Constraining EFTs with Machine Learning

↓

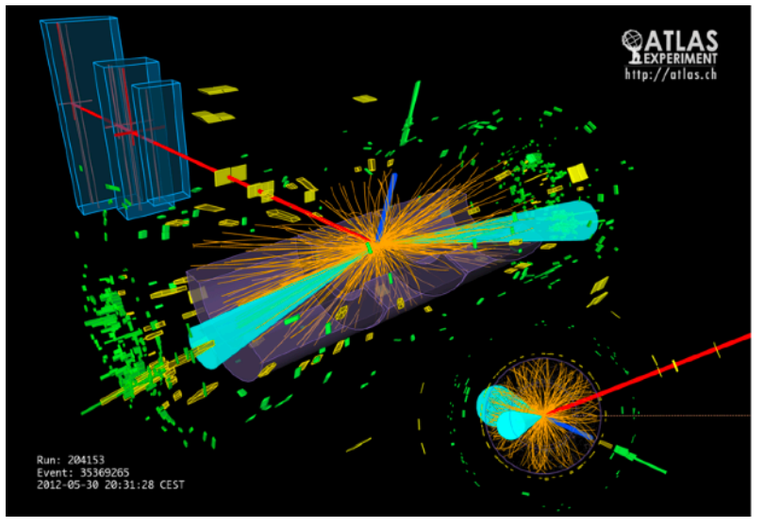
MadMiner

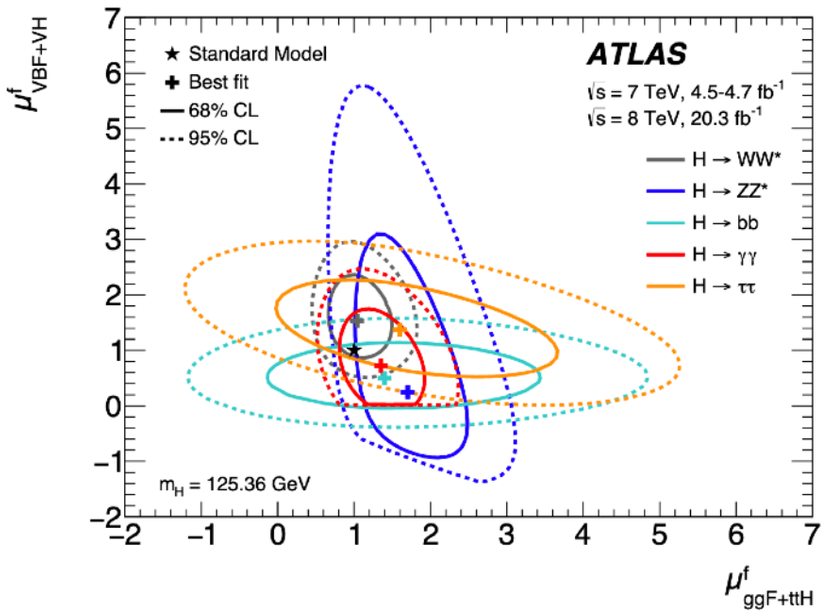
Felix Kling

SLAC NATIONAL ACCELERATOR LABORATORY

2-4 October 2019
Argonne National Laboratory
In Search of new Physics using SMEFT

Motivation





Era of Data

- large statistics at LHC
- lots of kinematic distributions
- high dimensional data: **x**

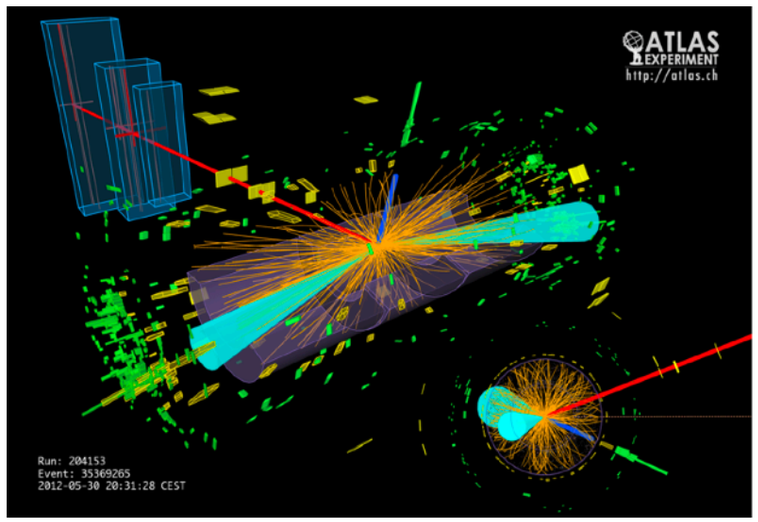
Complex Theories

- large number of theory parameters: θ
- predict lots of kinematic features
- Example: SMEFT $\mathcal{L} = \sum \theta_i \mathcal{O}_i$

Motivation

likelihood function $p(x|\theta)$:
likelihood of an observation x as a function of the theory parameter θ

likelihood ratio $r(x|\theta_{ref},\theta)=p(x|\theta)/p(x|\theta_{ref})$
"how much more likely is data x described by theory θ than θ_{ref}"

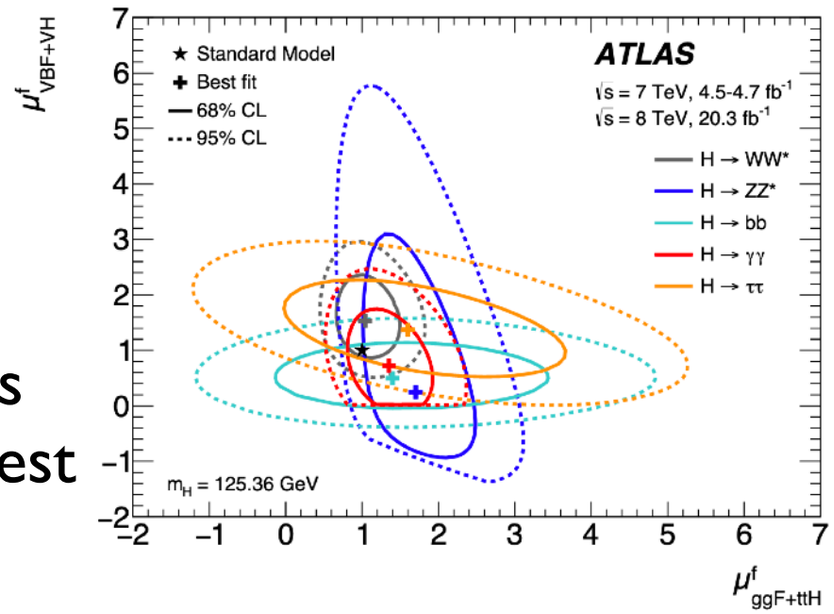


Estimator for
likelihood function $p(x|\theta)$

→

Inference

maximum-likelihood estimators
limits based on likelihood ratio test



Era of Data

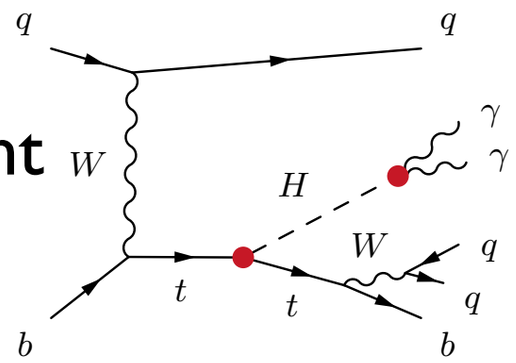
- large statistics at LHC
- lots of kinematic distributions
- high dimensional data: **x**

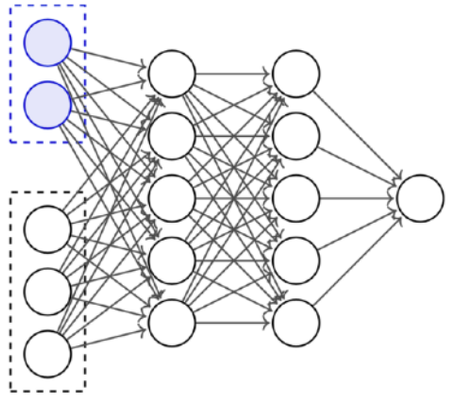
Complex Theories

- large number of theory parameters: θ
- predict lots of kinematic features
- Example: SMEFT $\mathcal{L} = \sum \theta_i \mathcal{O}_i$

Motivation

Matrix Element Information





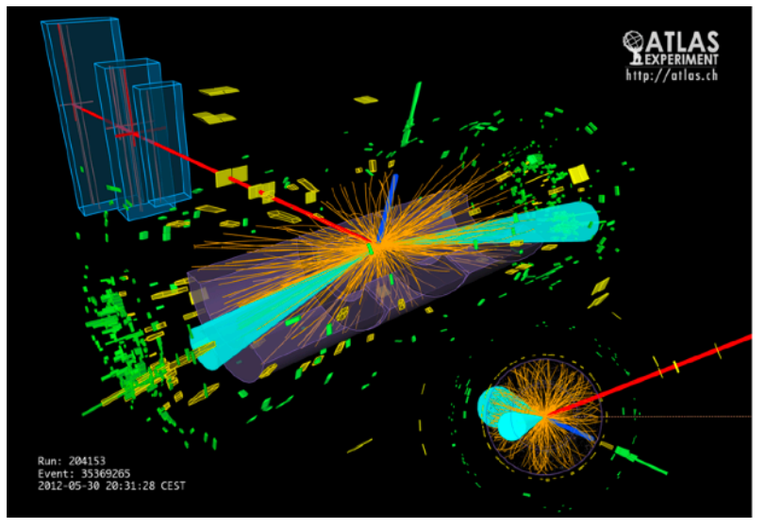
Machine Learning

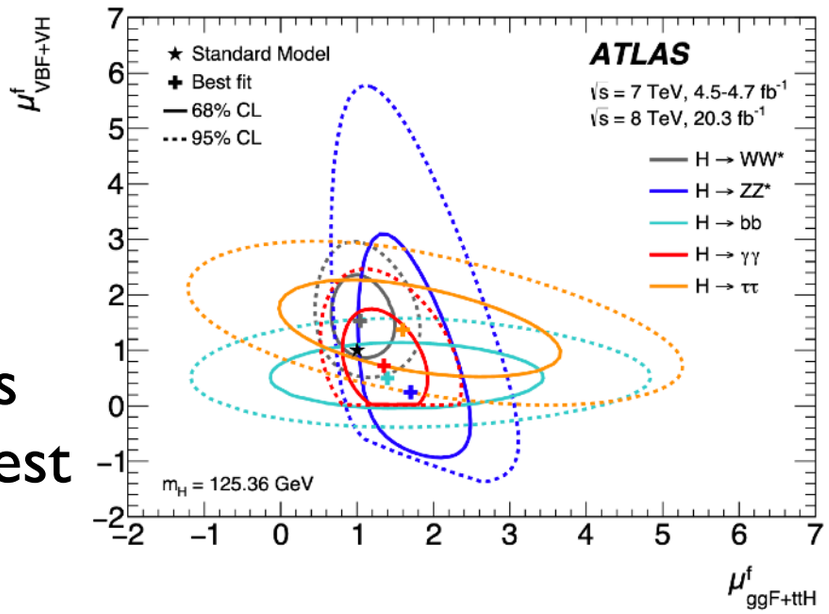
MadMiner

Estimator for likelihood function $p(x|\theta)$

Inference

maximum-likelihood estimators
limits based on likelihood ratio test





Era of Data

- large statistics at LHC
- lots of kinematic distributions
- high dimensional data: **x**

Complex Theories

- large number of theory parameters: θ
- predict lots of kinematic features
- Example: SMEFT $\mathcal{L} = \sum \theta_i \mathcal{O}_i$

Outline

Inference

Particle Physics Simulations / Inference Techniques

MadMiner

Inference Method / Optimal Observables / Fisher Information / Tool

Physics Example

Probing SMEFT in tth

Summary and Conclusion

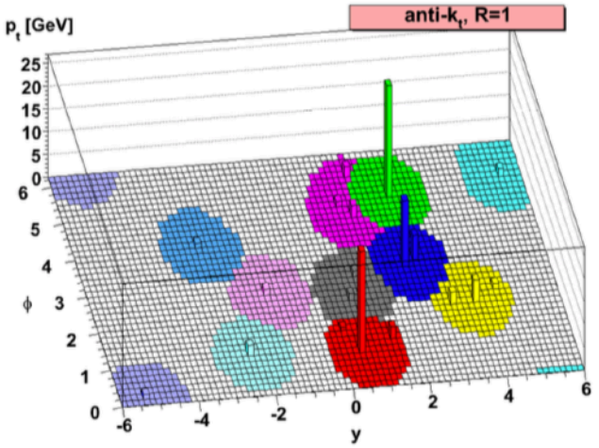
Inference

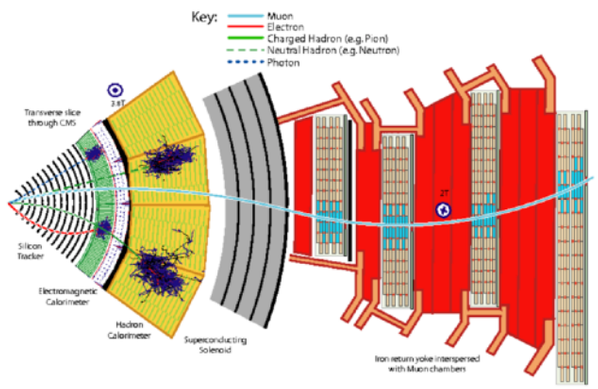
Particle Physics Simulations/ Inference Techniques

