Learning to constrain new physics

Johann Brehmer

New York University

University of Maryland, Elementary Particle Seminar

August 27, 2018

The legacy of the LHC

- What's hiding in the electroweak sector?

  $\Rightarrow$ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, …)

The legacy of the LHC

- What's hiding in the electroweak sector?

  ⇒ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, …)

- These measurements are difficult!

  1. Many parameters

$$S = \int \mathrm{d}^4 x \left[\mathcal{L}_{\mathrm{SM}} + \frac{f_{\phi,2}}{\Lambda^2} \, \frac{1}{2} \, \partial_\mu (\phi^\dagger \phi) \, \partial^\mu (\phi^\dagger \phi) + \frac{f_{\phi,3}}{\Lambda^2} \, \frac{1}{3} (\phi^\dagger \phi)^3 \right.$$
$$+ \frac{f_{GG}}{\Lambda^2} \, (\phi^\dagger \phi) \, G^a_{\mu\nu} \, G^{\mu\nu\, a} - \frac{f_{BB}}{\Lambda^2} \, \frac{g'^2}{4} \, (\phi^\dagger \phi) \, B_{\mu\nu} \, B^{\mu\nu} - \frac{f_{WW}}{\Lambda^2} \, \frac{g^2}{4} \, (\phi^\dagger \phi) \, W^a_{\mu\nu} \, W^{\mu\nu\, a}$$
$$+ \frac{f_B}{\Lambda^2} \, \frac{\mathrm{i} g'}{2} \, (D^\mu \phi)^\dagger \, D^\nu \phi \, B_{\mu\nu} + \frac{f_W}{\Lambda^2} \, \frac{\mathrm{i} g}{2} \, (D^\mu \phi)^\dagger \, \sigma^a \, D^\nu \phi \, W^a_{\mu\nu}$$
$$+ \frac{f_\ell}{\Lambda^2} \, (\phi^\dagger \phi) \, \overline{L}_L \phi \, \ell_R + \frac{f_u}{\Lambda^2} \, (\phi^\dagger \phi) \, \overline{Q}_L \tilde{\phi} \, u_R + \frac{f_d}{\Lambda^2} \, (\phi^\dagger \phi) \, \overline{Q}_L \phi \, d_R$$
$$\left. + \frac{f_{G\widetilde{G}}}{\Lambda^2} \, (\phi^\dagger \phi) \, G^a_{\mu\nu} \, \widetilde{G}^{\mu\nu\, a} - \frac{f_{B\widetilde{B}}}{\Lambda^2} \, \frac{g'^2}{4} (\phi^\dagger \phi) \, B_{\mu\nu} \, \widetilde{B}^{\mu\nu} - \frac{f_{W\widetilde{W}}}{\Lambda^2} \, \frac{g^2}{4} (\phi^\dagger \phi) \, W^a_{\mu\nu} \, \widetilde{W}^{\mu\nu\, a} \right]$$

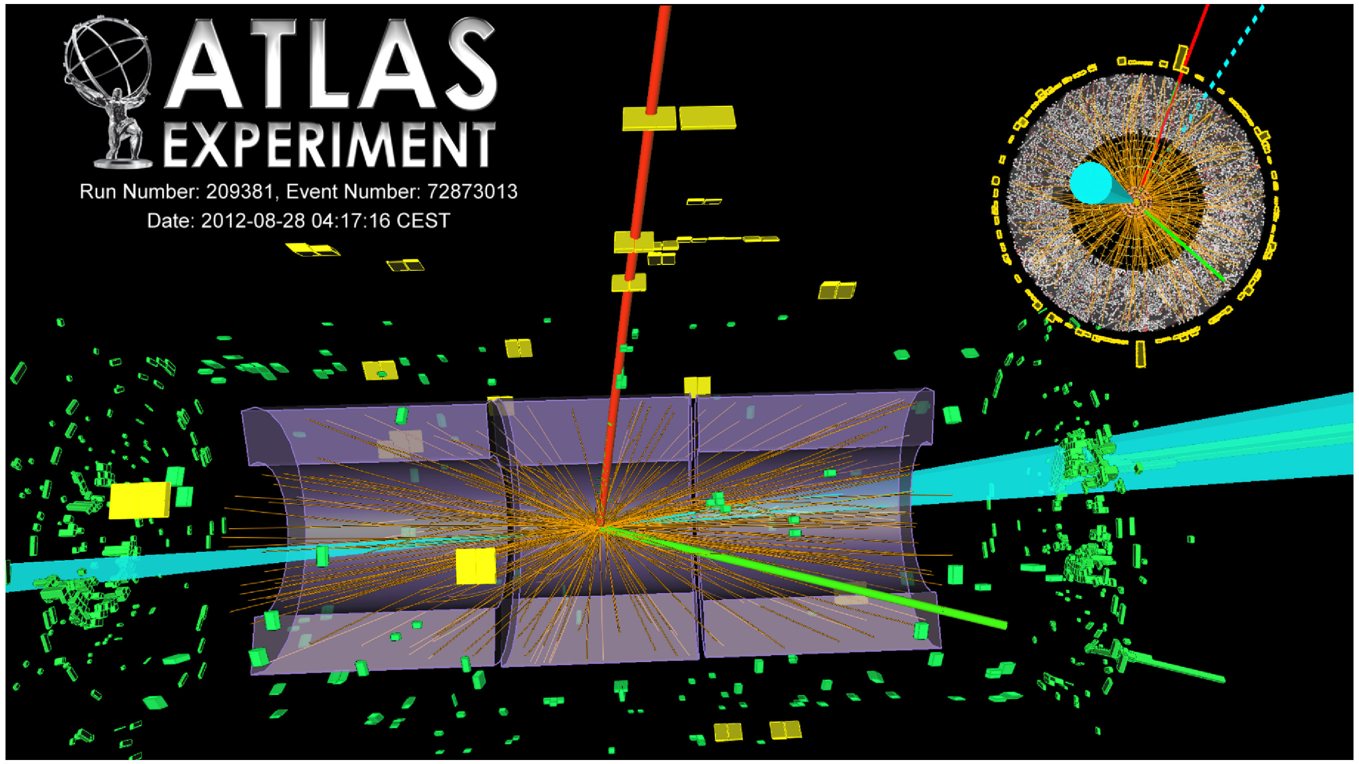
The legacy of the LHC

- What's hiding in the electroweak sector?

  ⇒ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, …)

- These measurements are difficult!

  1. Many parameters
  2. Many observables



[ATLAS 1501.04943]

The legacy of the LHC

- What's hiding in the electroweak sector?

  ⇒ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, …)

- These measurements are difficult!

  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects

The legacy of the LHC

- What's hiding in the electroweak sector?

  ⇒ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, …)

- These measurements are difficult!

  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects

[M. Yao, idea for analogy: K. Cranmer]

The legacy of the LHC

- What's hiding in the electroweak sector?

  ⇒ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, …)

- These measurements are difficult!

  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects
  4. The likelihood function of high-dimensional observables cannot be calculated

The legacy of the LHC

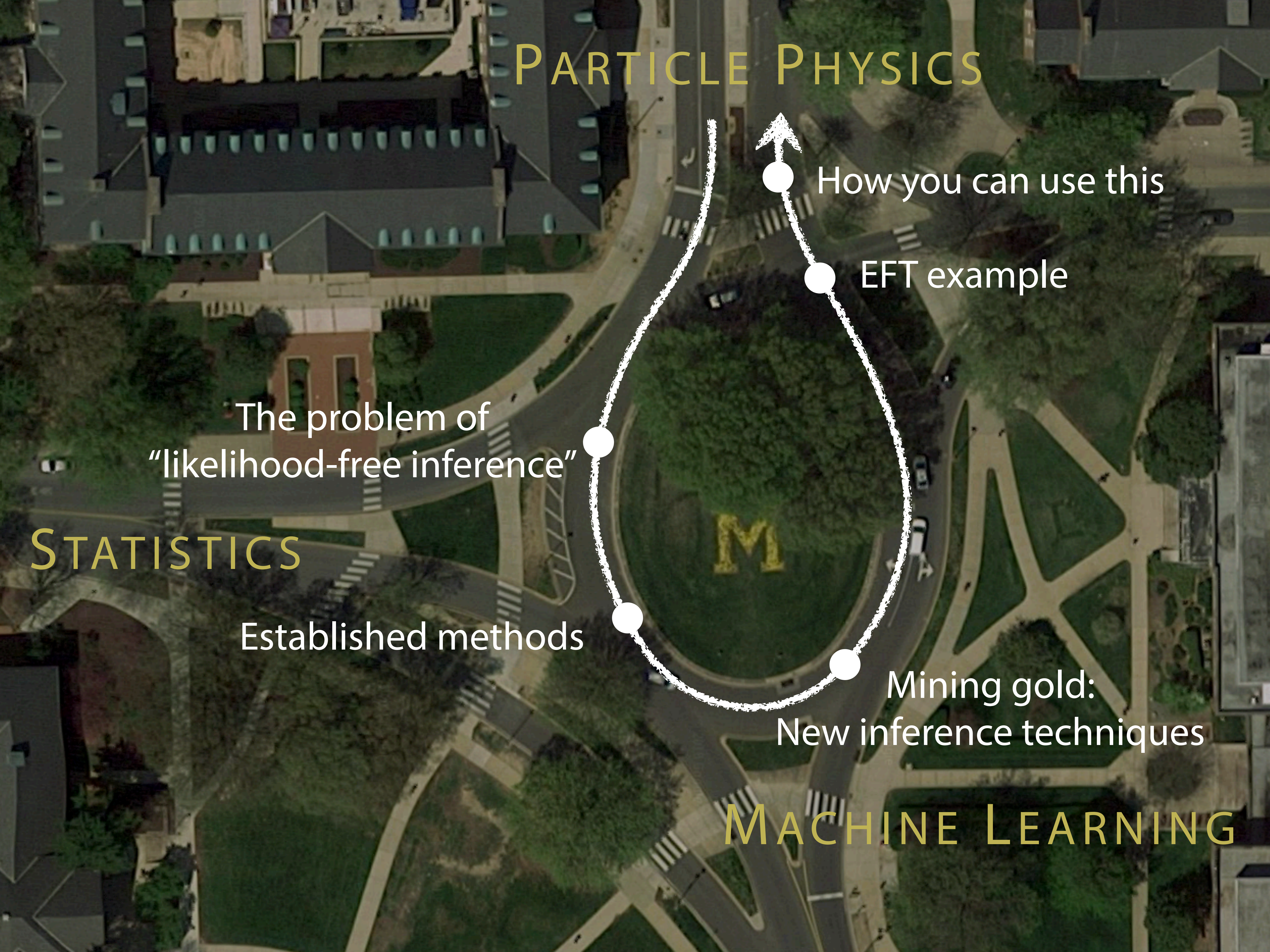
- What's hiding in the electroweak sector?

