Designing Experiments for Robust Parameter Estimation

Mohan Krishnamoorthy
Sven Leyffer, Holger Schulz, Stephen Mrenna,
Juliane Müller, Moritz Schulze

January 14, 2020

All-hands Meeting 2020

Tuning Monte Carlo Event Generators

$$\min_{\boldsymbol{p}\in\Omega} \sum_{b} \frac{(MC_b(\boldsymbol{p}) - \mathcal{R}_b)^2}{\sigma_b^2}.$$

- However, performing MC simulations can be very expensive

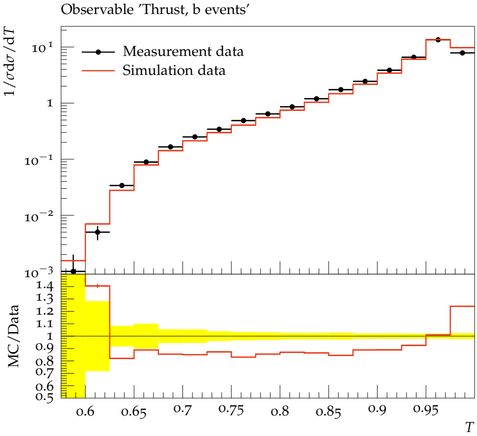
$$MC_b(\boldsymbol{p}) \approx r_b(\boldsymbol{p}) \quad \forall b.$$

$$\min_{\boldsymbol{p}\in\Omega} \sum_{b} \frac{(r_b(\boldsymbol{p}) - \mathcal{R}_b)^2}{\sigma_b^2}.$$

- Goal is to tune the parameter $\boldsymbol{p}$ for a large number of observables

$$\min_{\boldsymbol{p}\in\Omega} \sum_{\mathcal{O}\in\theta} w_{\mathcal{O}} \sum_{b\in\mathcal{O}} \frac{(r_b(\boldsymbol{p}) - \mathcal{R}_b)^2}{\sigma_b^2}.$$

Tuning Monte Carlo Event Generators



Robust Optimization Formulation

$$\underset{\boldsymbol{w}\in[0,1],\boldsymbol{p}\in\Omega}{\text{minimize}} \sum_{\mathcal{O}\in\theta} w_{\mathcal{O}} \sum_{b\in\mathcal{O}} \underset{d_b\in\mathcal{U}_b}{\text{maximize}} \left(r_b(\boldsymbol{p}) - d_b\right)^2 .$$

$$\underset{\boldsymbol{t},\boldsymbol{w}\in[0,1],\boldsymbol{p}\in\Omega}{\text{minimize}} \sum_{\mathcal{O}\in\theta} \sum_{b\in\mathcal{O}} t_b$$

subject to

$$t_b \geq w_{\mathcal{O}} \left(r_b(\boldsymbol{p}) - d_b\right)^2 \quad \forall d_b \in \mathcal{U}_b, \forall b \in \mathcal{O}, \forall \mathcal{O} \in \theta.$$

Robust Optimization Formulation

$$\underset{\boldsymbol{t},\boldsymbol{w}\in[0,1],\boldsymbol{p}\in\Omega}{\text{minimize}} \sum_{\mathcal{O}\in\theta}\sum_{b\in\mathcal{O}} t_b$$

subject to

$$t_b \geq w_{\mathcal{O}}\left(r_b(\boldsymbol{p}) - (\mathcal{R}_b - \sigma_b)\right)^2 \quad \forall b \in \mathcal{O}, \forall \mathcal{O} \in \theta$$

$$t_b \geq w_{\mathcal{O}}\left(r_b(\boldsymbol{p}) - (\mathcal{R}_b + \sigma_b)\right)^2 \quad \forall b \in \mathcal{O}, \forall \mathcal{O} \in \theta.$$