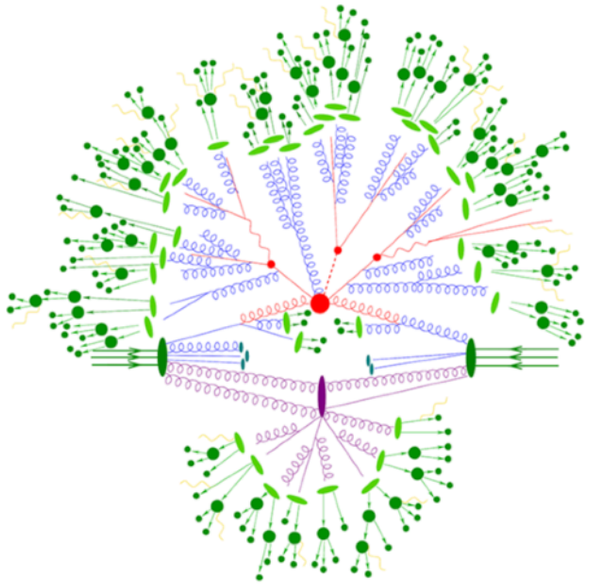
Particle Physics Simulations

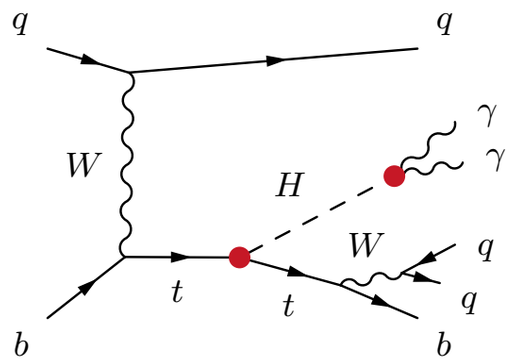
Latent Variables z: internal, not observable

Observables	Detector Effects	Shower Hadronization	Parton-Level Momenta	Theory Parameters
x	z_d	z_s	z_p	θ









Particle Physics Simulations

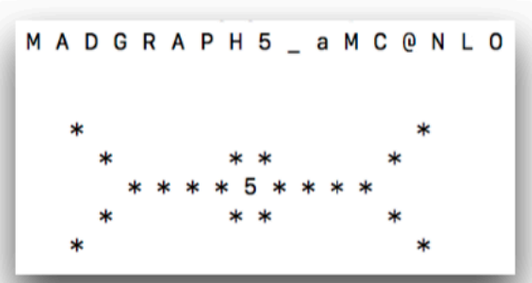
Latent Variables z: internal, not observable

Observables	Detector Effects	Shower Hadronization	Parton-Level Momenta	Theory Parameters
x	z_d	z_s	z_p	θ

$$p(x|\theta) = \int dz_d \; dz_s \; dz_p \quad p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

obtain MC event sample

sampling performed by MC tools:



Prediction / Simulation

Particle Physics Simulations

Latent Variables z: internal, not observable

Observables	Detector Effects	Shower Hadronization	Parton-Level Momenta	Theory Parameters
x	z_d	z_s	z_p	θ

$$p(x|\theta) = \int dz_d \; dz_s \; dz_p \quad p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

likelihood function needed

It's infeasible to calculate the integral over this enormous latent space

↓

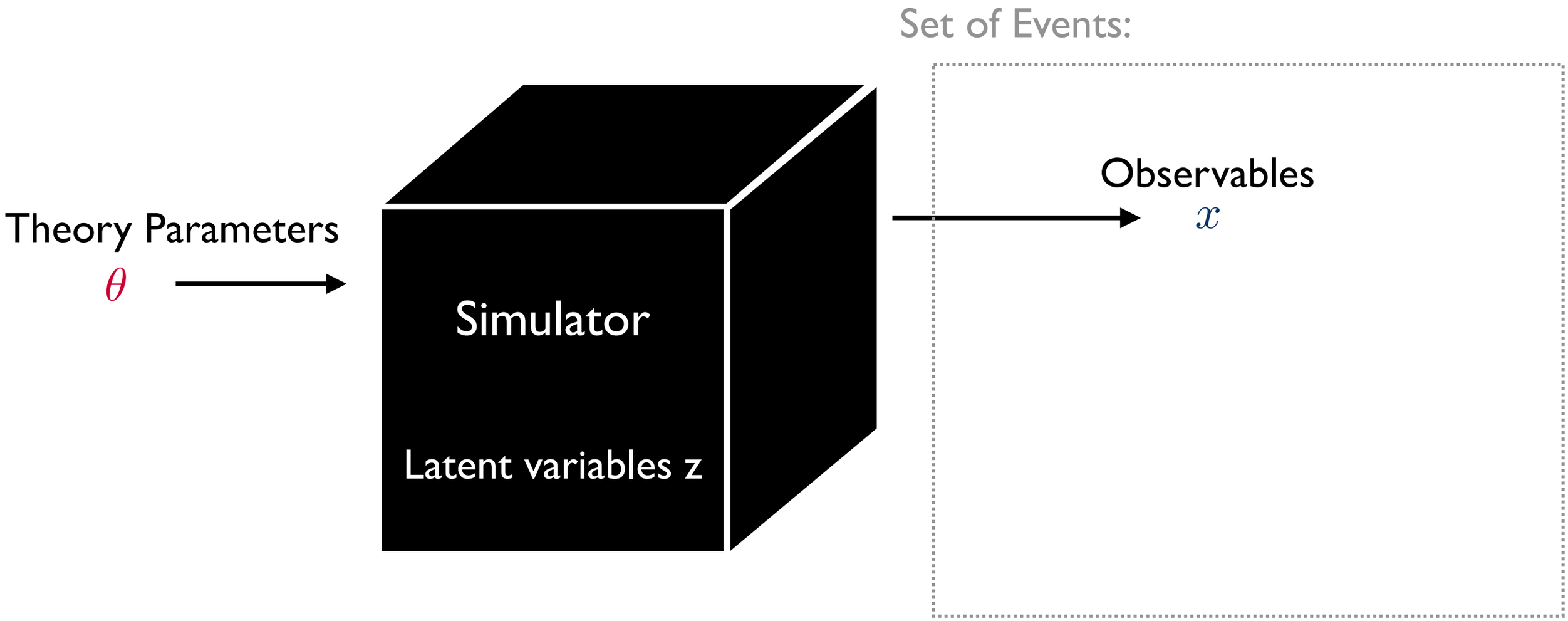
likelihood function is intractable

↓

use estimator $\hat{p}(x|\theta)$

Inference

Particle Physics Simulations



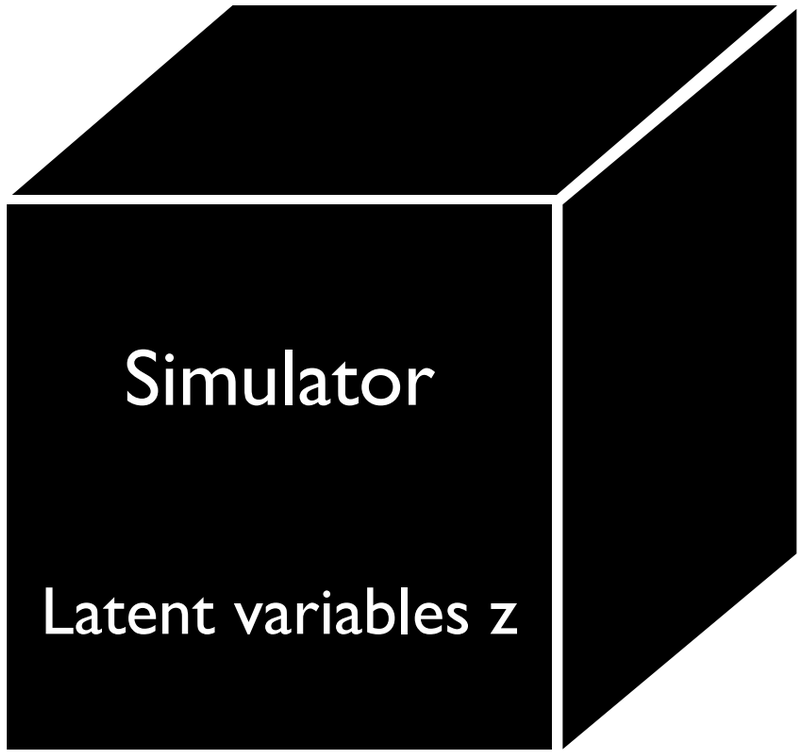
Prediction: - well understood mechanistic model
- simulator can generate samples

Inference: - likelihood function $p(x|\theta)$ is intractable
- use inference based estimator $\hat{p}(x|\theta)$

Particle Physics Simulations

Theory Parameters

θ



Set of Events:

Observables

x

Joint likelihood ratio

$r(x, z|\theta)$

given by matrix elements

$|\mathcal{M}(z_p|\theta)|^2 / |\mathcal{M}(z_p|\theta_{\text{ref}})|^2$

(use morphing)

Prediction: - well understood mechanistic model
- simulator can generate samples

Inference: - likelihood function $p(x|\theta)$ is intractable
- use inference based estimator $\hat{p}(x|\theta)$

Inference Techniques

Traditional

- use summary statistics
  * hand-picked observables x'
  * estimate $p(x'|\theta)$
- information loss
- problem dependent

Examples:

- rate only (cut and count)
- histograms
- Approximate Bayesian Computation
- STXS

Machine Learning

- multivariate analysis
- works great for S vs BG
- struggles with S' vs S
  * large number of S'
  * very similar S', S

Examples:

- Neural Density Estimator
- ML Classifier

Matrix Element Based

- multivariate analysis
- uses $p(x|\theta) \sim |M(x|\theta)|^2$
- works great at parton level
  * S' vs S is easy
- requires approximations in reality
  * S vs BG can be hard

Examples:

- Matrix Element Method
- Optimal Observables

power of machine learning

physics insight of matrix element information

MadMiner

[J. Brehmer, K. Cranmer, G. Louppe, J. Pavez 1805.00013, 1805.00020,1805.12244]

[J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

MadMiner

Inference Method / Optimal Observables / Fisher Information / Tool

Inference Method

Problem Setup:

We want: Likelihood-ratio $r(x|\theta_{ref},\theta)$ (data ↗ x, theory ↖ θ)

Generators give us: Joint Likelihood-ratio $r(x,z|\theta_{ref},\theta)$

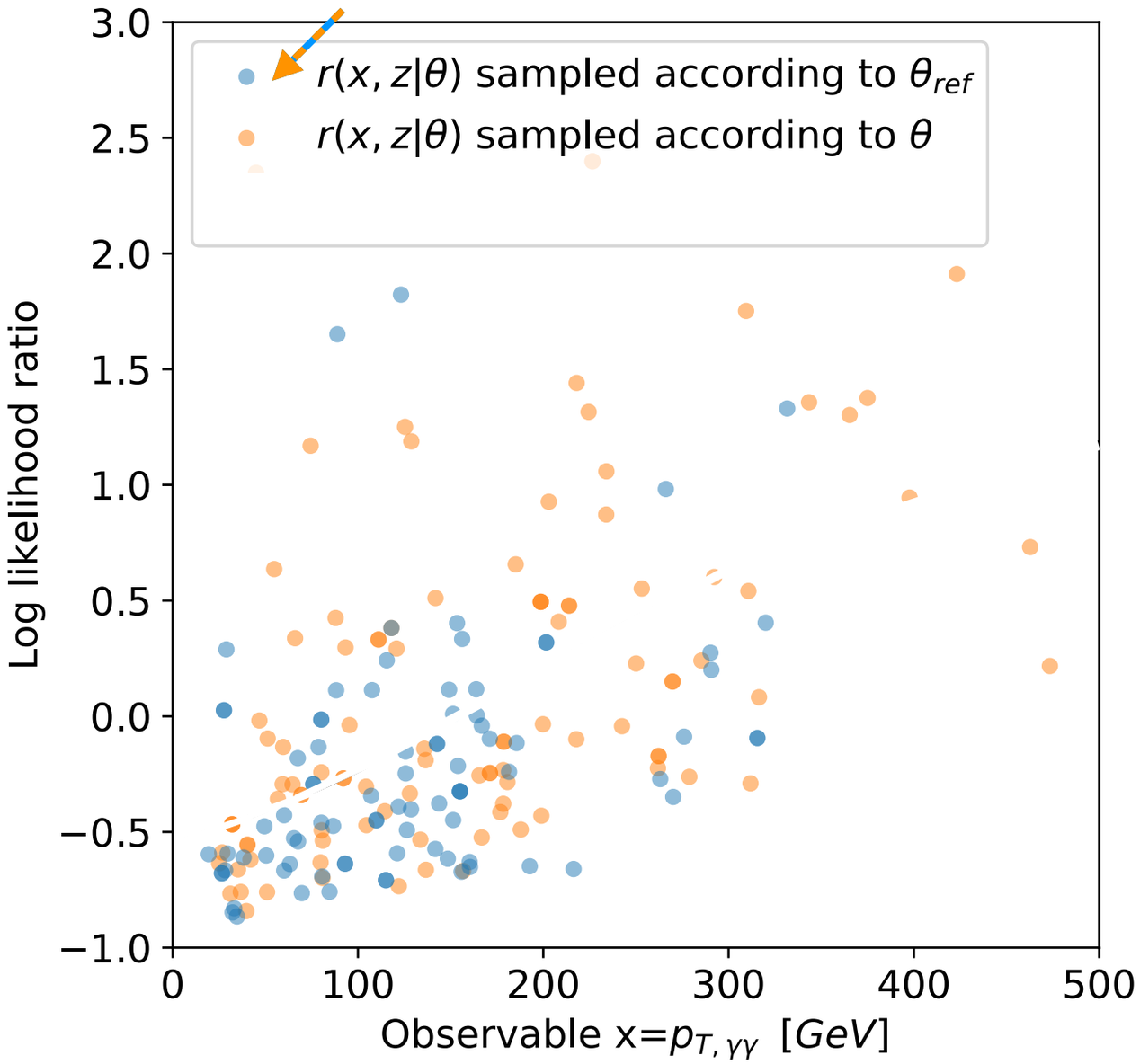
A Pedagogical Example for Illustration

- processes
  - * signal: fully leptonic tth $pp \to t\bar{t}h \to \left(b\ell^+\right)\left(\bar{b}\ell^-\right)(\gamma\gamma)E_T^{\text{miss}}$
  - * background : ttγγ continuum
- theory
  - * SMEFT model: $\mathcal{L} = \mathcal{L}_{SM} + c_G\mathcal{O}_G$ with $\mathcal{O}_G = g_s^2/m_W^2\left(H^\dagger H\right)G^a_{\mu\nu}G_a^{\mu\nu}$
  - * theory parameter: $\theta = 100 \times c_G$
- simulation: MadGraph5 + Pythia8 + Delphes3 + HL-LHC setup [HL-LHC WG: 1902.00134]
- one observable: $x = p_{T,\gamma\gamma}$
- latent variables:
  - * neutrino pz
  - * jet charge+flavor
  - * all the hadron momenta
  - * random numbers in detector simulation
  - * all other observables

Inference Method

How is Likelihood Estimated?