  ⇒ precision constraints on dimension-6 EFT operators
  (or Pseudo-Observables, non-linear EFT, ...)

- These measurements are difficult!

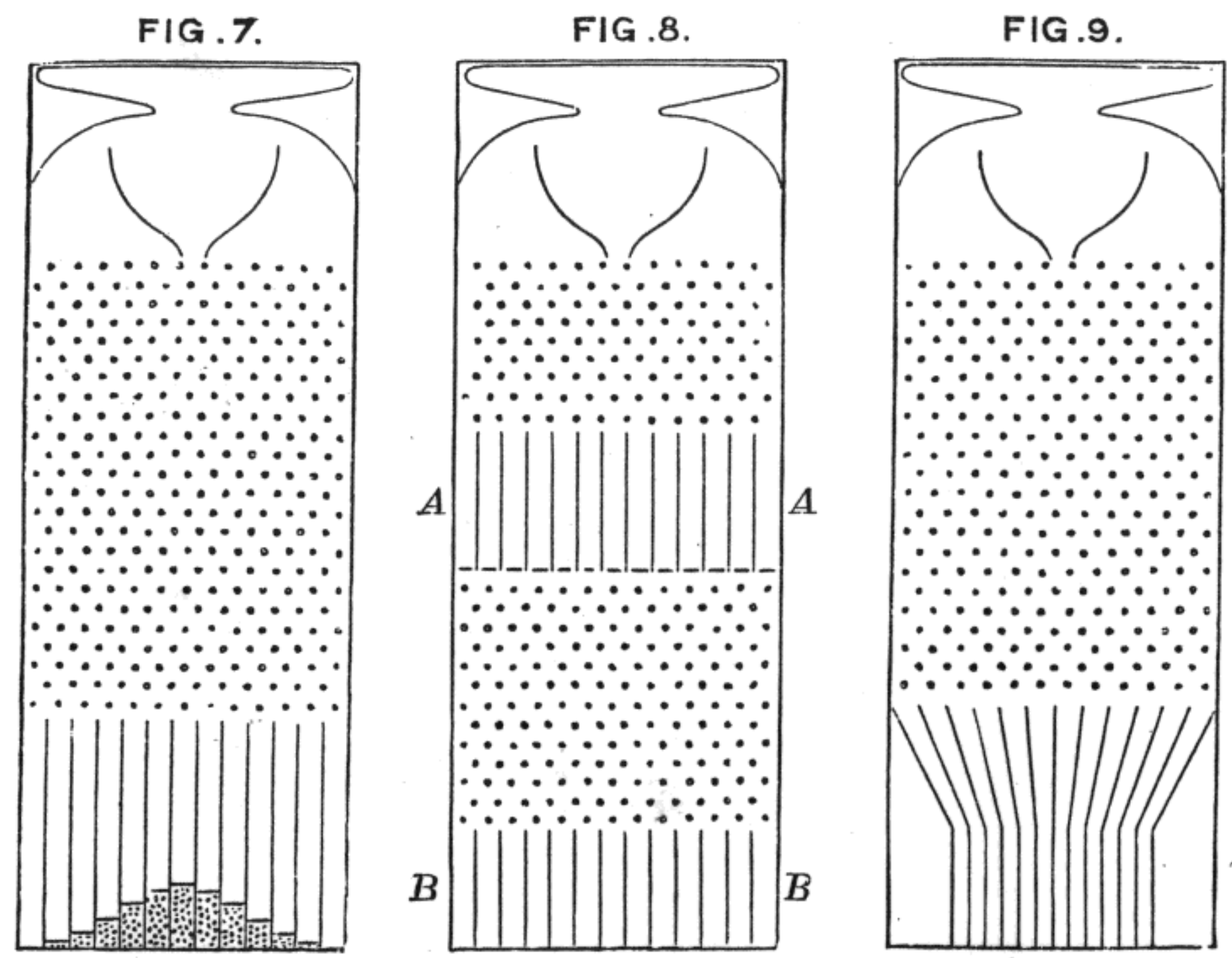
  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects
  4. The likelihood function of high-dimensional observables cannot be calculated

- Established data analysis methods struggle...
  Can new analysis techniques improve the sensitivity?


Particle Physics
How you can use this
EFT example
The problem of
“likelihood-free inference”
Statistics
Established methods
Mining gold:
New inference techniques
Machine Learning

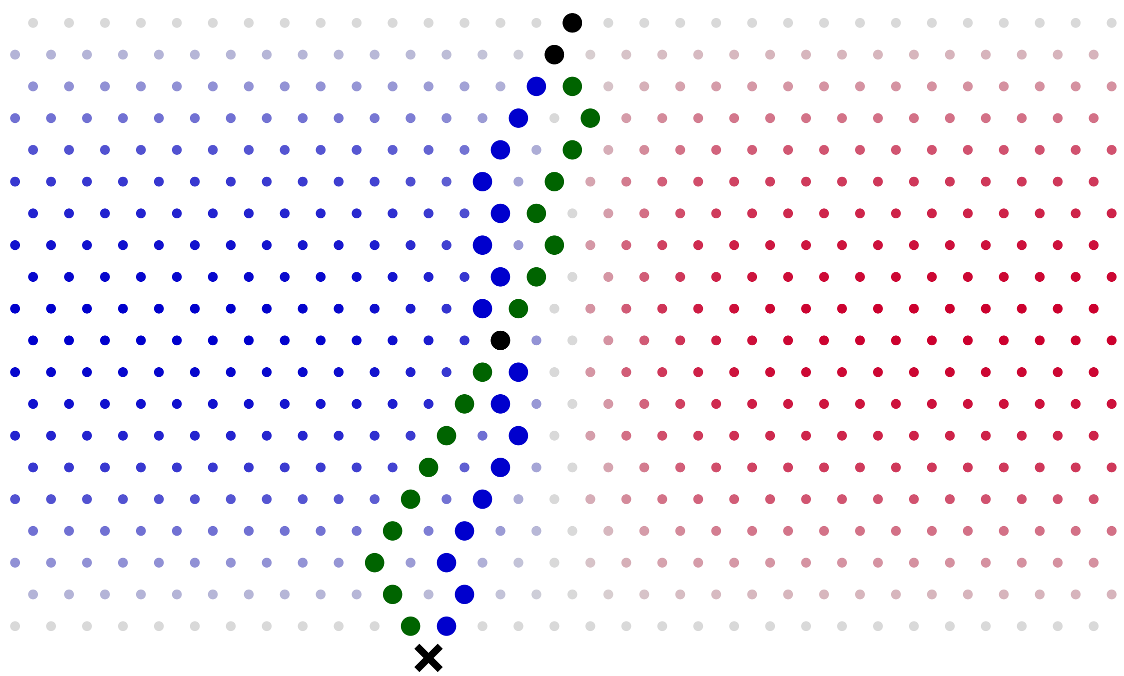
Likelihood-free inference

The Galton board



[Source: F. Galton 1889]

Probabilities



Probability of ending in bin x:

$$p(x) = \int \mathrm{d}z \; p(x, z)$$

Sum over
all trajectories
("latent variables")

Probability of
each path z
from start to x

The generalized Galton board

What if probability to go left at a nail is not always 0.5, but some (known) function of some parameters θ?

- Prediction: given θ, generate samples of observations $\{x_i\}$.

  Simple: just drop balls!

The generalized Galton board

What if probability to go left at a nail is not always 0.5, but some (known) function of some parameters θ?

- Prediction: given θ, generate samples of observations $\{x_i\}$.

  Simple: just drop balls!

