Simulation Based Inference

Gregory Ottino
290E Particle Physics Seminar
April 20th, 2022

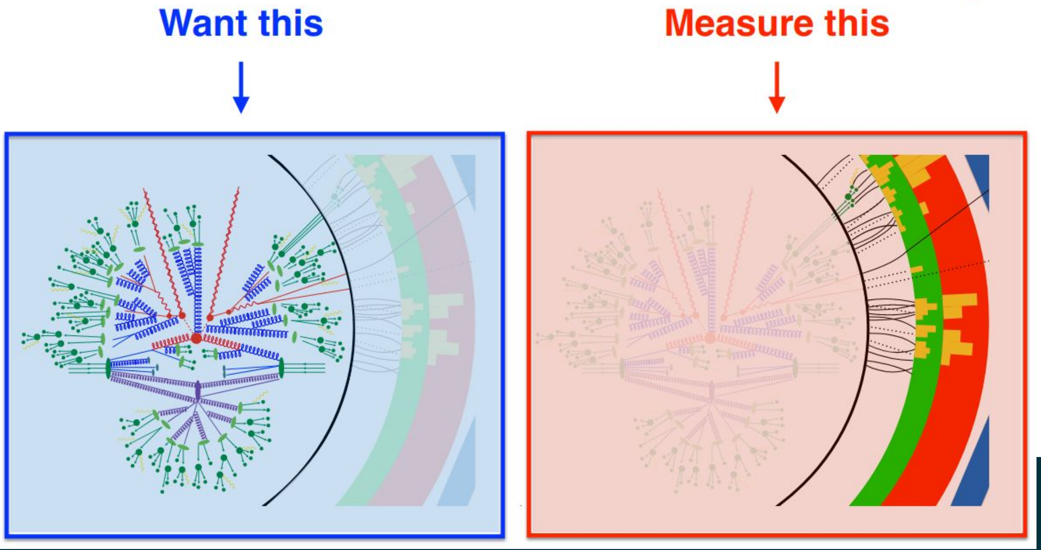
Overview

- Review of standard inference in particle physics
- Beyond simple summary statistics
- Learning the likelihood (ratio)
- Integration and augmentation
- Back to summary statistics
- Systematics
- Conclusion

Inspired by a shorter introduction in Ben Nachman's talk

What is experimental particle physics?

- What we want:
  - Fundamental physics parameters, generally expressed through a quantum field theory
  - Particle masses, coupling strengths
- What we have:
  - Signals in a detector eg hits in a tracker or calorimeter



Dr. Nachman's 290E talk

Full likelihoods in physics analysis

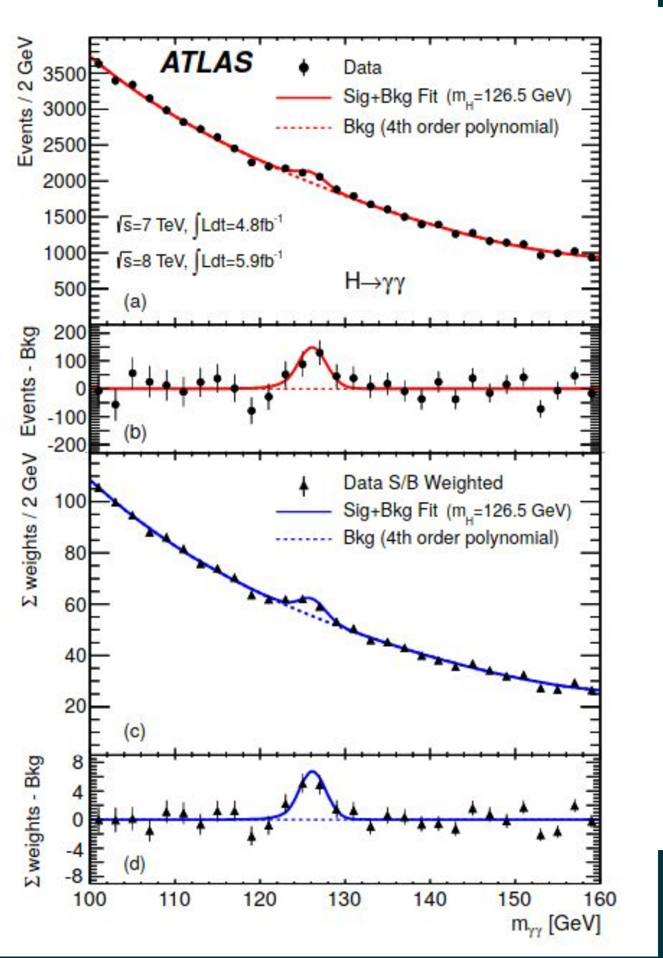
- Particle physics is a strongly predictive model
  - Can in principle compare data to theory in a statistical way
    - Ideally through a likelihood (ratio)
- What makes this challenging?
  - Matrix element -> Hadronization -> Detector interaction -> Readout
    - Factorizable components, but each nontrivial
  - Many final state objects at many energy scales
  - Full likelihood is practically not feasible to compute