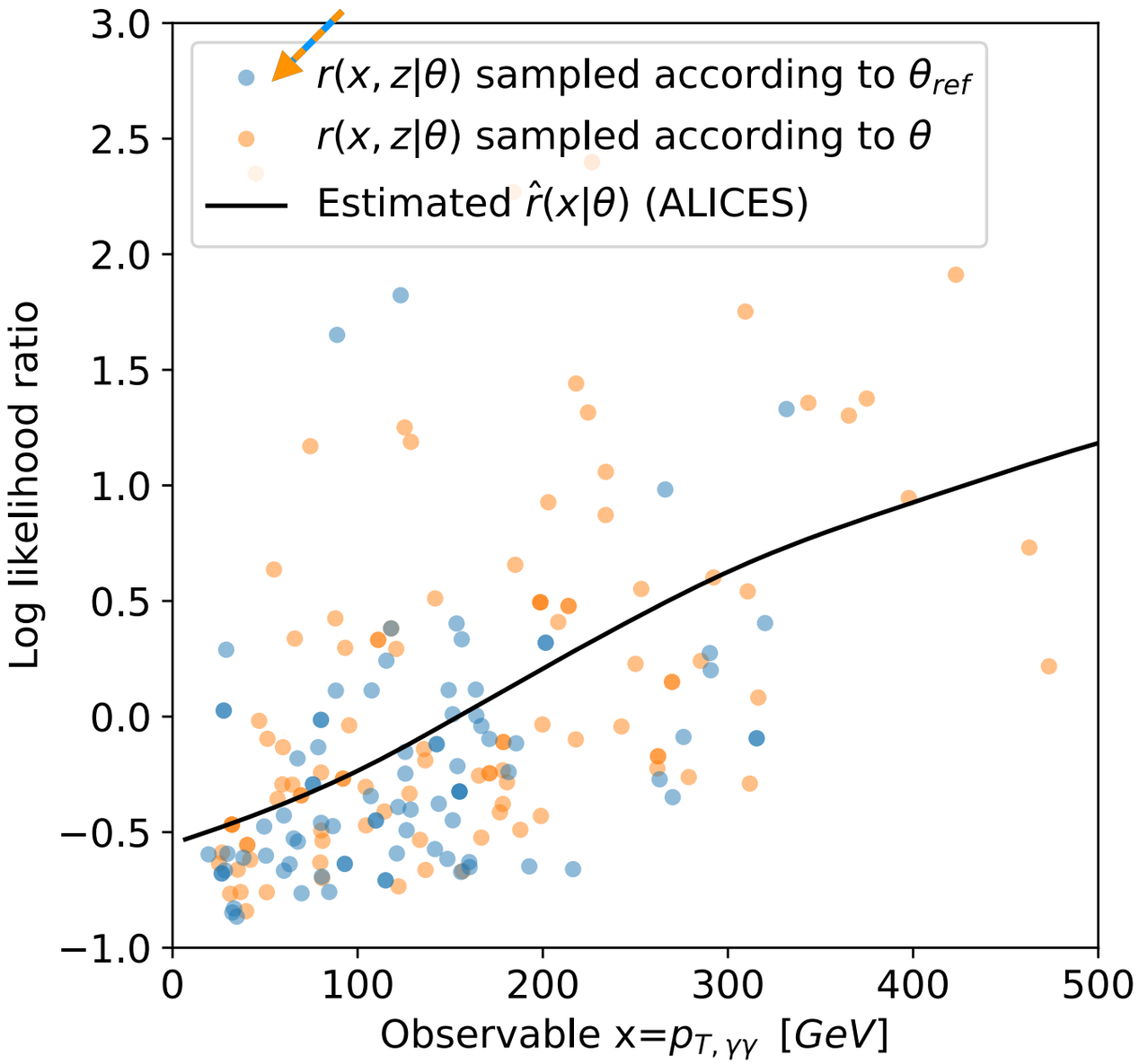
- consider two models BSM (θ=1) vs SM (θ_{ref}=0)
  * sampled events according to p(x|SM) , p(x|BSM)
  * y-axis joint likelihood ratio r(x,z|BSM,SM)



Inference Method

How is Likelihood Estimated?

- consider two models BSM (θ=1) vs SM (θ_{ref}=0)
  * sampled events according to p(x|SM) , p(x|BSM)
  * y-axis joint likelihood ratio r(x,z|BSM,SM)



- define functional (loss function)

$$L\left[\hat{r}(x|\theta)\right] \sim \sum |r(x|\theta) - r(x,z|\theta)|^2$$

and minimize it

$$r(x|\theta) = \arg\min_{\hat{r}(x|\theta)} L_r\left[\hat{r}(x|\theta)\right]$$

neural network — loss function — stochastic gradient descent

→ obtain "best fit"

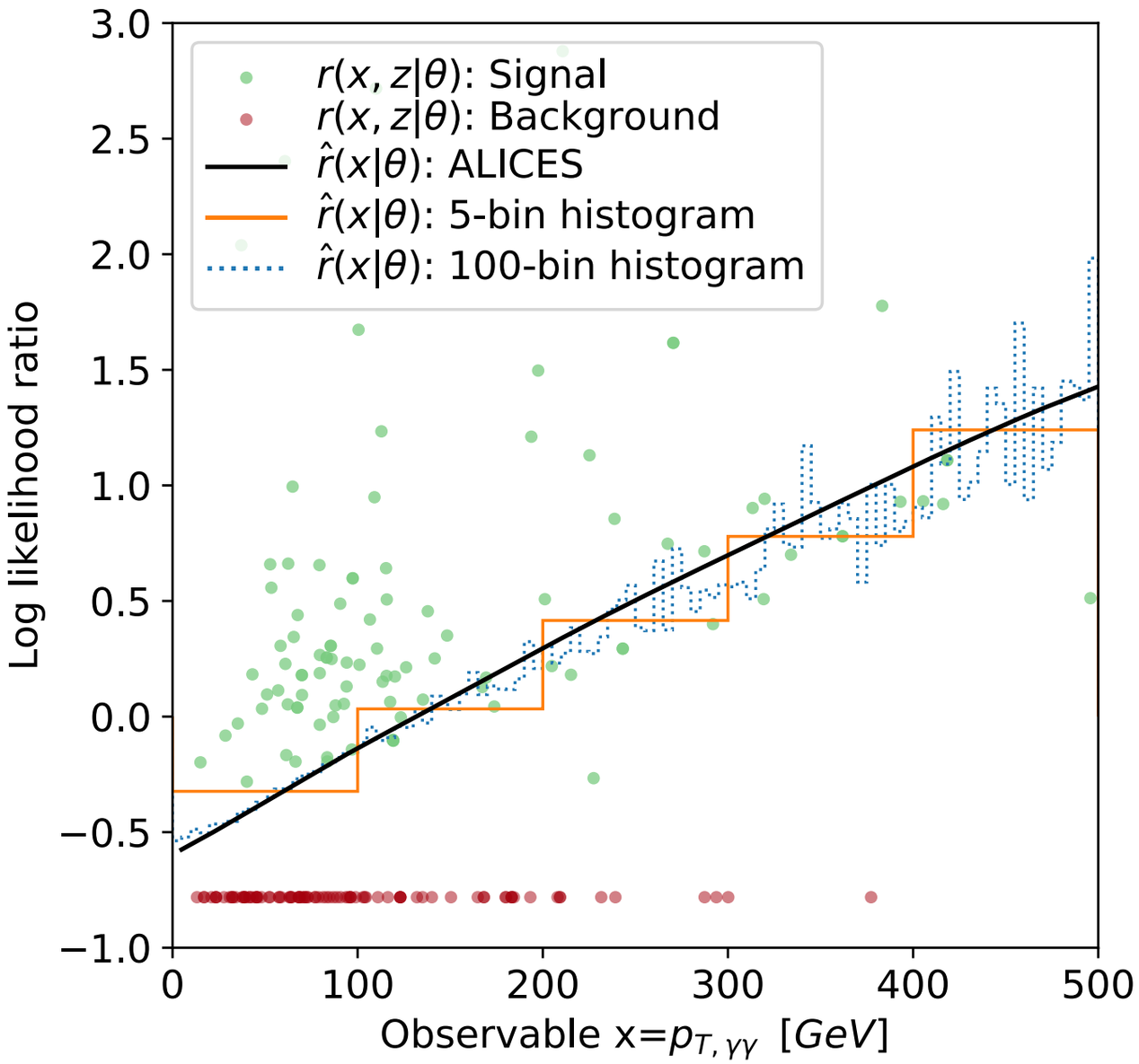
A sufficiently expressive network, efficiently trained in this way with enough data will learn the likelihood ratio function r(x|θ)!

[Proof: J. Brehmer, K. Cranmer, G. Louppe, J. Pavez 1805.00020]

Inference Method

How about Backgrounds?

- consider two models: BSM (θ=1) vs SM (θ_{ref}=0)
- include ttH signal and tt$\gamma\gamma$ background



- additional latent variable z: process label

- points: joint likelihood ratio
  * constant for background:
    $r(x,z|\theta) = \sigma(\theta_{ref})/\sigma(\theta)$
    ↓ "best fit"

- estimate $r(x, \theta)$

- compare to LLR obtained using histogram
  * agrees well!
  * ML "continuum limit of for large # of bins"

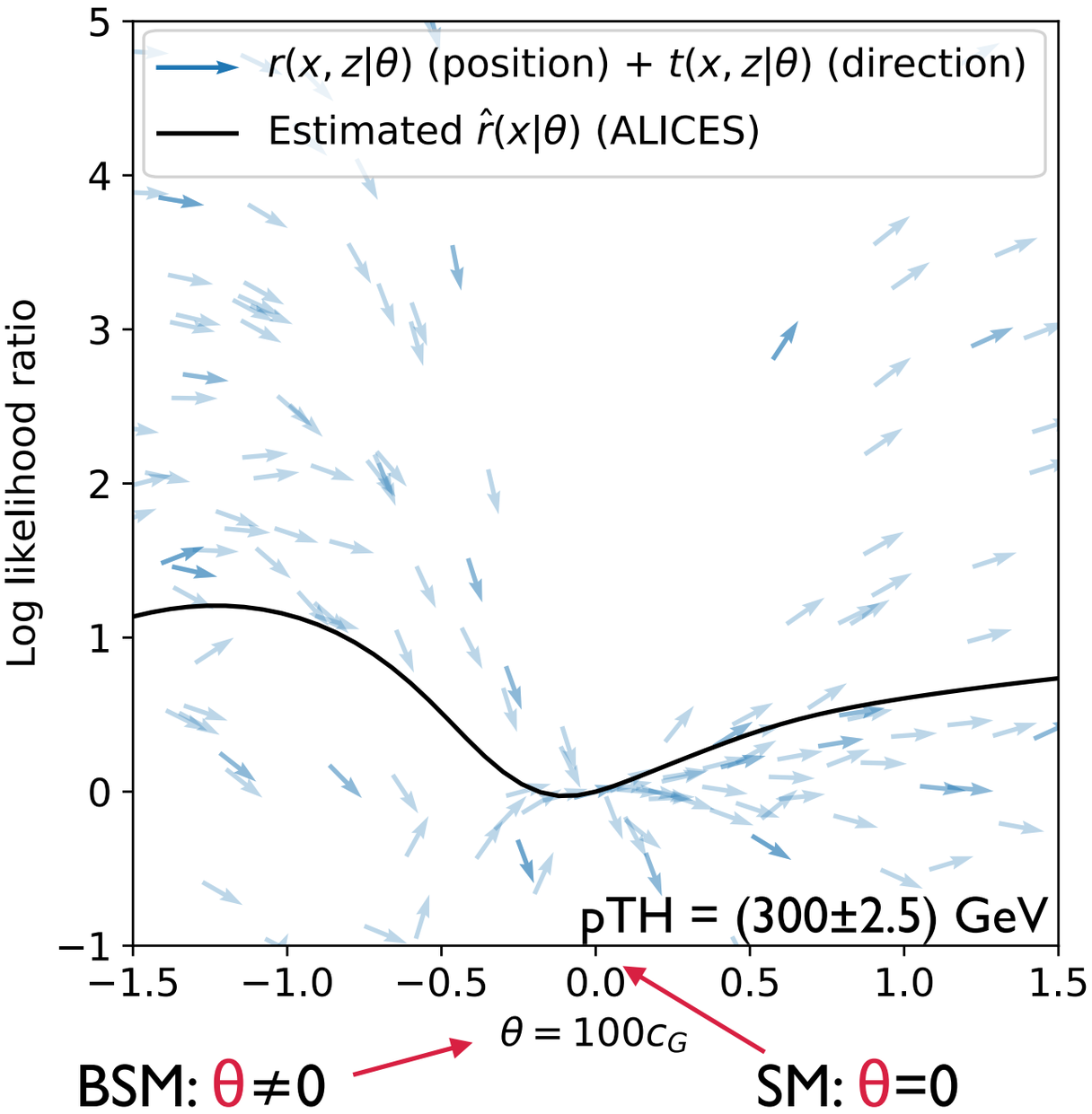
- realistic problem: x and θ high dimensional

Inference Method

Useful: the Score!