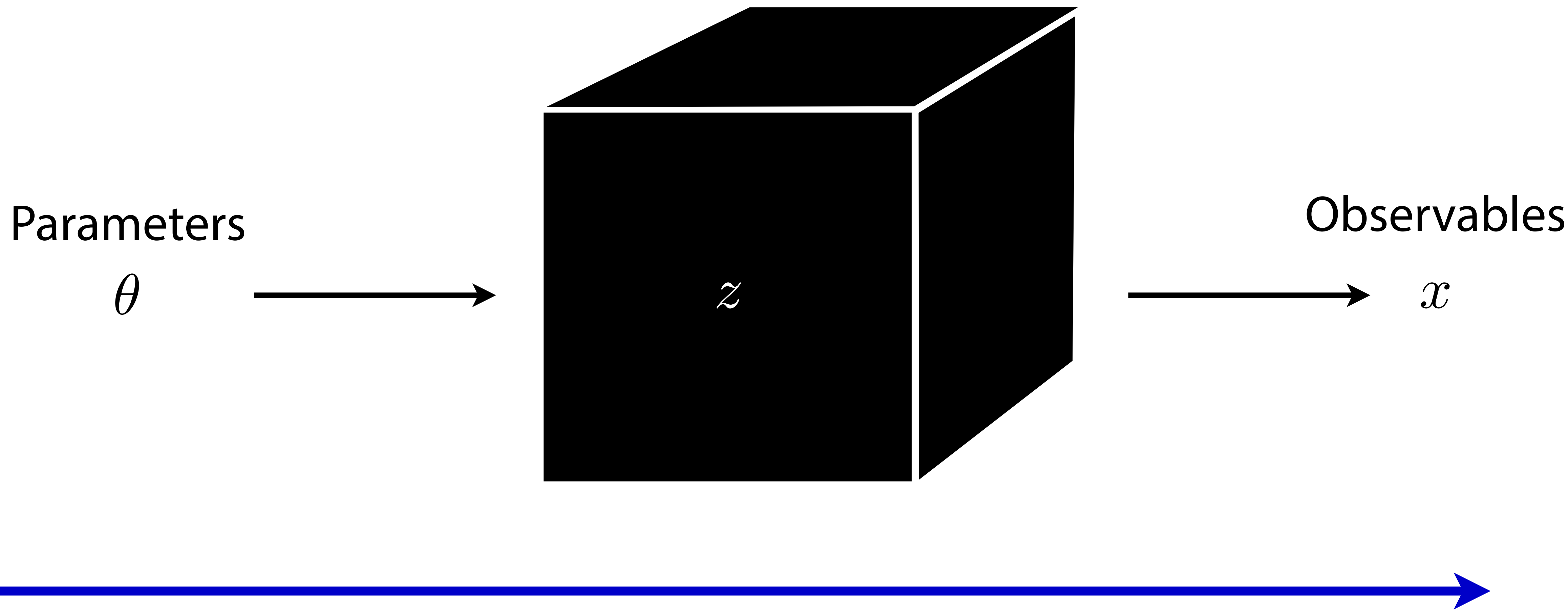
- Inference: given observations $\{x_i\}$, what are the most likely values for θ?

  “Easy” problem if we can evaluate likelihood $p(x_i|\theta)$. But

$$p(x|\theta) = \int \mathrm{d}z \, p(x, z|\theta)$$

  The number of possible paths z can be huge, and it becomes impossible to calculate the integral!

“Likelihood-free inference”

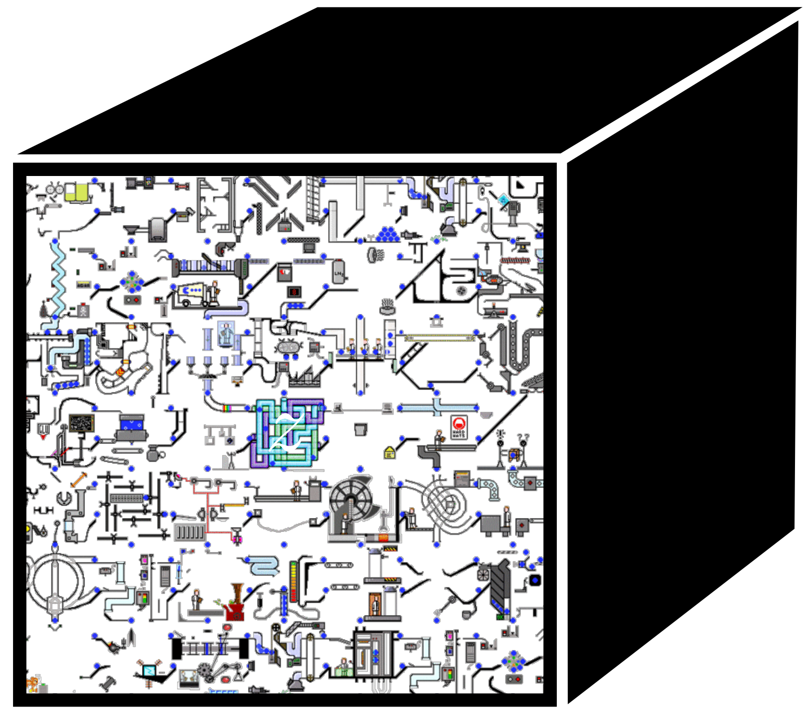


Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

“Likelihood-free inference”

Parameters
θ

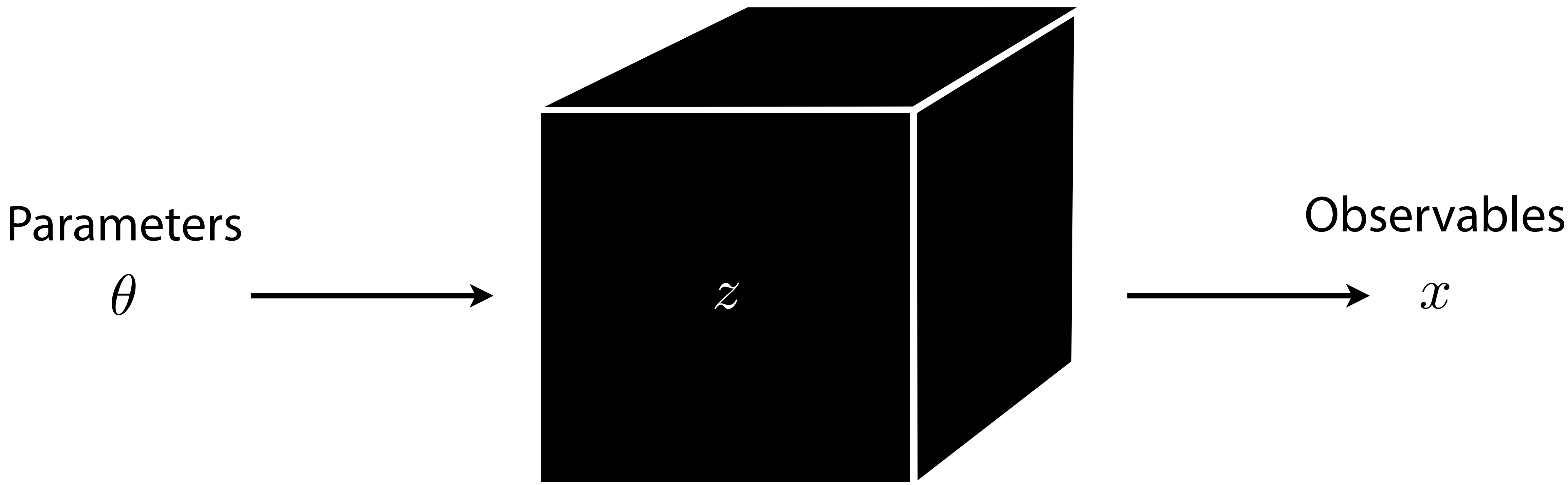


Observables
x

Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

"Likelihood-free inference"



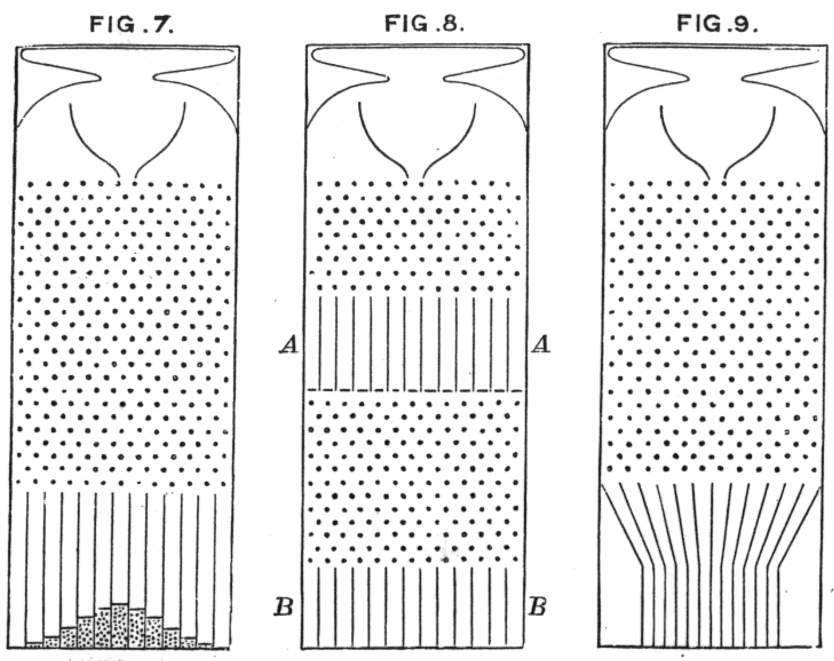
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

