional” analyses for e.g. region/binning optimization
  - Domain adaptation: ensure marginalized observables are modeled well

neos: N. Simpson, L. Heinrich



Differentiable analysis

- Rather than θ up/down variation, compute value and gradient
  - Auto-diff vs. finite-diff performance
- Higher order derivatives? How analytic are these things?
  - Need second order to get asymmetric (and it probably does not extrapolate well :)

$$\lambda(\theta) = \frac{\sigma}{N} \sum_{x_i \sim P(x)}^{N} w(x_i, \theta) 1(x_i \in \text{bin})$$

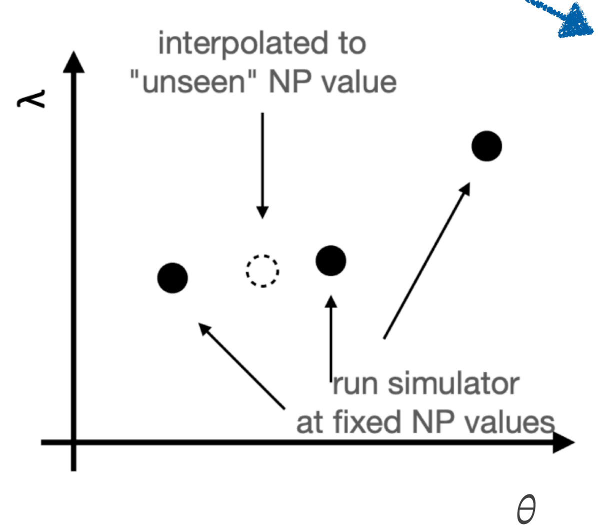
$$\approx \lambda(\theta_0) + \frac{d\lambda}{d\theta}\Big|_{\theta=\theta_0} (\theta - \theta_0) + \cdots$$

(reweight, e.g. efficiency) 👍

$$\lambda(\theta) = \frac{\sigma}{N} \sum_{x_i \sim P(x)}^{N} 1(x_i + \Delta(x_i, \theta) \in \text{bin})$$