Robust Optimization Formulation

$$\underset{\boldsymbol{t},\boldsymbol{w}\in[0,1],\boldsymbol{p}\in\Omega}{\text{minimize}} \sum_{\mathcal{O}\in\theta}\sum_{b\in\mathcal{O}} t_b$$

subject to

$$t_b \geq w_{\mathcal{O}}\left(r_b(\boldsymbol{p}) - (\mathcal{R}_b - \sigma_b)\right)^2 \quad \forall b \in \mathcal{O}, \forall \mathcal{O} \in \theta$$

$$t_b \geq w_{\mathcal{O}}\left(r_b(\boldsymbol{p}) - (\mathcal{R}_b + \sigma_b)\right)^2 \quad \forall b \in \mathcal{O}, \forall \mathcal{O} \in \theta$$

$$e^T w \geq (|\theta|\mu)/100.$$

Envelope Filtering

$$FMIN_b = \underset{\boldsymbol{p}\in\Omega}{\text{minimize}}\,[r_b(\boldsymbol{p})]\,. \quad FMAX_b = \underset{\boldsymbol{p}\in\Omega}{\text{maximize}}\,[r_b(\boldsymbol{p})]\,.$$

$$\begin{cases} \text{Keep bin b,} & \text{if } FMIN_b \leq d_b \leq FMAX_b \\ \text{Discard bin b,} & \text{otherwise.} \end{cases}$$

Hypothesis Filtering

$$\chi^2_{\{b_i,\ldots,b_j\}} = \sum_{b \in \{b_i,\ldots,b_j\}} \frac{(r_b(\boldsymbol{p}) - d_b)^2}{\sigma_b^2}.$$

For this test statistic, we hypothesize that:

H_0: There is no significant difference between $r_b(\boldsymbol{p})$ and d_b or that the data d_b can be appropriately described by $r_b(\boldsymbol{p})$.

H_1: H_0 is rejected i.e., there is a significant difference between $r_b(\boldsymbol{p})$ and d_b.

$$\rho_{\{b_i,\ldots,b_j\}} = |\{b_i, \ldots, b_j\}| - |\boldsymbol{p}|.$$

$$\chi^2_{c,\{b_i,\ldots,b_j\}} = f(\alpha, \rho_{\{b_i,\ldots,b_j\}}).$$

Hypothesis Filtering

$$\chi^2_{\{b_i,\ldots,b_j\}} = \sum_{b\in\{b_i,\ldots,b_j\}} \frac{(r_b(\boldsymbol{p}) - d_b)^2}{\sigma_b^2}.$$

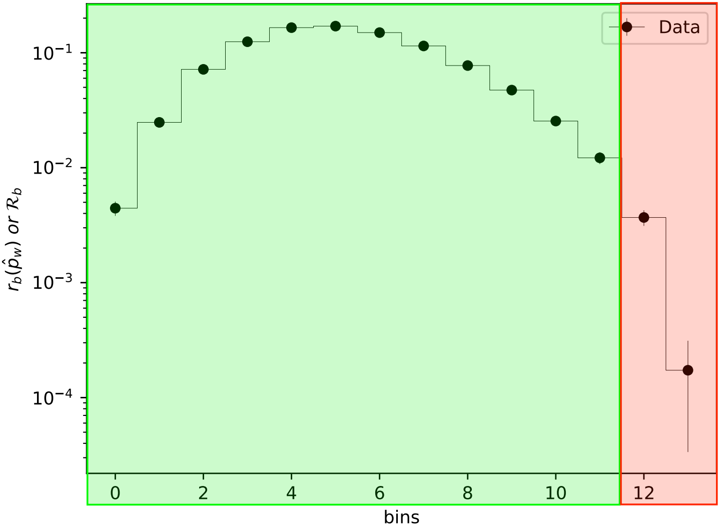
For this test statistic, we hypothesize that:

H_0: There is no significant difference between $r_b(\boldsymbol{p})$ and d_b or that the data d_b can be appropriately described by $r_b(\boldsymbol{p})$.

