Constraining EFTs with Machine Learning

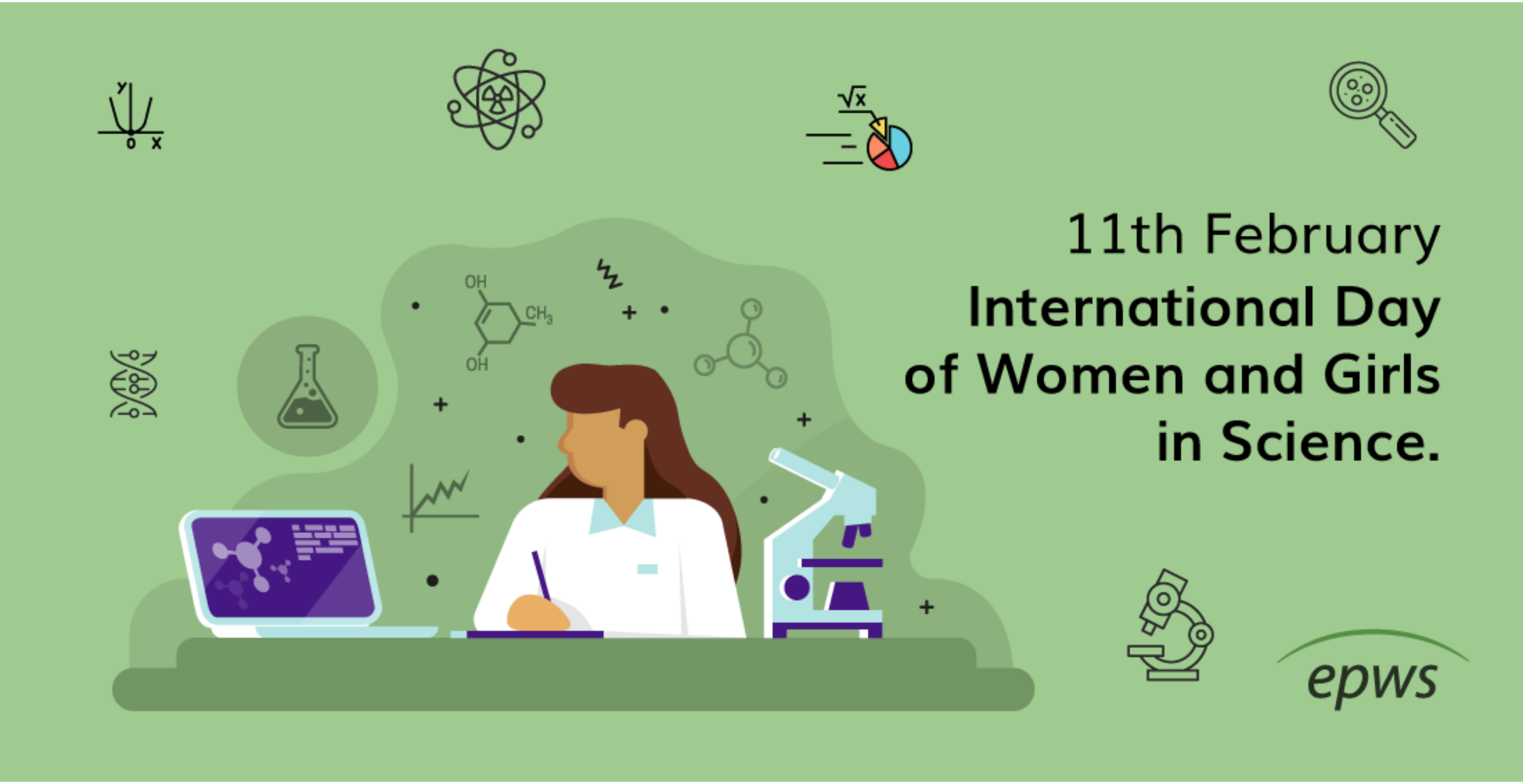
$\downarrow$

MadMiner

Felix Kling

SLAC NATIONAL ACCELERATOR LABORATORY

11 February 2020
UC Davis


11th February
International Day
of Women and Girls
in Science.
epws

Motivation

Status of the Field:

- Higgs discovery: Standard Model complete
- no discovery of new physics yet

Motivation

Status of the Field:

- Higgs discovery: Standard Model complete
- no discovery of new physics yet

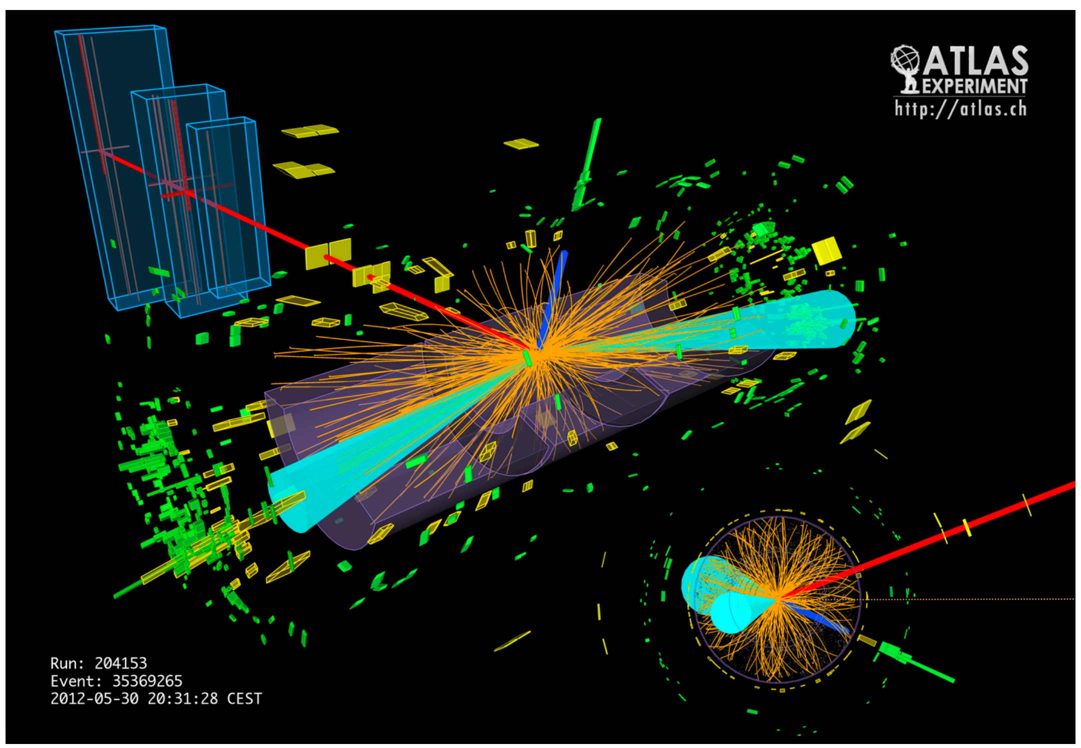
- maybe BSM physics will be obvious (di-X resonance at XXX GeV)
- maybe BSM physics is more subtle (Higgs couplings in dim6 SMEFT)

Motivation

Status of the Field:

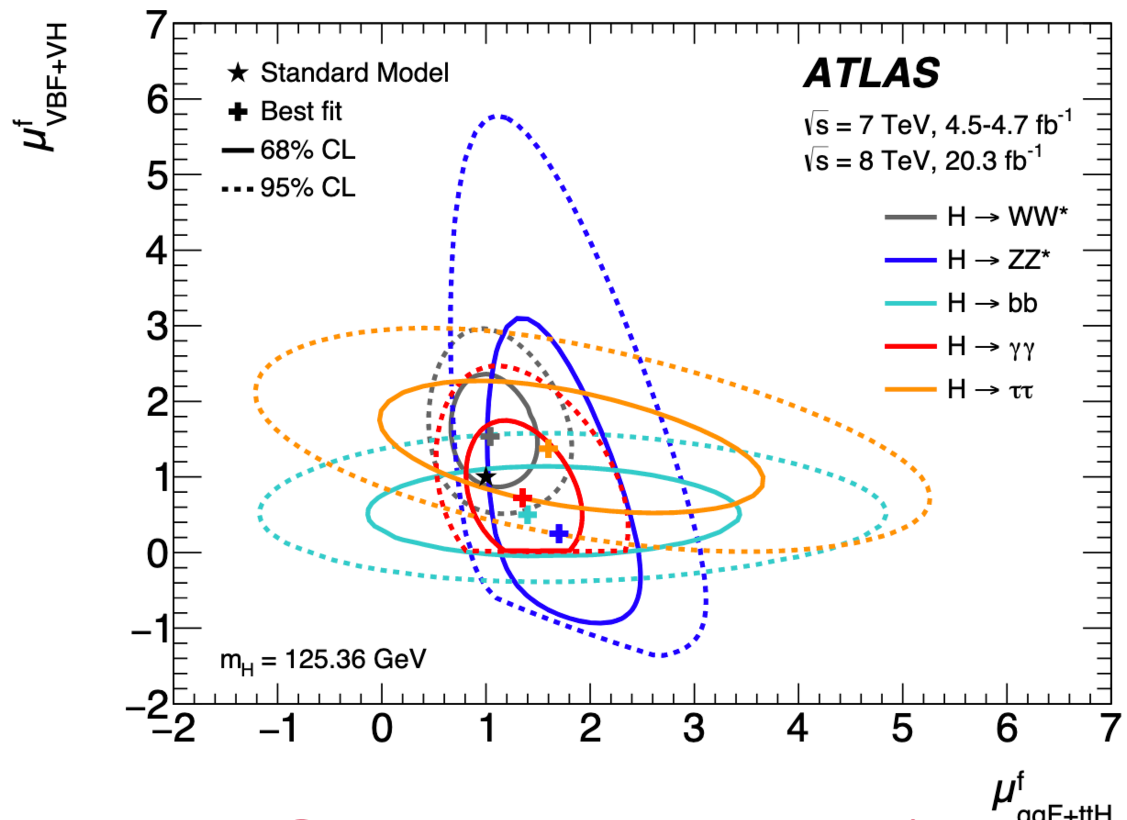
- Higgs discovery: Standard Model complete
- no discovery of new physics yet

- maybe BSM physics will be obvious (di-X resonance at XXX GeV)
- maybe BSM physics is more subtle (Higgs couplings in dim6 SMEFT)



Era of Data

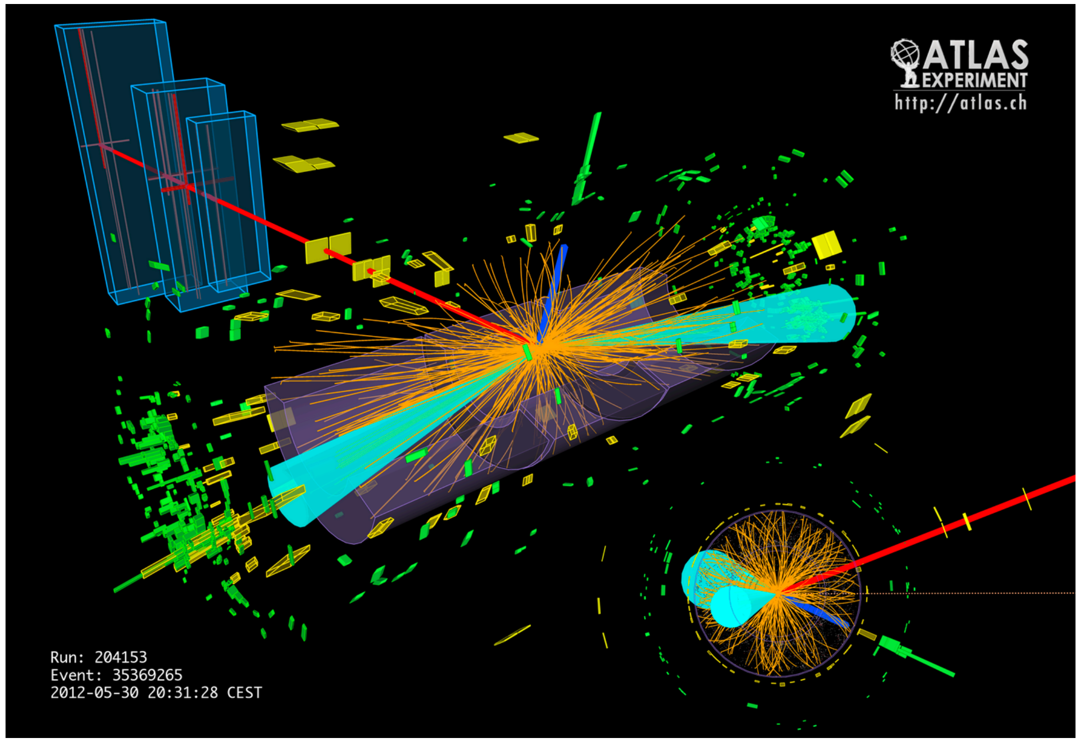
high statistics, many distributions, multivariate analyses



Complex Theories

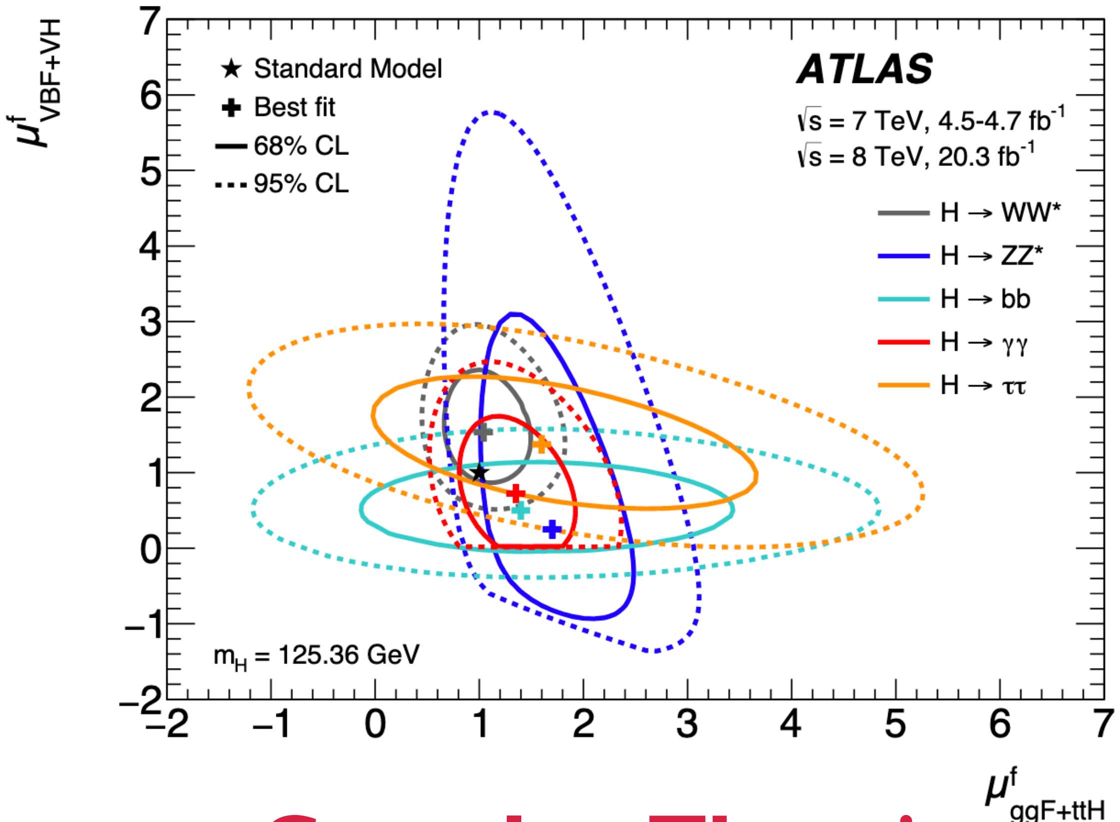
large number of theory parameters, predict subtle kinematic features

How to test Theories in an Era of Data?



Era of Data

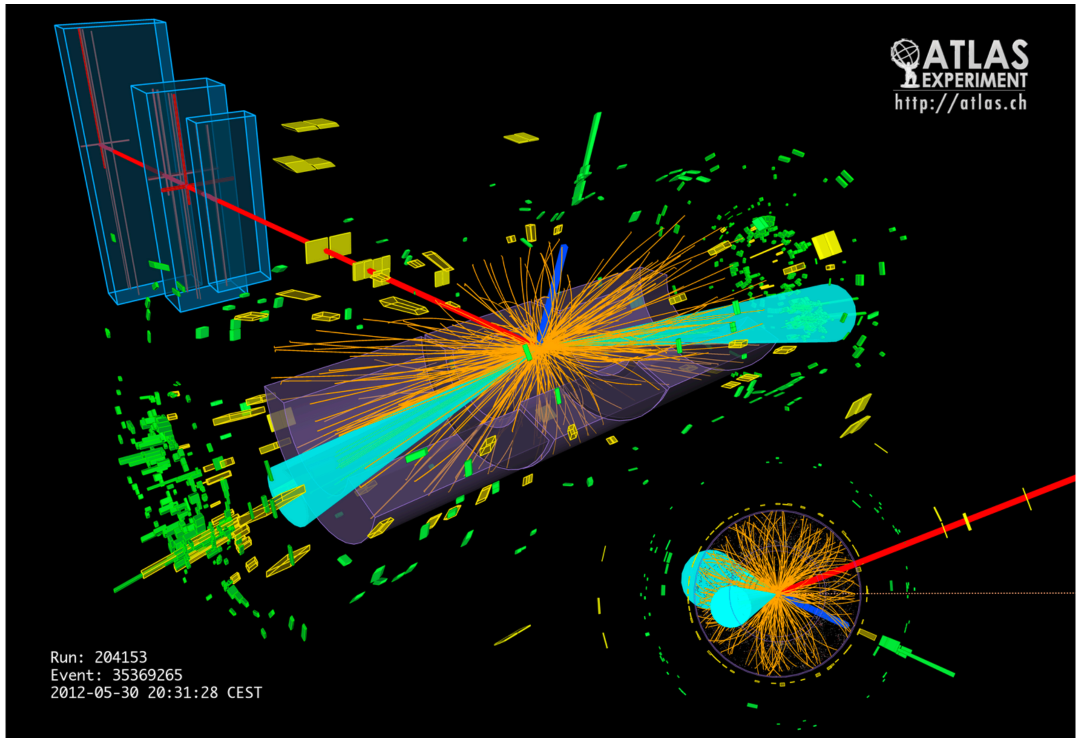
high statistics, many distributions, multivariate analyses

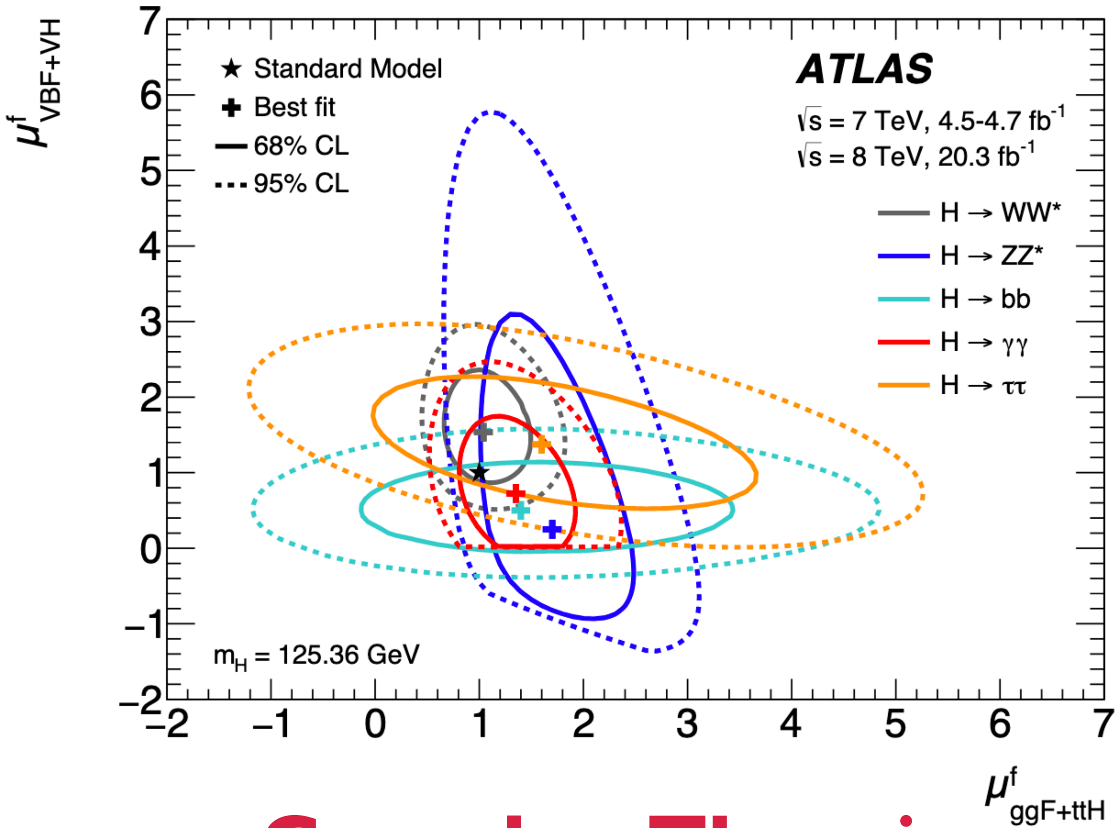


Complex Theories

large number of theory parameters, predict subtle kinematic features

How to extract all Information from the Data?





Era of Data

high statistics, many distributions, multivariate analyses

Complex Theories

large number of theory parameters, predict subtle kinematic features

Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

Optimal Observables and Fisher Information
This will turn out to be useful.

The MadMiner Tool
Using these methods is super easy!

A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

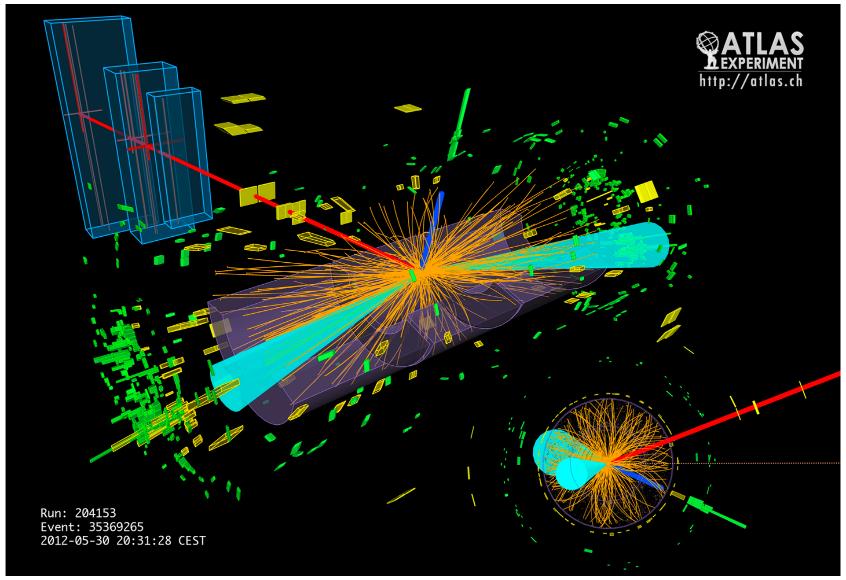
Optimal Observables and Fisher Information
This will turn out to be useful.