- knowing the derivative often helps: "*How does data x change, when theory θ is changed?*"

  ⟶ Score: $t(x|\theta) = d \log p(x|\theta)/d\theta$



- position: joint likelihood ratio
  log r(x,z|θ) as function of θ
- slope: joint score
  t(x,z|θ) as function of θ

obtain "best fit"

- estimate r(x, θ) and t(x, θ)

Likelihood Ratio Estimator (ALICES)

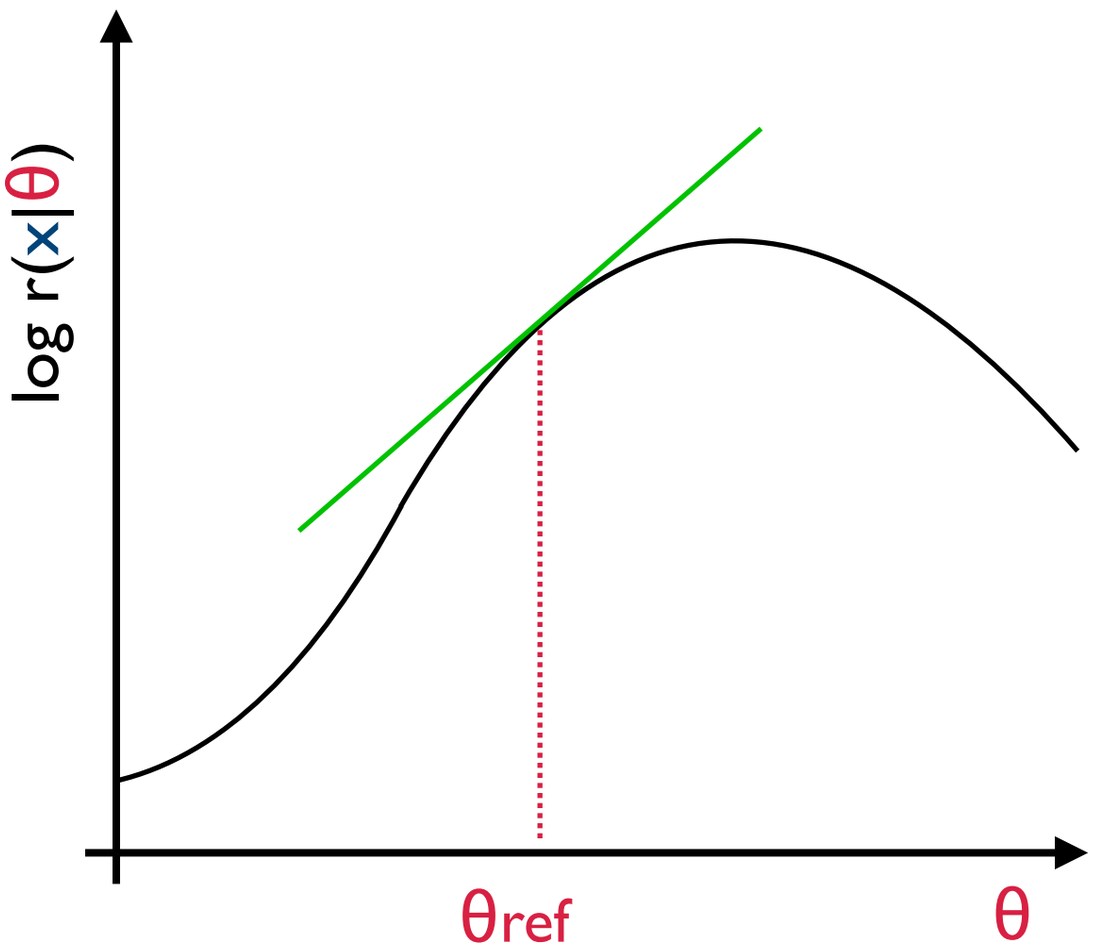
- learn LLR as function of x and θ
- use r(x,z|θ) and t(x,z|θ) as input

$$\mathrm{NN} : (x, \theta) \to \hat{r}(x|\theta) \approx p(x|\theta)/p(x|\theta_{\mathrm{ref}})$$

Optimal Observables

Scores are Optimal Observables

- expand LLR around θ_{ref}
  * $\log r(x|\theta) = \log r(x|\theta_{\text{ref}}) + t(x)|_{\theta_{\text{ref}}} \cdot (\theta - \theta_{\text{ref}}) + \ldots$



- score: $t(x|\theta) = d \log p(x|\theta)/d\theta$
- close to θ_{ref}
  * score is sufficient statistics
  * knowing $t(x)|_{\theta_{\text{ref}}}$ is as powerful as knowing $r(x|\theta)$
  * $t(x)|_{\theta_{\text{ref}}}$ are optimal observables
- in SMEFT: $t(x)|_{\text{SM}}$ is sensitive to interference

Score Estimator (SALLY)
- learn score as function of x at θ_{ref}
- $t(x,z|\theta_{\text{ref}})$ as input

$$\text{NN} : x \to \hat{t}(x) \approx \nabla_\theta \log(x|\theta)\big|_{\theta_{\text{ref}}}$$

Fisher Information

Fisher Information

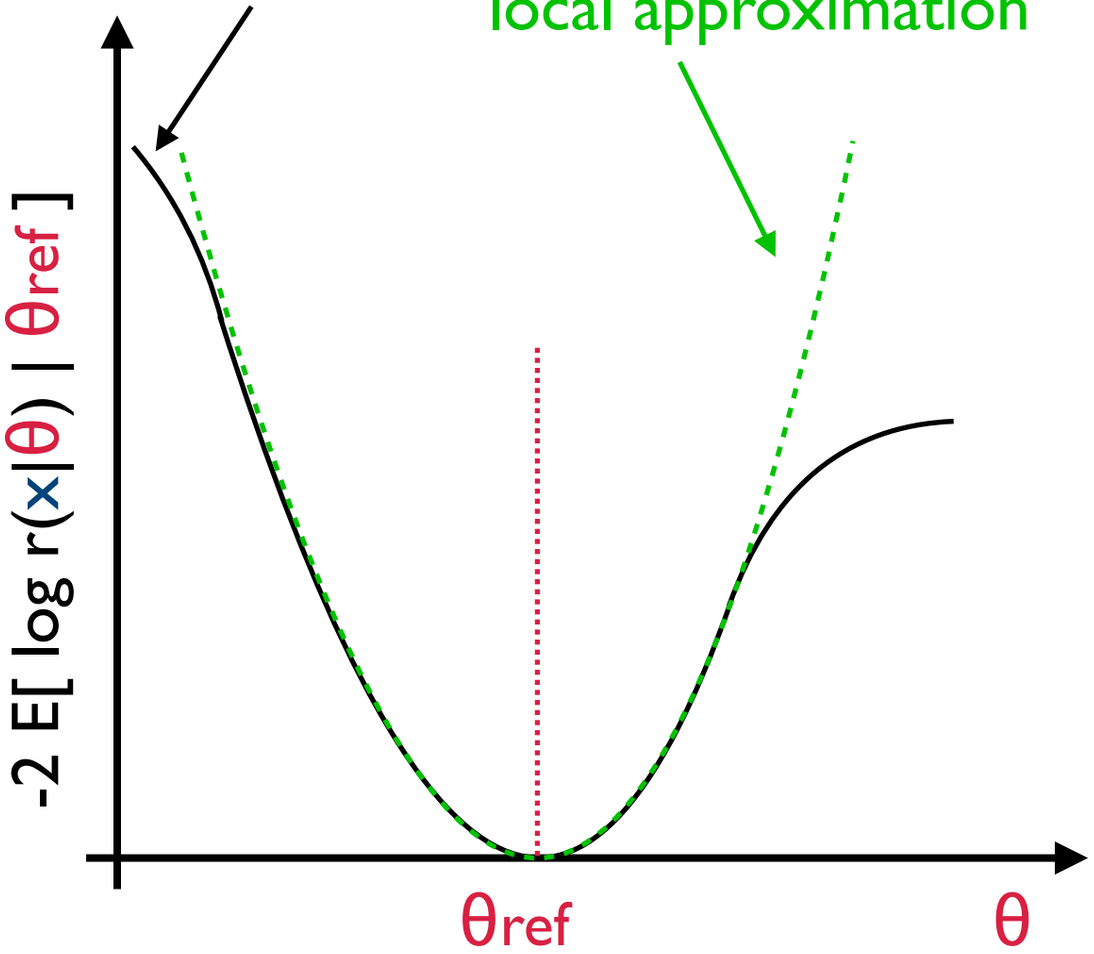
See [J. Brehmer, K. Cranmer, FK, T. Plehn 1612.05261]

- expand expected LLR around θ_{ref}

$$-2\mathbb{E}[\log r(x|\theta)|\theta_{\text{ref}}] \approx I_{ij}(\theta_{\text{ref}}) \times (\theta-\theta_{\text{ref}})_i(\theta-\theta_{\text{ref}})_j + \dots$$

full results

local approximation



scores

- Fisher Information:

$$I_{ij} = \mathcal{L}\frac{\partial_i\sigma(\theta)\partial_j\sigma(\theta)}{\sigma(\theta)} + \frac{\mathcal{L}\,\sigma(\theta)}{n}\sum_{x\sim p(x|\theta_{\text{ref}})} t_i(x)t_j(x)$$

- useful properties:
  - * simple: n×n matrix (for n theory parameters)
  - * Cramer-Rao bound: $\text{cov}[\hat{\theta}|\theta_0] \leq I_{ij}^{-1}(\theta_0)$
  - * independent of parameterizations of x
  - * covariant under $\theta \to \theta'$
  - * additive between experiments / phase-space
  - * easy to include systematics
  - * defines metric on theory parameter space

The Fisher information encodes the maximum sensitivity of observables to theory parameters for a given experiment

MadMiner: The Tool

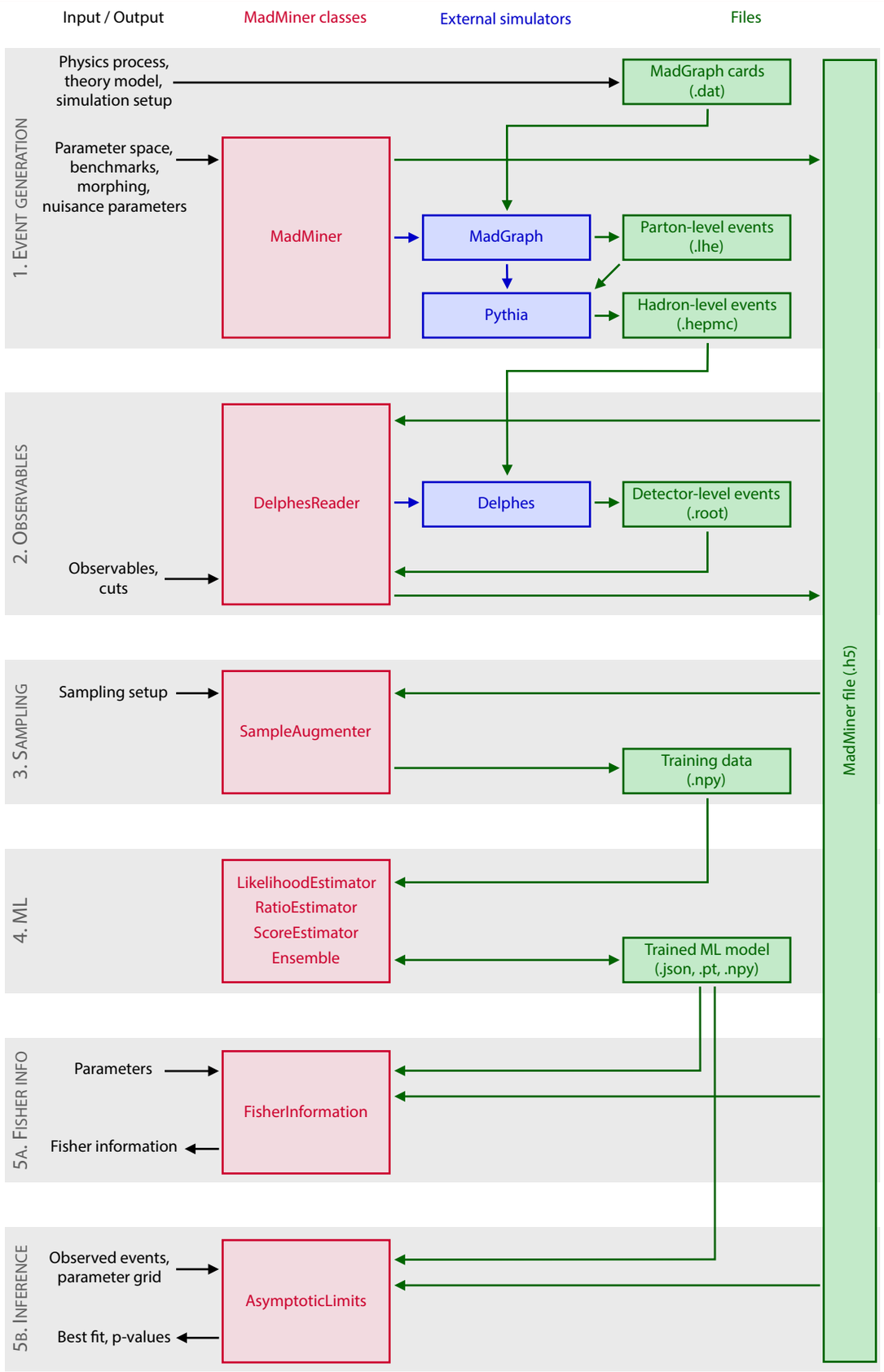
MadMiner [J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

- automizes these techniques
- straightforward to apply them to LHC problems

- out of the box: Pheno-level analysis
  * MadGraph, Pythia, Delphes
  * backgrounds
  * PDF/scale uncertainties
  * ML uncertainties
  * morphing
  * many inference techniques (SALLY, ALICES …)

- scalable to state-of-the-art experimental tools

- python package
  * modular interface
  * extensive documentation
  * on GitHub
    github.com/diana-hep/madminer
  * easy to install
    `pip install madminer`



Physics Example

Probing SMEFT in tth

Probing SMEFT in tth

The Physics Process

- processes
  * signal: fully leptonic tth $pp \to t\bar{t}h \to \left(b\ell^+\right)\left(\bar{b}\ell^-\right)(\gamma\gamma)E_T^{\text{miss}}$
  * background: ttγγ continuum
  * 24.5 tth and 33.6 ttγγ events

Disclaimer: This is not the most sensitive process, but it demonstrates many features in MadMiner.