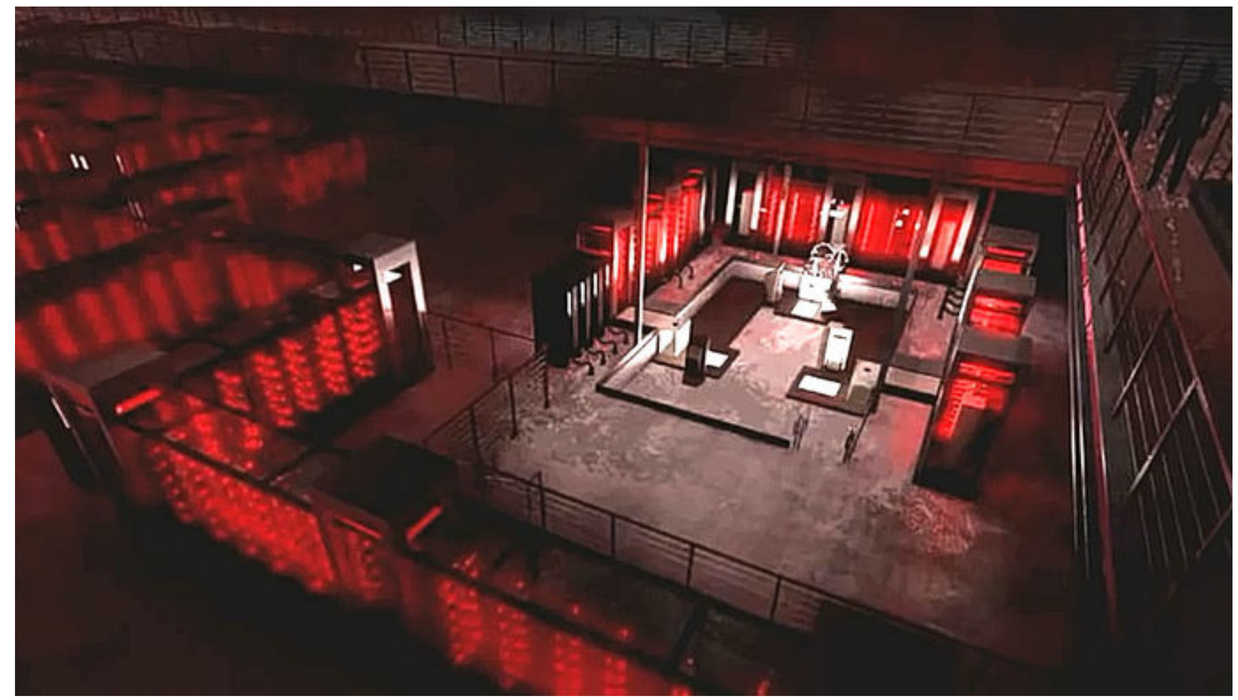
Inference:

- Likelihood function $p(x|\theta)$ is intractable
- Inference needs estimator $\hat{p}(x|\theta)$

Galton board: metaphor for simulator-based science



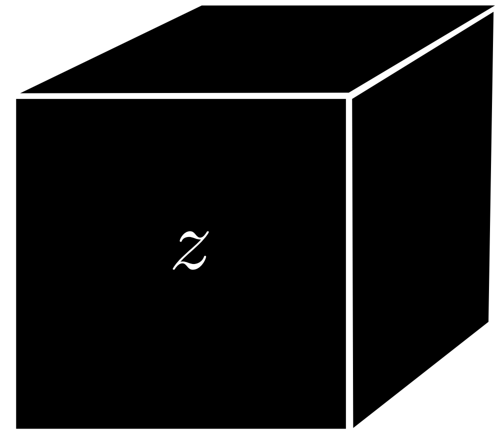
[Source: F. Galton 1889]



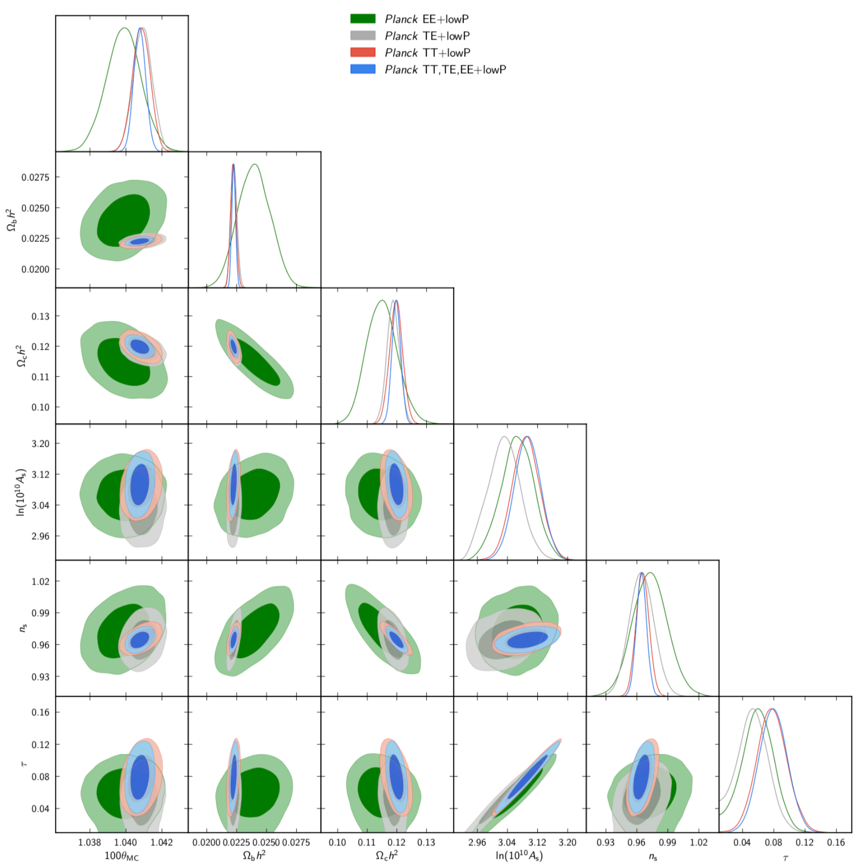
[Source: HBO 2018]

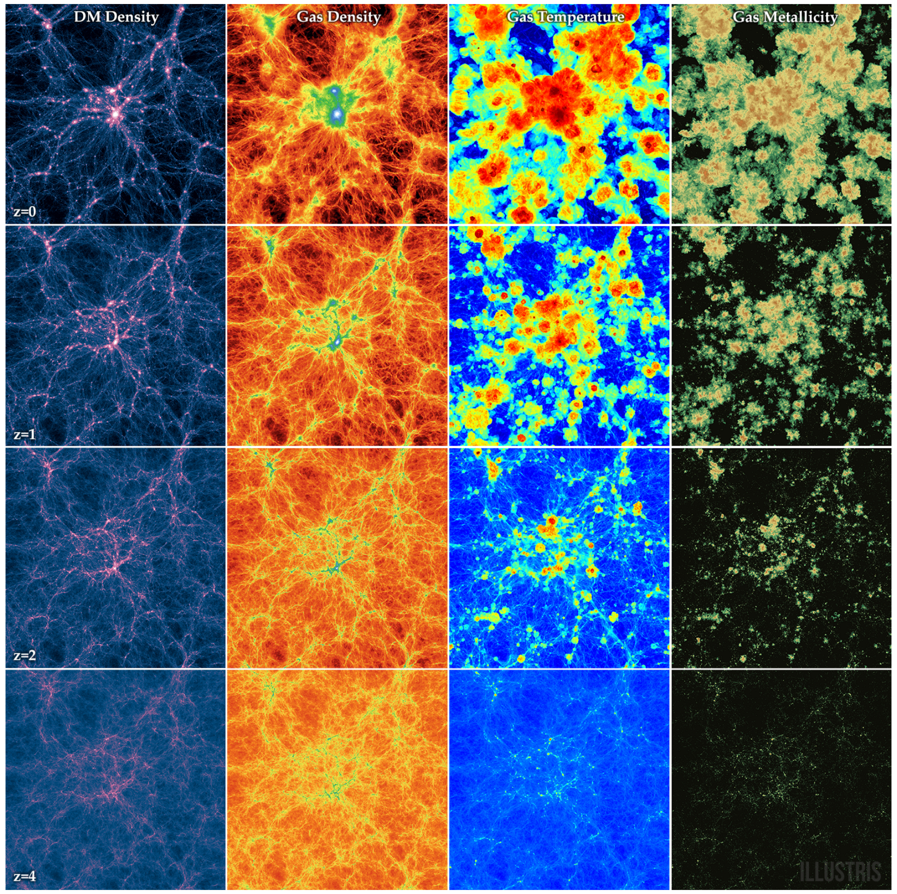
Galton board device	$\longrightarrow$	Computer simulation
Parameters θ	$\longrightarrow$	Parameters of interest θ
Bins x	$\longrightarrow$	Observables x
Path z	$\longrightarrow$	Latent variables z (stochastic execution trace through simulator)