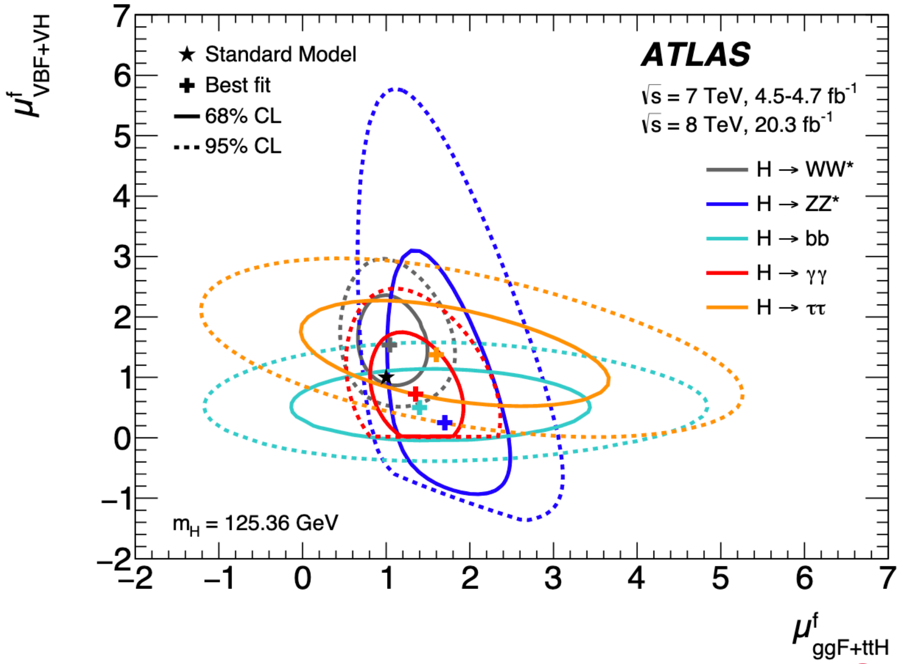
The MadMiner Tool
Using these methods is super easy!

A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

Terminology





Observables x

anything that can be measured

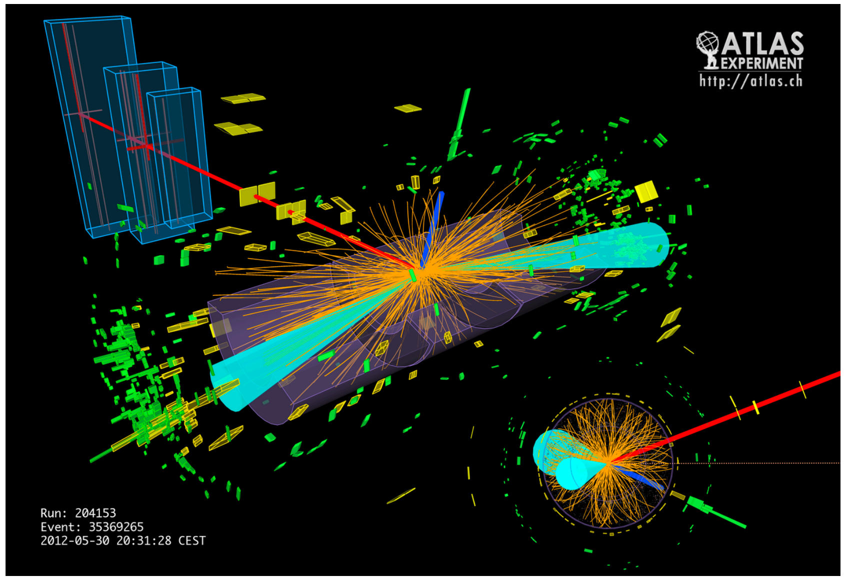
Example: x={E, pT, m, Δφ, … }.

Theory Parameters θ

describes the theory

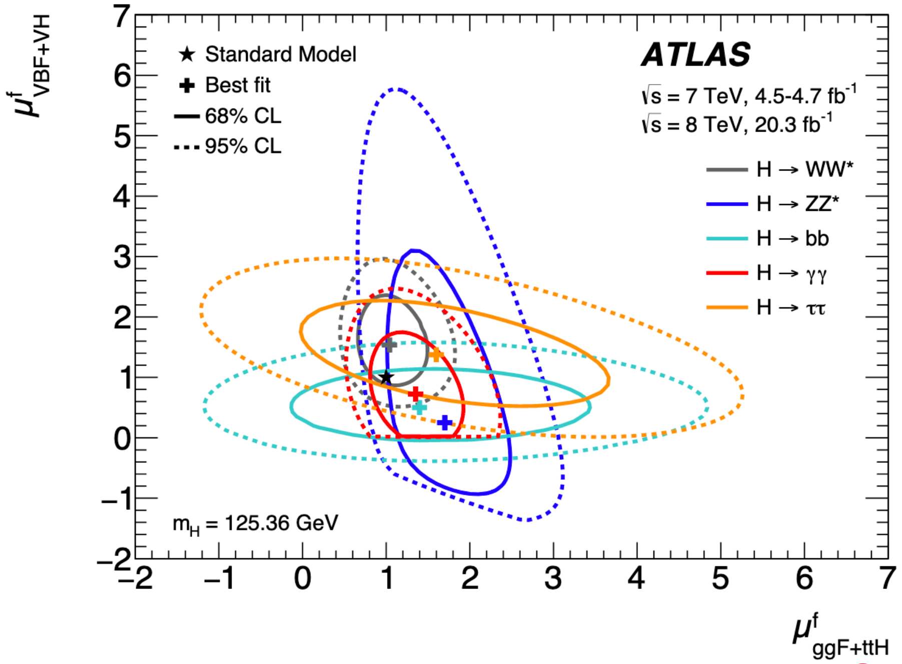
Example: $\mathcal{L} = \sum \theta_i \mathcal{O}_i$.

Terminology



Prediction

Inference



Observables x

anything that can be measured

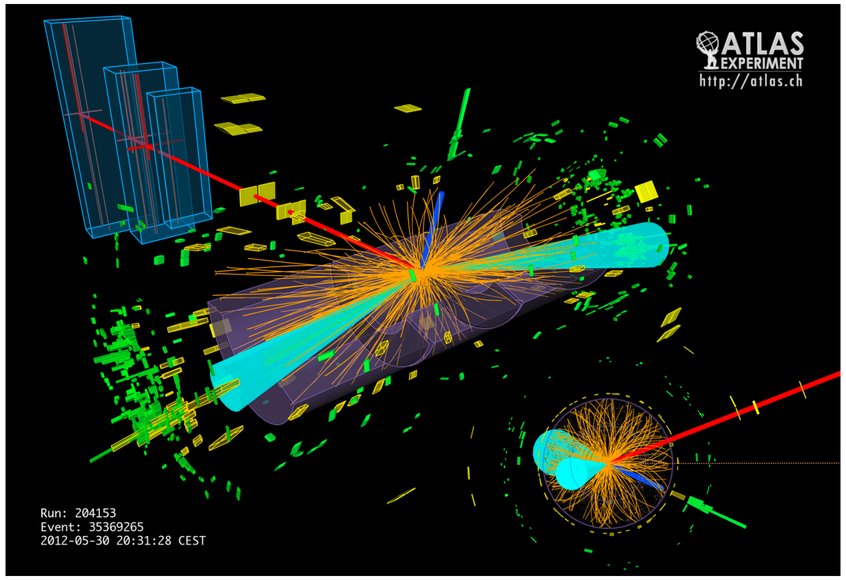
Example: x={E, pT, m, Δφ, … }.

Theory Parameters θ

describes the theory

Example: $\mathcal{L} = \sum \theta_i \mathcal{O}_i$.

Terminology



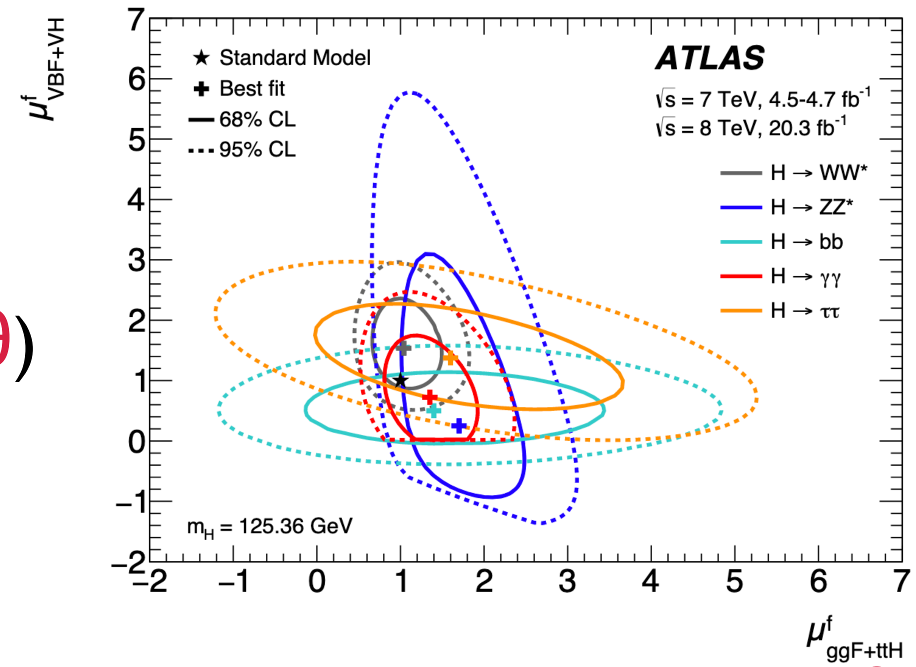
Prediction

←

likelihood function p(x|θ)

→

Inference



Observables x

anything that can be measured

Example: x={E, pT, m, Δφ, … }.

Theory Parameters θ

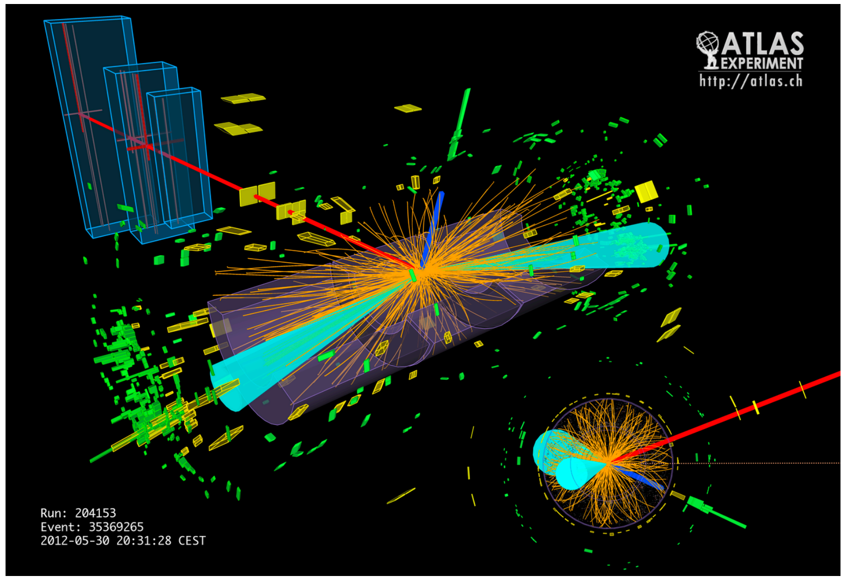
describes the theory

Example: $\mathcal{L} = \sum \theta_i \mathcal{O}_i$.

Likelihood Function p(x|θ)

likelihood of an observation x as a function of the theory parameter θ

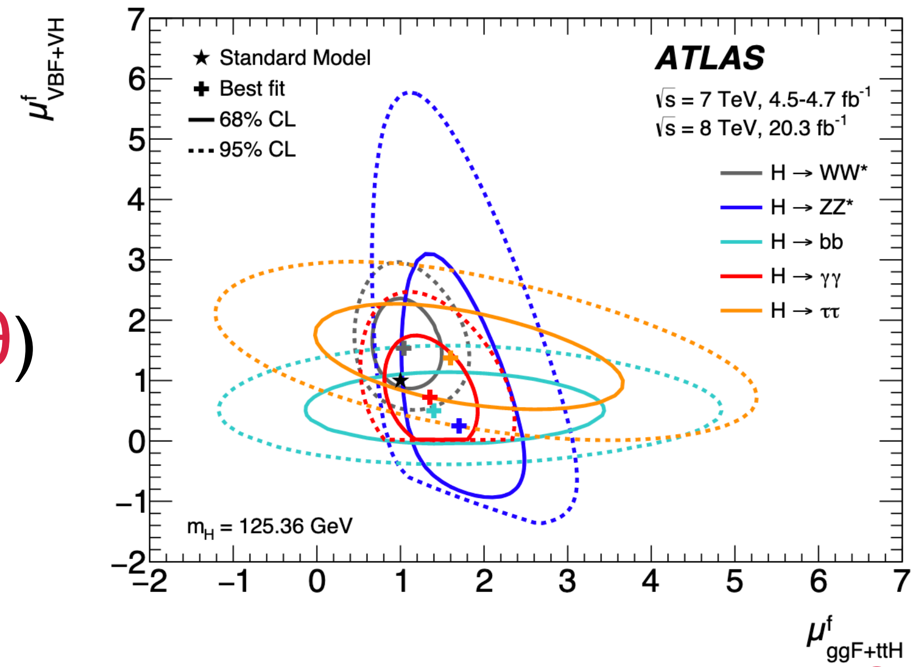
Terminology



Prediction

← likelihood function p(x|θ) →

Inference



Observables x

anything that can be measured

Example: x={E, pT, m, Δφ, … }.

Theory Parameters θ

describes the theory

Example: $\mathcal{L} = \sum \theta_i \mathcal{O}_i$.

Likelihood Function p(x|θ)

likelihood of an observation x as a function of the theory parameter θ

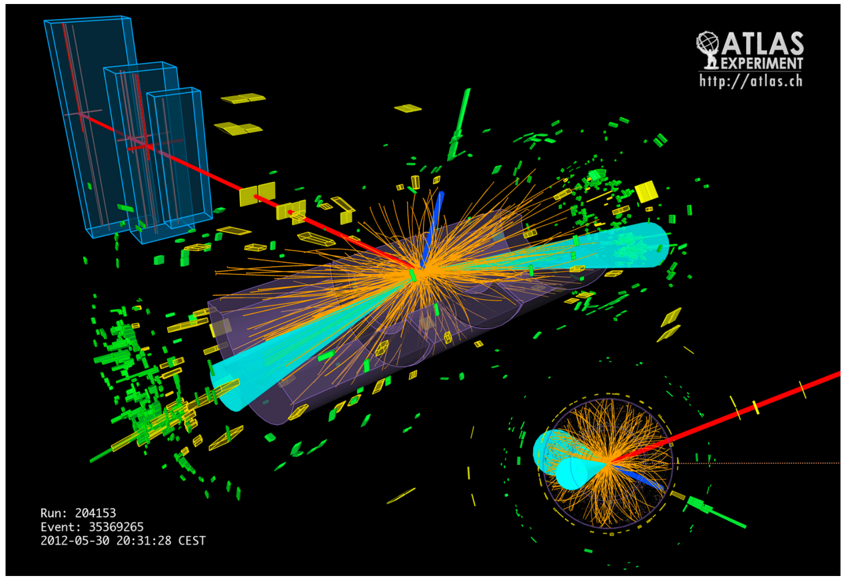
Likelihood Ratio $r(x|\theta_{ref},\theta)=p(x|\theta)/p(x|\theta_{ref})$

"how much more likely is data x described by theory θ than θ_{ref}"

Neyman-Pearson Lemma:

The log-likelihood ratio log $r(x|\theta_{ref},\theta)$ is the most powerful test statistic to discriminate between two hypotheses θ_{ref} and θ.

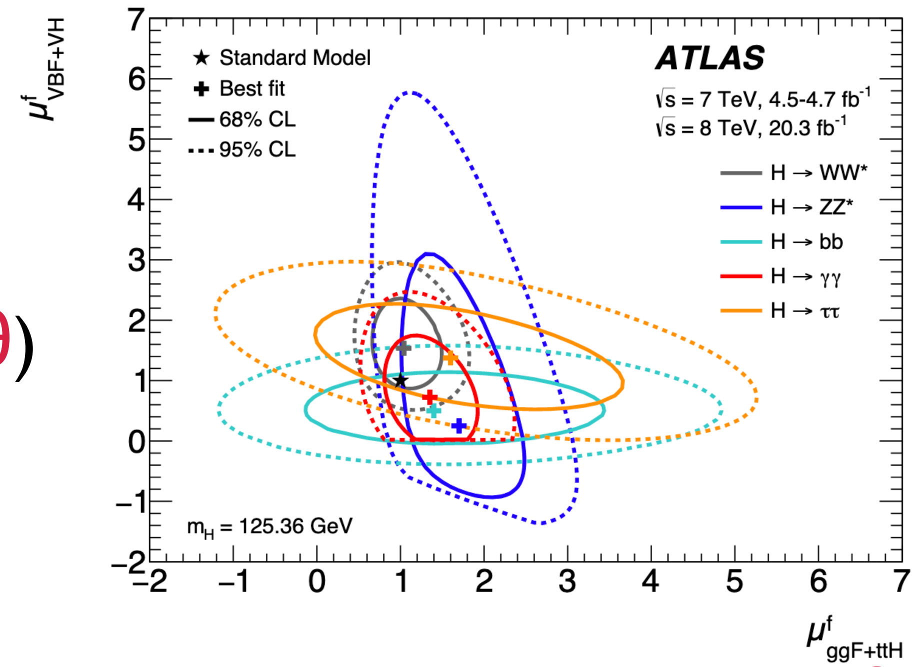
Terminology



Prediction

likelihood function p(x|θ)

Inference



Observables x

anything that can be measured

Example: x={E, pT, m, Δφ, … }.

Theory Parameters θ

describes the theory

Example: $\mathcal{L} = \sum \theta_i \mathcal{O}_i$.

Likelihood Function p(x|θ)

likelihood of an observation x as a function of the theory parameter θ

How can we obtain p(x|θ)?

Particle Physics Simulations

Observables	Detector Effects	Shower Hadronization	Parton Level	Theory Parameters

$$x \longleftarrow z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$

$$p(x|z_d) \qquad p(z_d|z_s) \qquad p(z_s|z_p) \qquad p(z_p|\theta)$$

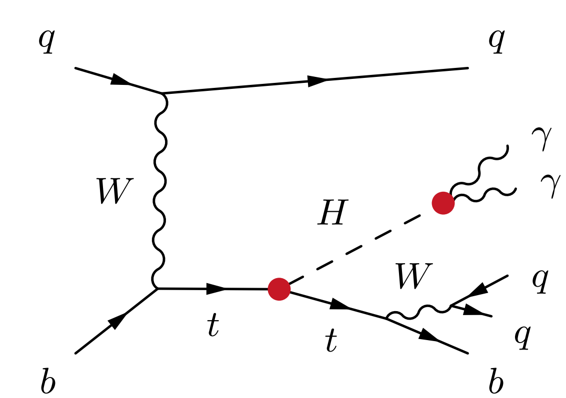
$\longleftarrow$ Prediction / Simulation

Particle Physics Simulations

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

$x \leftarrow z_d \leftarrow z_s \leftarrow z_p \leftarrow \theta$

$p(x|z_d)$ $\quad p(z_d|z_s)$ $\quad p(z_s|z_p)$ $\quad p(z_p|\theta)$



MADGRAPH5_aMC@NLO

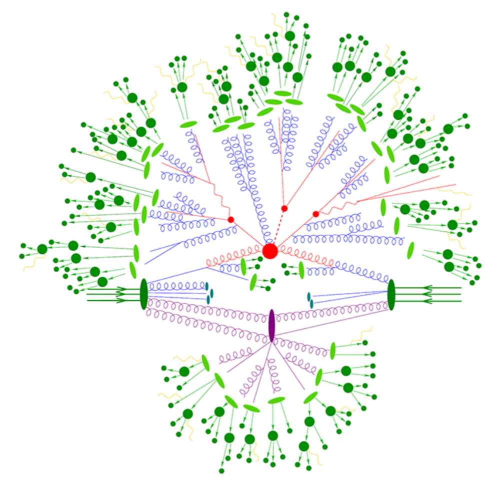
Prediction / Simulation

Particle Physics Simulations

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

$$x \leftarrow z_d \leftarrow z_s \leftarrow z_p \leftarrow \theta$$

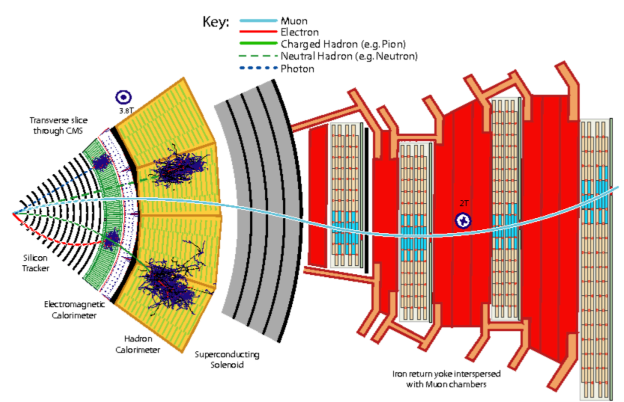
$p(x|z_d)$ $\quad p(z_d|z_s)$ $\quad p(z_s|z_p)$ $\quad p(z_p|\theta)$



Prediction / Simulation

Particle Physics Simulations

Observables		Detector Effects	Shower Hadronization	Parton Level	Theory Parameters
x	$\leftarrow$	z_d $\leftarrow$	z_s $\leftarrow$	z_p $\leftarrow$	θ
	$p(x\|z_d)$	$p(z_d\|z_s)$	$p(z_s\|z_p)$	$p(z_p\|\theta)$	



DELPHES
fast simulation

GEANT4
A SIMULATION TOOLKIT

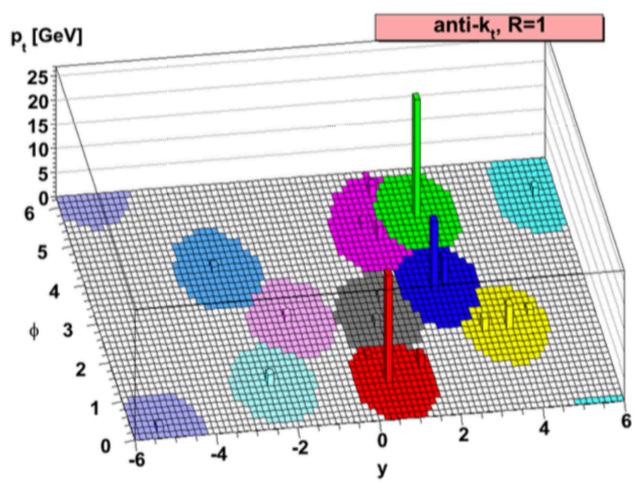
Prediction / Simulation

Particle Physics Simulations

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

$$x \leftarrow z_d \leftarrow z_s \leftarrow z_p \leftarrow \theta$$

$p(x|z_d)$ $\quad$ $p(z_d|z_s)$ $\quad$ $p(z_s|z_p)$ $\quad$ $p(z_p|\theta)$



Prediction / Simulation

Particle Physics Simulations

Latent Variables z: internal, not observable

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

$$x \leftarrow z_d \leftarrow z_s \leftarrow z_p \leftarrow \theta$$

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

obtain MC event sample

Prediction / Simulation

Particle Physics Simulations

Latent Variables z: internal, not observable

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

$$x \longleftarrow z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$

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

likelihood function needed

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

$\downarrow$

likelihood function is *intractable*

$\downarrow$

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

Inference $\longrightarrow$

Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

Optimal Observables and Fisher Information
This will turn out to be useful.