- theory
  * SMEFT model: $\mathcal{L} = \mathcal{L}_{SM} + c_u\mathcal{O}_u + c_G\mathcal{O}_G + c_{uG}\mathcal{O}_{uG}$

$$\mathcal{O}_u = -\frac{1}{v^2}\left(H^\dagger H\right)\left(H^\dagger \bar{Q}_L\right)u_R$$ - rescales top Yukawa: $y_t \to y_t\,(1+3/2\,c_u)$

$$\mathcal{O}_G = \frac{g_s^2}{m_W^2}\left(H^\dagger H\right)G^a_{\mu\nu}G_a^{\mu\nu}$$ - Higgs-gluon coupling: $g_{ggh} \to g_{ggh}(1 + 192\pi^2/g^2\,c_G)$

$$\mathcal{O}_{uG} = -\frac{4g_s}{m_W^2}y_u\left(H^\dagger \bar{Q}_L\right)\gamma^{\mu\nu}T_a u_R G^a_{\mu\nu}$$ - chromo-dipole moment: gtt, ggtt, gtth, ggtth

- simulation: MadGraph5 + Pythia8 + Delphes3 + HL-LHC setup
  * PDF4LHC: scale + PDF uncertainties

- 48 observable: x={px, py, pz, E, pT, η, ΔΦ, MET, m_{ij}, …}

Probing SMEFT in tth

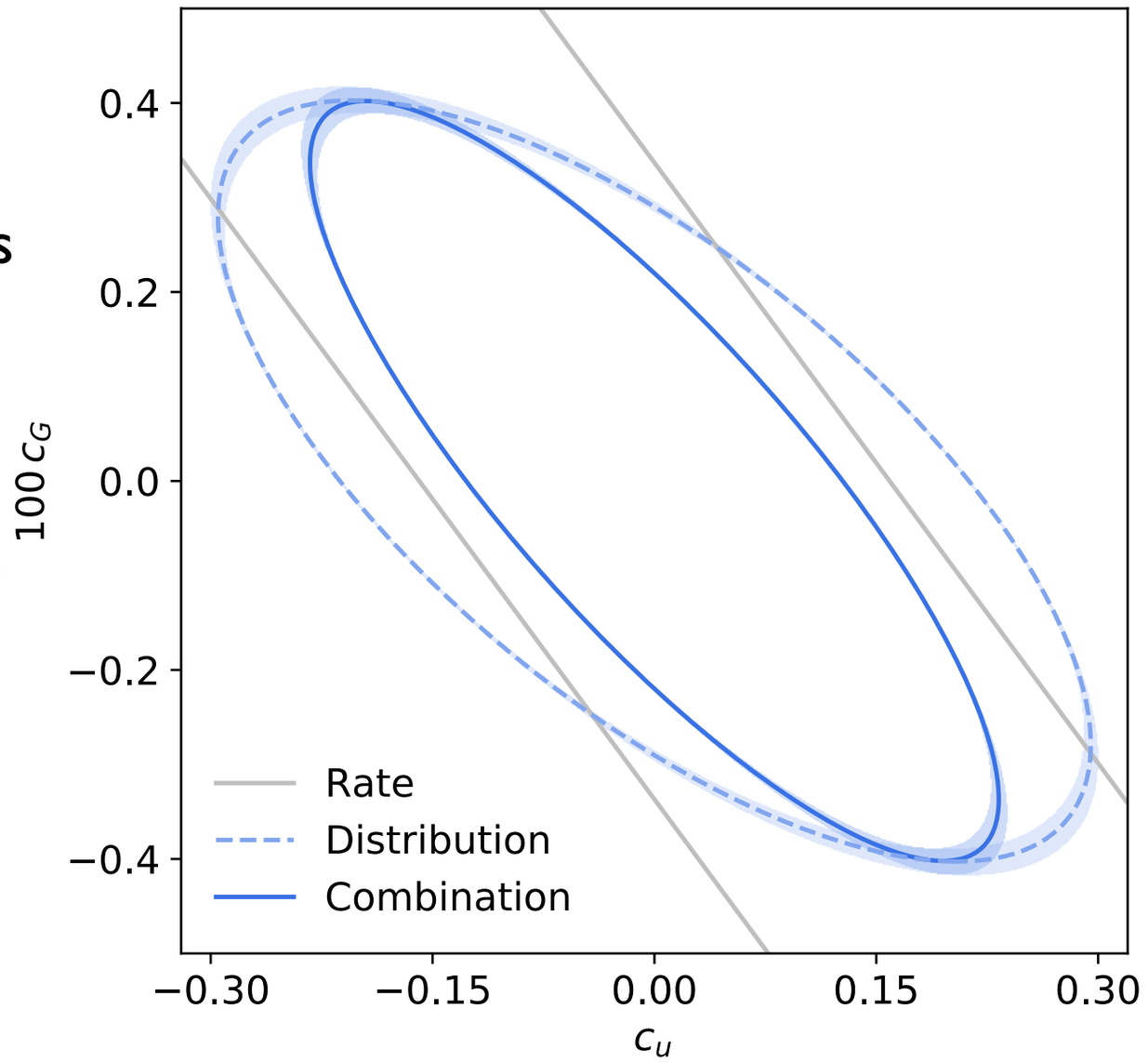
Fisher Information Analysis:

- Fisher Information matrix:
  * simple: 3x3 matrix
  * describes all operators simultaneously
  * only sensitive to interference effects

$$I_{ij} = \begin{array}{c} \begin{matrix} c_u & 100c_G & 100c_{uG} \end{matrix} \\ \begin{pmatrix} 140.5 & 68.1 & 170.6 \\ 68.1 & 47.1 & 105.7 \\ 170.6 & 105.7 & 283.3 \end{pmatrix} \end{array} \begin{matrix} c_u \\ 100c_G \\ 100c_{uG} \end{matrix}$$

- translate into limits
  * 68% CL limits
  * separate information in rate+distributions

- results
  * rate: constrains only one direction
  * distributions: complementary information
  * shaded band: 2σ ML ensemble variance

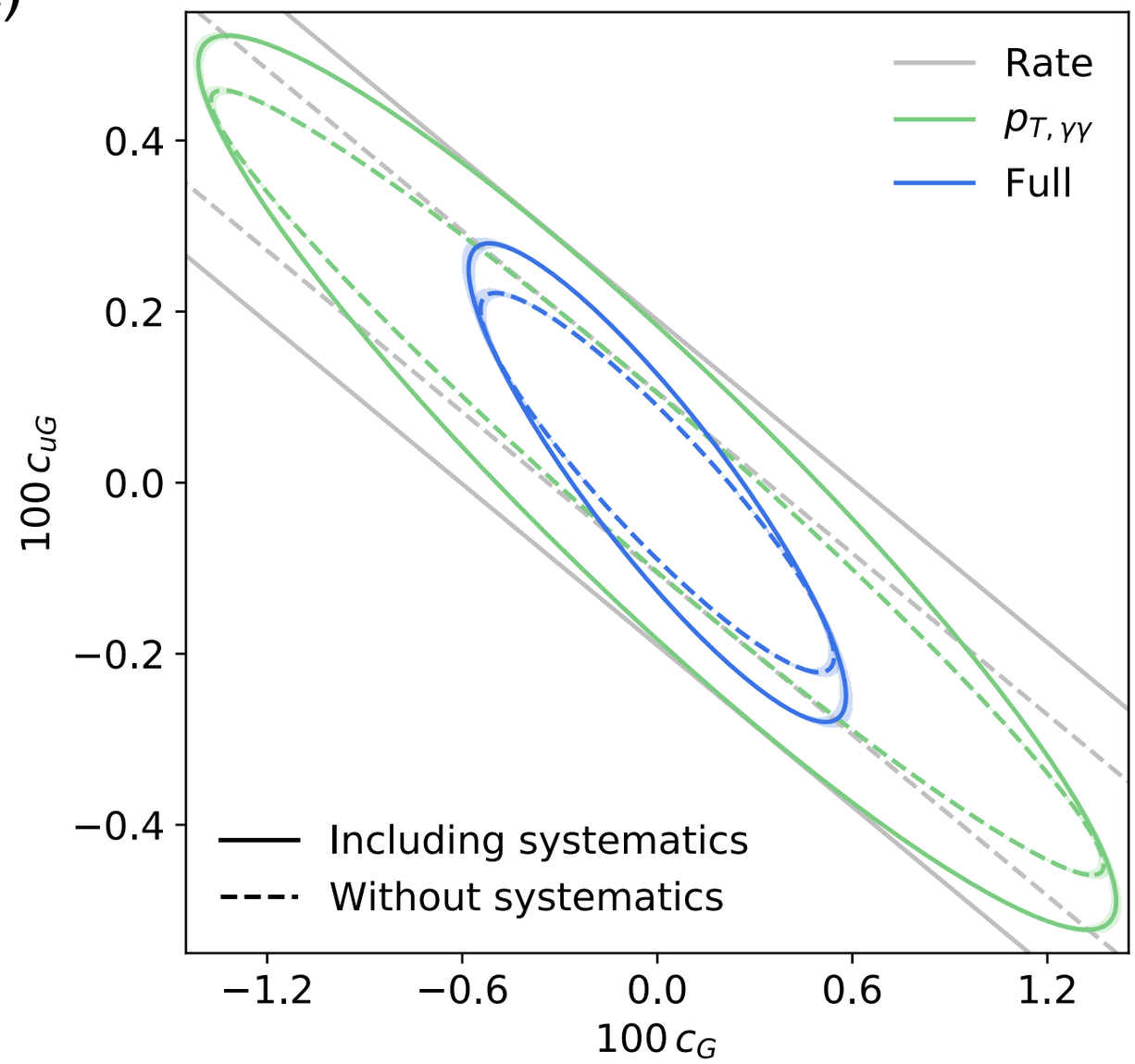


Probing SMEFT in tth

How about Systematics?

- strategy
  * introduce nuisance parameter ν for scale (2) and PDF (30) uncertainties
  * replace r(x|θ)→r(x|θ,ν)
  * learn score, obtain Fisher Info (3+32 dim.)
  * profile over ν

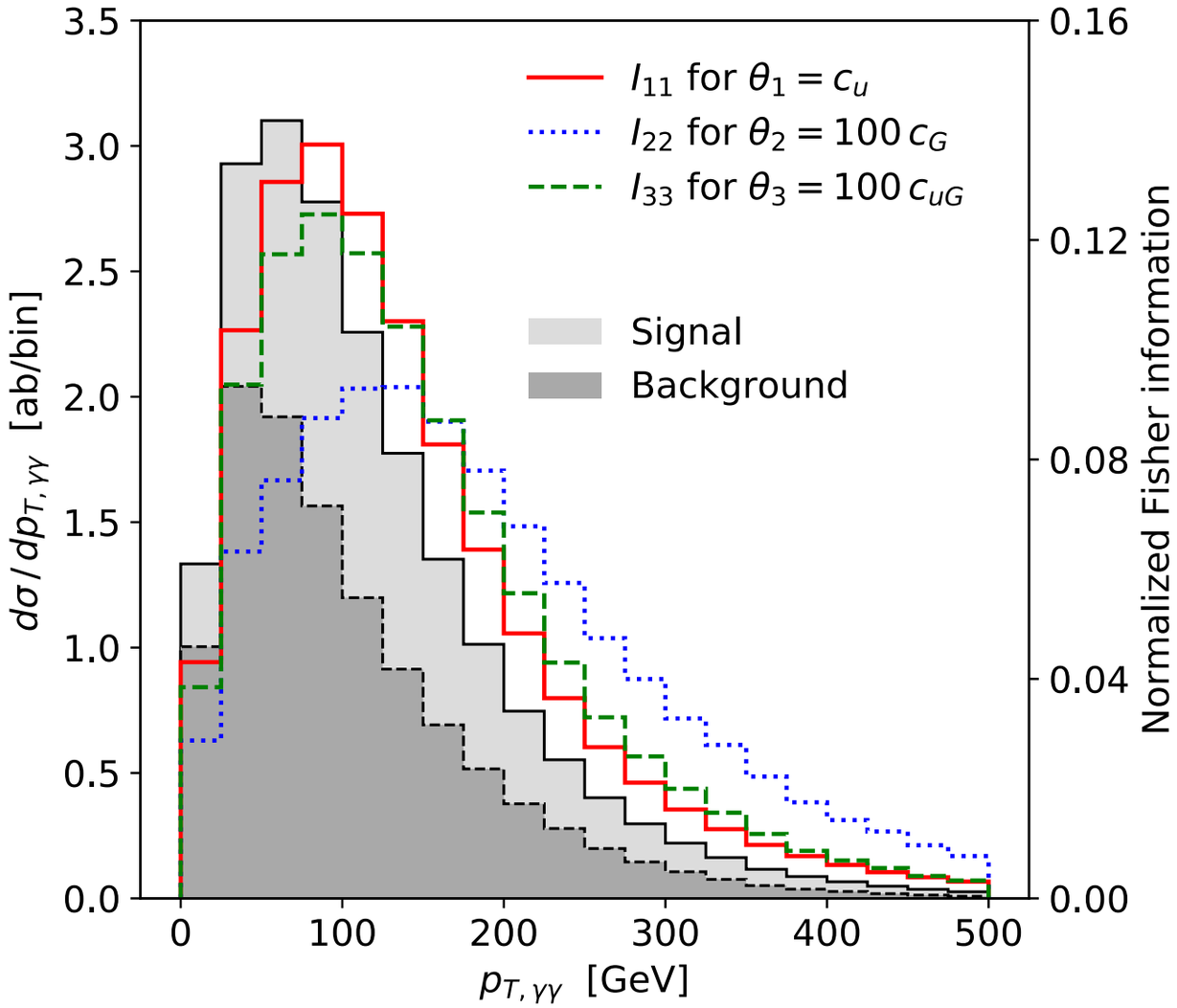
- results:
  * mainly scale uncertainties
  * systematic reduce reach in rate-sensitive direction
  * multivariate analysis less affected