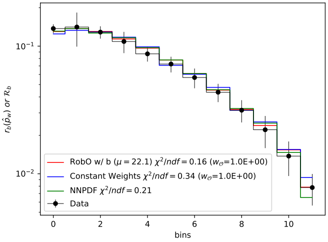
H_1: H_0 is rejected i.e., there is a significant difference between $r_b(\boldsymbol{p})$ and d_b.

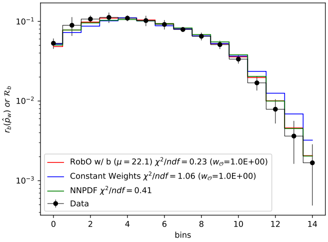
$$\begin{cases} \text{Keep bins } \{b_i,\ldots,b_j\}, & \text{if } \chi^2_{\{b_i,\ldots,b_j\}} \leq \chi^2_{c,\{b_i,\ldots,b_j\}} \\ \text{Discard bins } \{b_i,\ldots,b_j\}, & \text{otherwise.} \end{cases}$$

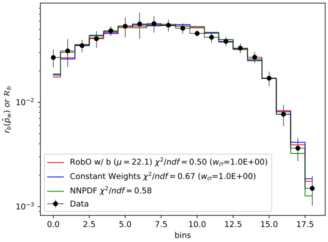
Hypothesis Filtering

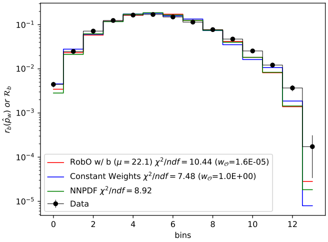


Histogram Plots of 4 observables from A14* dataset









* A14 data is used in the ATLAS note *ATL-PHYS-PUB-2014-021* titled ATLAS Pythia 8 tunes to 7 TeV data.

Summary and What's next

- Summary
  - Optimization approach for robust parameter estimation
  - Filtering to remove bins that cannot be accurately described by the MC simulation
- Next...
  - Evaluating the usefulness of filtering in terms of quality of results and computational efficiency
  - Design of experiments paper
  - Develop stochastic optimization algorithm to improve efficiency of the bi-level optimization approaches
  - Tuning parameters directly for MC simulations instead of for rational approximations

Extra Slides

Pairwise Comparison of Multiple Approaches

Say that

$$R_o(\boldsymbol{p}) = \sum_{b \in o} \frac{(r_b(\boldsymbol{p}) - \mathcal{R}_b)^2}{\sigma_b^2}$$

Then

$$C_\epsilon(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)}) \coloneqq \left| \left\{ o | R_o(\boldsymbol{p}^{(i)}) \leq R_o(\boldsymbol{p}^{(j)}) - \epsilon \right\} \right|, \quad o \in \theta$$

We say that $\boldsymbol{p}^{(i)}$ is ϵ-better than $\boldsymbol{p}^{(j)}$ iff

$$C_\epsilon(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)}) \geq C_\epsilon(\boldsymbol{p}^{(j)}, \boldsymbol{p}^{(i)}) + 1$$

Pairwise Comparison of Approaches for A14 dataset

	1	2	3	4	5
1	–	206	233	210	205
2	200	–	219	206	203
3	173	187	–	205	198
4	196	200	201	–	206
5	201	203	208	200	–

Legend

1. Bi-level optimization approach using FIM with bound, $\mu = 400.0$ (data used in optimization is **unfiltered**)
2. Robust Opt with bound, $\mu = 22.1$ (data used in optimization is **unfiltered**)
3. Bi-level optimization approach using FIM with bound, $\mu = 215.0$ (data used in optimization is **hypothesis filtered**)
4. Robust Opt with bound, $\mu = 34.0$ (data used in optimization is **hypothesis filtered**)
5. NNPDF (Tune published in the ATLAS note corresponding to the leading-order parton density function **NNPDF23LO**)