The MadMiner Tool
Using these methods is super easy!

A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

$$\log r_{\text{full}}(\{x\}|\theta, \theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta, \theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta, \theta_{\text{ref}})$$

likelihood function

$$p_{\text{full}}(\{x\}|\theta) = \text{Pois}(n|L\sigma(\theta)) \times \prod_{i=1}^{n} p(x_i|\theta)$$

full likelihood
(all events)

rate
(event number)

kinematics
(single event)

Inference Techniques

$$\log r_{\mathrm{full}}(\{x\}|\theta,\theta_{\mathrm{ref}}) = \log r_{\mathrm{rate}}(n|\theta,\theta_{\mathrm{ref}}) + \cancel{\sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\mathrm{ref}})}$$

ignore this term

Option 1: Rate Only

- only consider rate: "cut and count"
- $\log r_{\mathrm{full}}(n|\theta,\theta_{\mathrm{ref}}) = \log r_{\mathrm{rate}}(n|\theta,\theta_{\mathrm{ref}}) = \mathrm{Pois}(n|L\sigma(\theta))/\mathrm{Pois}(n|L\sigma(\theta_{\mathrm{ref}}))$
- all additional kinematic information lost

Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

Histogram

Option 1: Rate Only
- only consider rate: "cut and count"
- $\log r_{\text{full}}(n|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) = \text{Pois}(n|L\sigma(\theta))/\text{Pois}(n|L\sigma(\theta_{\text{ref}}))$
- all additional kinematic information lost

Option 2: Summary Statistics
- few hand-picked observables x'
- estimate p(x'|θ)
- information loss
- problem dependent
- Example:
  * histograms
  * STXS

Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

Histogram

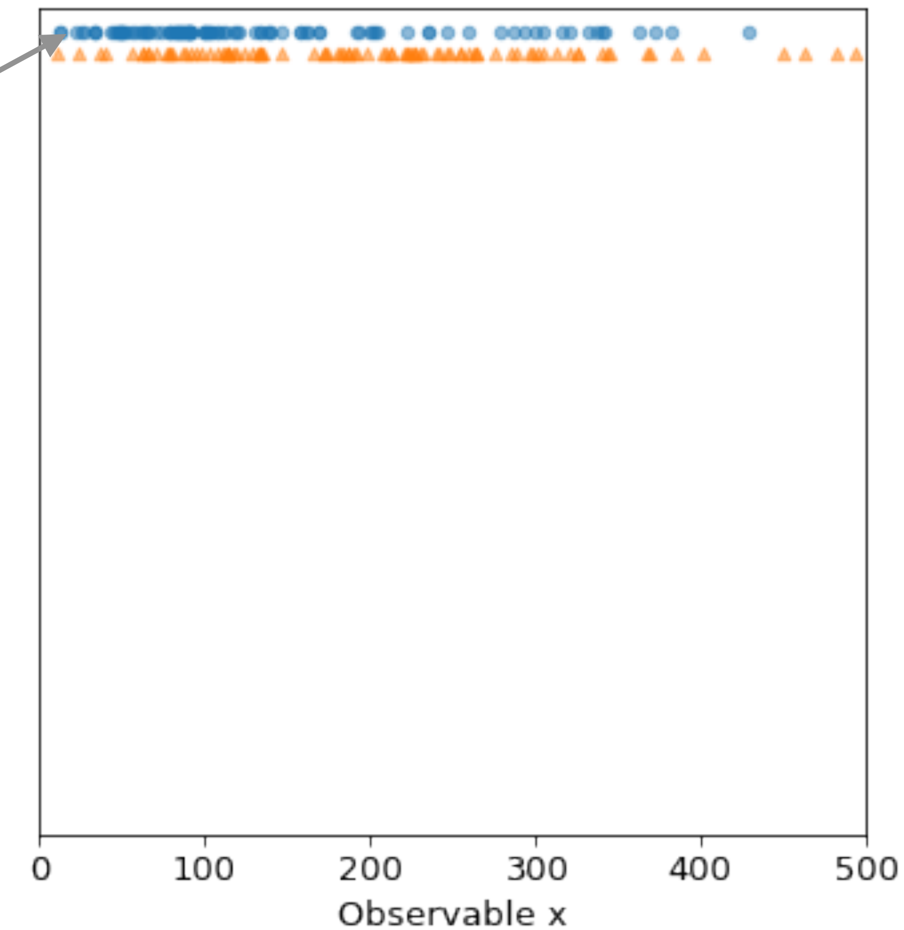
Option 1: Rate Only

- only consider rate: "cut and count"
- $\log r_{\text{full}}(n|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) = \text{Pois}(n|L\sigma(\theta))/\text{Pois}(n|L\sigma(\theta_{\text{ref}}))$
- all additional kinematic information lost

Option 2: Summary Statistics

- few hand-picked observables x'
- estimate p(x'|θ)
- information loss
- problem dependent
- Example:
  * histograms
  * STXS

get event samples



Inference Techniques

$$\log r_{\mathrm{full}}(\{x\}|\theta,\theta_{\mathrm{ref}}) = \log r_{\mathrm{rate}}(n|\theta,\theta_{\mathrm{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\mathrm{ref}})$$

Histogram

Option 1: Rate Only

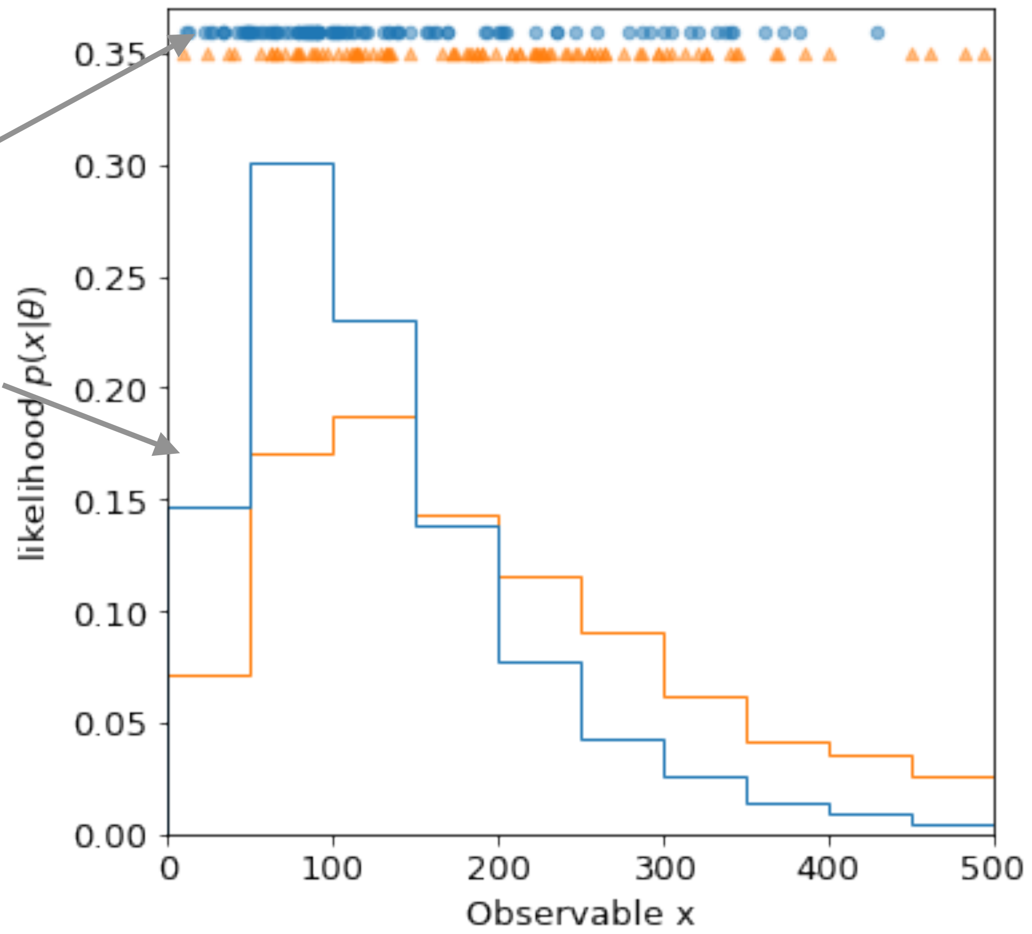
- only consider rate: "cut and count"
- $\log r_{\mathrm{full}}(n|\theta,\theta_{\mathrm{ref}}) = \log r_{\mathrm{rate}}(n|\theta,\theta_{\mathrm{ref}}) = \mathrm{Pois}(n|L\sigma(\theta))/\mathrm{Pois}(n|L\sigma(\theta_{\mathrm{ref}}))$
- all additional kinematic information lost

Option 2: Summary Statistics

- few hand-picked observables x'
- estimate p(x'|θ)
- information loss
- problem dependent
- Example:
  * histograms
  * STXS

get event samples

fill histograms



Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta, \theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta, \theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta, \theta_{\text{ref}})$$

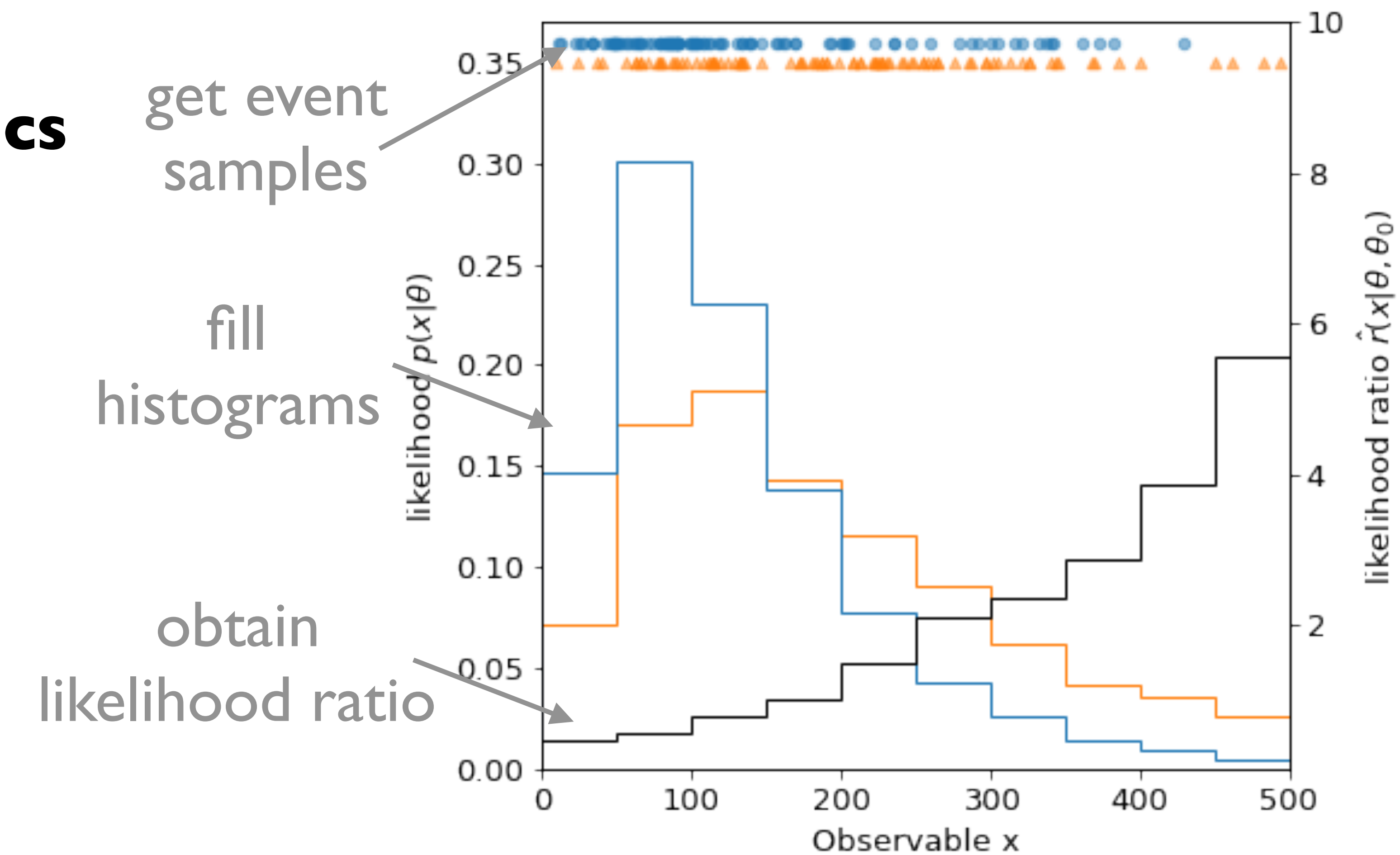
use Histogram

Option 1: Rate Only

- only consider rate: "cut and count"
- $\log r_{\text{full}}(n|\theta, \theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta, \theta_{\text{ref}}) = \text{Pois}(n|L\sigma(\theta))/\text{Pois}(n|L\sigma(\theta_{\text{ref}}))$
- all additional kinematic information lost

Option 2: Summary Statistics

- few hand-picked observables x'
- estimate p(x'|θ)
- information loss
- problem dependent
- Example:
  * histograms
  * STXS



$$\log r_{\text{full}}(\{x\}|\theta, \theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta, \theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta, \theta_{\text{ref}})$$

use ML Classifier

Option 3: Machine Learning

- estimate r(x|θ) from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier

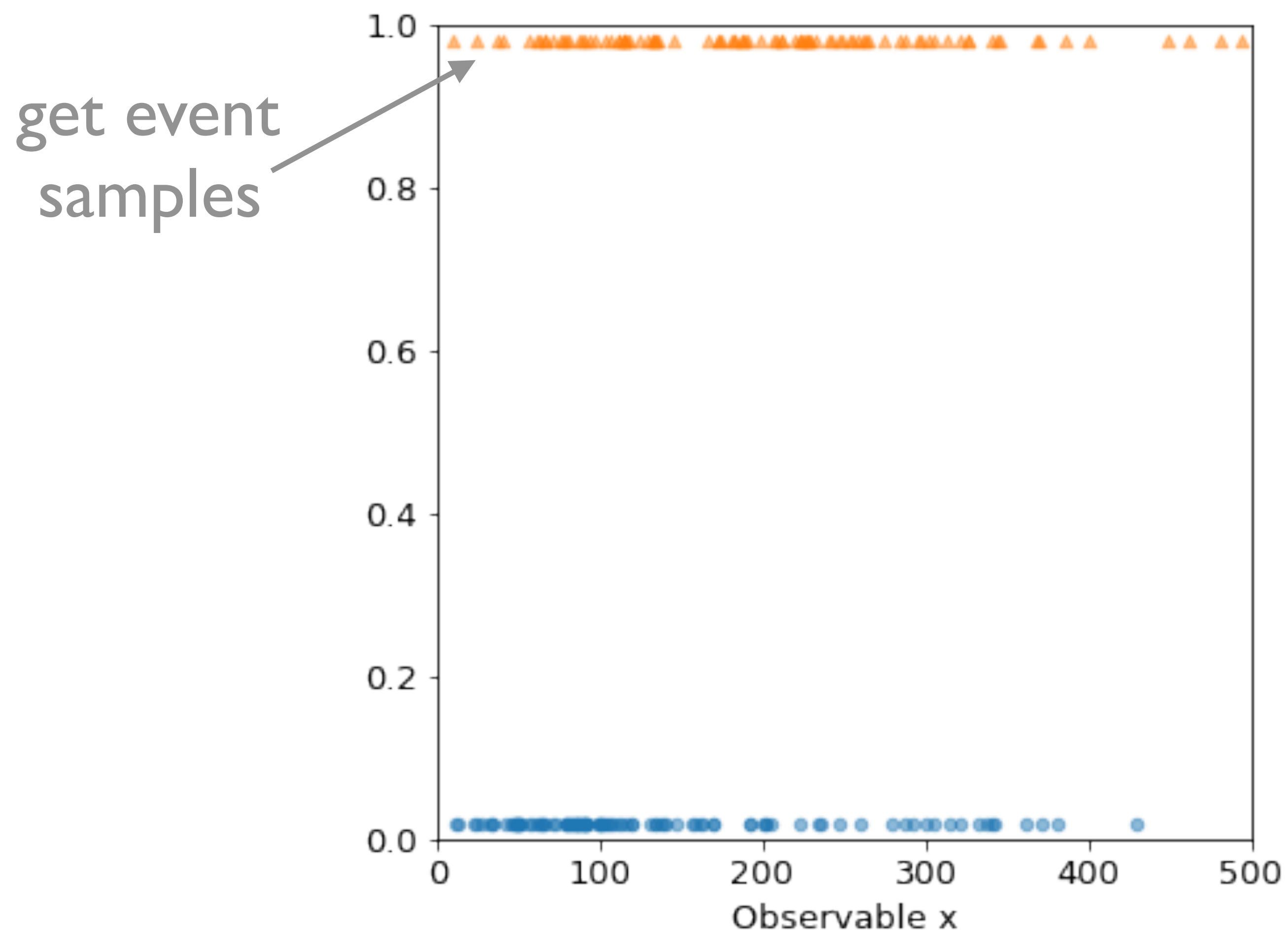
Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

use ML Classifier

Option 3: Machine Learning

- estimate r(x|θ) from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier



Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

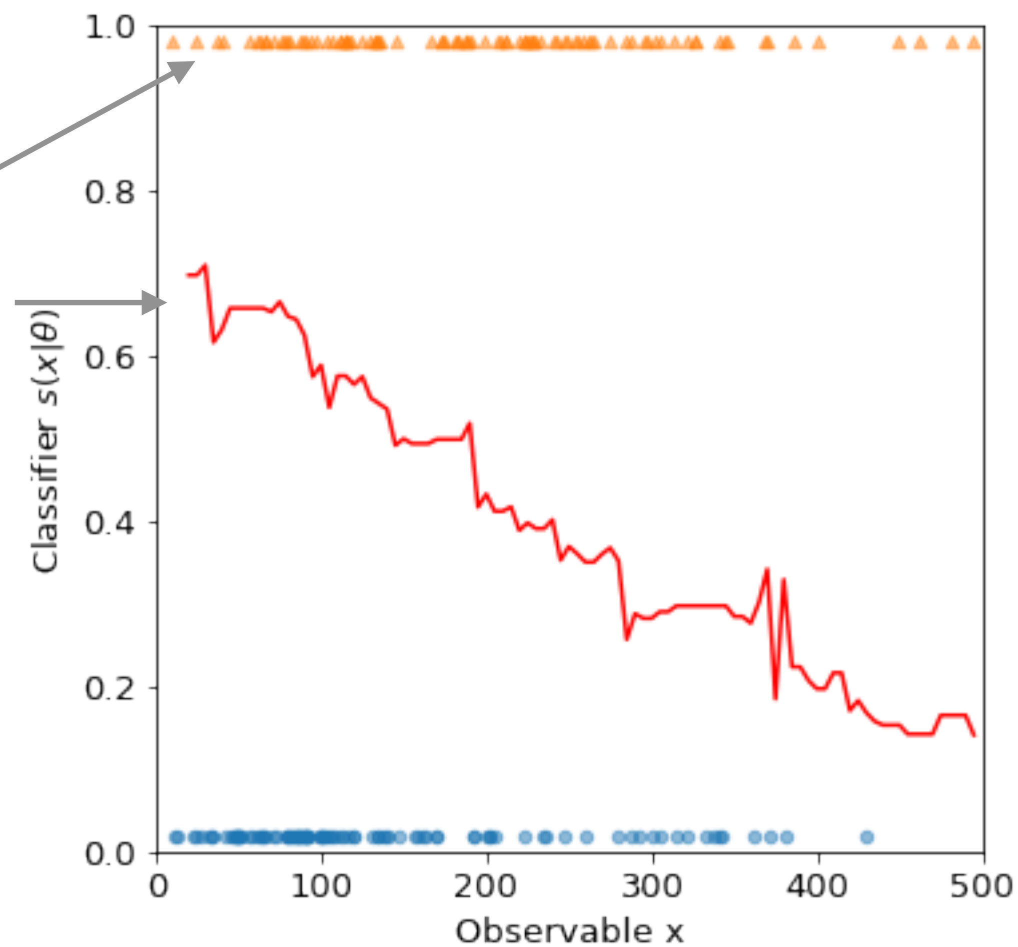
use ML Classifier

Option 3: Machine Learning

- estimate r(x|θ) from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier

get event samples

train classifier



$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

use ML Classifier

Option 3: Machine Learning

- estimate r(x|θ) from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier

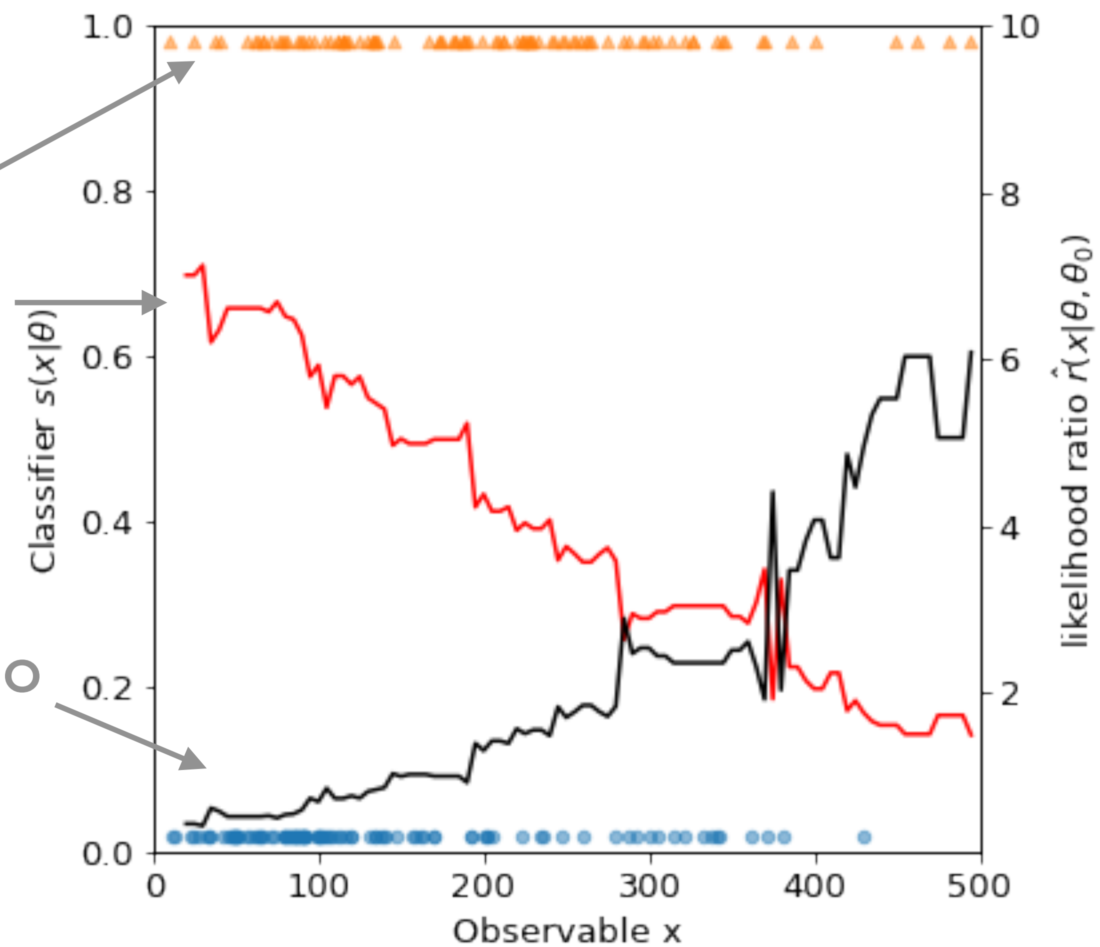
- optimal decision function:

$$s(x|\theta) = \frac{p(x|\theta_{\text{ref}})}{p(x|\theta) + p(x|\theta_{\text{ref}})}$$

get event samples

train classifier

obtain likelihood ratio



Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

use ML Classifier

Option 3: Machine Learning

- estimate $r(x|\theta)$ from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier