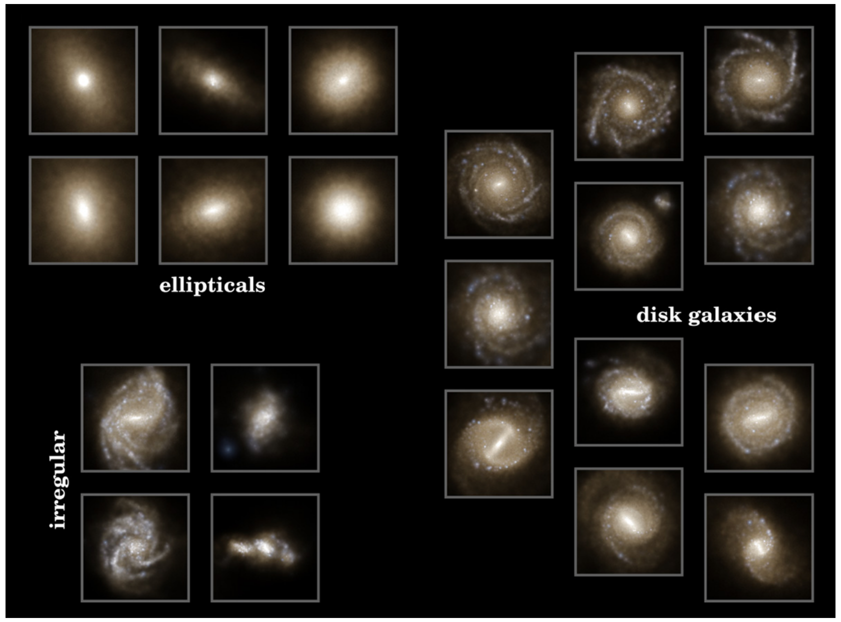
Cosmological N-body simulations

Parameters θ → 

→ Observables x



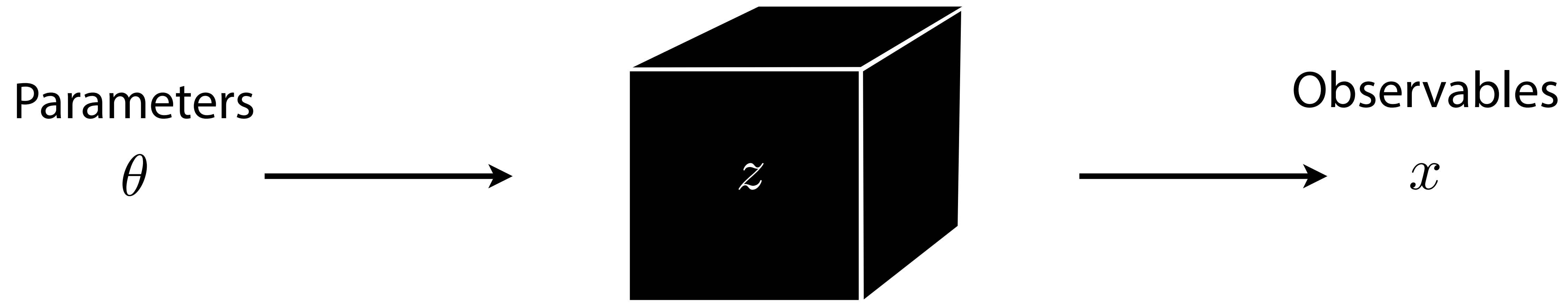


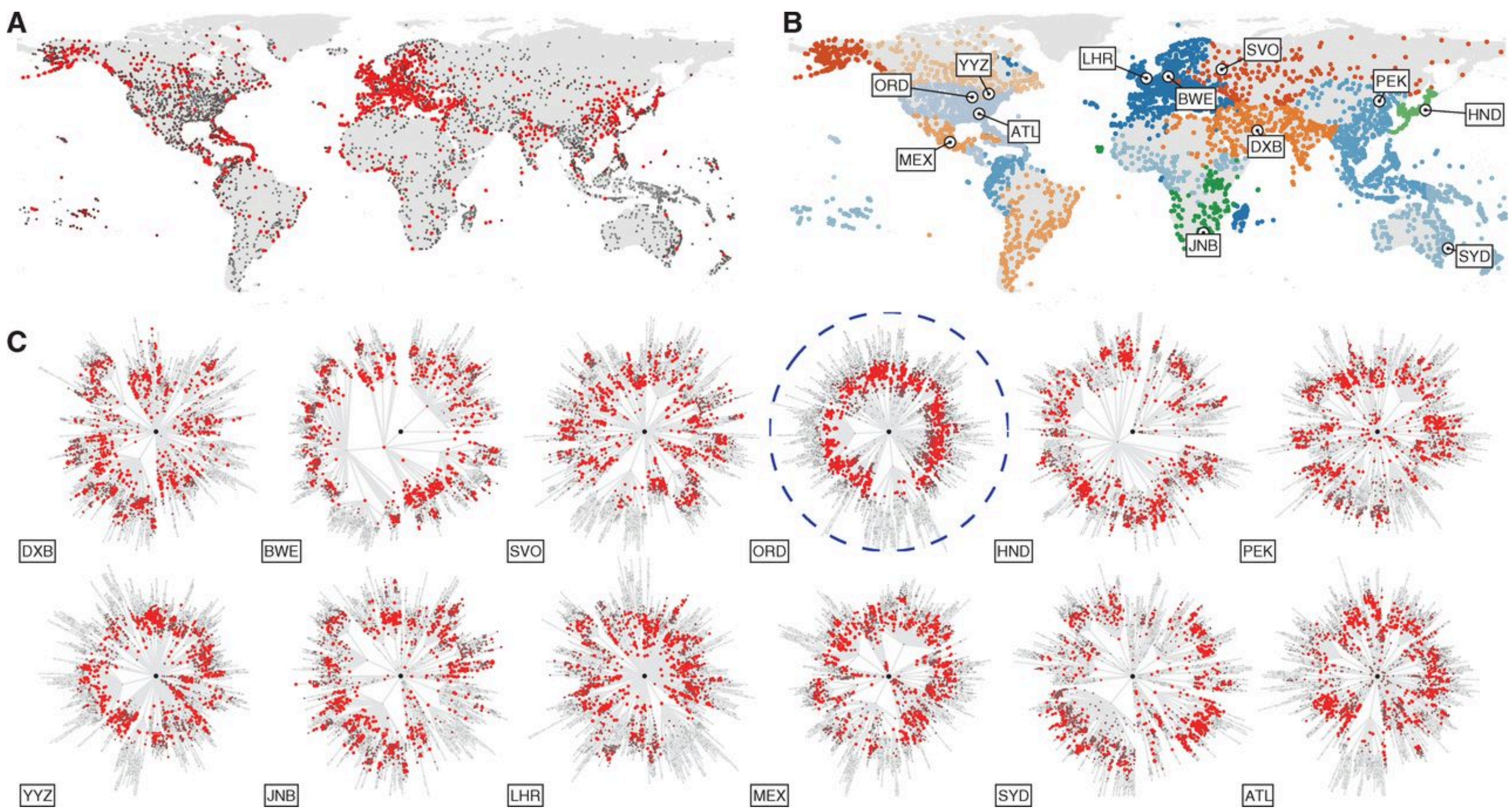


[Source: Planck 1502.01589]

[Source: Illustris 1405.2921]

Epidemiology





[Source: D. Brockmann, D. Helbing 2013]