$$p_{\text{full}}(\mathcal{D}|\theta) = \text{Pois}(n|\epsilon\, L\, \sigma(\theta)) \prod_i p(x_i|\theta)$$

Summary statistics

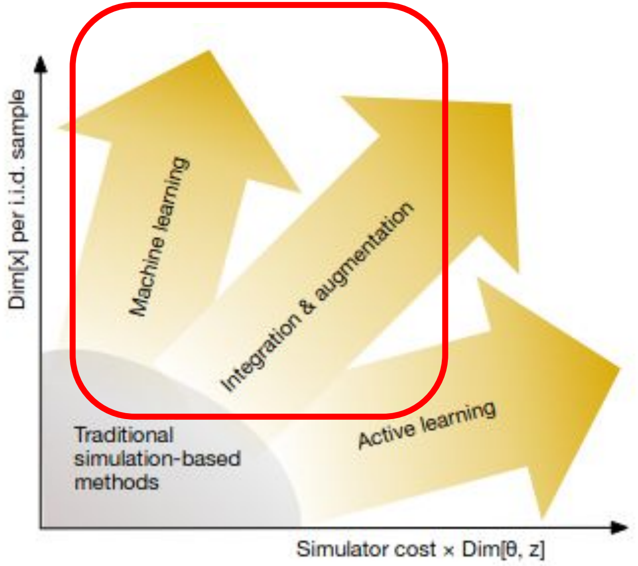
Higgs discovery in a mass summary statistic from ATLAS in 2012

- Standard strategy to get around intractable computations
  - Summary statistics like mass, pT, other kinematic variables
  - Reduce dimensions of problem to 1
- Pros
  - Use physics knowledge to distill massive integral over many dimensions to simple 1-D histograms
  - Appealing to physical intuition
- Cons
  - Some information is nearly always lost



Why now?

- Machine learning has advanced greatly in recent years
  - Ability to learn surrogates for likelihood (ratio) and/or powerful summary statistics directly from the higher dimensional data
  - Iterative learning allows for higher sample efficiency
  - New pipelines combining simulation and inference can improve both sampling efficiency and quality of inference



Surrogate learning I: likelihood

- General likelihood estimation requires estimating a density
- Three steps in the process
  - Run simulation chain for $p(x|\theta)$ for variariety of points θ
  - Train a neural density estimator where θ itself is also an input into the model
    - This training treats θ as a continuous parameter in a smooth space, leveraging nearby points to generate a conditional probability which is valid for general values of θ
  - Once trained, the model can be evaluated for various data and parameter points to define best fit points or exclusion limits
- Amortization = High upfront cost, but once $p(x|\theta)$ is determined, it can be evaluated much more quickly for large numbers of events

Surrogate learning II: likelihood ratio

- Likelihood ratios are simpler than likelihoods, and can be trained via classifiers as opposed to a density estimation
  - Simulating data at some standard point $p_{ref}(x)$ (eg the SM) and for a variety of θ's, $p(x|\theta)$
  - A binary classifier, $\hat{s}(x|\theta)$ will converge to: $p_{ref}(x)/[p(x|\theta) + p_{ref}(x)]$
  - which is simply related to likelihood ratio: $\hat{r}(x|\theta) = [1 - \hat{s}(x|\theta)]/\ \hat{s}(x|\theta)$

- A large advantage here is that likelihood ratios can then be treated with standard statistical tools and are generally more computationally efficient than full likelihoods

