Approximate Bayesian Computation: Some thoughts from a high energy physicist

Josh Bendavid (Caltech)

Caltech

November 10, 2015
Data Science @ LHC 2015 Workshop

Some Important Points to Keep In Mind

- Tempting to map "computer model" $f(\theta)$ from Richard's talk to ATLAS/CMS full generation + simulation + reconstruction chain
- Worst case scenario: Evaluating metric distance for each set of parameter values requires generating $O(10^6)$ full-sim MC events (tens of thousands of CPU hours)
- A few possible ways this kind of technique can still be useful:
  - Unfold data to generator level (or similarly produce generator $\rightarrow$ reconstructed level response matrices which can be applied quickly to generator level MC)$\rightarrow$
  - Extract reduced set of parameters from data using one or a few full Monte Carlo samples, then perform ABC-type method with a much simpler model (e.g. Bayesian integration over Higgs couplings in Higgs combination)
  - Realize model parameter variations as **reweighting** of one or a few full Monte Carlo samples

A few more thoughts

- Many common approaches in HEP could be mapped onto "summary statistics" (e.g. energy energy scale/resolution study for $Z \to ee$ with crystal ball parameters from CB $\oplus$ BW fits to mass distribution as summary statistics)
- "Post-hoc regression adjustments" could be applicable/computationally tractable in a number of cases (similar to what is already done "calibrating" top mass extraction from kinematic fits vs Monte Carlo samples)
- Mapping to other statistical approaches: "summary statistics" could just be event yields in histogram bins, metric distance could be χ^2 or binned likelihood?
- In some cases worth understanding advantages/disadvantages/correspondence with more likelihood-based approaches

Existing Closely Related HEP Example

- Special case where reweighting of full-sim Monte Carlo samples is "exact": parton distribution function variations
- In case where PDF variations are already provided as "MC replicas" (NNPDF, PDF4LHC15), these can be treated as samples from a Bayesian prior
- Method from http://arxiv.org/abs/hep-ph/9803393: Instead of rejection sampling, evaluate compatibility of each replica with data in some observable(s) and just reweight replica with weight

$$w_k = \frac{e^{-\chi_k^2/2}}{\sum_i e^{-\chi_i^2/2}}$$

- This procedure can probably be considered a variation of ABC methods
- Recent example of this procedure studying PDF constraints with Tevatron Drell-Yan Forward Backward Asymmetry:

http://arxiv.org/abs/1507.02470