- optimal decision function:

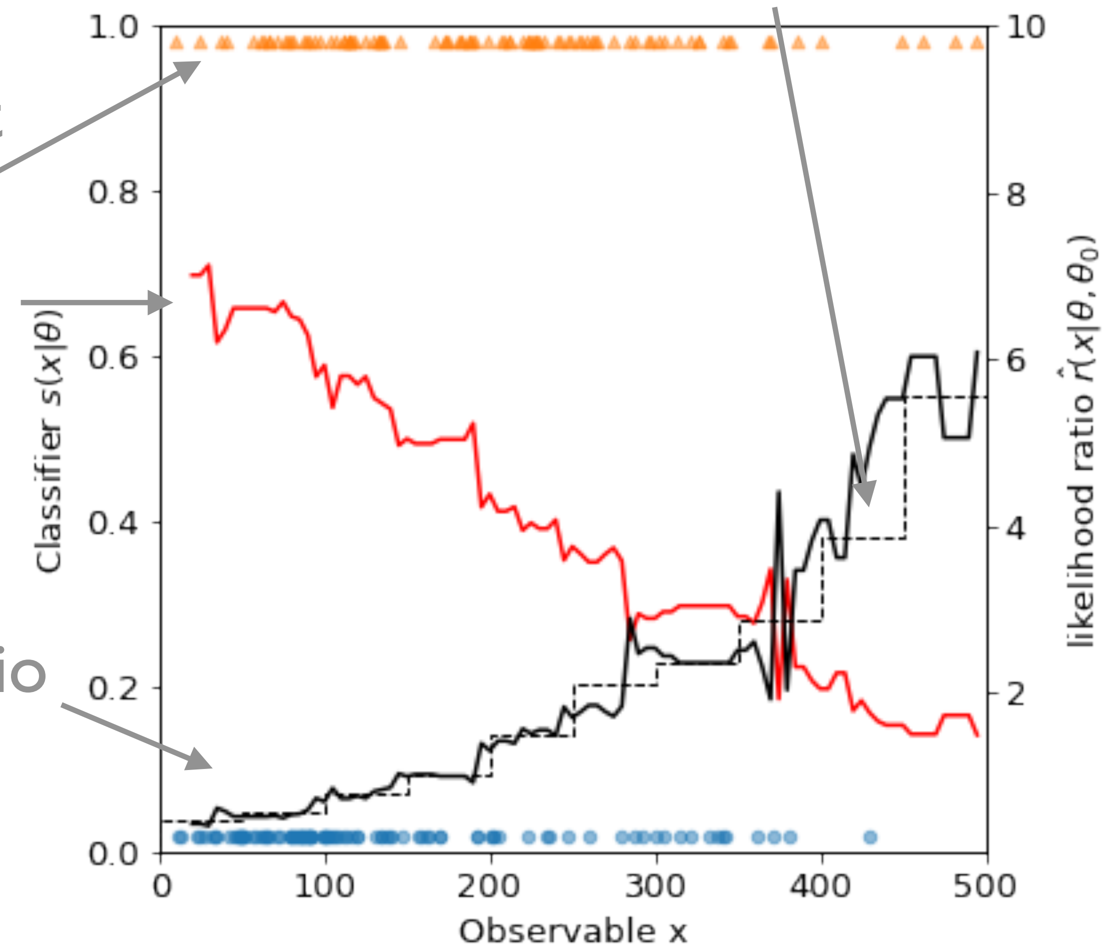
$$s(x|\theta) = \frac{p(x|\theta_{\text{ref}})}{p(x|\theta) + p(x|\theta_{\text{ref}})}$$

agrees with histogram …

get event samples

train classifier

obtain likelihood ratio



Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

use Matrix Element

Option 3: Machine Learning

- estimate r(x|θ) from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier

Option 4: Matrix Element Based

- uses $p(x|\theta)\sim|M(x|\theta)|^2$
- multivariate analysis, direct physics insight
- requires approximations of detector response: $x=z_p$
- works great at parton level: S' vs S is easy
- S vs BG can be hard
- Example: Matrix Element Method (MEM), Optimal Observables (OO)

Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

Option 3: Machine Learning

- estimate r(x|θ) from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S'   * very similar S', S
- physics not always clear
- Example: ML Classifier

Can we have MEM at detector level?

Option 4: Matrix Element Based

- uses $p(x|\theta) \sim |M(x|\theta)|^2$
- multivariate analysis, direct physics insight
- requires approximations of detector response: $x=z_p$
- works great at parton level: S' vs S is easy
- S vs BG can be hard
- Example: Matrix Element Method (MEM), Optimal Observables (OO)

Inference Techniques

$$\log r_{\text{full}}(\{x\}|\theta,\theta_{\text{ref}}) = \log r_{\text{rate}}(n|\theta,\theta_{\text{ref}}) + \sum_{i=1}^{n} \log r(x_i|\theta,\theta_{\text{ref}})$$

Option 3: Machine Learning

- estimate $r(x|\theta)$ from multivariate analysis
- works great for S vs BG
- struggles with S' vs S:
  *large number of S' * very similar S', S
- physics not always clear
- Example: ML Classifier

Option 4: Matrix Element Based

- uses $p(x|\theta) \sim |M(x|\theta)|^2$
- multivariate analysis, direct physics insight
- requires approximations of detector response: $x=z_p$
- works great at parton level: S' vs S is easy
- S vs BG can be hard
- Example: Matrix Element Method (MEM), Optimal Observables (OO)

power of
machine learning

↓

MadMiner

↑

physics insight of
matrix element information

Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

Optimal Observables and Fisher Information
This will turn out to be useful.

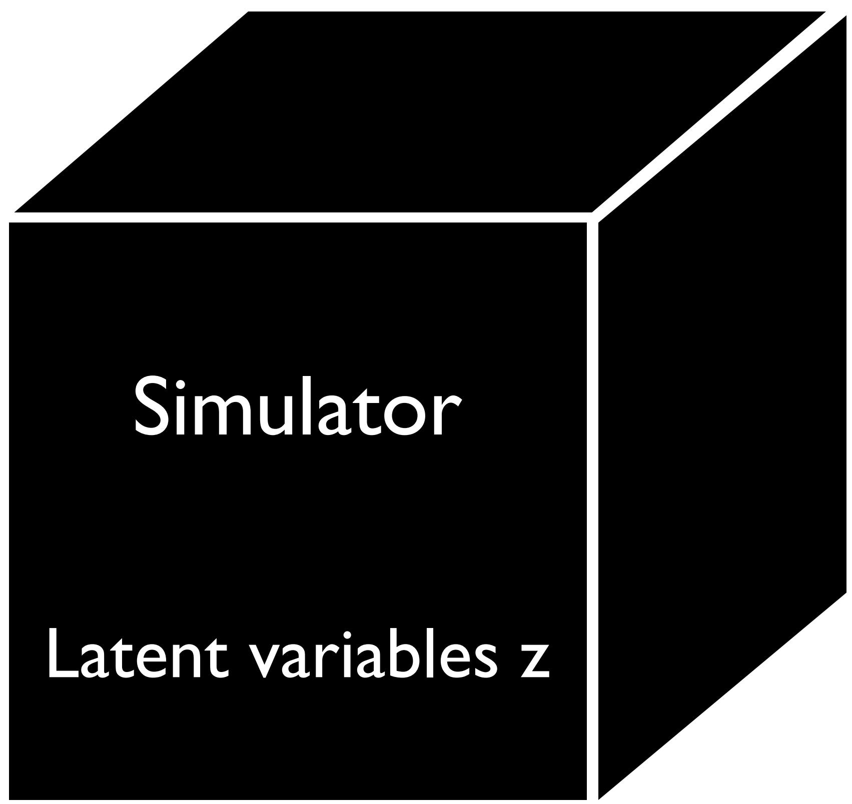
The MadMiner Tool
Using these methods is super easy!

A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

Theory Parameters

θ



Set of Events:

Observables

x

matrix elements

$|\mathcal{M}(z_p|\theta)|^2$

joint likelihood ratio

$$r(x, z|\theta) = \frac{|\mathcal{M}(z_p|\theta)|^2}{|\mathcal{M}(z_p|\theta_{\text{ref}})|^2}$$

The MadMiner Approach

Problem Setup:

We want: Likelihood-ratio $r(x|\theta)$

data → x, theory → θ

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

The MadMiner Approach

Problem Setup:

We want: Likelihood-ratio $r(x|\theta)$ — data ↗ ↖ theory

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

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]

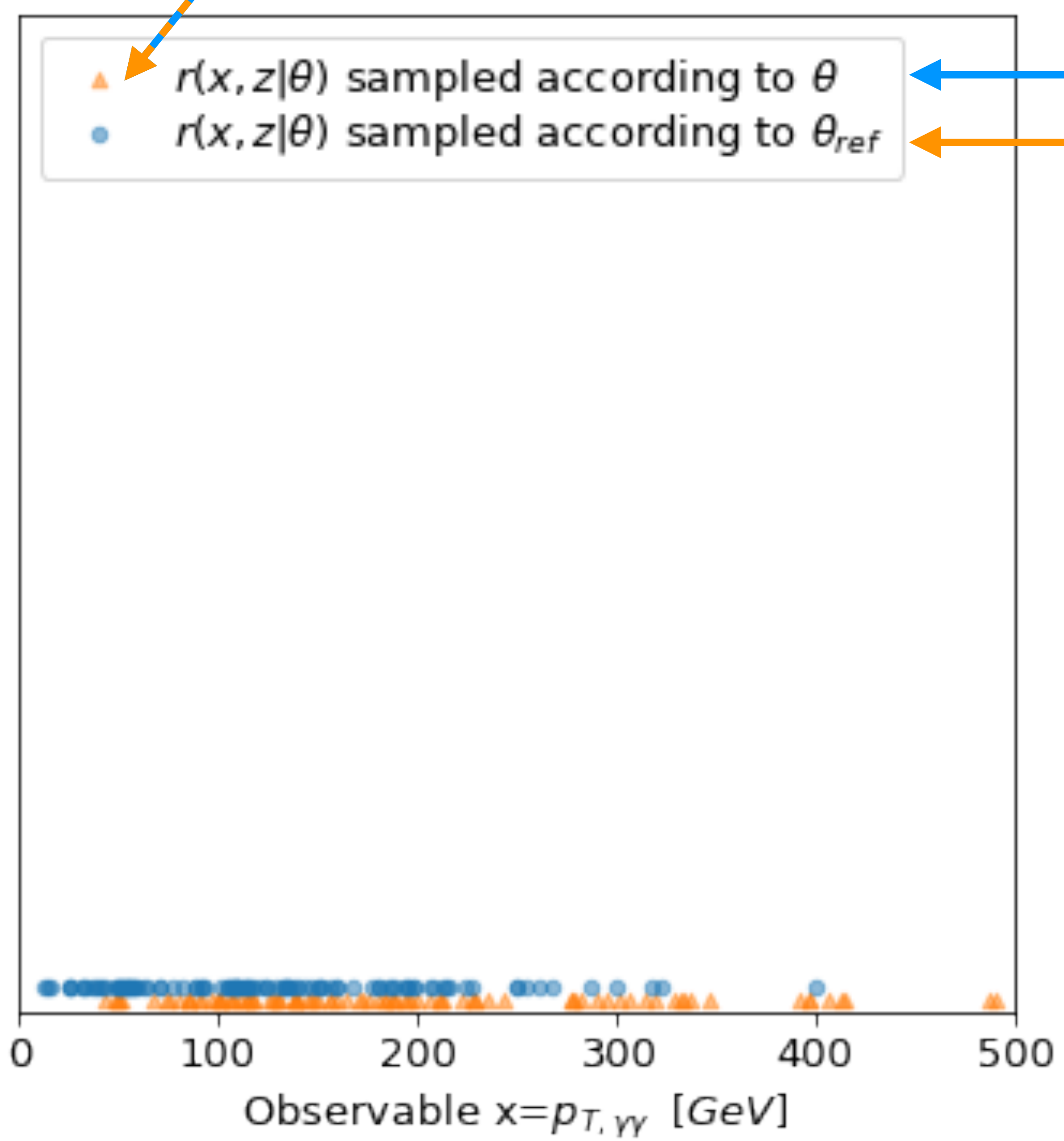
- one observable: $x = p_{T,\gamma\gamma}$

- latent variables:
  * neutrino pz
  * jet charge+flavor
  * all the hadron momenta
  * random numbers in detector simulation
  * all other observables

The MadMiner Approach

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)



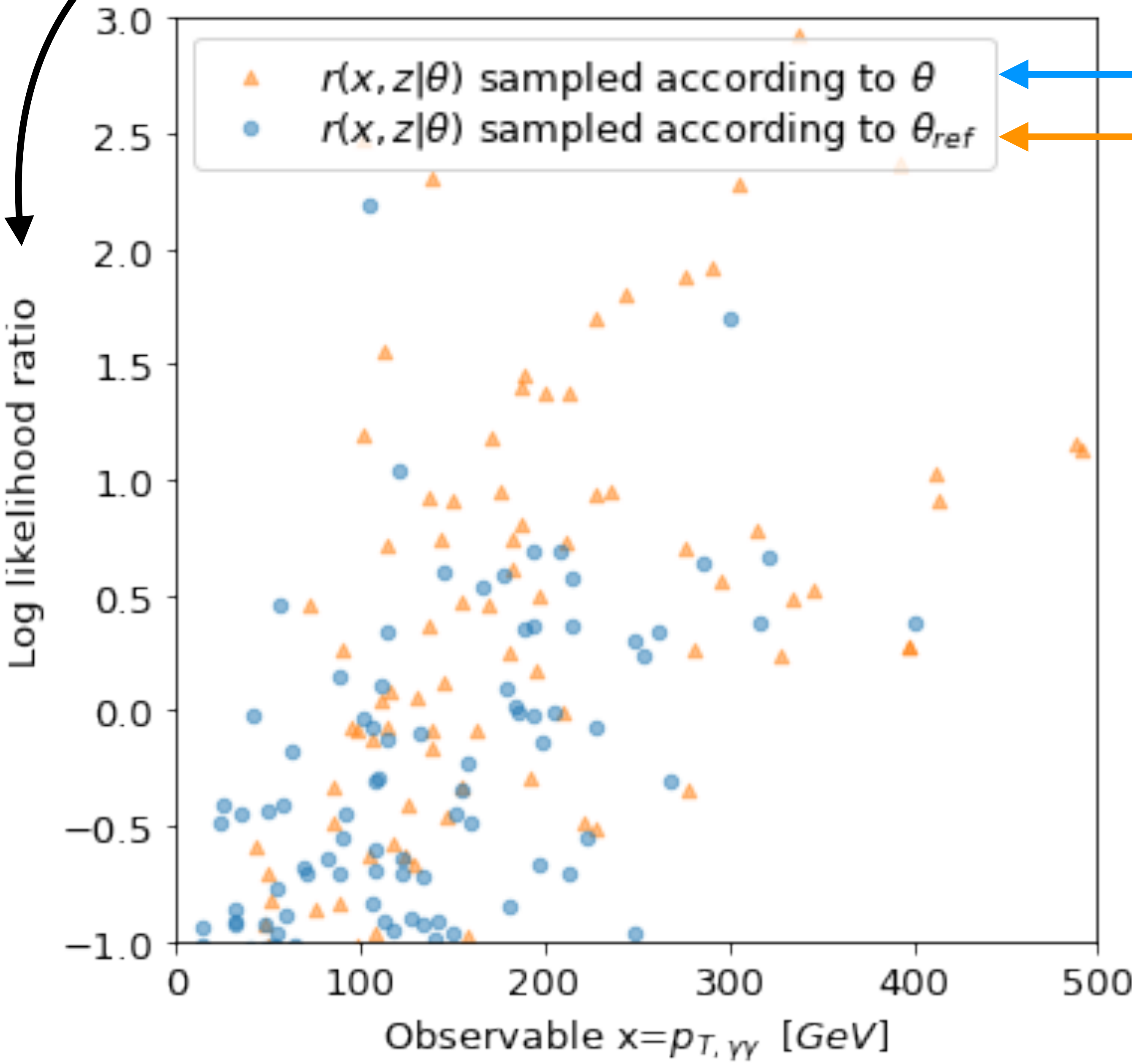
sampled events according to p(x|SM)

sampled events according to p(x|BSM)

- more BSM points at high pT

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)



sampled events according to p(x|SM)

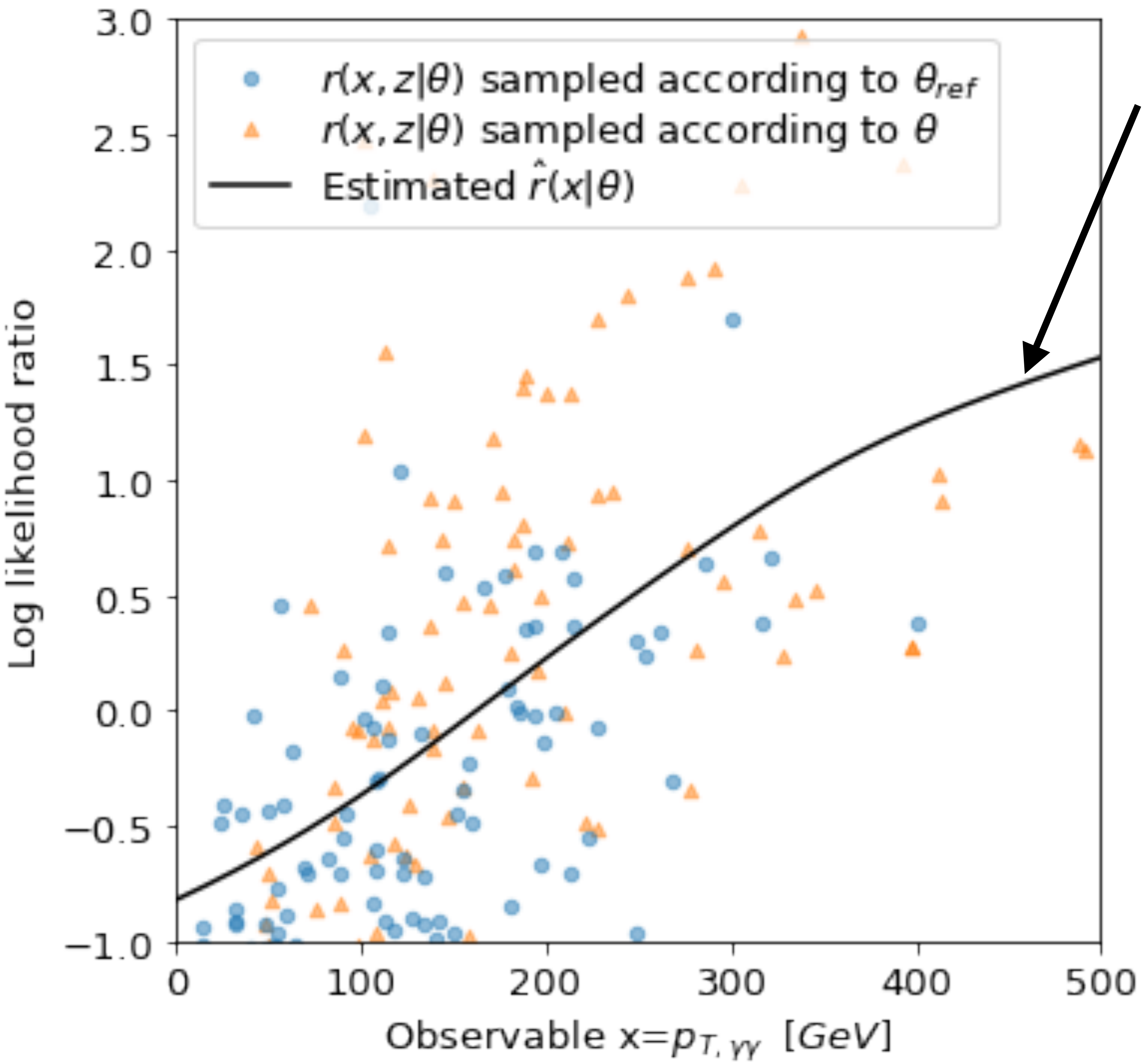
sampled events according to p(x|BSM)

- more BSM points at high pT
  → BSM more likely than SM at high pT

The MadMiner Approach

How is Likelihood Estimated?