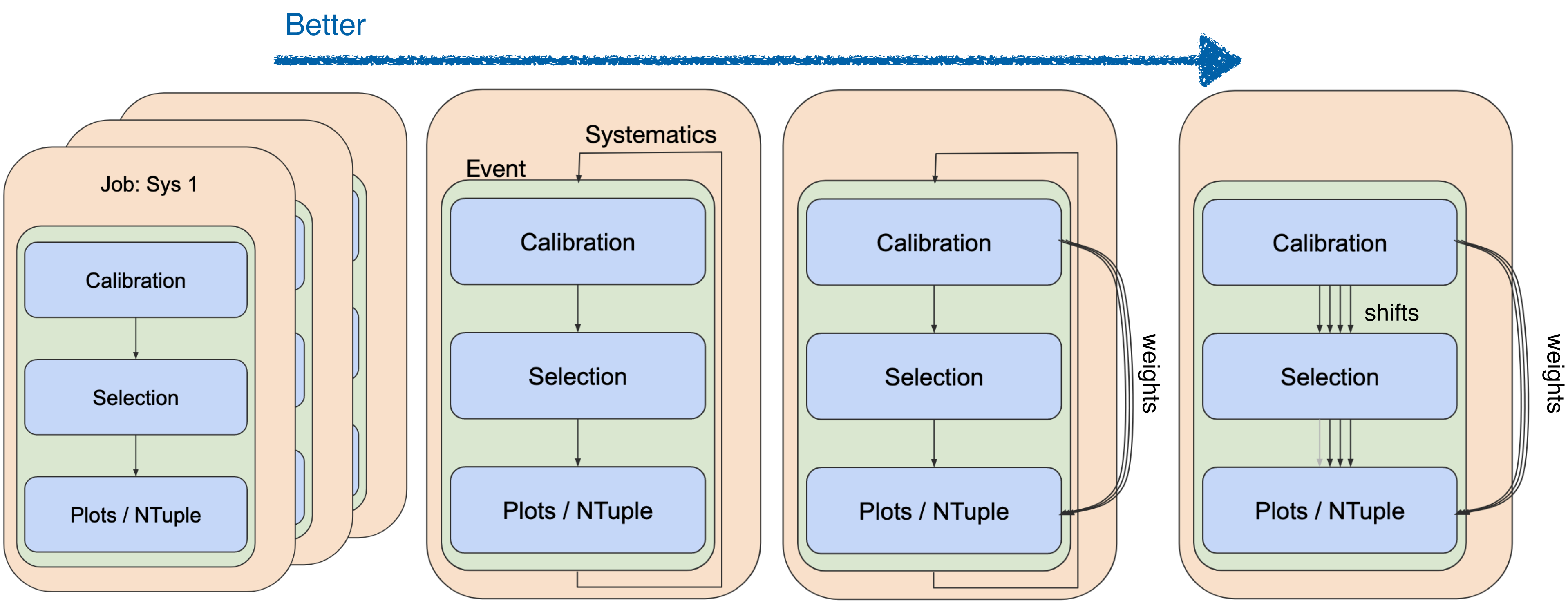
(shift, e.g. energy scale) 🤔

What happens here?



Analysis task graph

- Simplest solution: re-run everything with alternate ν
- Better: loop over event while in-memory (likely CPU cache)
  - Why? Because IO is very expensive
- Best: compute all weights, compute shifts only as necessary



S. Hageboeck

Reduced data formats

- Goal: maximize usability, minimize disk space
  - Keep minimal subset of observables x
- Tradeoff with functions $w(x,\theta)$ and $\Delta(x,\theta)$:
  - Large subset of x needed to evaluate: better to save output for θ_0, θ_1, θ_2
  - Small subset of x needed to evaluate: better to save those inputs, evaluate "on-the-fly"
  - Overlap with what is needed to identify the bin ➜ more likely on-the-fly
- CMS NanoAOD: calibrated objects, very few systematics
  - Keep only those too difficult to parameterize
    - Unclustered energy Δ for MET: per-PF candidate species energy scale uncertainty
  - ATLAS DAOD_PhysLite: similar goals
- Other considerations
  - CMS MiniAOD: lossy compression of track covariance matrices
  - Common weight trick: store 1-w with reduced-precision mantissa

On-the-fly evaluation

- Often calibrations and systematics go hand-in-hand
  - Redefine $f(y';\nu') = f(y + (y' - y); \nu' - y)$ so auxiliary measurement is "spot-on"
- In CMS, corrections+uncertainty have long been parameterized
  - Lately, move towards standardizing to reduce proliferation of (often poorly-designed) serialization formats and (often slow) evaluation frameworks
- Correctionlib
  - A well-structured JSON data format for a wide variety of ad-hoc correction factors encountered in a typical HEP analysis and a companion evaluation tool suitable for use in C++ and python programs.
  - Development started Nov. 2020, all CMS analysis-stage corrections now compatible
  - Presented at PyHEP '22: youtube