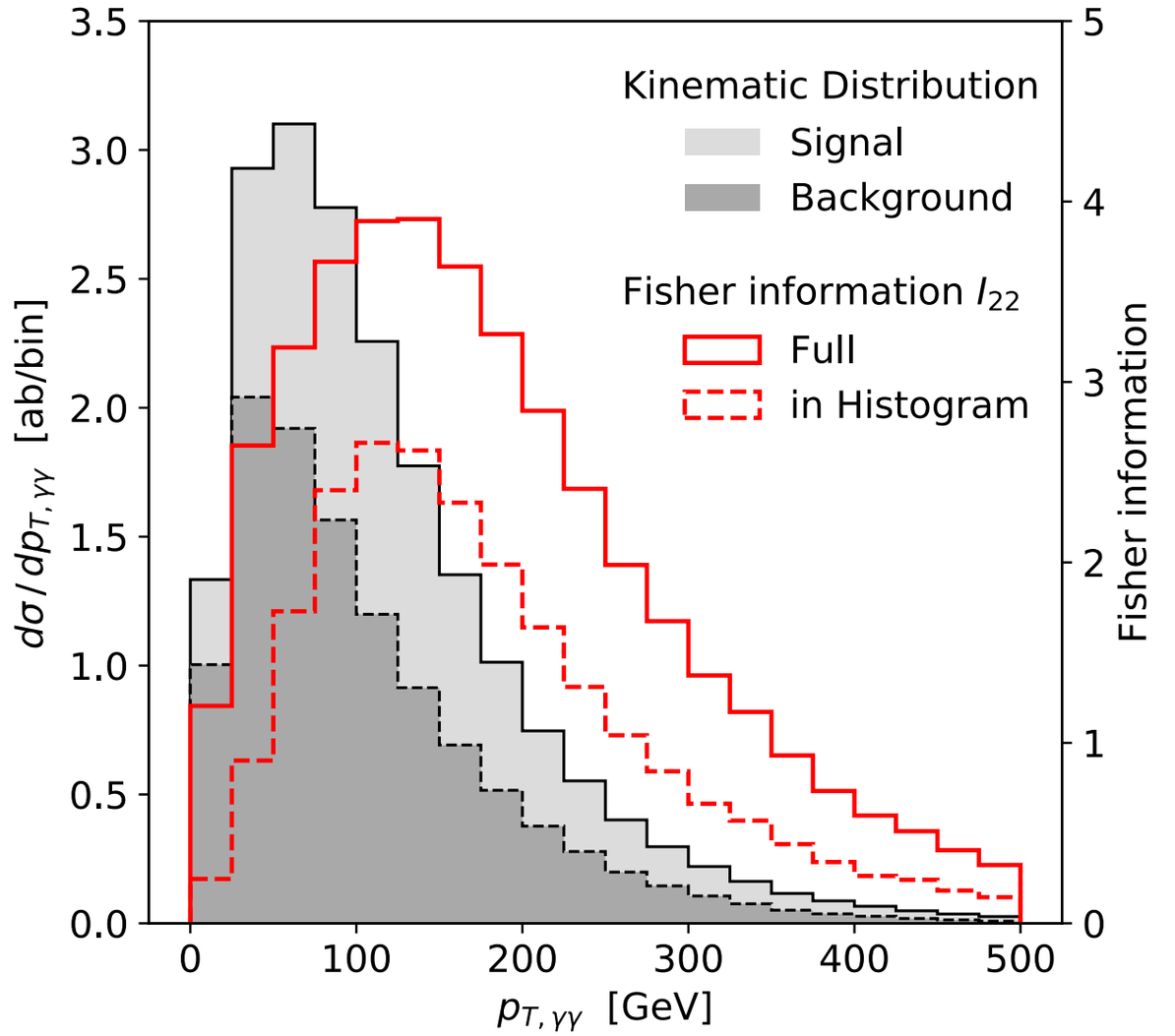


Probing SMEFT in tth

Kinematic distribution of the Fisher information

- identify the important phase-space regions





Probing SMEFT in tth

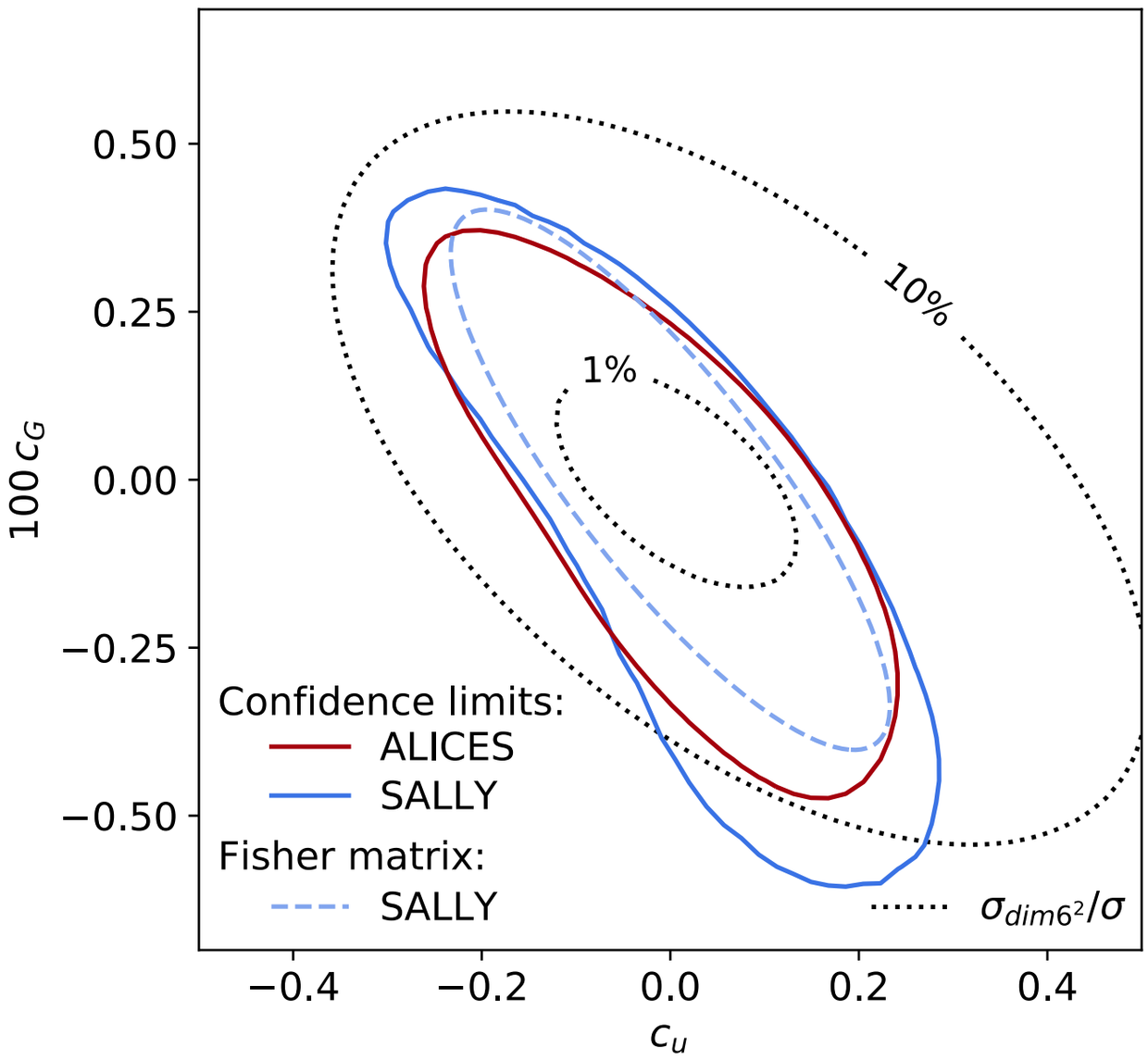
How good is local approximation?

- Fisher Information matrix
  * estimate
  * only sensitive to interference effects by construction
  * symmetric limits (ellipses)

- SALLY
  * estimates score $t(x)|_{SM}$
  * use score as optimal observable (filled in histograms)
  * optimal only close to SM

- ALICES:
  * estimate $r(x|\theta)$
  * optimal limits in whole parameter space

- this example analysis:
  * few data, weak constraints
  * dim6 squared terms important in probed parameter space
  * local / full limits differ