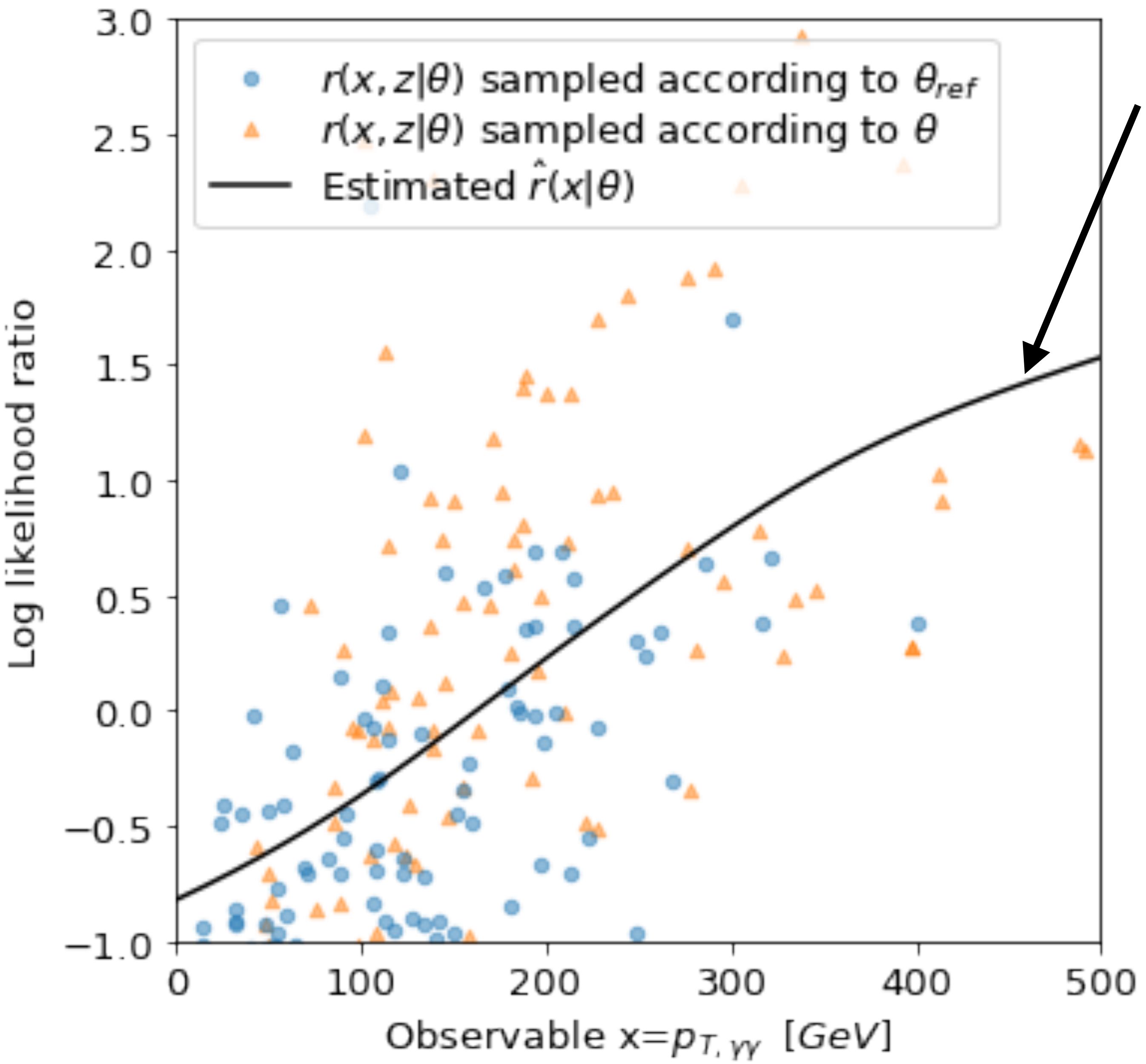
- consider two models BSM (θ=1) vs SM (θ_{ref}=0)
  * sampled events according to p(x|SM) , p(x|BSM)
  * y-axis: joint likelihood ratio r(x,z|BSM,SM)



- estimated r(x|SM,BSM)
  → obtain "best fit"

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)



- estimated r(x|SM,BSM)
  → obtain "best fit"
- define functional (loss function)

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

and minimize it

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

neural network

loss function

stochastic gradient descent

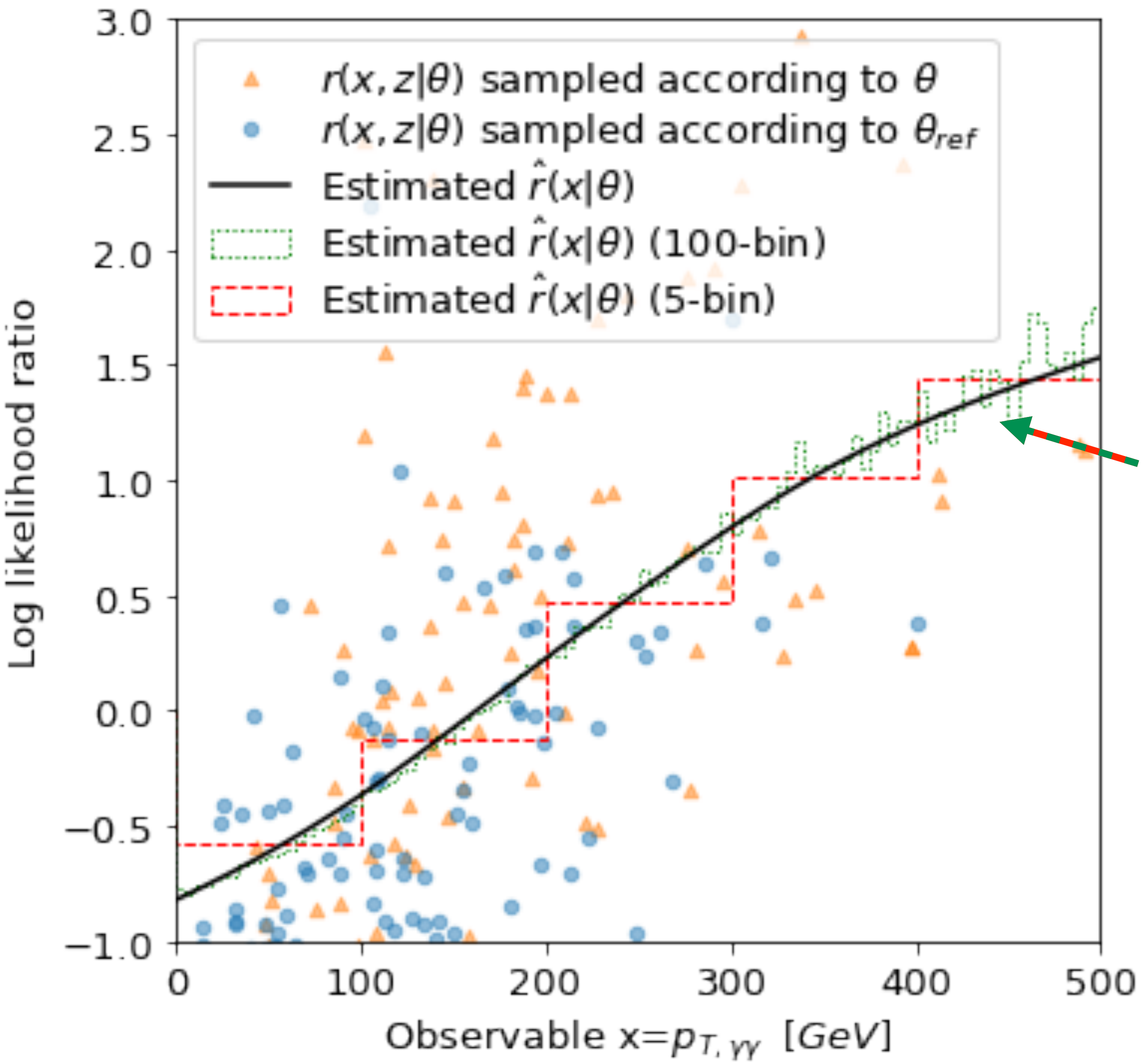
A sufficiently expressive network, efficiently trained in this way with enough data will learn the likelihood ratio function r(x|θ)!

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

The MadMiner Approach

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)



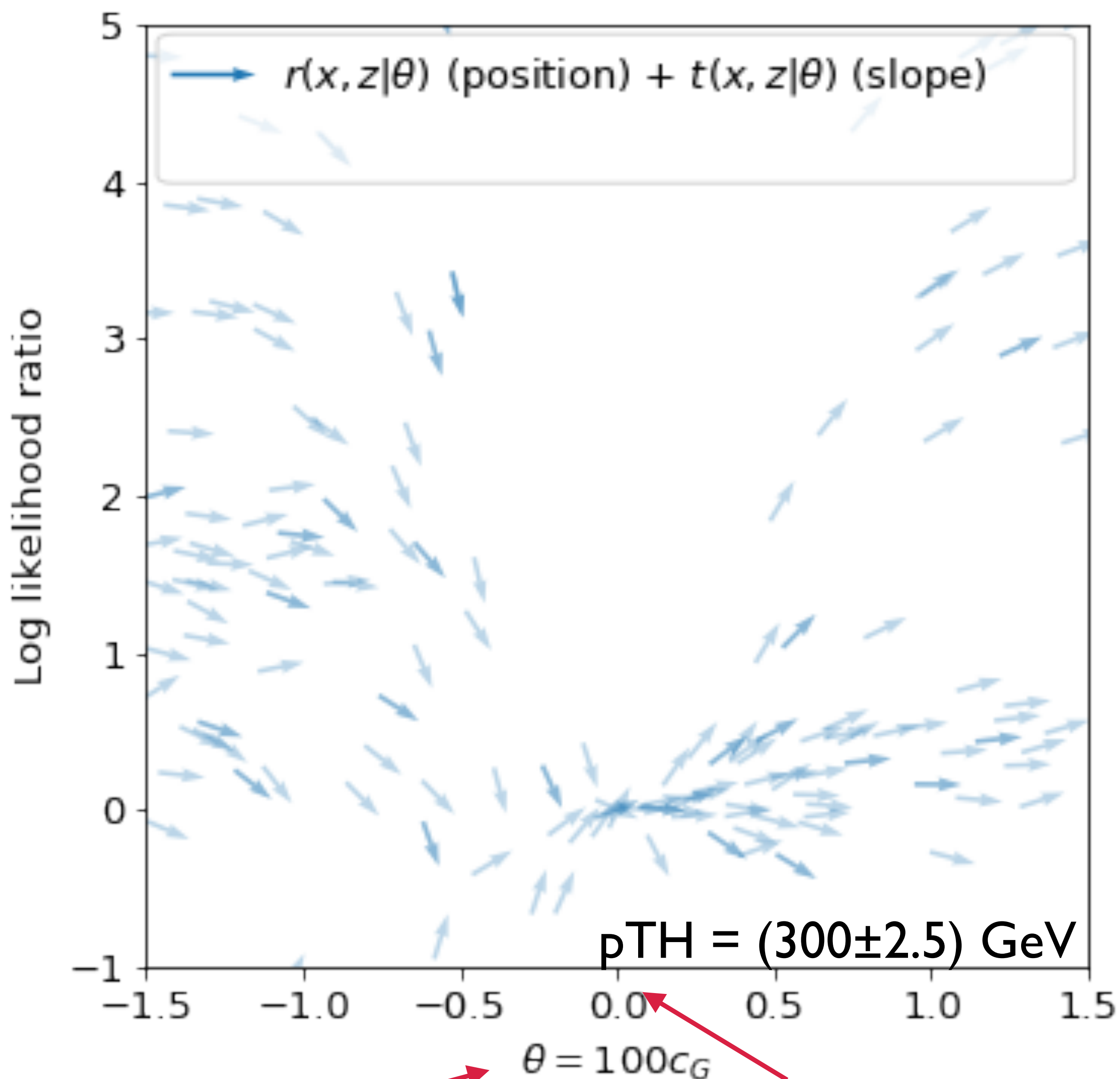
- estimated r(x|SM,BSM)
  → obtain "best fit"

- LLR obtained using histogram
  → agrees well :)
  → "continuum limit of for large # of bins"

Useful: the Score!

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

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



BSM: $\theta \neq 0$

SM: $\theta = 0$

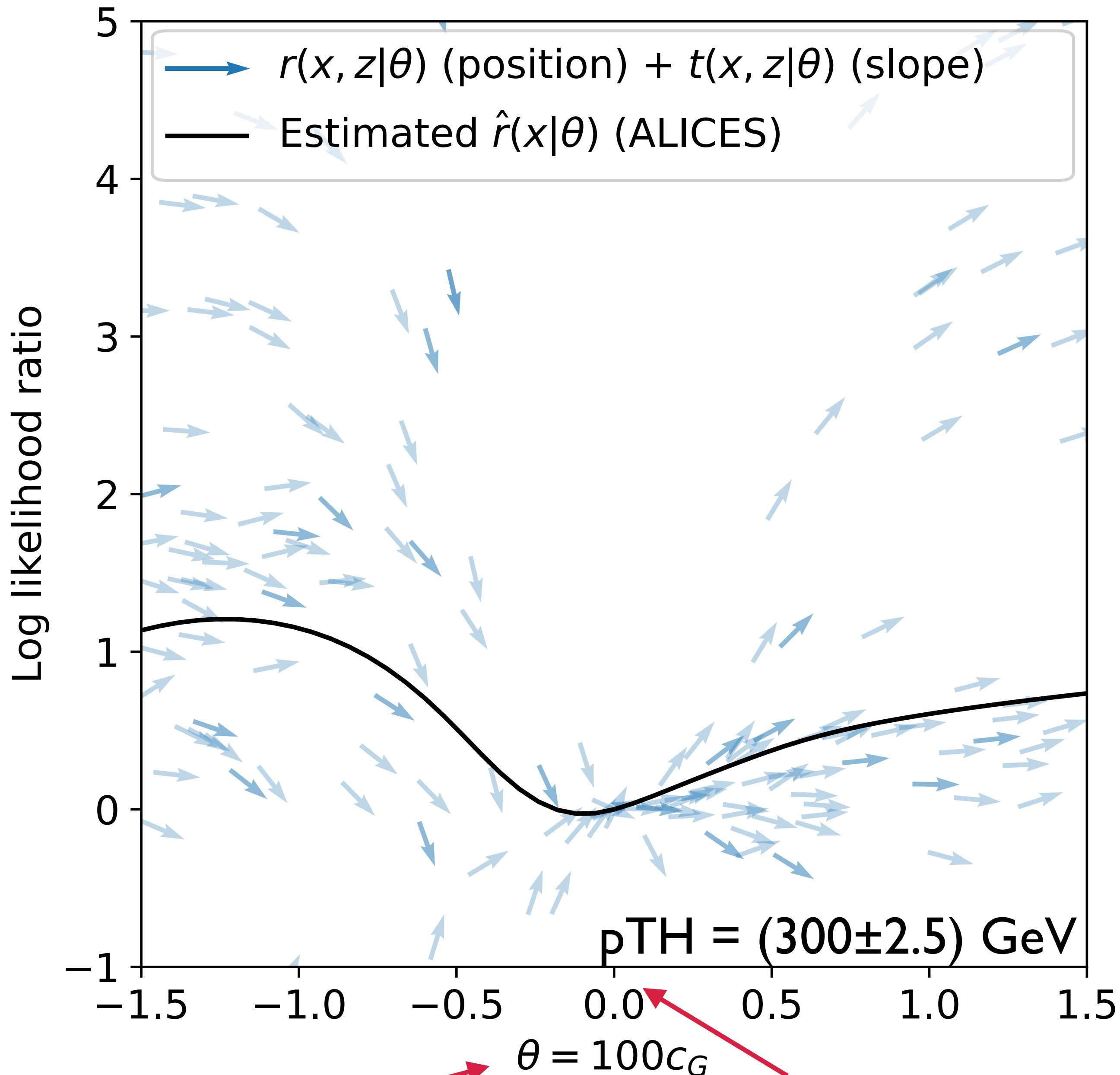
- position: joint likelihood ratio
  $\log r(x,z|\theta)$ as function of θ
- slope: joint score
  $t(x,z|\theta)$ as function of θ

The MadMiner Approach

Useful: the Score!

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

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



BSM: $\theta \neq 0$

SM: $\theta = 0$

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

↓ obtain "best fit"

- estimate $r(x, \theta)$ and $t(x, \theta)$

Likelihood Ratio Estimator (ALICES)

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

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

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

Optimal Observables and Fisher Information
This will turn out to be useful.

The MadMiner Tool
Using these methods is super easy!

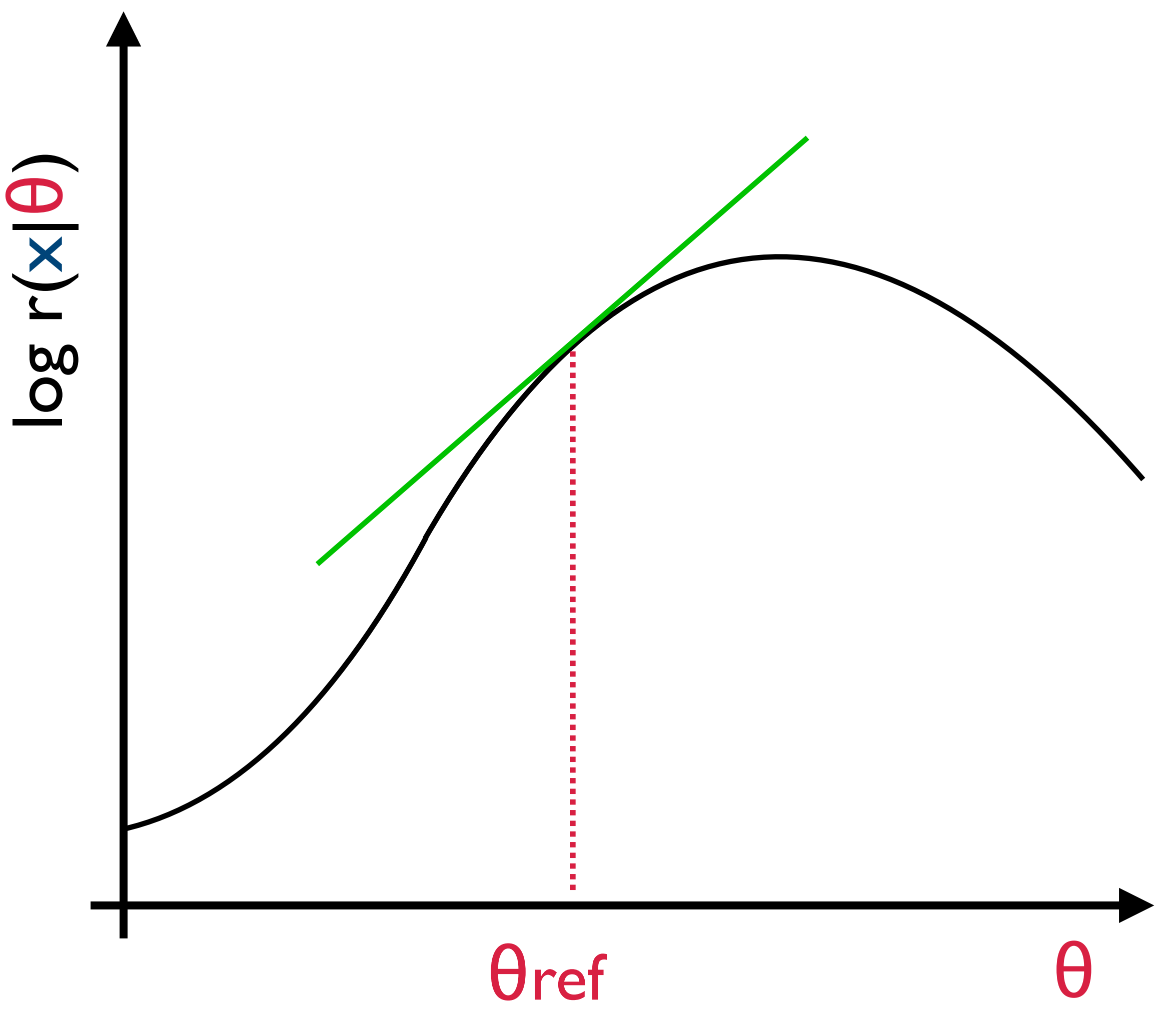
A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

Optimal Observables

Scores are Optimal Observables

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



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

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

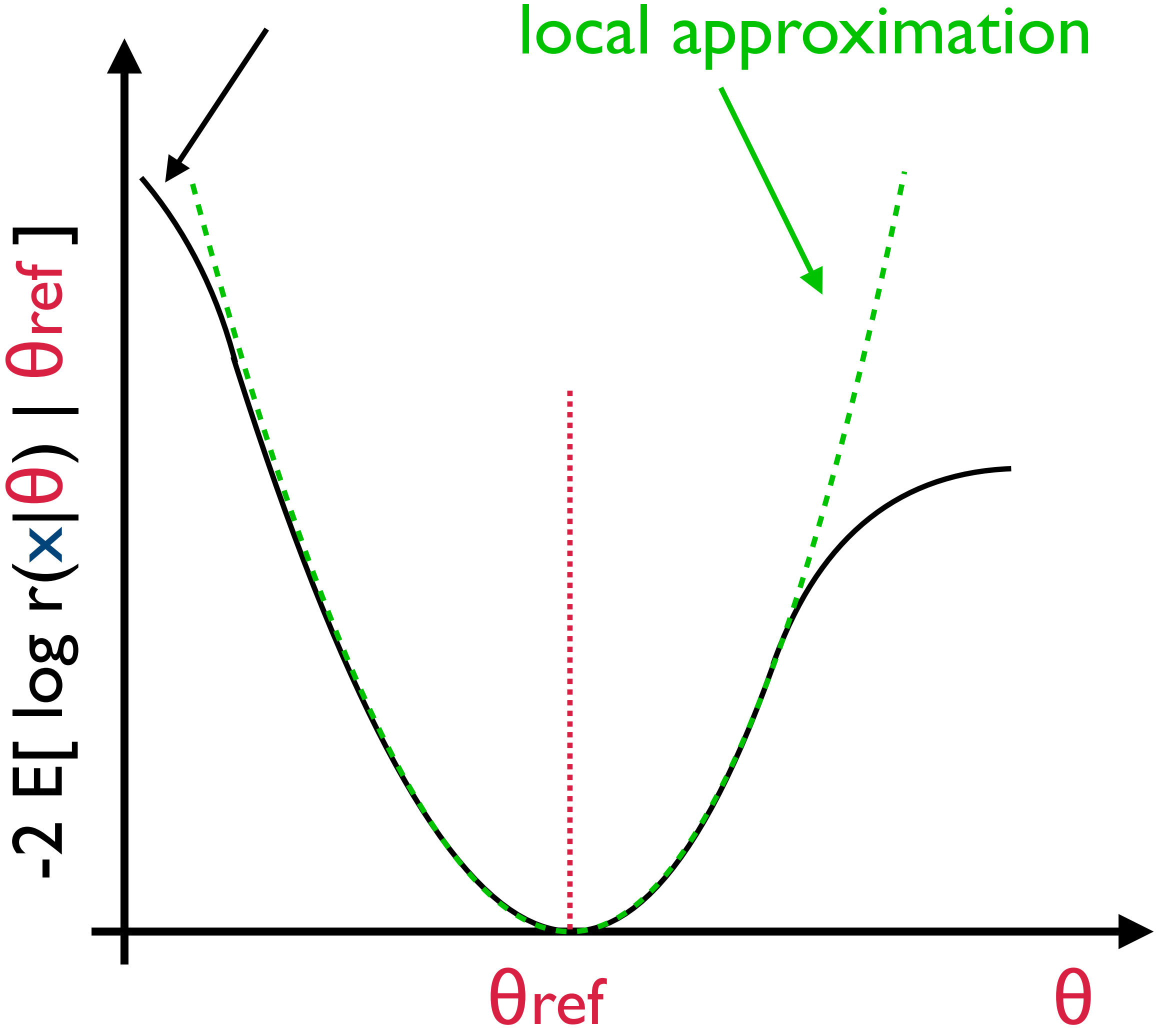
$$\mathrm{NN} : x \to \hat{t}(x) \approx \nabla_\theta \log(x|\theta)\Big|_{\theta_{\mathrm{ref}}}$$

Fisher Information

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

- expand expected LLR around θref

$$\mathbb{E}[-2\log r_{\text{full}}(x|\theta)|\theta_{\text{ref}}] = I_{ij}(\theta_{\text{ref}}) \times (\theta-\theta_{\text{ref}})_i(\theta-\theta_{\text{ref}})_j + \cdots$$



- Fisher Information:

scores

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

- useful properties:
  * simple: nxn 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 θ→ θ'
  * 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

Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

Optimal Observables and Fisher Information
This will turn out to be useful.