Particle physics

Parameters of interest

Theory parameters

θ

Evolution

Particle physics

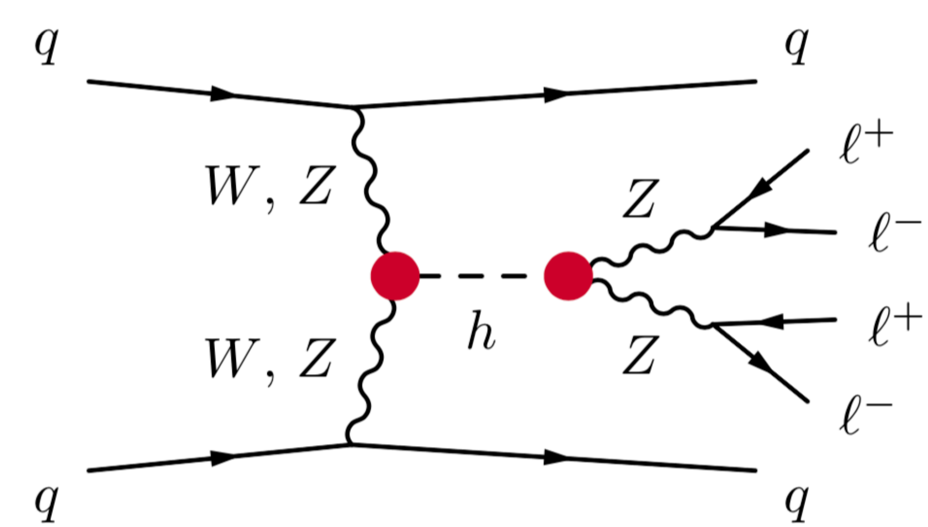
Latent variables

Parameters of interest

Parton-level momenta

Theory parameters

$$z_p \longleftarrow \theta$$



Evolution

Particle physics

Latent variables

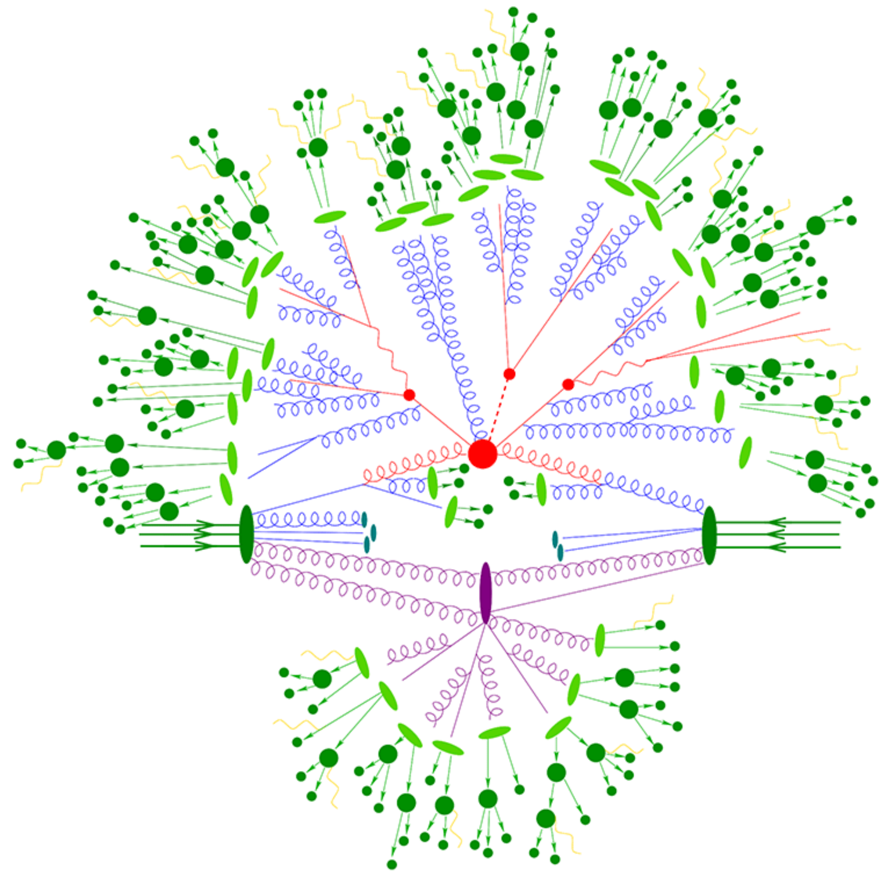
Parameters of interest

Shower splittings

Parton-level momenta

Theory parameters

$z_s \longleftarrow z_p \longleftarrow \theta$



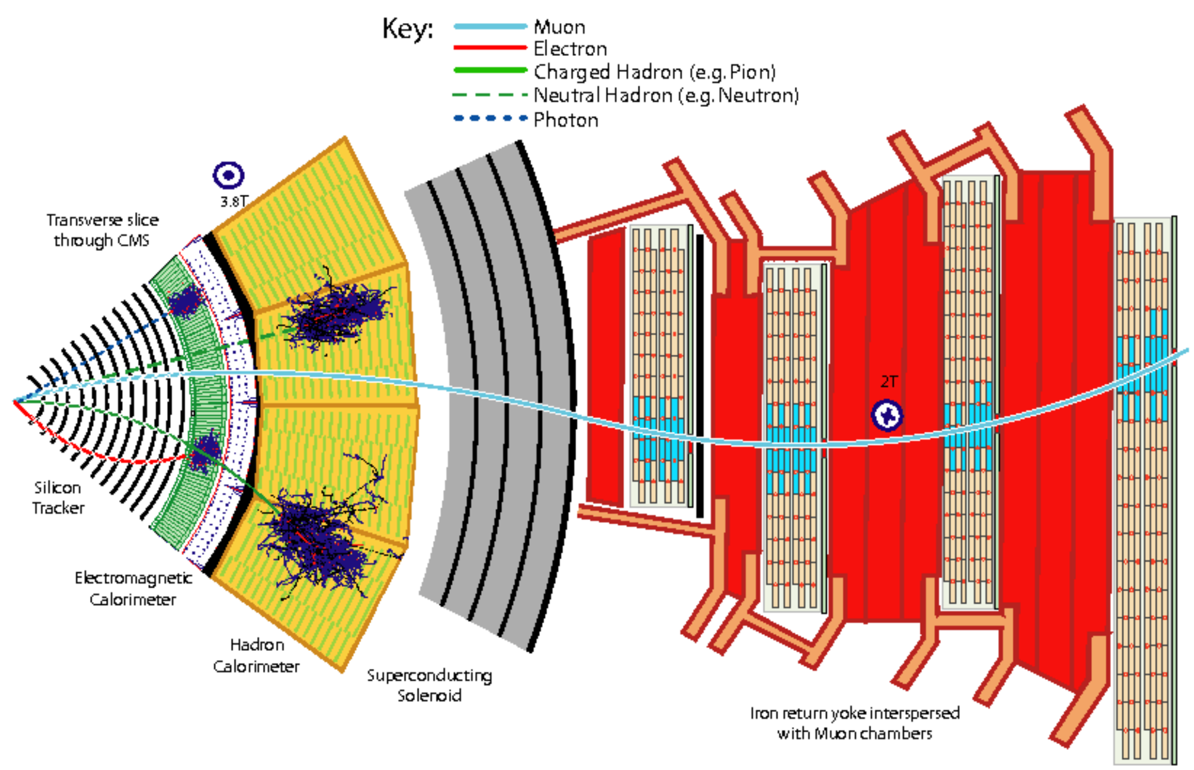
[Source: F. Krauss]

Evolution

Particle physics

	Latent variables		Parameters of interest
Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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



[Source: CMS]

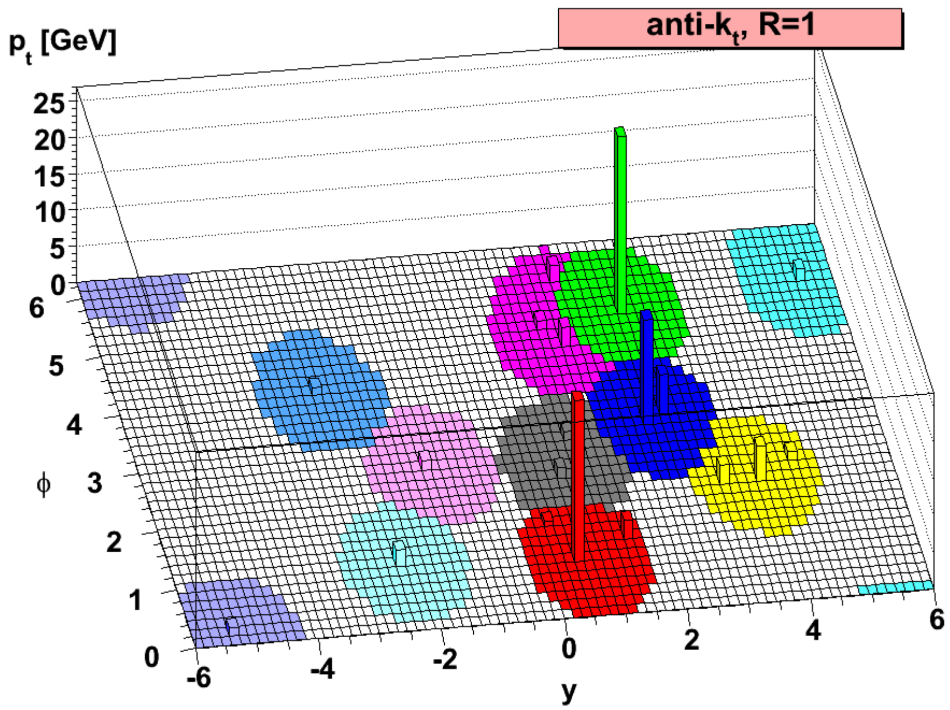
$\longleftarrow$ Evolution

Particle physics

Features | Latent variables | Parameters of interest

Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
x	z_d	z_s	z_p	θ

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



[Source: M. Cacciari, G. Salam, G. Soyez 0802.1189]

Evolution

Particle physics

Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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

Sample from $p(x|z_d)$ $p(z_d|z_s)$ $p(z_s|z_p)$ $p(z_p|\theta)$

Prediction (simulation)

Particle physics

Features | Latent variables | Parameters of interest

Observables | Detector interactions | Shower splittings | Parton-level momenta | Theory parameters

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

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \; p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

$\longrightarrow$

Inference

Particle physics

Features | Latent variables | Parameters of interest

Observables | Detector interactions | Shower splittings | Parton-level momenta | Theory parameters

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

$$p(x|\theta) = \boxed{\int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p}\, p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

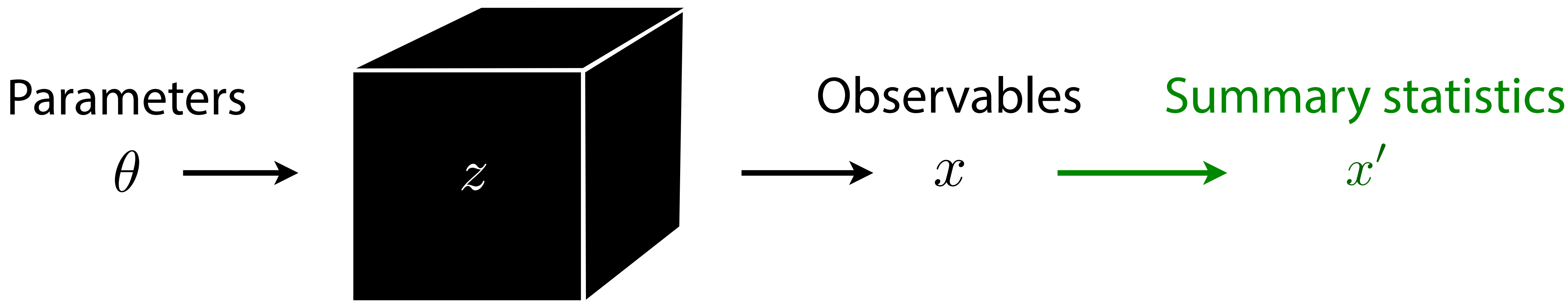
Infeasible to calculate the integral over this enormous space

Inference

Why has that not stopped us so far?