Probing SMEFT in tth

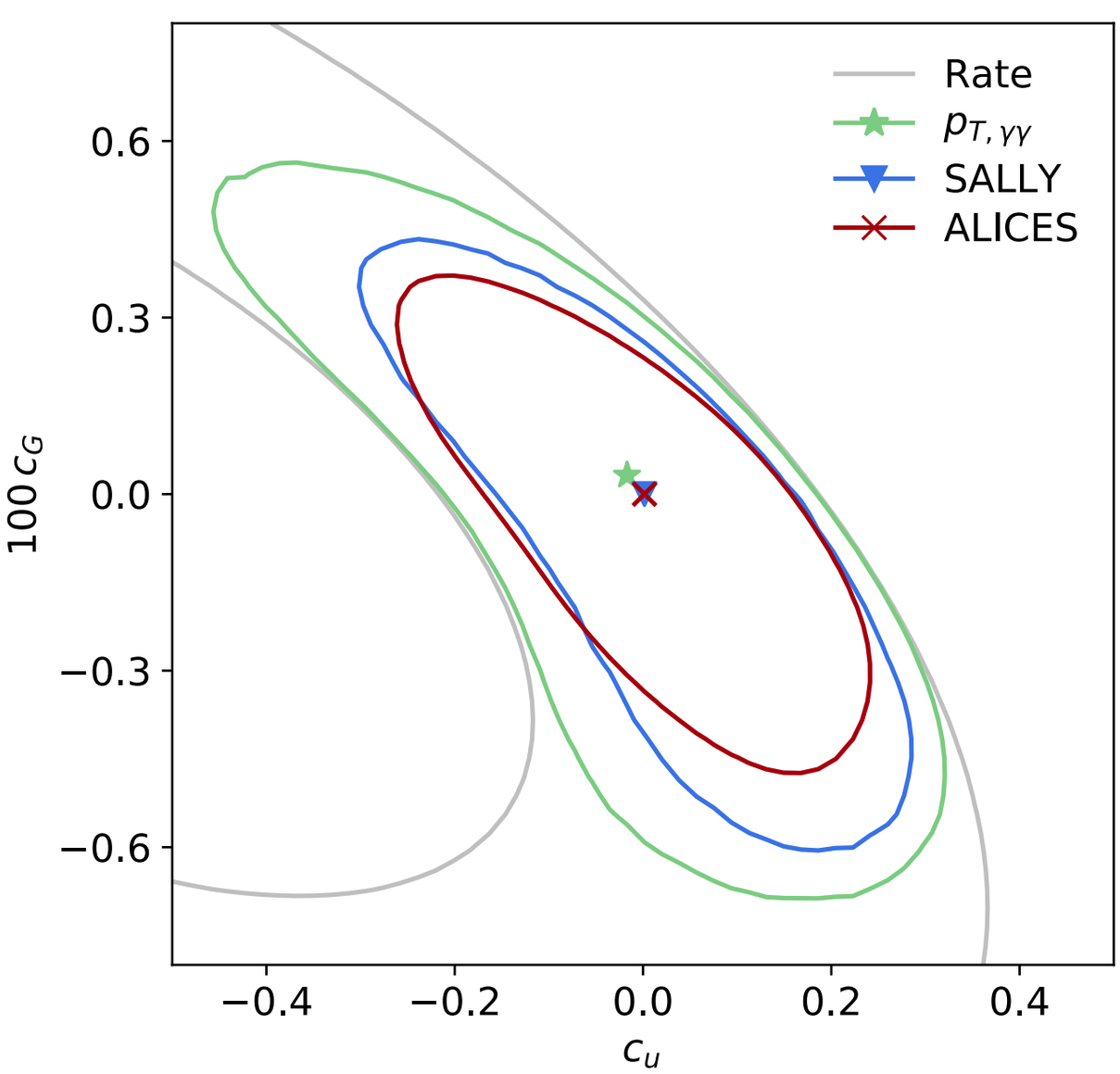
Full Results

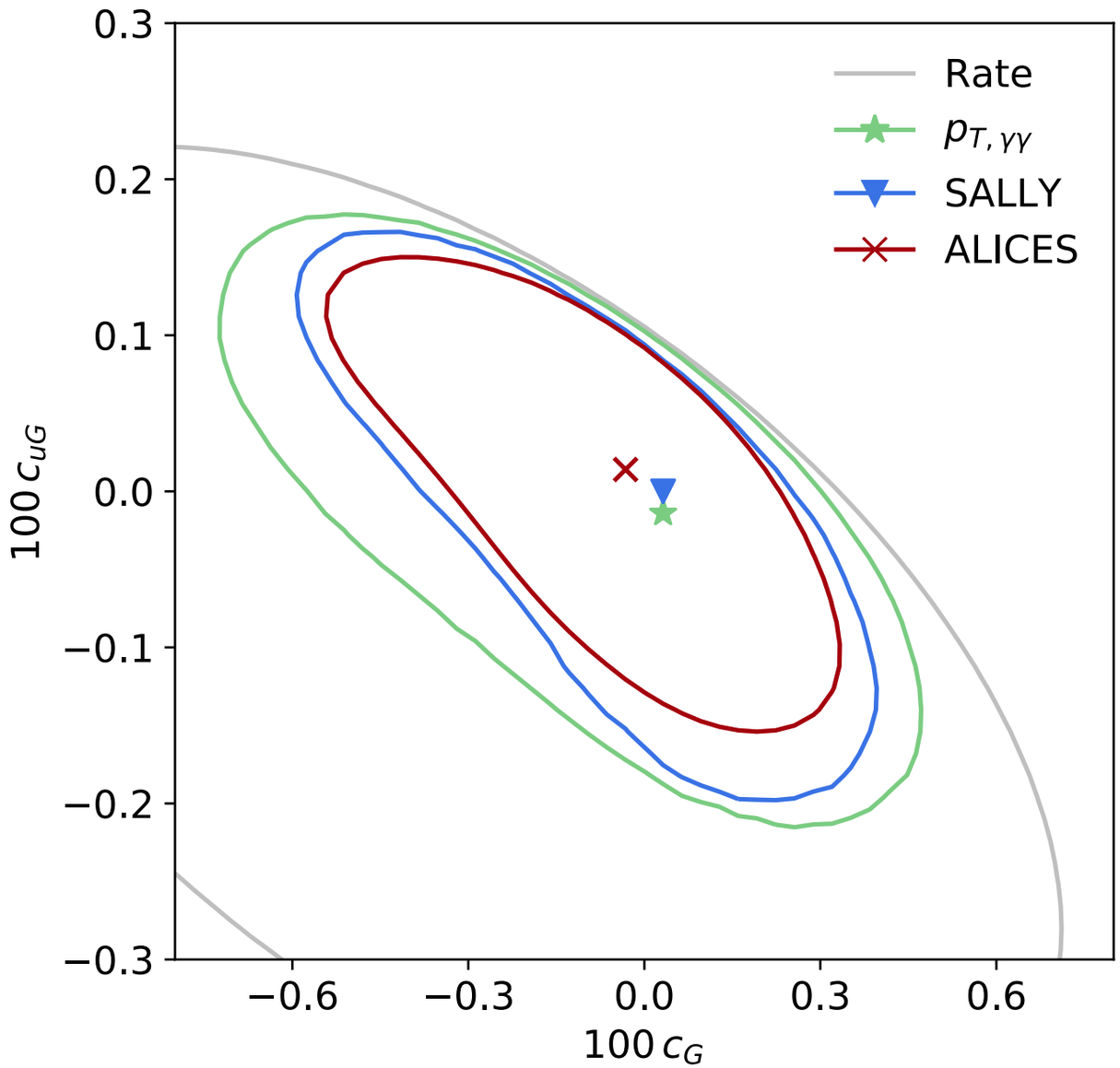
- inference methods

  total rate — SALLY: score $t(x)|_{SM}$ as locally optimal observables

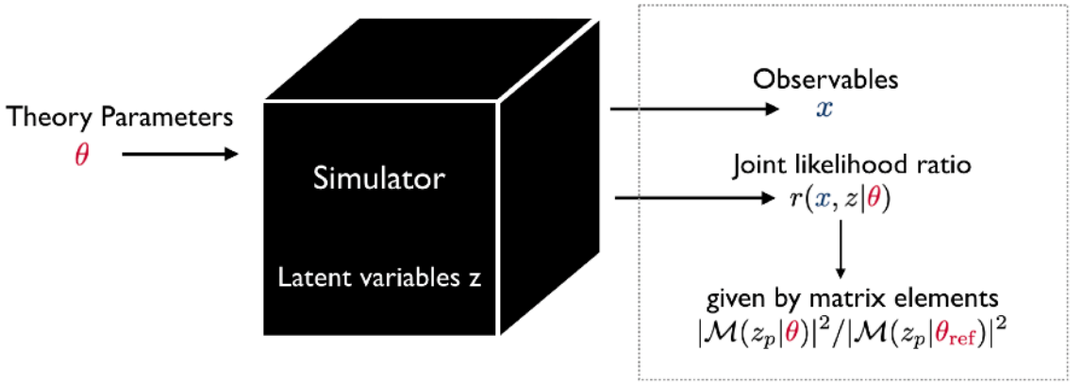
  pTH histogram — ALICES: use full $r(x|\theta)$

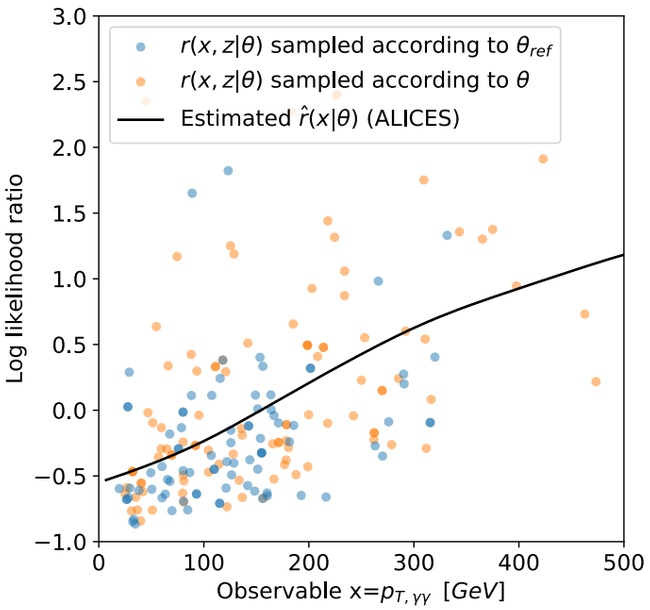
- multivariate methods significantly improve reach

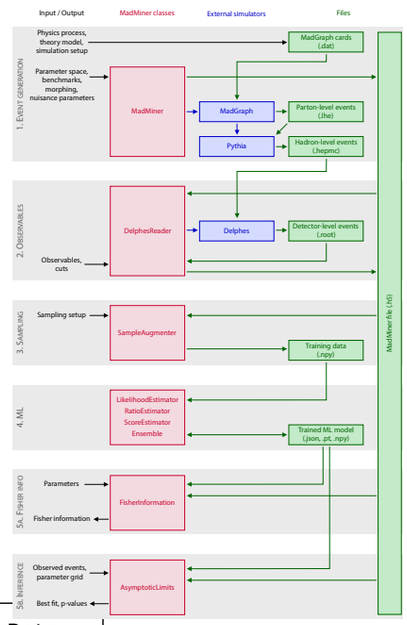


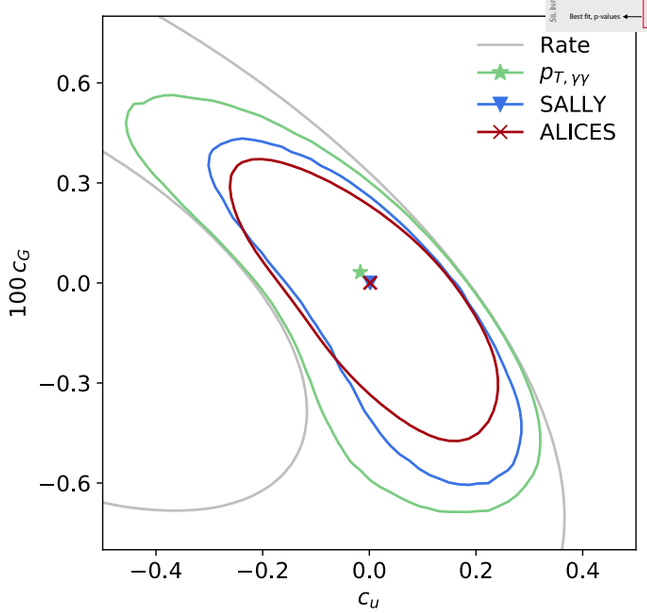


Summary and Conclusion









Motivation: LHC probe high-dimensional theory space θ with high dimensional data x

* task: determine likelihood function p(x|θ)

Method: uses multivariate inference techniques leveraging information in matrix elements and power of machine learning

* estimate full likelihood ratio including correlations + systematics
* learn optimal summary statistics
* ideal for SMEFT measurements

Tool: MadMiner [Brehmer, Cranmer, Espejo, FK 1907.10621]

* python package
* MadGraph add-on
* automizes all steps of analysis
* already used for Pheno studies

[Brehmer, Dawson, Homiller FK, Plehn, 1908.06980]