Integration and augmentation

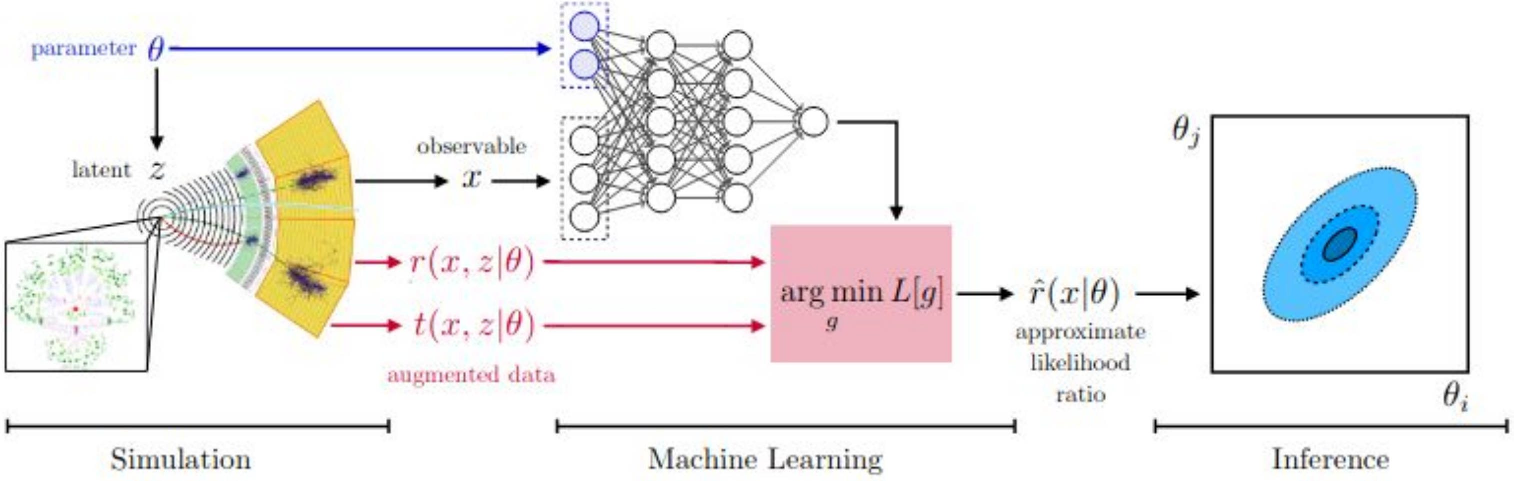
- Likelihood (ratio) estimation is powerful in that it reduces data problems in HEP to purely statistical problems
- However, we know more about the latent states in the simulation chain than is utilized in a purely black box approach, and that can be leveraged in ML approaches to simulation based inference
- In particular, 2 useful quantities can be computed and factorized:

$$r(x,z|\theta) \equiv \frac{p(x,z|\theta)}{p_{\text{ref}}(x,z)} = \frac{p(x|z_d)\,p(z_d|z_s)\,p(z_s|z_p)\,p(z_p|\theta)}{p(x|z_d)\,p(z_d|z_s)\,p(z_s|z_p)\,p_{\text{ref}}(z_p)} = \frac{|\mathcal{M}|^2(z_p|\theta)}{|\mathcal{M}|^2_{\text{ref}}(z_p)}\,\frac{\sigma_{\text{ref}}}{\sigma(\theta)}$$

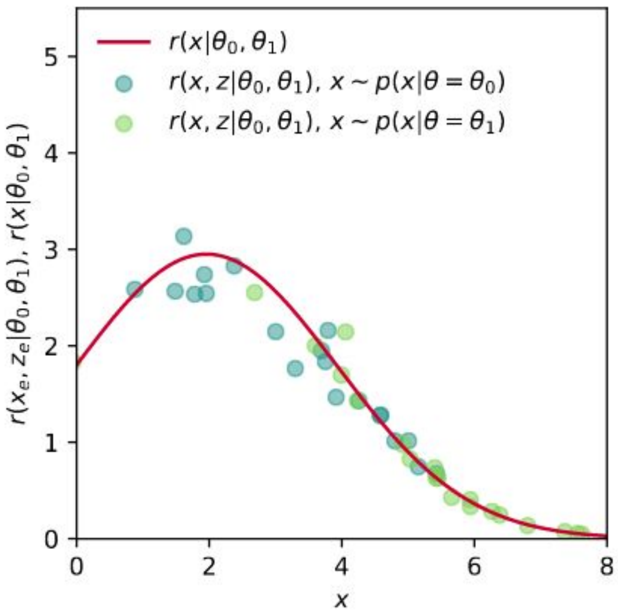
$$t(x,z|\theta) \equiv \nabla_\theta \log p(x,z|\theta) = \frac{p_x(x|z_d)\,p_d(z_d|z_s)\,p_s(z_s|z_p)\,\nabla_\theta p_p(z_p|\theta)}{p_x(x|z_d)\,p_d(z_d|z_s)\,p_s(z_s|z_p)\,p_p(z_p|\theta)} = \frac{\nabla_\theta|\mathcal{M}|^2(z_p|\theta)}{|\mathcal{M}|^2(z_p|\theta)} - \frac{\nabla_\theta\sigma(\theta)}{\sigma(\theta)}\,.$$