Why has that not stopped us so far?

- The physicist's way:

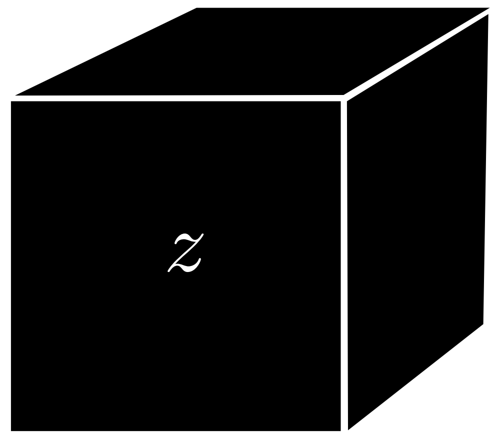


Why has that not stopped us so far?

- The physicist's way:

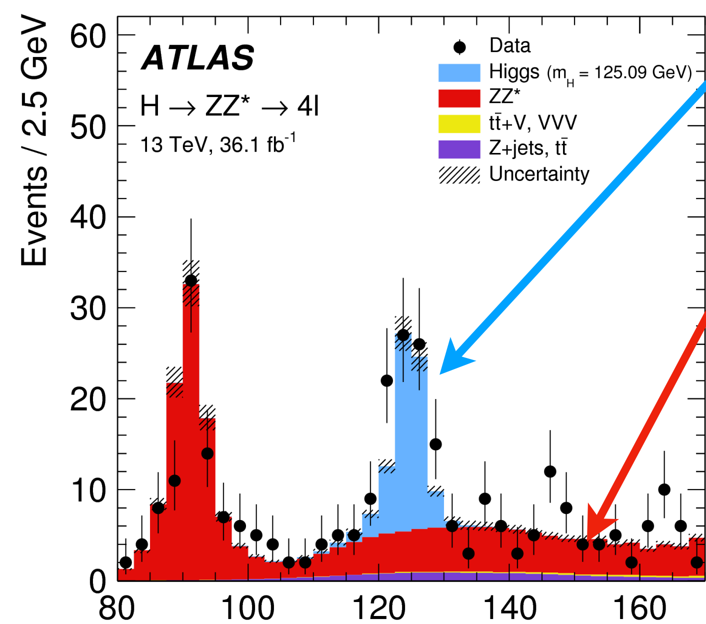
Parameters

$\theta \longrightarrow$



$\longrightarrow$ Observables x $\longrightarrow$ Summary statistics x'

$$\hat{p}(x|\theta) = p(x'|\theta) =$$



x'

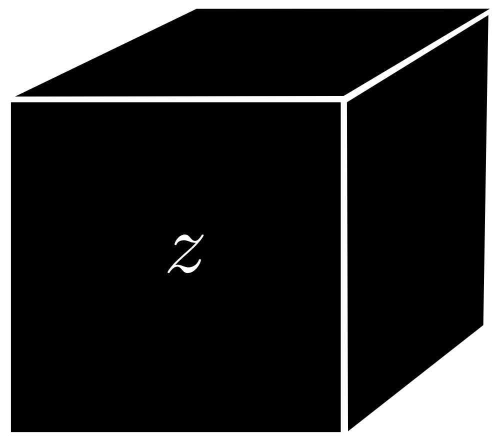
[ATLAS 1712.02304]

Run simulator for different θ, fill histograms

Why has that not stopped us so far?

- The physicist's way:

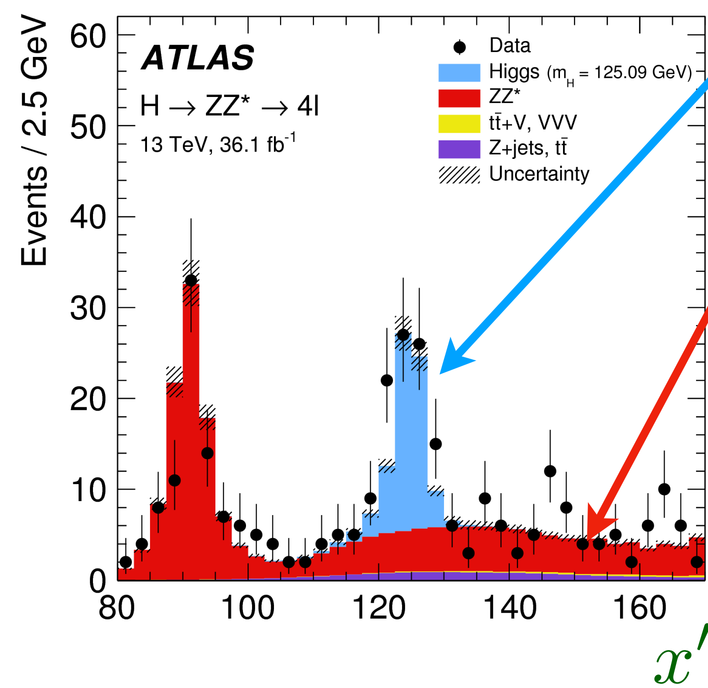
Parameters θ $\longrightarrow$



$\longrightarrow$ Observables x $\longrightarrow$ Summary statistics x'

Run simulator for different θ, fill histograms

$$\hat{p}(x|\theta) = p(x'|\theta) =$$



[ATLAS 1712.02304]

- But how to choose x'? Standard variables lose information, analysis of high-dimensional x (fully differential cross section, including correlations) often more powerful

Short interlude: Information geometry

- Fisher information

$$I_{ij}[\theta] = -\,\mathbb{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta_i\,\partial\theta_j}\middle|\theta\right]$$

describes maximal sensitivity to parameters according to Cramér-Rao bound

$$\mathrm{cov}[\hat{\theta}|\theta_0]_{ij} \geq I^{-1}_{ij}(\theta_0)$$

Expected error
of best-fit parameter

Inverse of
Fisher information

Short interlude: Information geometry

- Fisher information

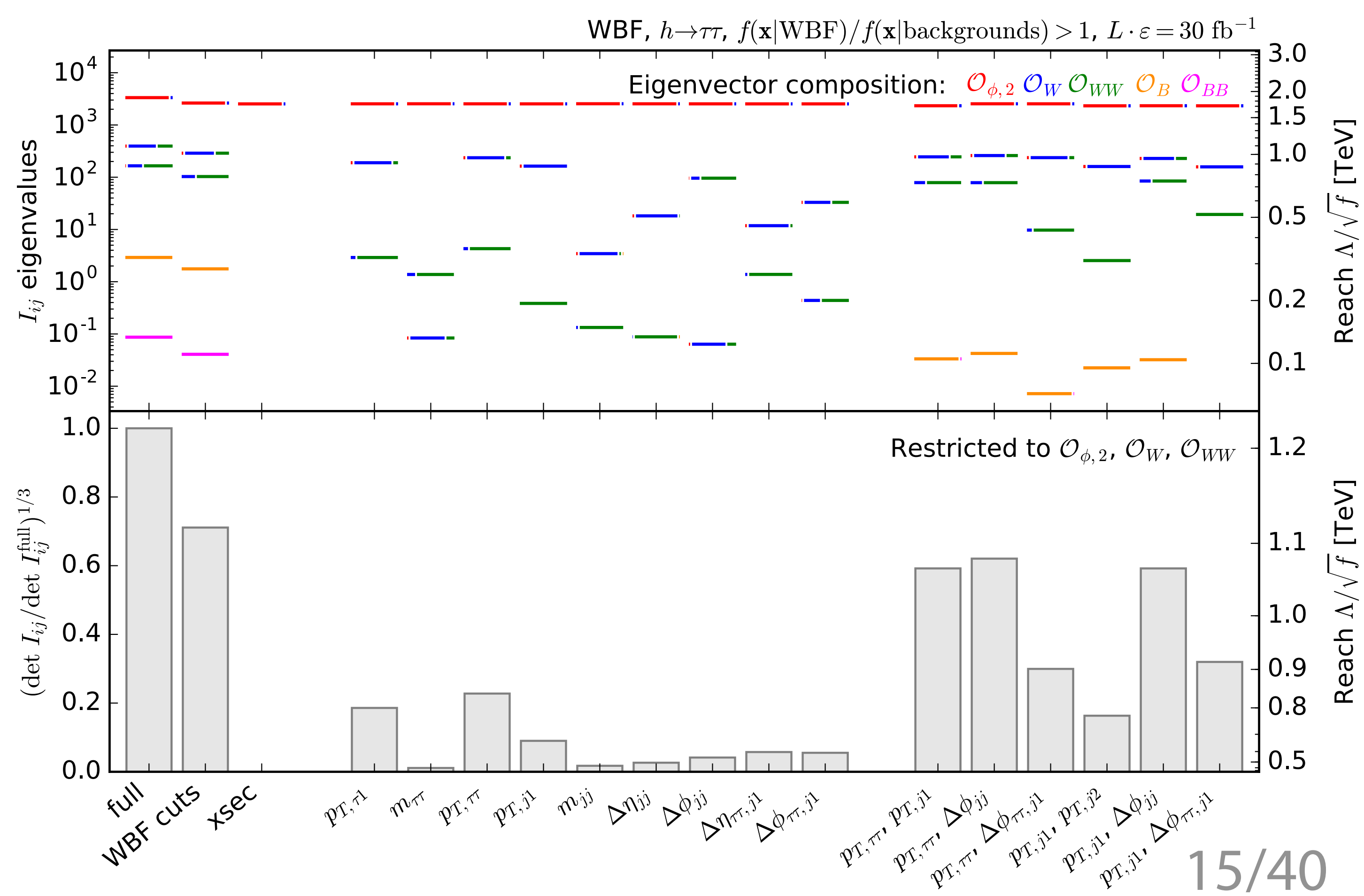
$$I_{ij}[\theta] = -\,\mathbb{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta_i\,\partial\theta_j}\middle|\theta\right]$$