Summary and Conclusion

MadMiner Team

Johann Brehmer

Kyle Cranmer

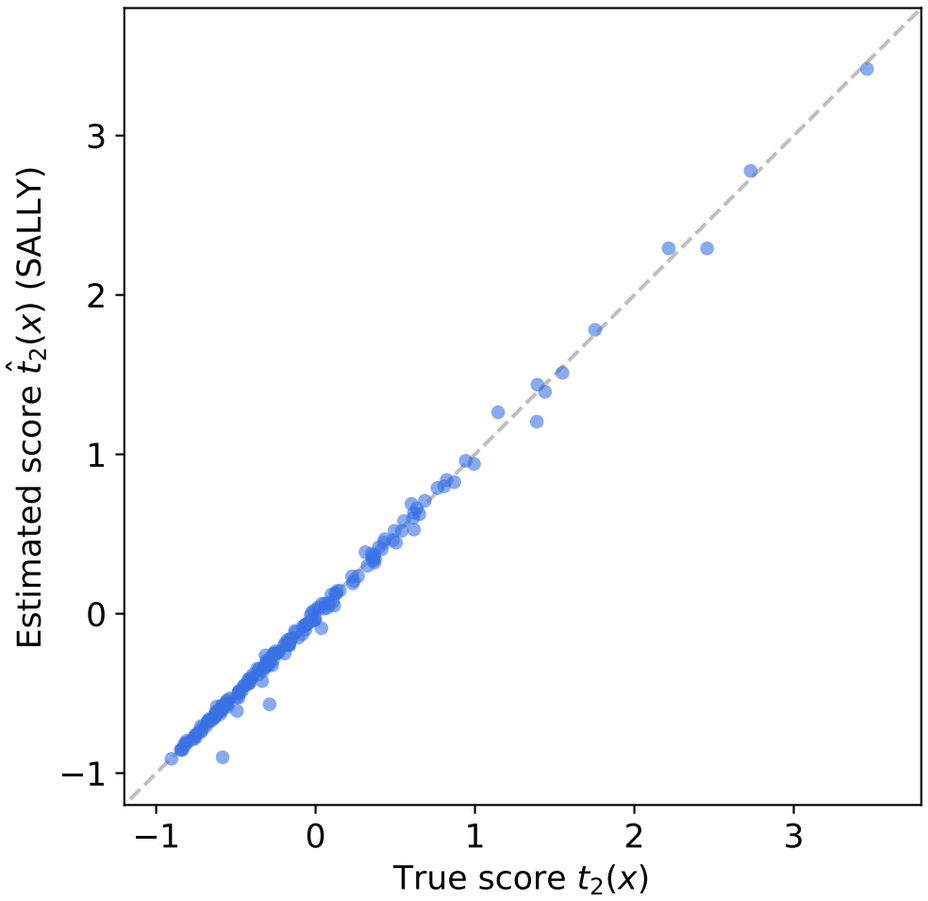
Irina Espejo

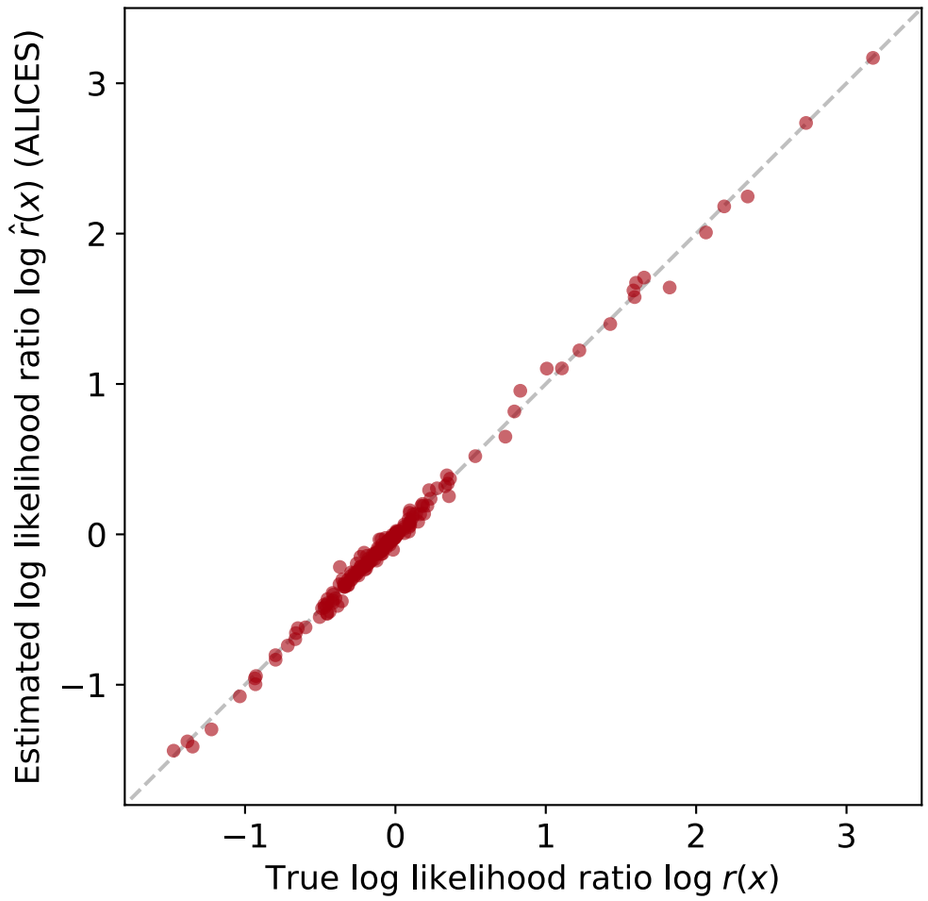
Felix Kling

Already used by many people

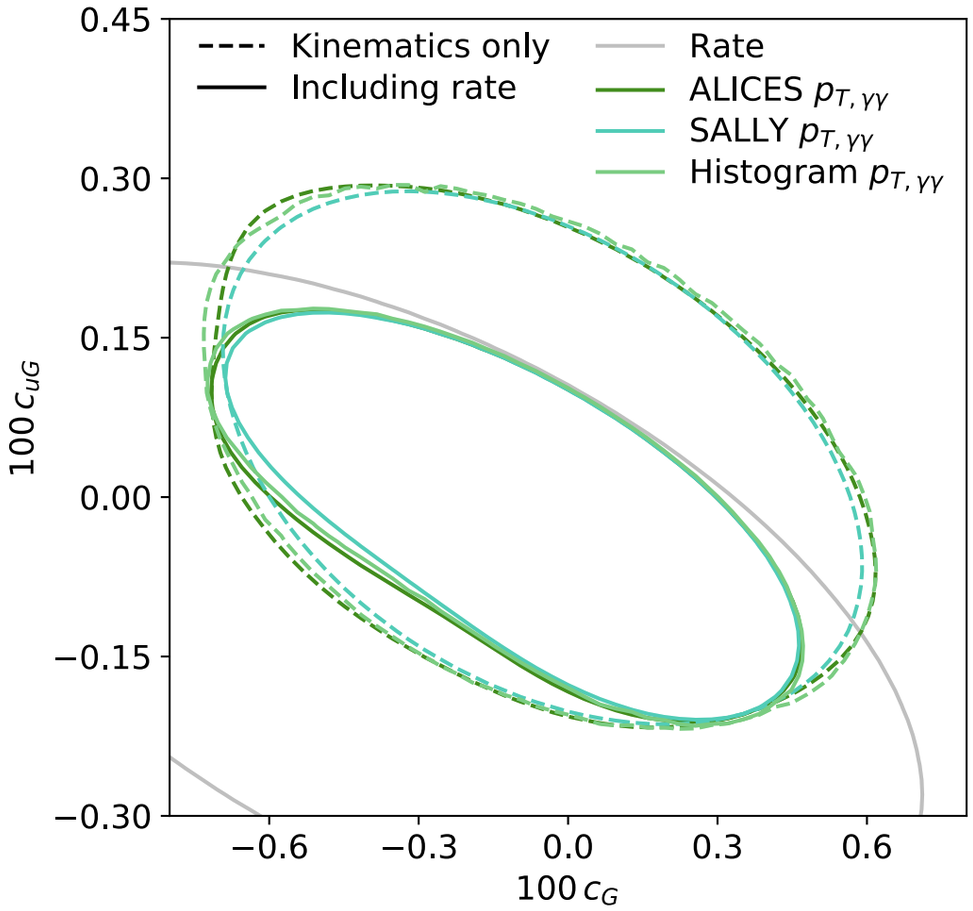
Thank you a lot for testing!

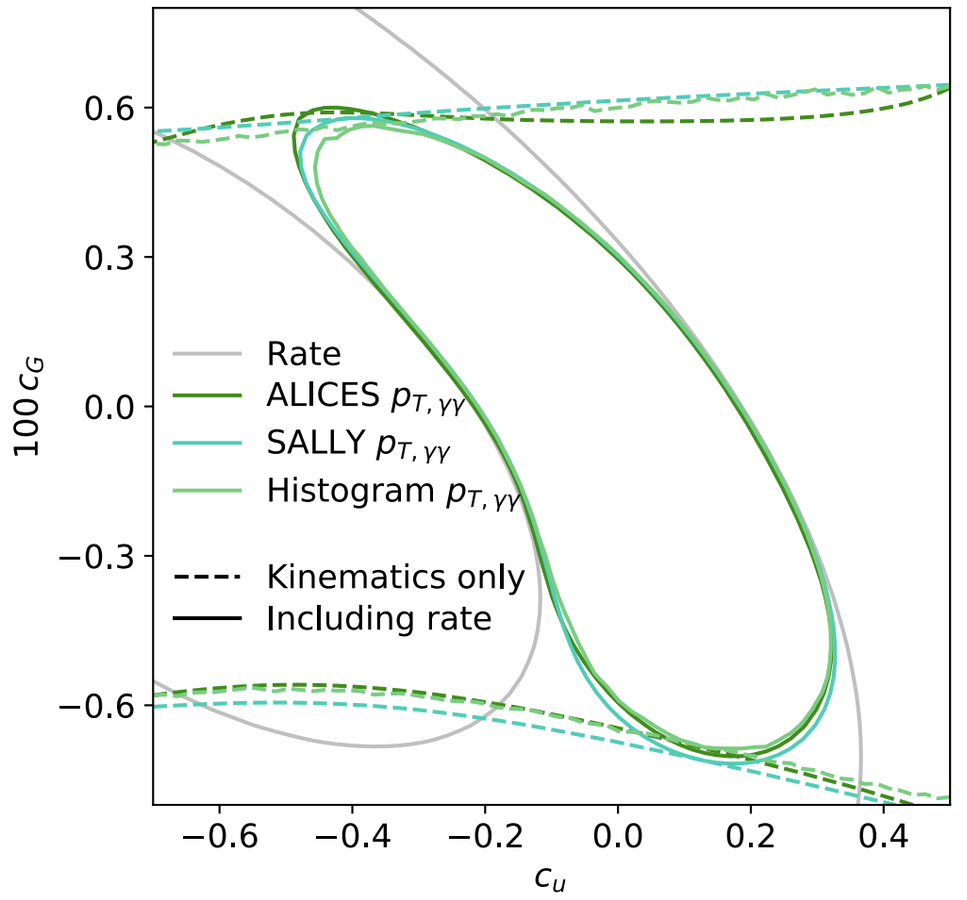
Backup


Estimated score $\hat{t}_2(x)$ (SALLY)
True score $t_2(x)$
−1
0
1
2
3

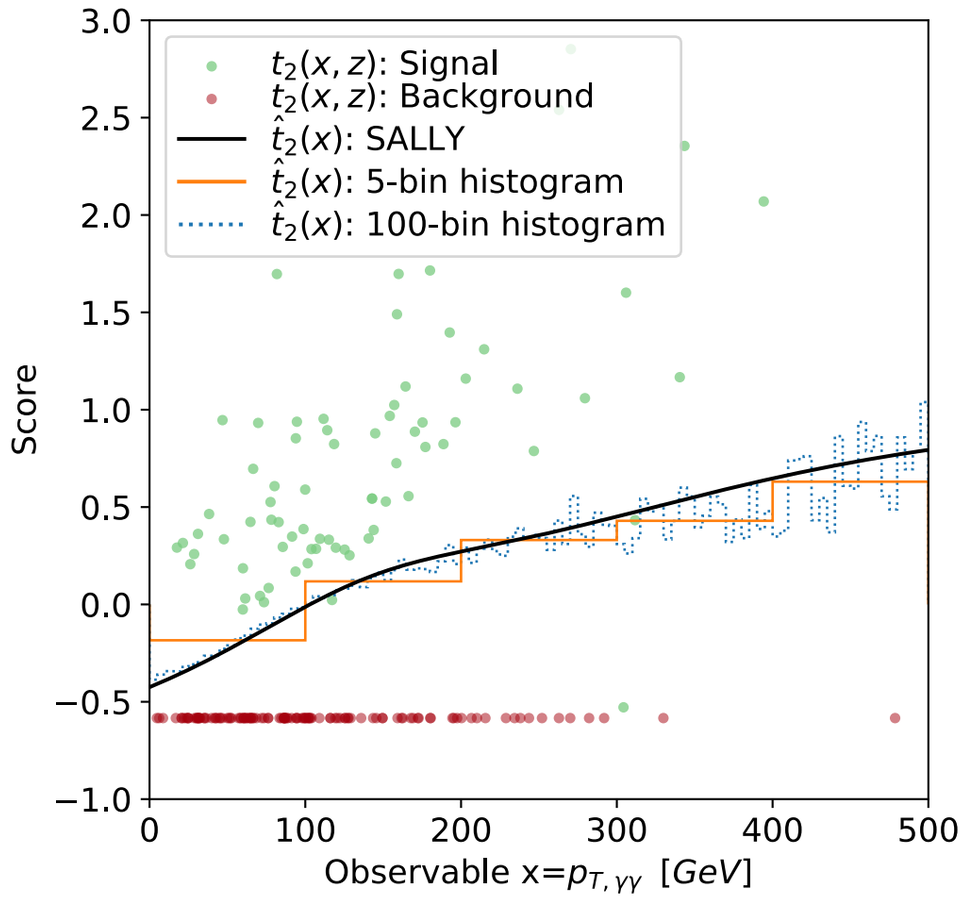

Estimated log likelihood ratio log $\hat{r}(x)$ (ALICES)
True log likelihood ratio log $r(x)$
−1
0
1
2
3

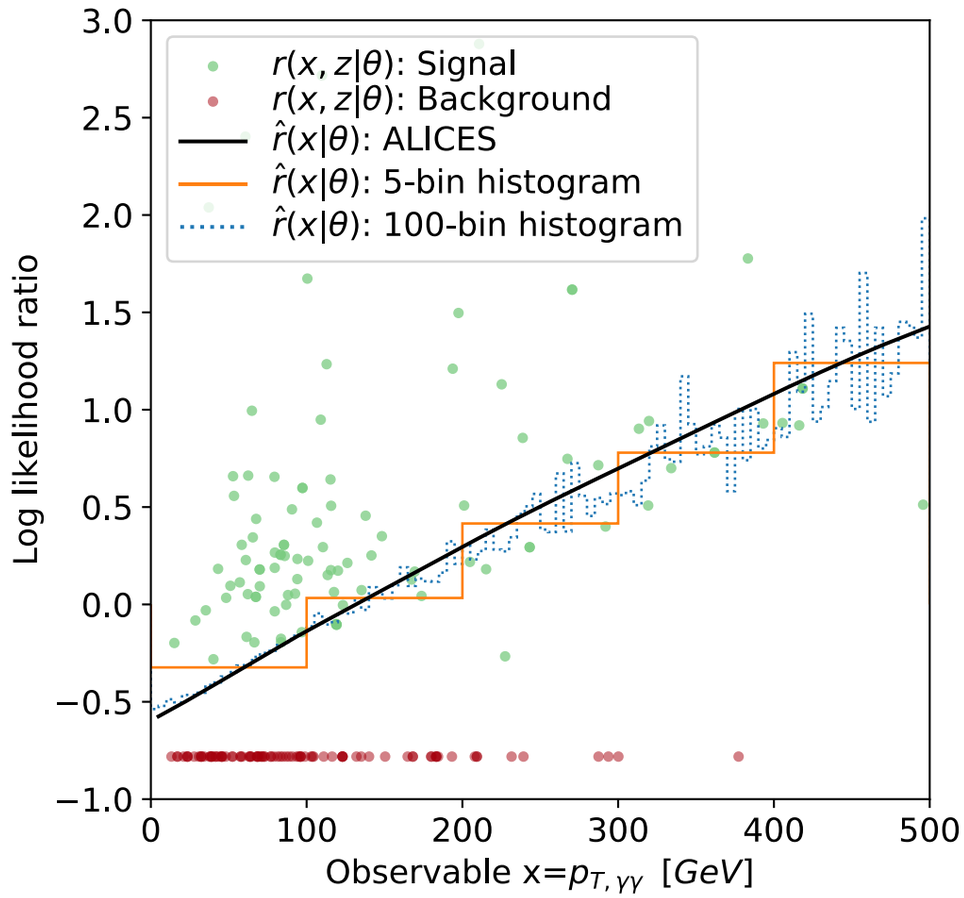
Backup


Kinematics only
Including rate
Rate
ALICES $p_{T,\gamma\gamma}$
SALLY $p_{T,\gamma\gamma}$
Histogram $p_{T,\gamma\gamma}$
0.45
0.30
0.15
0.00
−0.15
−0.30
$100\,c_{uG}$
−0.6
−0.3
0.0
0.3
0.6
$100\,c_G$


Rate
ALICES $p_{T,\gamma\gamma}$
SALLY $p_{T,\gamma\gamma}$
Histogram $p_{T,\gamma\gamma}$
Kinematics only
Including rate
0.6
0.3
0.0
−0.3
−0.6
$100\,c_G$
−0.6
−0.4
−0.2
0.0
0.2
0.4
c_u

Backup


$t_2(x,z)$: Signal
$t_2(x,z)$: Background
$\hat{t}_2(x)$: SALLY
$\hat{t}_2(x)$: 5-bin histogram
$\hat{t}_2(x)$: 100-bin histogram
Score
Observable x=$p_{T,\gamma\gamma}$ [GeV]


$r(x,z|\theta)$: Signal
$r(x,z|\theta)$: Background
$\hat{r}(x|\theta)$: ALICES
$\hat{r}(x|\theta)$: 5-bin histogram
$\hat{r}(x|\theta)$: 100-bin histogram
Log likelihood ratio
Observable x=$p_{T,\gamma\gamma}$ [GeV]