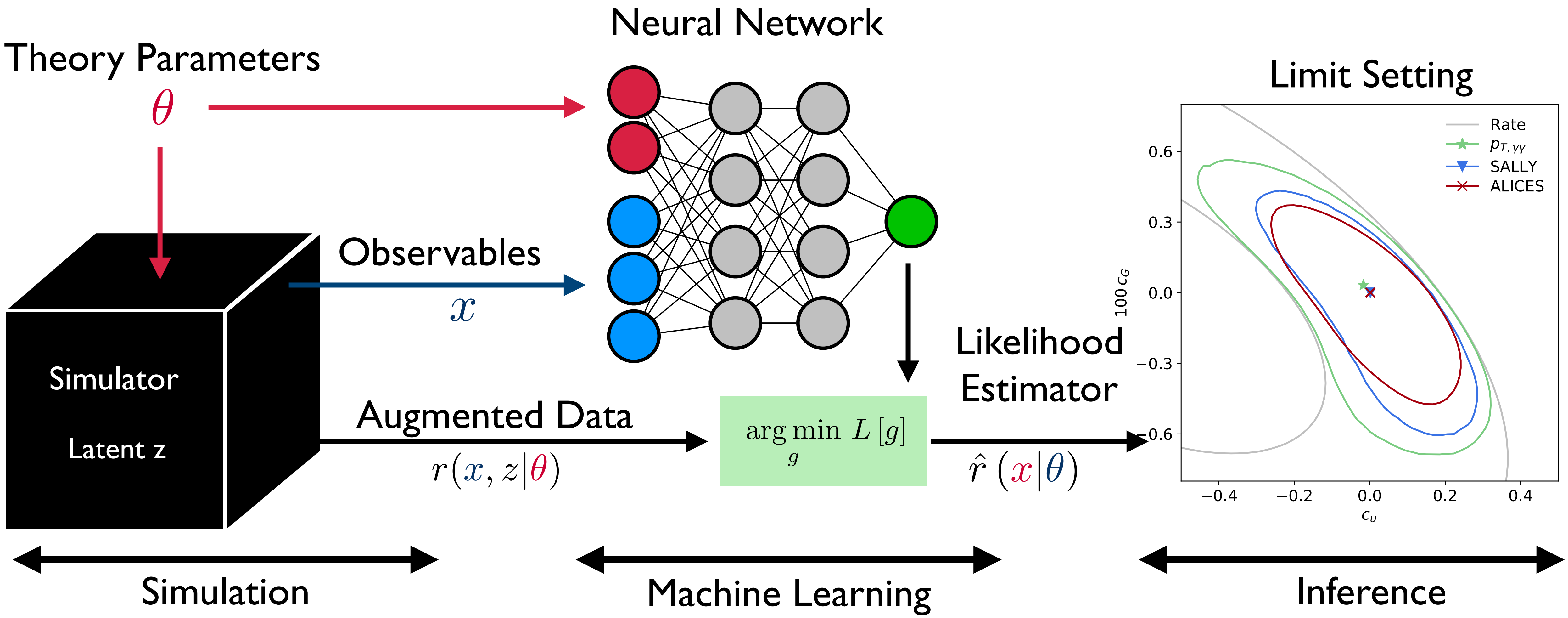
The MadMiner Tool
Using these methods is super easy!

A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

MadMiner: The Tool

A short summary



MadMiner: The Tool

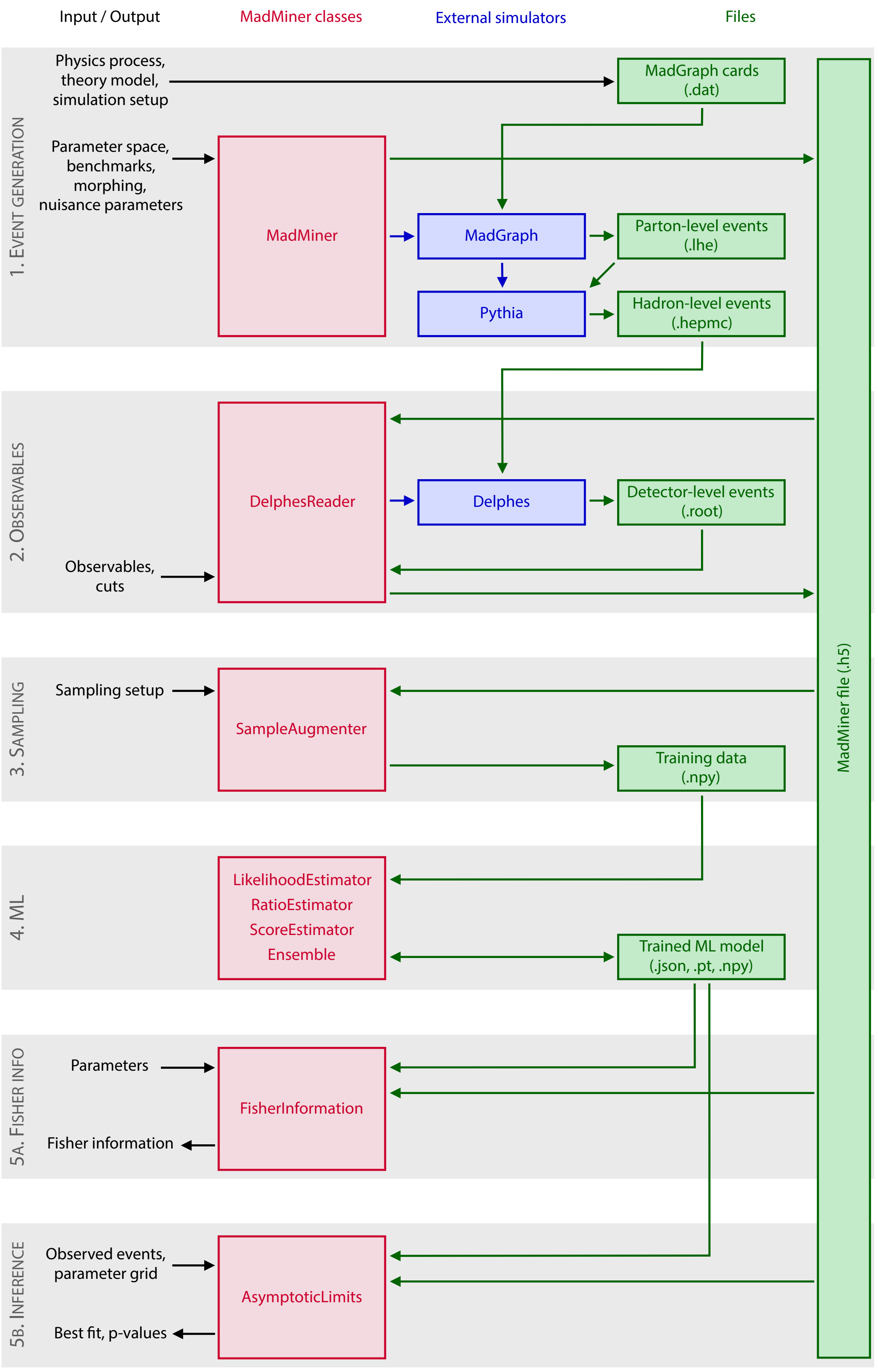
MadMiner [J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

- 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`



MadMiner: The Tool

MadMiner Team

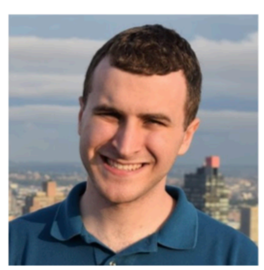
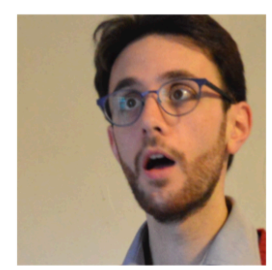
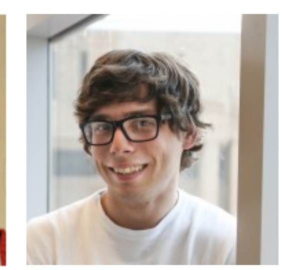
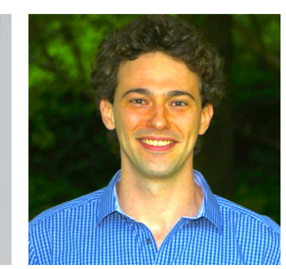

Johann Brehmer


Kyle Cranmer


Irina Espejo

Felix Kling

Already used by many people

Thank you a lot for testing!

Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

Optimal Observables and Fisher Information
This will turn out to be useful.

The MadMiner Tool
Using these methods is super easy!

A Realistic Physics Example
Probing SMEFT in tth

Summary and Conclusion

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: $y_t \to y_t\ (1+3/2\ c_u)$

$$\mathcal{O}_G = \frac{g_s^2}{m_W^2}\left(H^\dagger H\right) G_{\mu\nu}^a G_a^{\mu\nu}$$

- Higgs-gluon coupling: $g_{ggh} \to g_{ggh}(1 + 192\pi^2/g^2\ c_G)$

$$\mathcal{O}_{uG} = -\frac{4 g_s}{m_W^2} y_u \left(H^\dagger \bar{Q}_L\right) \gamma^{\mu\nu} T_a u_R G_{\mu\nu}^a$$

- chromo-dipole moment: gtt, ggtt, gtth, ggtth

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

- 48 observable: $x=\{p_x, p_y, p_z, E, p_T, \eta, \Delta\Phi, \text{MET}, m_{ij}, \ldots\}$

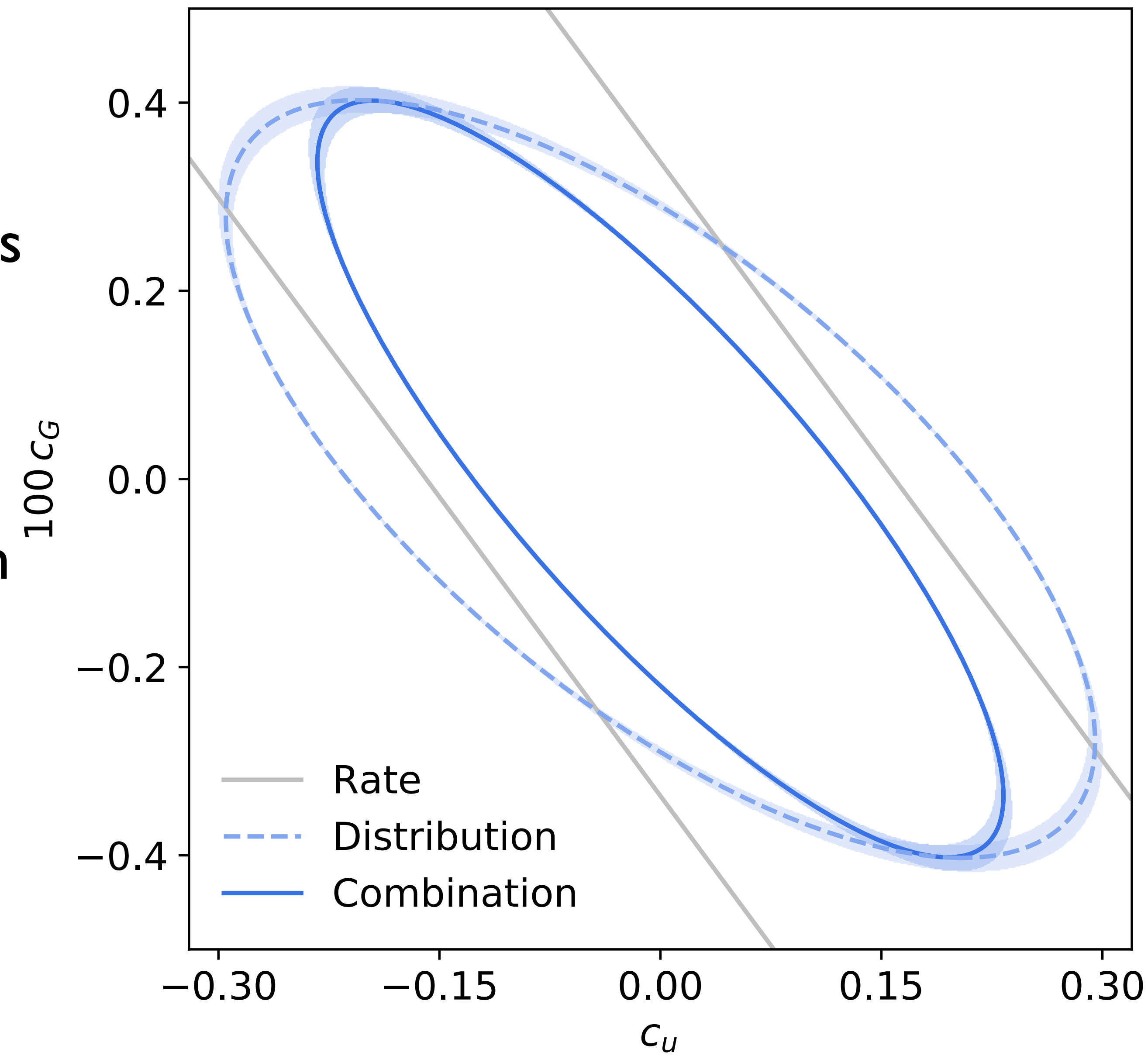
Fisher Information Analysis:

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

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

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

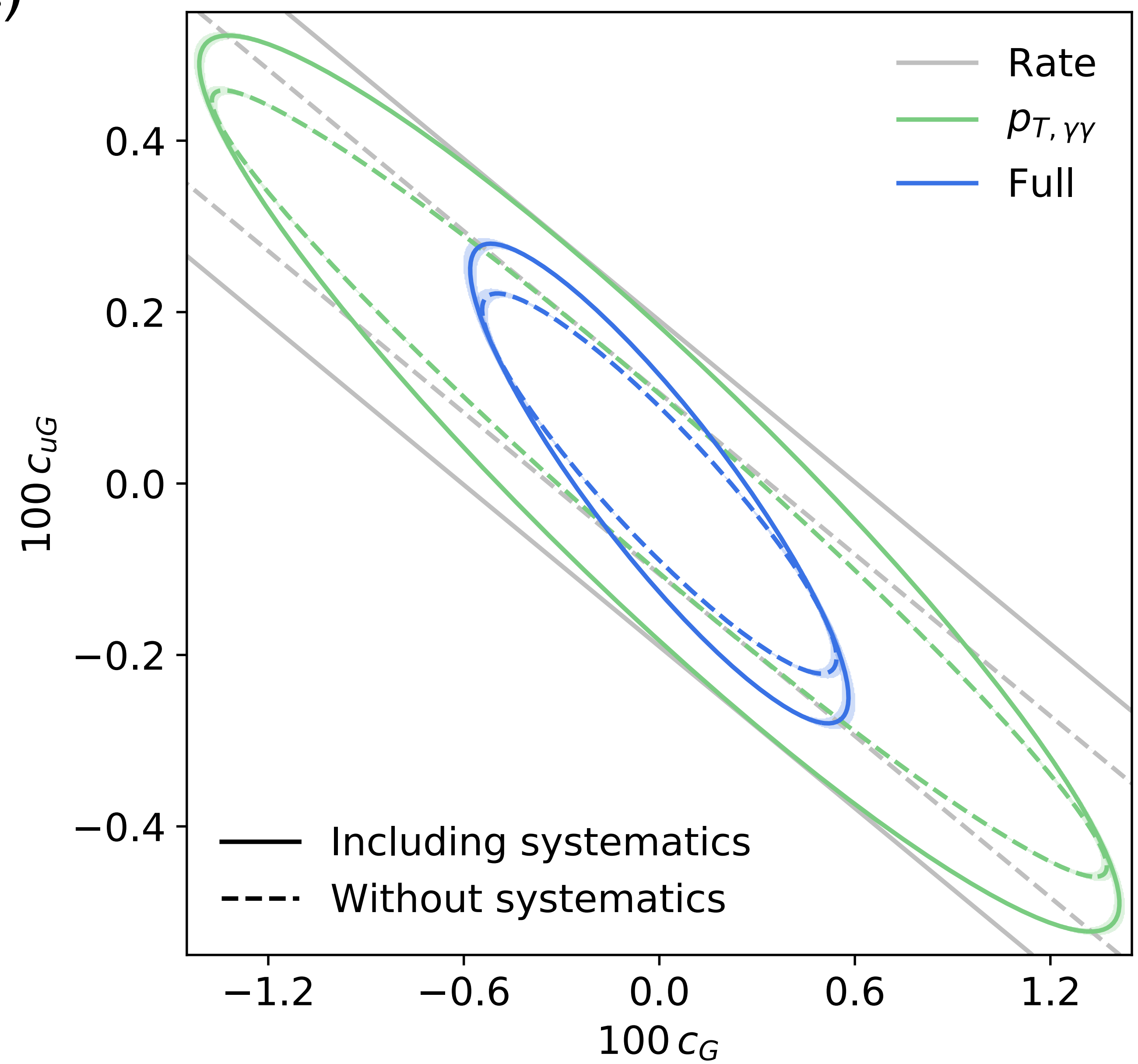
- results
  * rate: constrains only one direction
  * distributions: complementary information
  * shaded band: 2σ ML ensemble variance



How about Systematics?

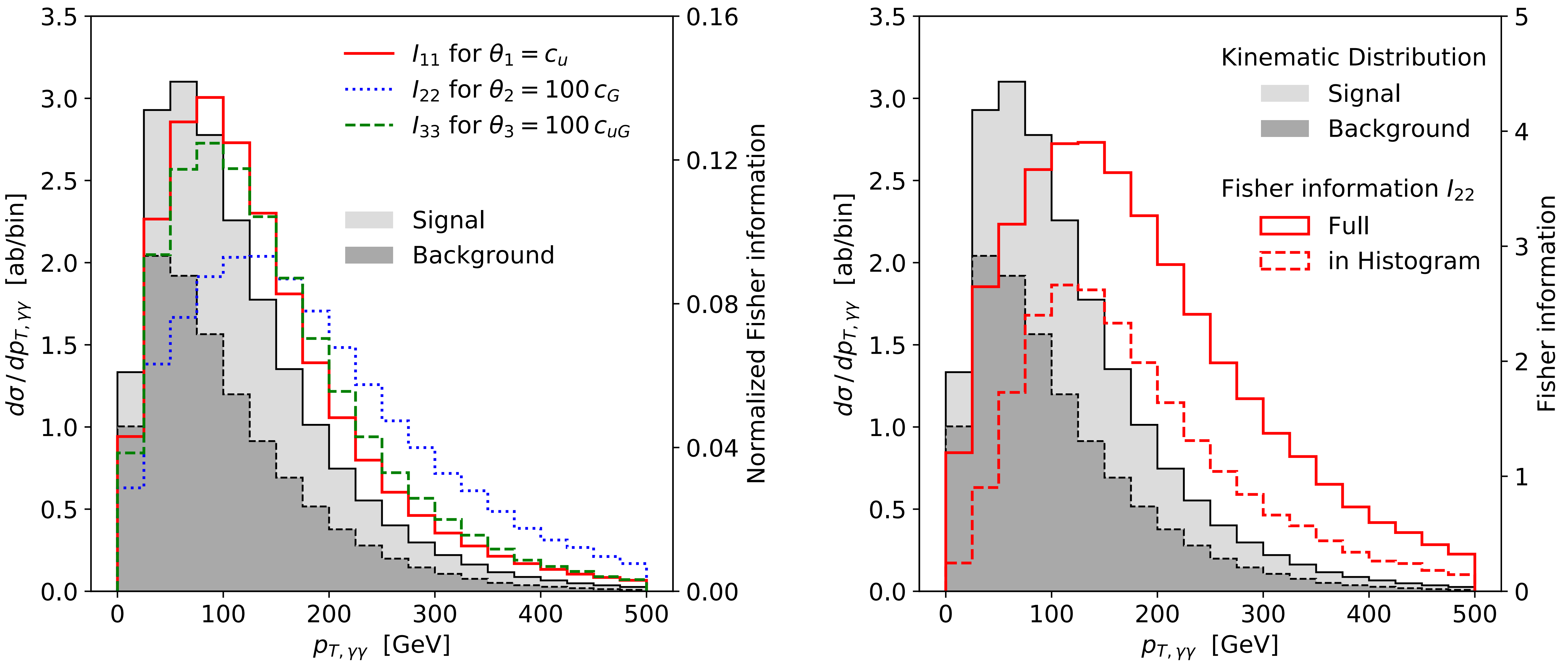
- strategy
  * introduce nuisance parameter ν for scale (2) and PDF (30) uncertainties
  * replace $r(x|\theta) \rightarrow r(x|\theta,\nu)$
  * learn score, obtain Fisher Info (3+32 dim.)
  * profile over ν

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



Kinematic distribution of the Fisher information

- identify the important phase-space regions



Probing SMEFT in tth

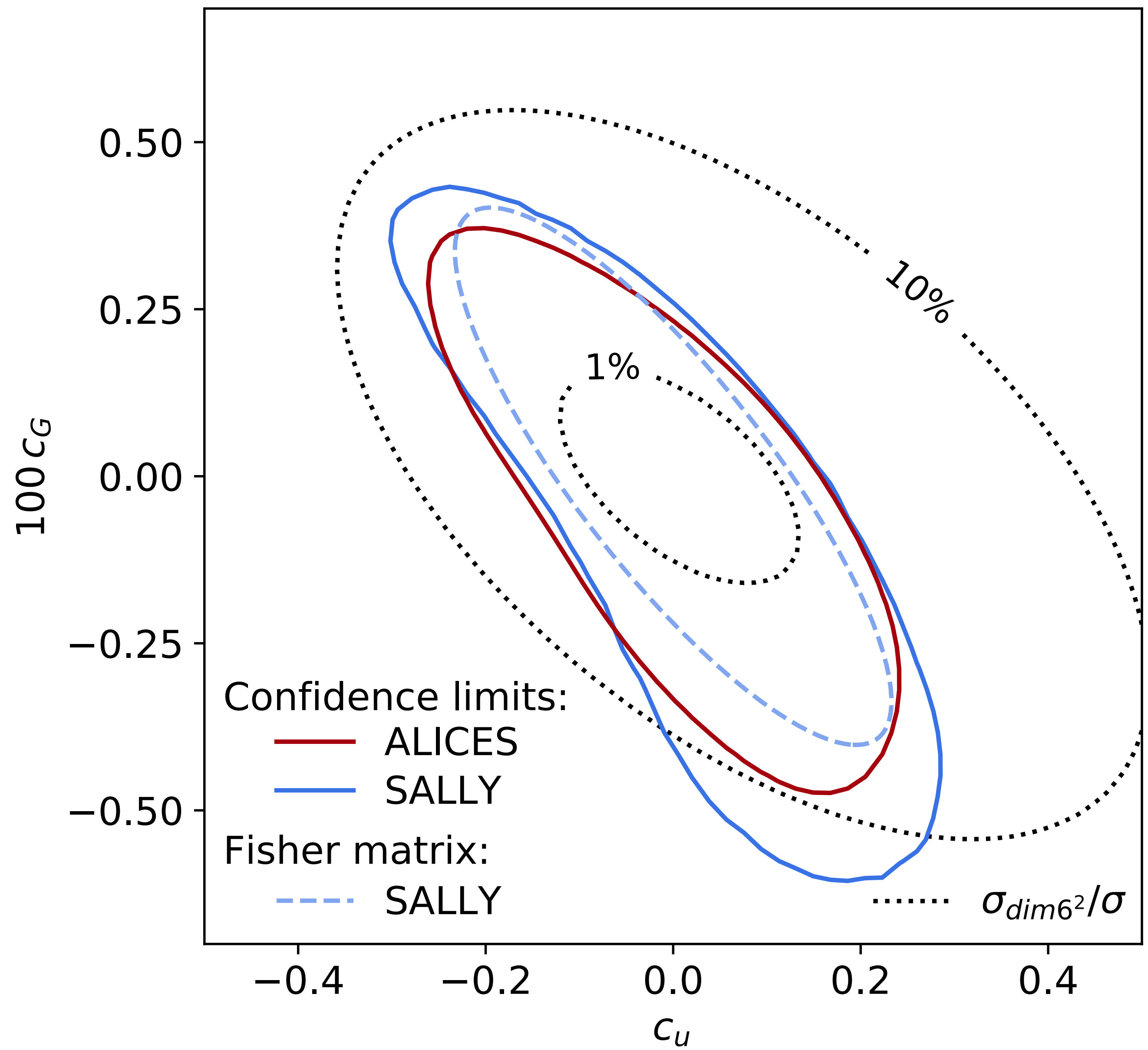
How good is local approximation?