Fermilab

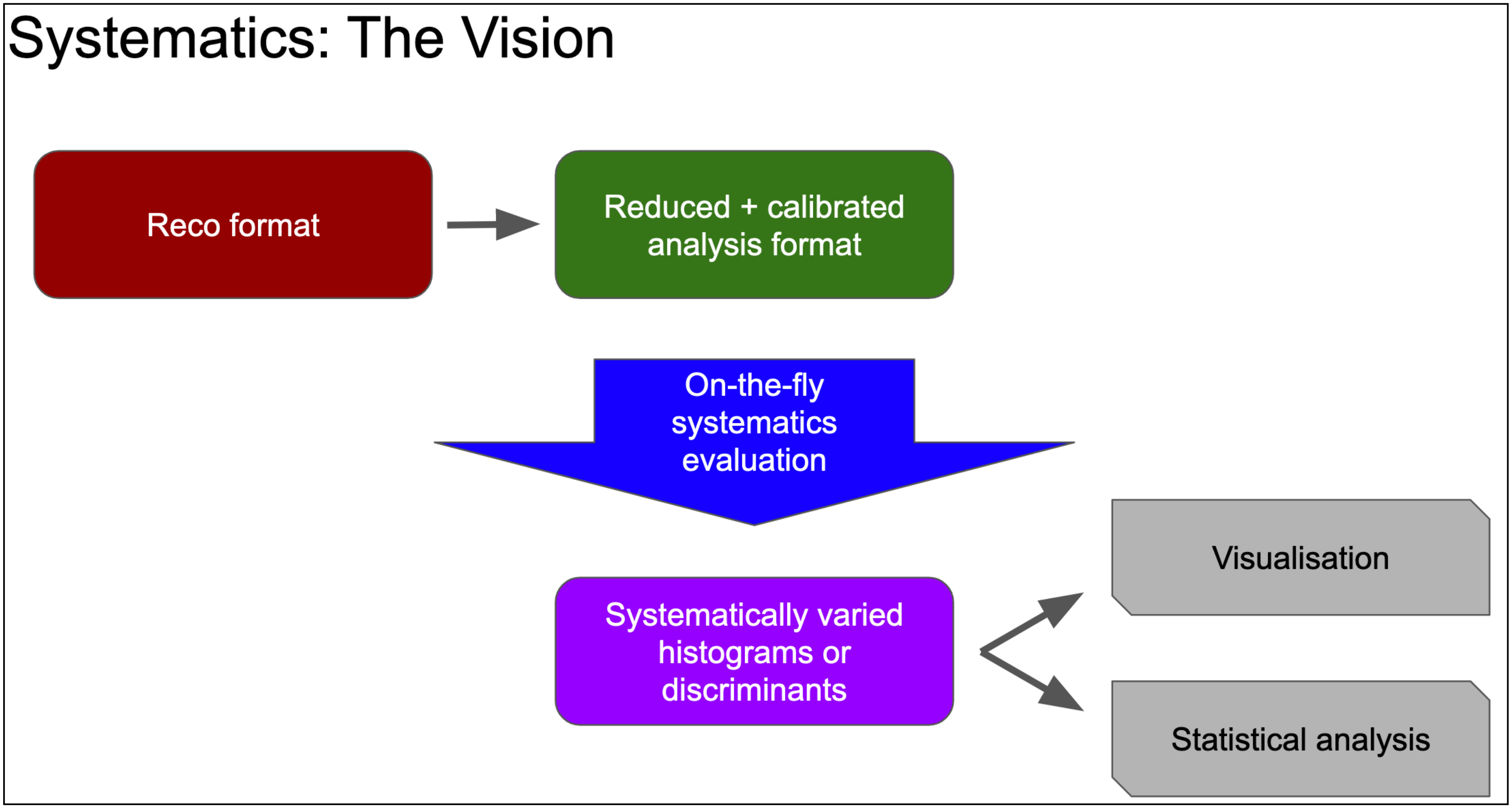
Metadata systems

- For reconstruction-level corrections *conditions* database is standard
  - Correctionlib json in database?
  - ML models in database?
- Book-keeping alternative samples
  - At least in CMS, no automated access to generation config at analysis stage
  - Most book-keeping by hand: key on dataset name
- Ongoing R&D here!

$$\int_{\text{bin}} P(x|\theta=\theta_1)\mathrm{d}x \approx \frac{\sigma}{N} \sum_{x_i \sim P(x|\theta=\theta_1)}^{N} 1(x_i \in \text{bin})$$