Integration and augmentation II

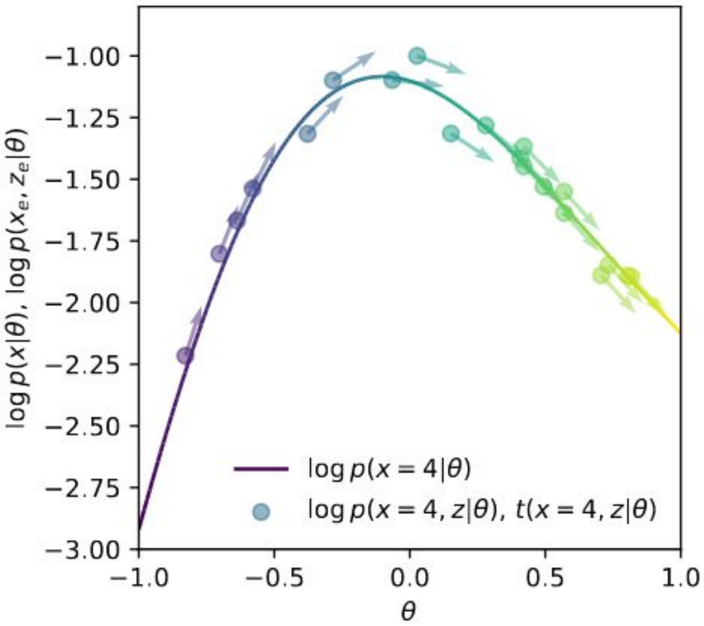
- Joint likelihood ratio and joint score can be reinserted into the training pipeline
- Additional information that can be used as labels in supervised training
  - Can reduce number of simulated events necessary for precision estimation of likelihood ratio



Examples of integration and augmentation



"the joint likelihood ratio provides noisy, but unbiased labels (green) for the likelihood ratio function to be learned (red)"



"the joint score adds noisy, but unbiased gradient information (arrows)"

Summary statistics (again!)

- Problem with aforementioned techniques: new analysis pipelines required
- But! Can use some of what we have learned from previous slides to systematize the construction of summary statistics
- Near some reference (SM) parameter value, can show the optimal observable are the likelihood score $t(x) = \nabla_\theta \log p(x|\theta)\Big|_{\theta_{\mathrm{ref}}}$

- Compute the joint score for each event, then a NN can be trained to minimize $|\hat{t}(x) - t(x, z)|^2$
- Can be shown this quantity asymptotes to the full likelihood score, and the NN discriminator can be used as the optimal summary statistics in the standard pipeline

A word on systematic uncertainty

- To trust these methods, you must trust two things
  - Simulation
  - ML surrogate
- For MC simulation, standard techniques such as profiling can be incorporated in to the surrogate of the likelihood
- Can compute likelihood (ratios) of toy MC which provides a conservative estimate guaranteed to not be more uncertain than the true data itself
- Can also try to apply some off the shelf NN uncertainty classification, not discussed at length here

Conclusions

- Exciting new statistical approach showing some promise on new ML focused statistical techniques for inference
- Pros:
  - Generic techniques can take advantage of advances in hardware and (ML) software to improve statistical inference
  - In principle, less loss of information than standard, summary statistic based methods in HEP
  - Guaranteed asymptotic behaviour for sufficiently large samples often available in HEP
  - Amortization can help with computational complexity
- Cons:
  - Less interpretable that standard summary statistics
  - Uncertainty guarantees can increase computational complexity, and understanding systematics more complex

References

- Mostly this paper:
  - https://arxiv.org/abs/2010.06439
- With additional resources from:
  - https://arxiv.org/abs/1805.00020
  - https://arxiv.org/abs/1805.00013
  - https://arxiv.org/abs/1911.01429
- And less technical introduction
  - https://www.nature.com/articles/s42254-021-00305-6
  -