- 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



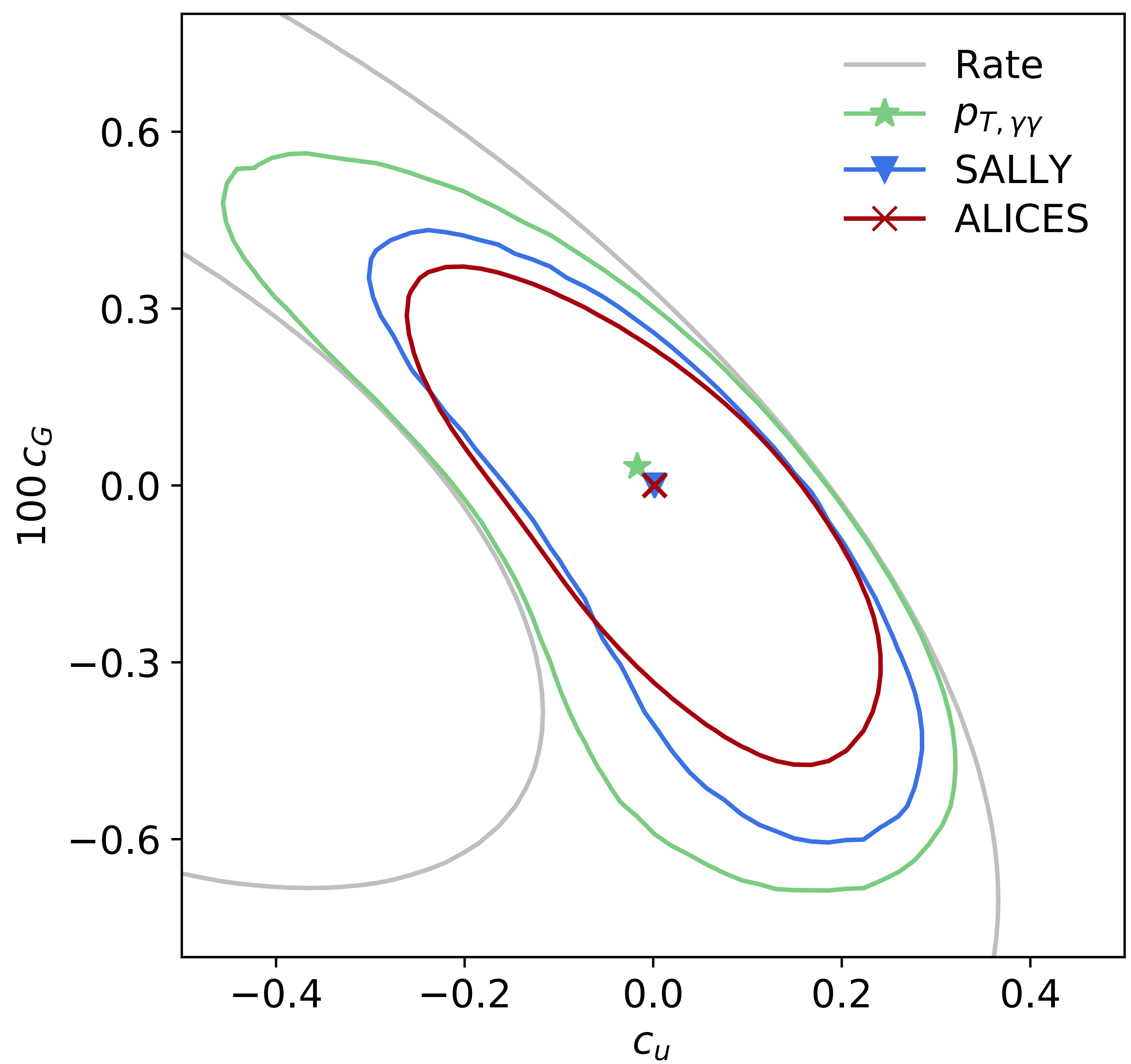
Full Results

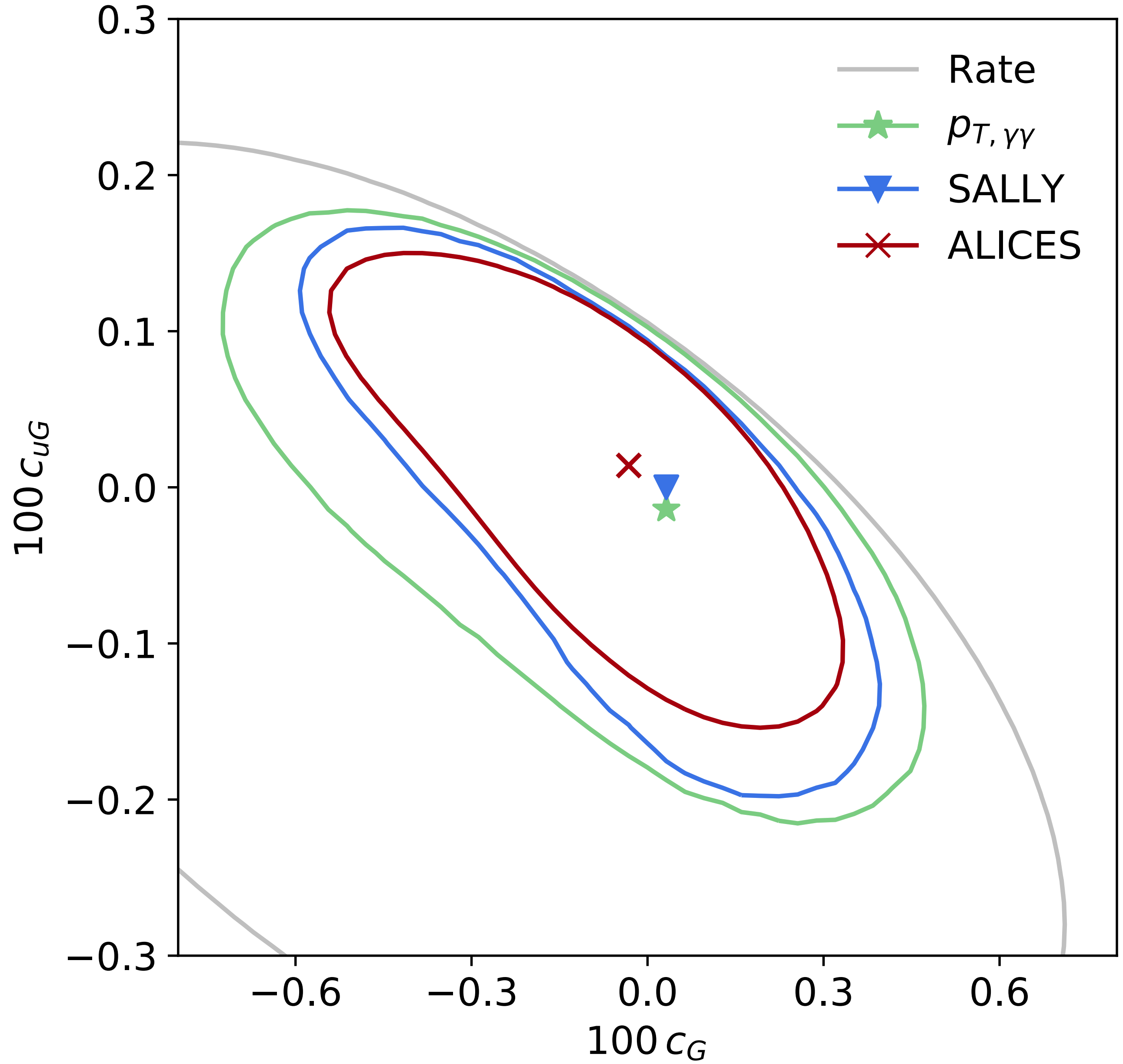
- inference methods

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

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

- multivariate methods significantly improve reach





Outline

Introduction: Inference
What's is the Problem?

Review: Inference Techniques
What did we do so far?

The MadMiner Approach
What do we do?

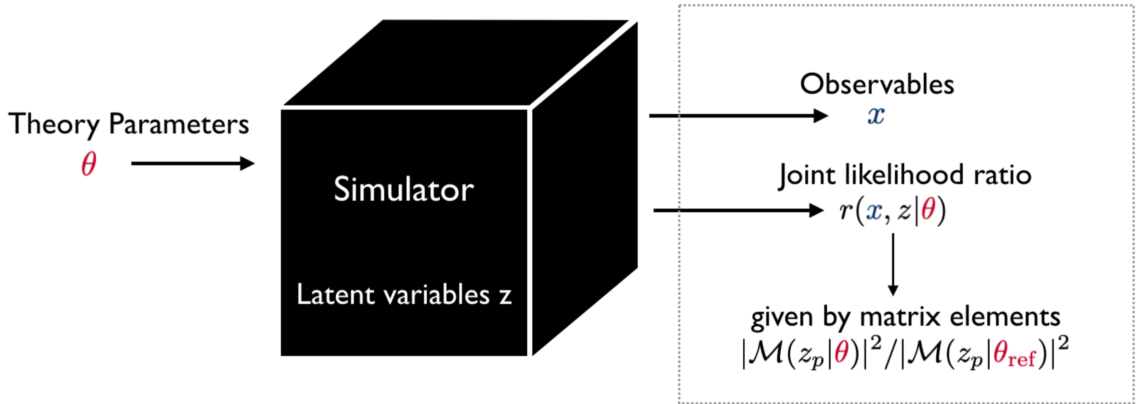
Optimal Observables and Fisher Information
This will turn out to be useful.

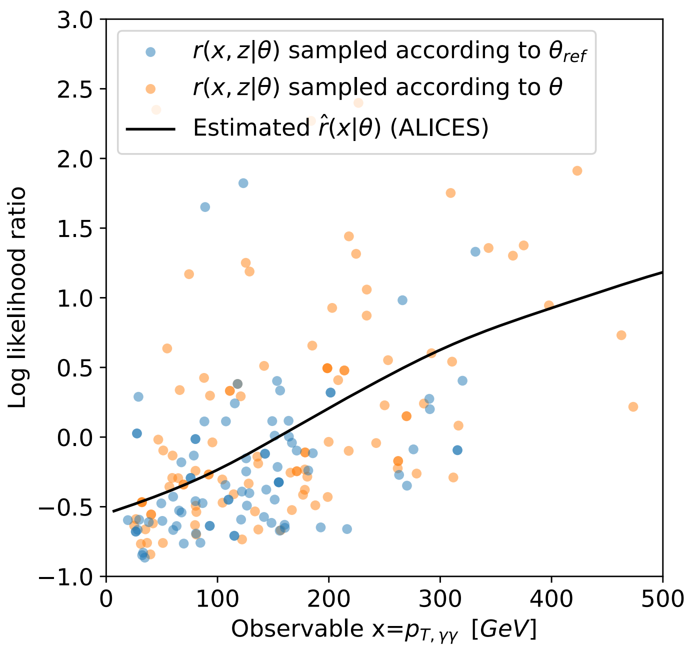
The MadMiner Tool
Using these methods is super easy!

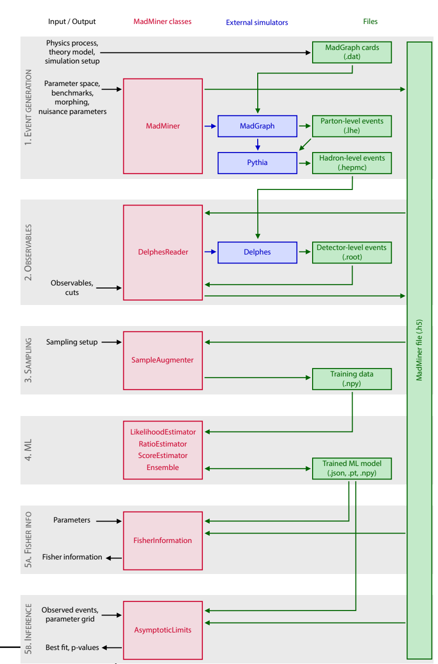
A Realistic Physics Example
Probing SMEFT in tth

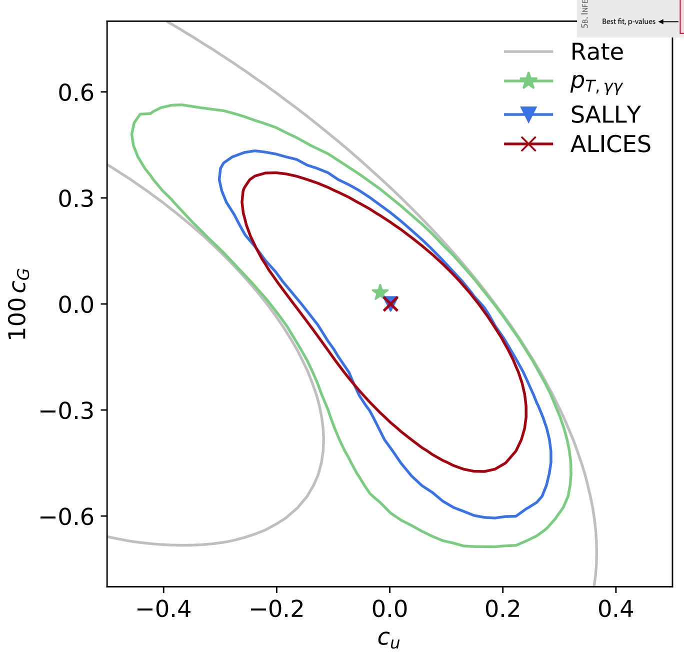
Summary and Conclusion

Summary and Conclusion









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

* task: determine likelihood function $p(x|\theta)$

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

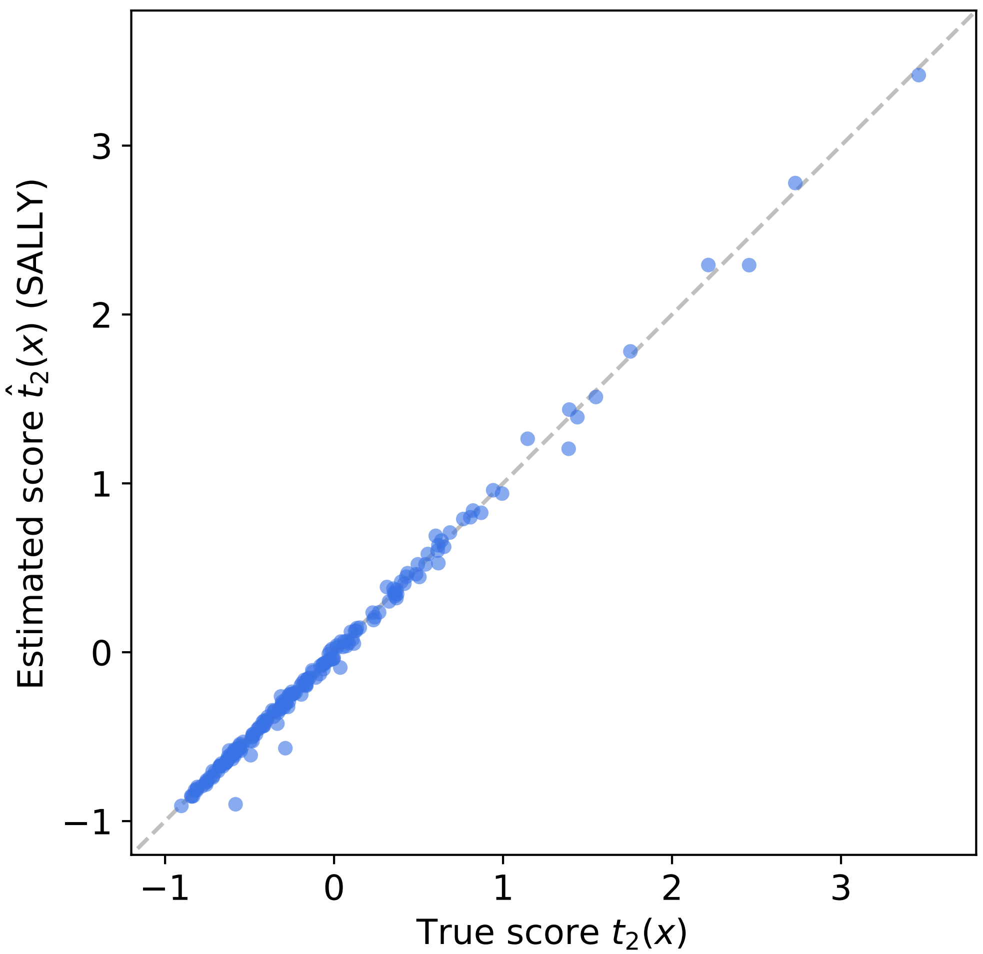
Tool: MadMiner [Brehmer, Cranmer, Espejo, FK 1907.10621]