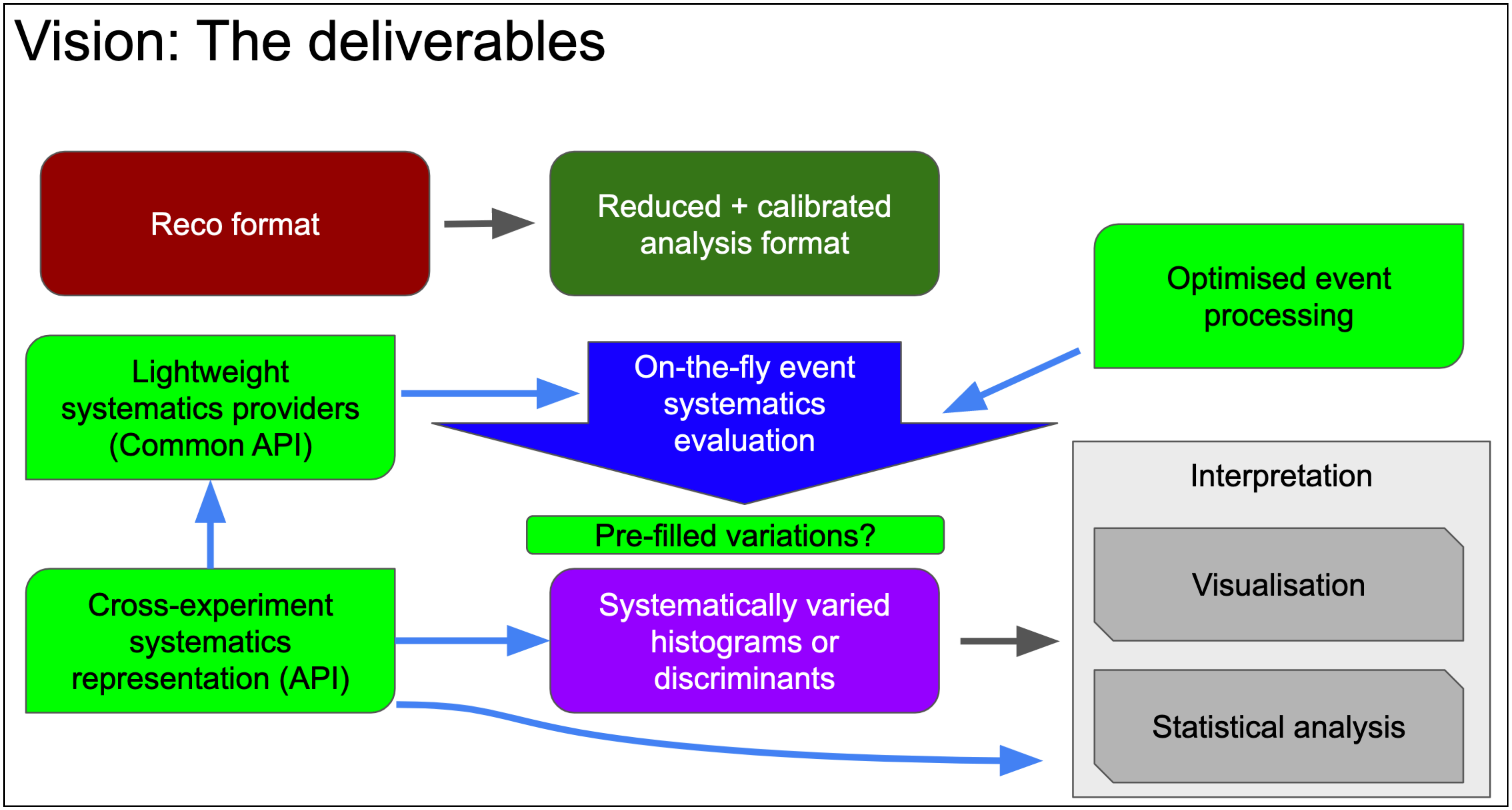
(alternative sample, e.g. 2-point)

A flowchart



P. Laycock, T.J. Khoo

A flowchart



P. Laycock, T.J. Khoo

Experimental prescriptions

- Non-trivial to agree on parameterization, but crucial for combinations
- Correlate (i.e. use same subset of θ for) common effects
  - Experimental effects (simplest: luminosity unc.)
  - Theory uncertainties for common processes
  - Etc.
  - Profit from increased sensitivity!
- CMS Higgs group: “datacards” (likelihood serialization format) are reviewed
  - Standard nuisances, naming conventions, sign, etc.
  - Simplifies combination later
- Is it worth establishing cross-experiment parameterization/nomenclature?

Template storage

- Multi-dimensional histograms: axis for systematic variation
- Filling histograms with weights vector
  - Save repeated bin lookup for same observables
  - Planned feature for boost::histogram boostorg/histogram#211
- Better to have serializable object tailored to our use case
  - RooDataFrame has part of the answer:

```
hx = ROOT.RDF.VariationsFor(nominal_hx)
hx["nominal"].Draw()          obtain all variations
hx["pt:down"].Draw("SAME")
```