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

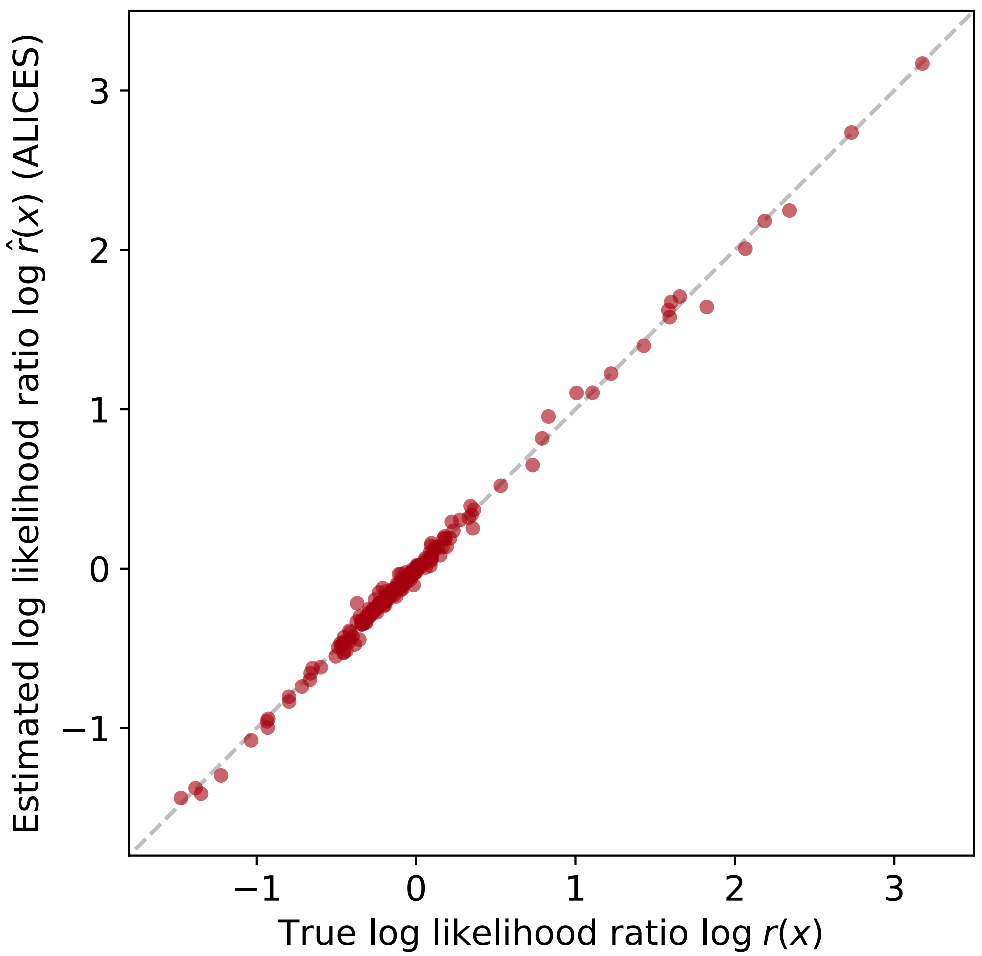
Backup

Cool stuff that didn't make it into the main part

Validation

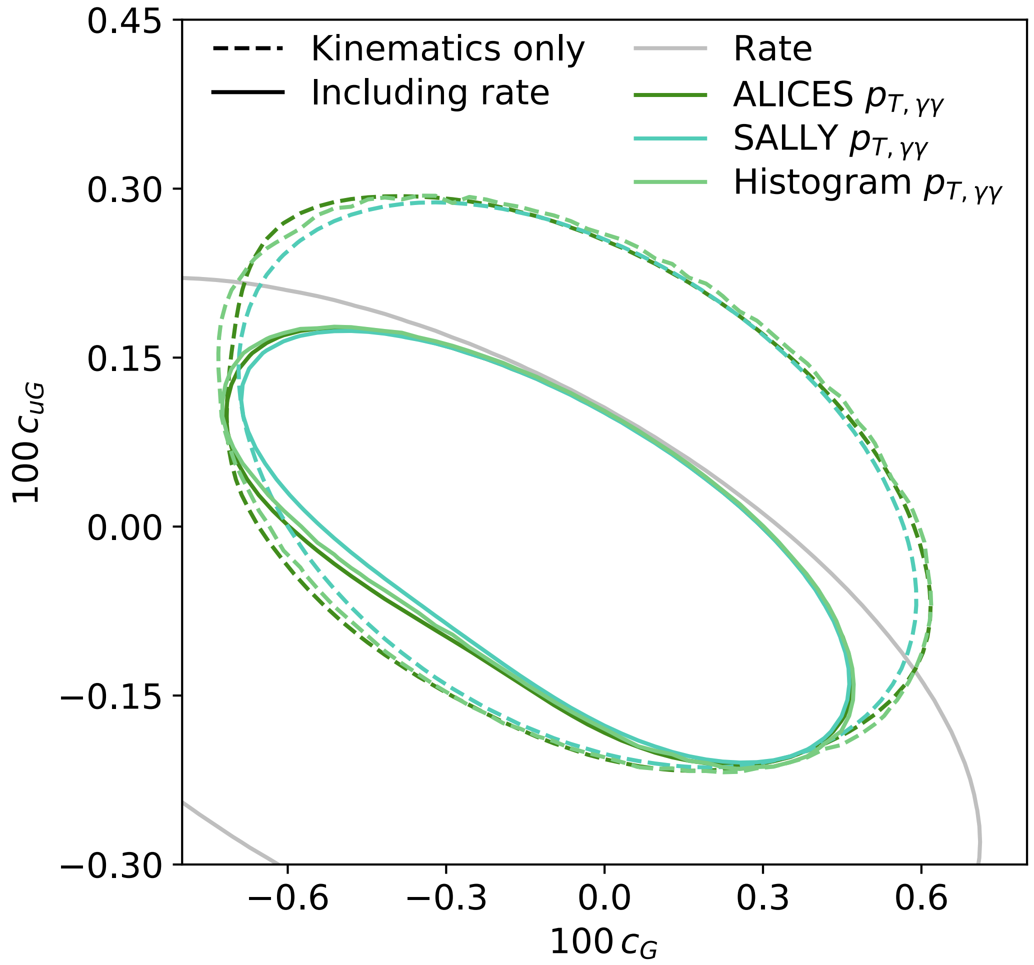


Estimated score $\hat{t}_2(x)$ (SALLY)
True score $t_2(x)$
−1
0
1
2
3

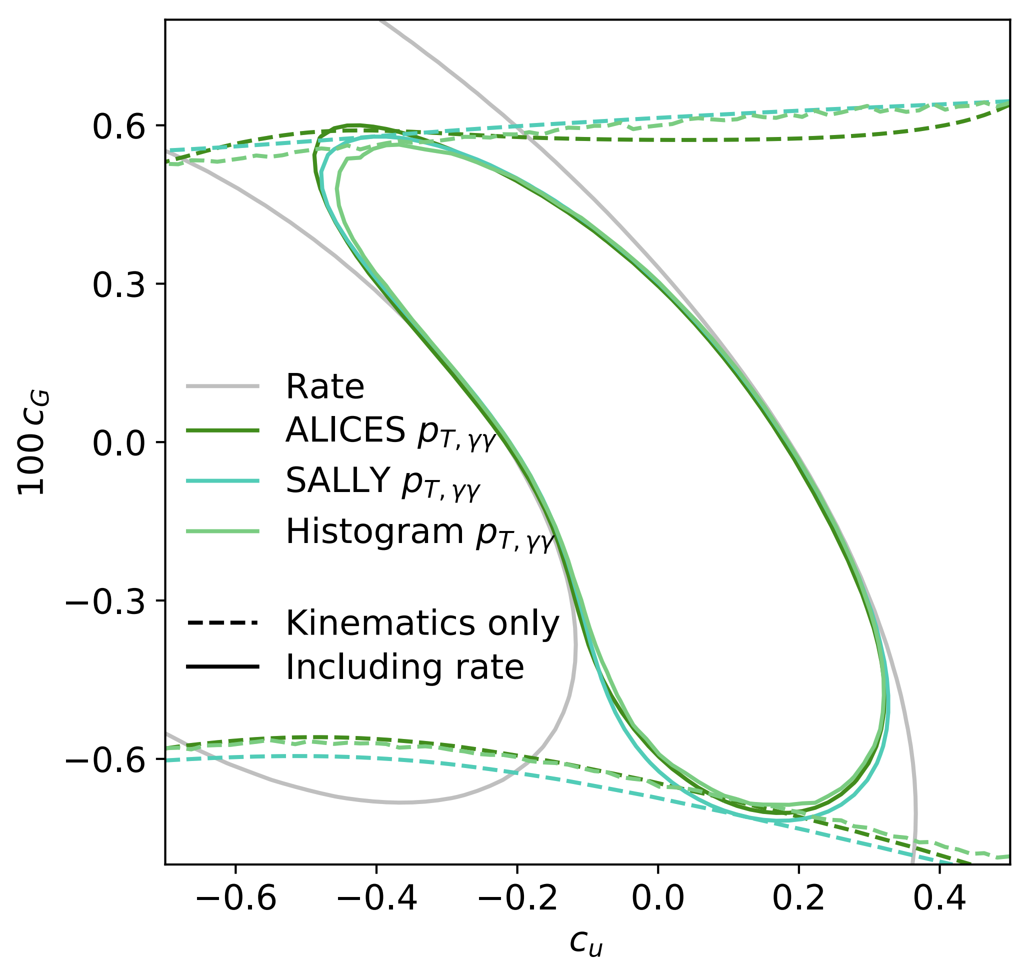


Estimated log likelihood ratio log $\hat{r}(x)$ (ALICES)
True log likelihood ratio log $r(x)$
−1
0
1
2
3

Validation

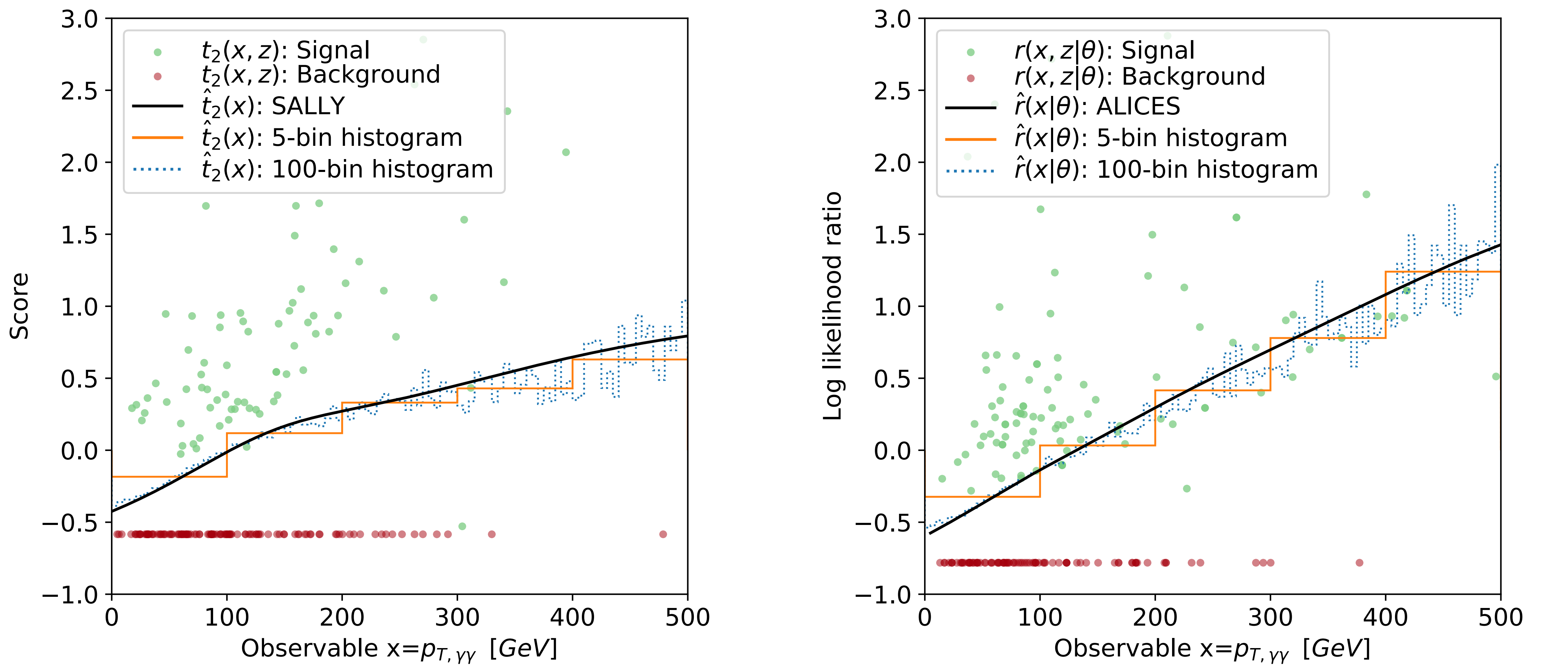


0.45
0.30
0.15
0.00
−0.15
−0.30
100 c_{uG}
−0.6
−0.3
0.0
0.3
0.6
100 c_G
Kinematics only
Including rate
Rate
ALICES $p_{T,\gamma\gamma}$
SALLY $p_{T,\gamma\gamma}$
Histogram $p_{T,\gamma\gamma}$



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
Rate
ALICES $p_{T,\gamma\gamma}$
SALLY $p_{T,\gamma\gamma}$
Histogram $p_{T,\gamma\gamma}$
Kinematics only
Including rate

Validation

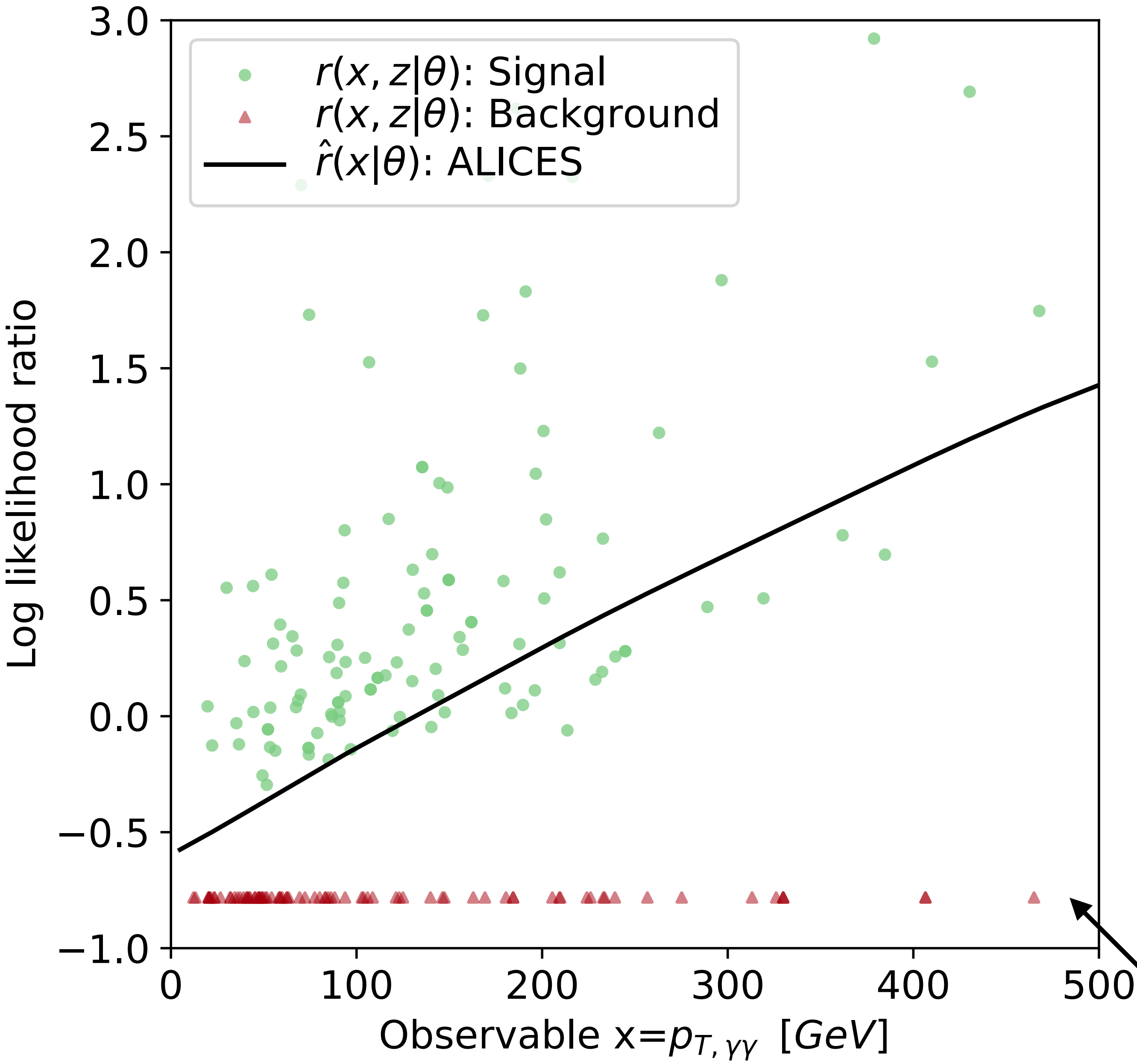


$t_2(x,z)$: Signal
$t_2(x,z)$: Background
$\hat{t}_2(x)$: SALLY
$\hat{t}_2(x)$: 5-bin histogram
$\hat{t}_2(x)$: 100-bin histogram
Score
Observable x=$p_{T,\gamma\gamma}$ $[GeV]$
$r(x,z|\theta)$: Signal
$r(x,z|\theta)$: Background
$\hat{r}(x|\theta)$: ALICES
$\hat{r}(x|\theta)$: 5-bin histogram
$\hat{r}(x|\theta)$: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

MadMiner with Backgrounds

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

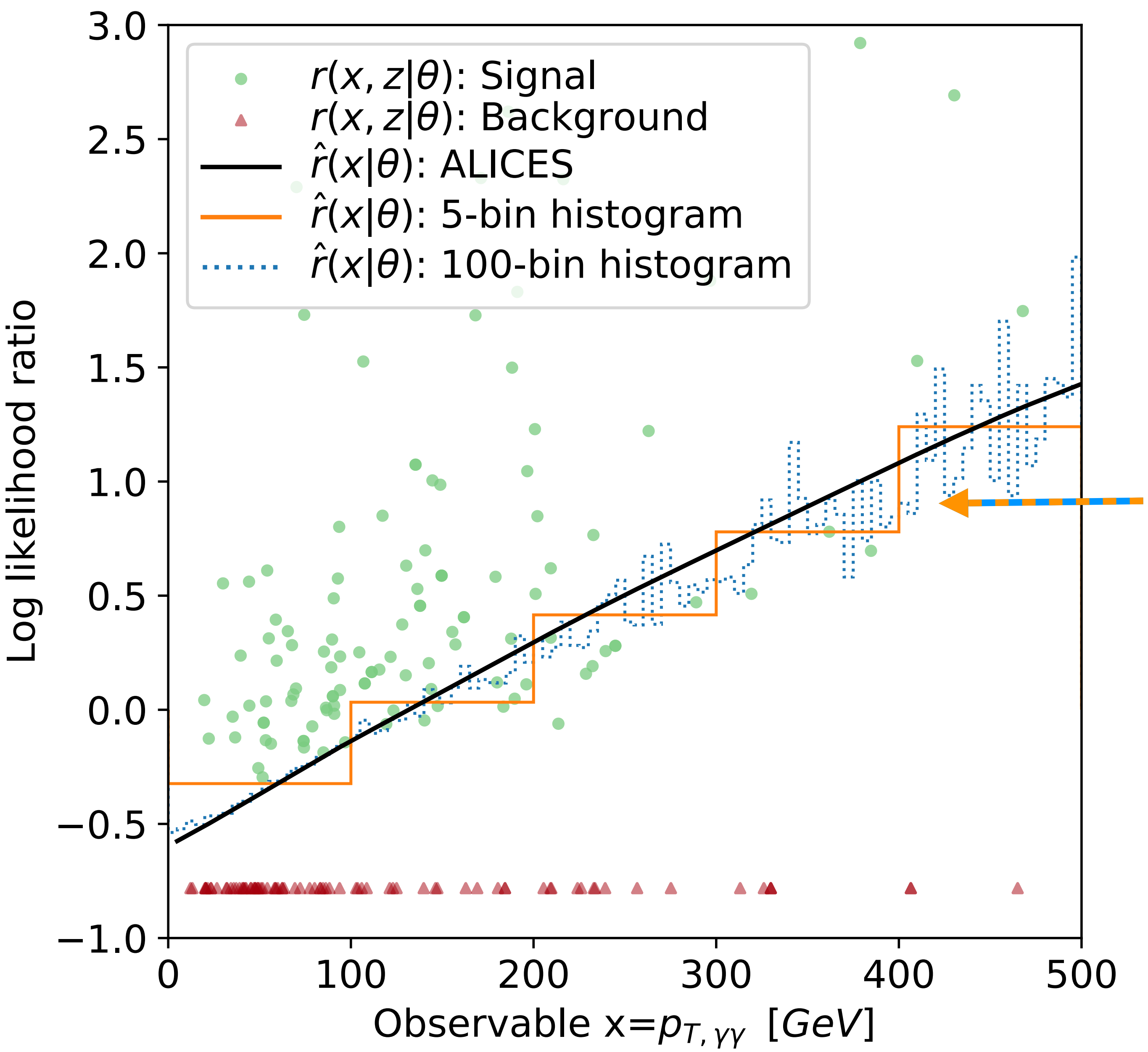
estimate r(x, θ): "best fit"

constant joint LLR for background:

$r(x,z|\theta) = \sigma(\theta_{ref})/\sigma(\theta)$

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

estimate r(x, θ): "best fit"

Comparison with Histogram

- LLR obtained using histogram
- agrees well :)
- ML "continuum limit of for large # of bins"

- realistic problem: x and θ high dimensional

Another example: WH production in SMEFT

- process WH>lvbb
- 3 operators contribute

$$\tilde{\mathcal{O}}_{HD} = (\phi^\dagger \phi) \Box (\phi^\dagger \phi) - \frac{1}{4} (\phi^\dagger D^\mu \phi)^* (\phi^\dagger D_\mu \phi)$$

$$\mathcal{O}_{HW} = \phi^\dagger \phi \, W^a_{\mu\nu} W^{\mu\nu a}$$

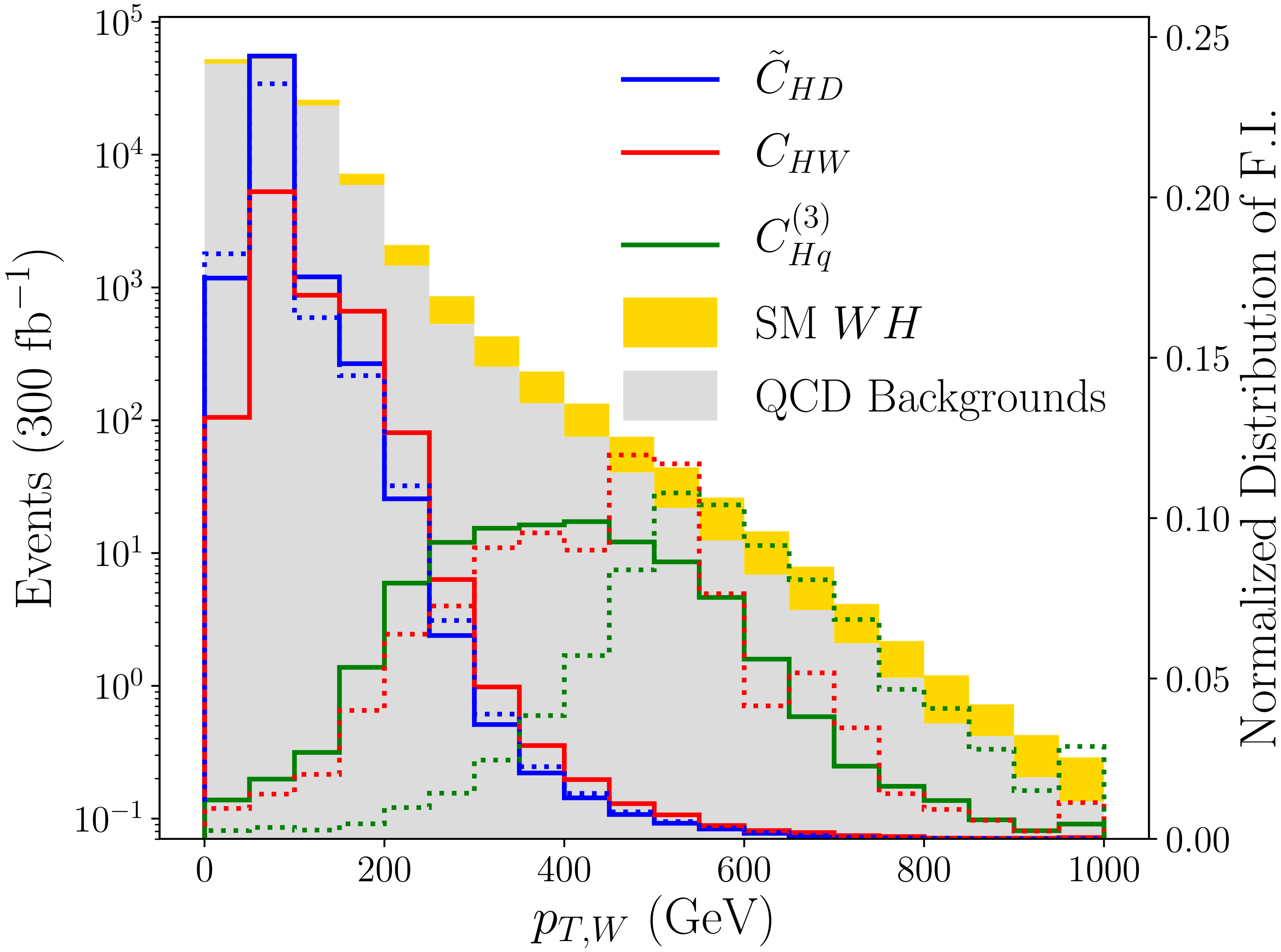
$$\mathcal{O}^{(3)}_{Hq} = (\phi^\dagger \overleftrightarrow{D}^a{}_\mu \, \phi) (\bar{Q}_L \sigma^a \gamma^\mu Q_L)$$

- 2 main observables: ptH and mT,tot
- where is the information
  * identify sensitive phase space regions

Distribution of Information:

— interference term

- - - - quadratic term



Another example: WH production in SMEFT

- process WH>lvbb
- 3 operators contribute
- 2 main observables: ptH and mT,tot

- how good can simplified template cross sections probe WH?
  * quantify performance using information geometry
  * compare to full information
  * additional high pT bin essential
  * include 2nd observable
  * multivariate analysis potentially much more powerful