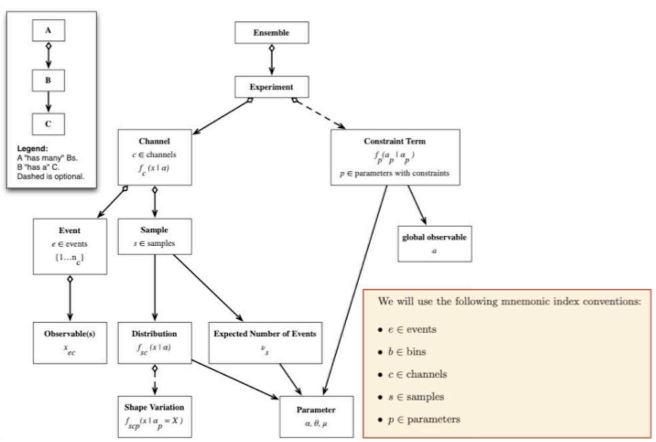
Template storage

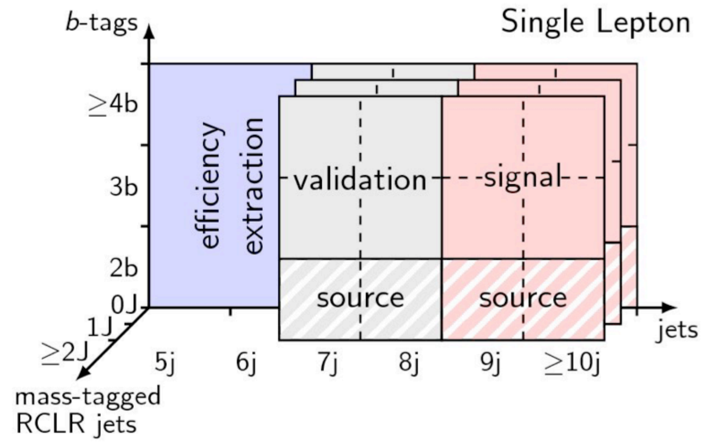
- [cabinetry](#) is a Python package to build and steer template fits

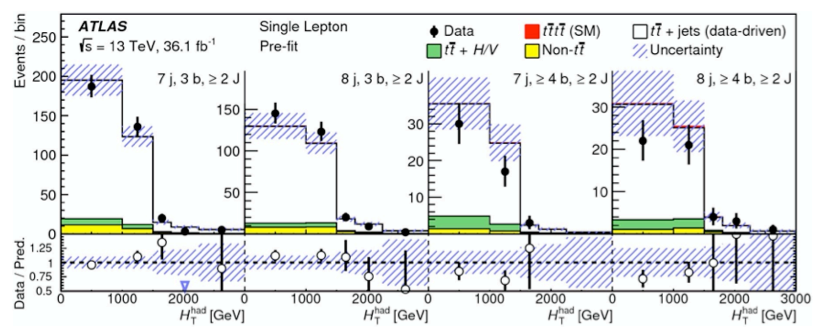
A point of convergence

Several aspects of Analysis Systems converge in a typical physics plot:

- Specification of signal / validation / control regions
- Specification of variables to be used for stat analysis
- Reduction to that format running on data and MC
- Management of MC samples, data driven backgrounds, etc.
- Management of systematic variations
- Feed reduced data (eg. histograms) into specification for statistical model / likelihood function
- Fitting & statistical tools
- Publishing results & derived data products
- Analysis preservation & gateways targeting reinterpretation







A. Held

Fermilab

Publishing statistical models

- More specifically the model $P(x, y; \theta, \nu) = P(x; \theta, \nu)P(y; \nu)$
- Why?
  - "The statistical models used to derive the results of experimental analyses are of incredible scientific value and are essential information for analysis preservation and reuse … [and] can enhance the short- and long-term impact of experimental results." (arxiv:2109.04981)
- A goal now 22 years old (K. Cranmer)
- Need a good data format, contenders:
  - pyhf JSON (HistFactory XML)
  - CMS combine datacard
  - RooWorkspace json: HS3
    - "A round-trip-capable, human-readable declarative format for statistical models was missing"

The end

Hopefully you have some idea now what this means

"An observed (expected) upper limit is placed on the signal strength μ, using the profile likelihood ratio test statistic, following the CL_s criterion, under asymptotic assumptions, and found to be …"

Scrap bin:

- Look-elsewhere effect
- Goodness of fit

Additional references

- Procedure for LHC Higgs combination http://cdsweb.cern.ch/record/1379837
- R. Cousins, Statistics in Theory https://arxiv.org/abs/1807.05996
- Asymptotic formulae for likelihood-based tests "CCGV" https://arxiv.org/abs/1007.1727
- Publishing statistical models https://arxiv.org/abs/2109.04981