describes maximal sensitivity to parameters according to Cramér-Rao bound

$$\mathrm{cov}[\hat{\theta}|\theta_0]_{ij} \geq I^{-1}_{ij}(\theta_0)$$

- We can calculate the Fisher information (with some approximations) on EFT operators in different processes and observables

⇒ In all cases, compressing to standard x' decreases information



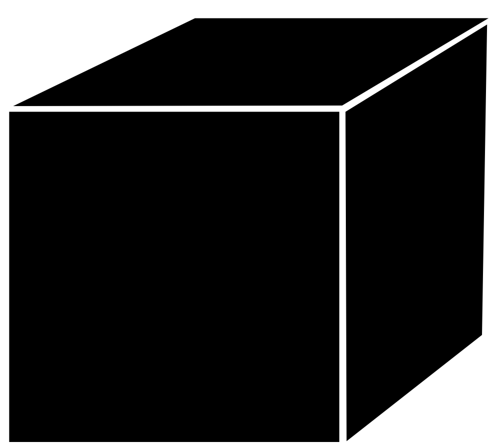
[JB, K. Cranmer, F. Kling, T. Plehn 1612.05261;
JB, F. Kling, T. Plehn, T. Tait 1712.02350]

Likelihood-free inference methods

Treat simulator as black box:

- Histograms of observables,
  Approximate Bayesian Computation
  Rely on summary statistics

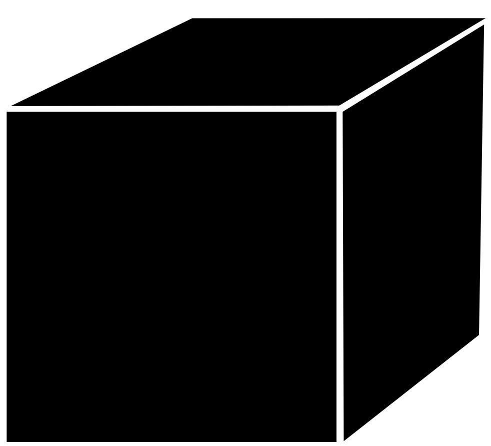
- Machine learning techniques
  Density networks, CARL, autoregressive models, normalizing flows, …



Likelihood-free inference methods

Treat simulator as black box:

- Histograms of observables, Approximate Bayesian Computation
  Rely on summary statistics
- Machine learning techniques
  Density networks, CARL, autoregressive models, normalizing flows, …



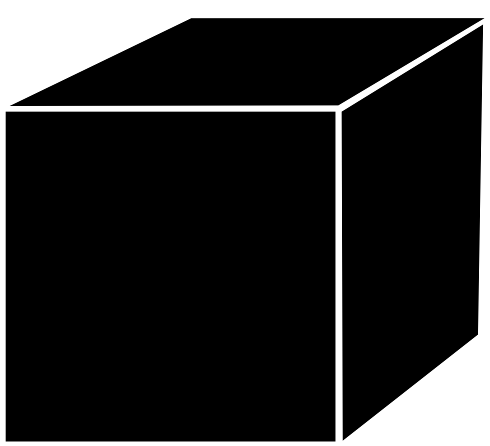
Use latent structure:

- Matrix Element Method, Optimal Observables, Shower Deconstruction
  Neglect or approximate shower + detector, explicitly calculate z integral

Likelihood-free inference methods

Treat simulator as black box:

- Histograms of observables, Approximate Bayesian Computation
  Rely on summary statistics
- Machine learning techniques
  Density networks, CARL, autoregressive models, normalizing flows, …



Use latent structure:

- Matrix Element Method, Optimal Observables, Shower Deconstruction
  Neglect or approximate shower + detector, explicitly calculate z integral

New!

- Mining gold from the simulator
  Leverage matrix-element information + machine learning

A new approach:
Mining gold from the simulator

[JB, K. Cranmer, G. Louppe, J. Pavez 1805.00013, 1805.00020, 1805.12244;
with M. Stoye 1808.00973; with F. Kling in progress]

What do we want?

- Neyman-Pearson lemma: likelihood ratio

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

observables model parameters

is ideal test statistics to tell θ_0 apart from θ_1

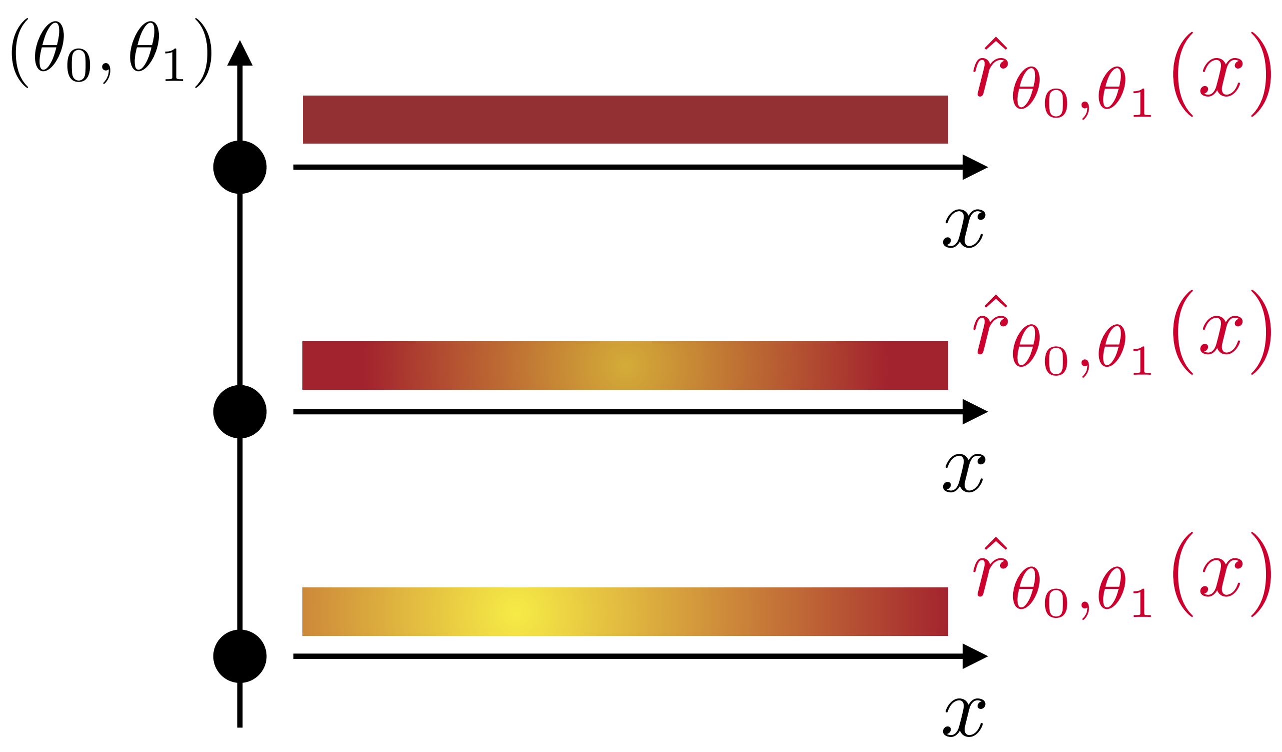
- That's intractable, so our goal is a precise estimator $\hat{r}(x|\theta_0, \theta_1)$

- With that, we can use standard procedures for frequentist limit setting
    - Neyman construction: p-values from toy experiments
    - Asymptotics [S. Wilks 1938; A. Wald 1943; G. Cowan, K. Cranmer, E. Gross, O. Vitells 1007.1727]

Two types of likelihood ratio estimators

A) Point by point:

- first, define grid of parameter points $\{(\theta_0, \theta_1)\}$
- for each combination (θ_0, θ_1) , create separate estimator $\hat{r}_{\theta_0,\theta_1}(x)$
- final results can be interpolated between grid points



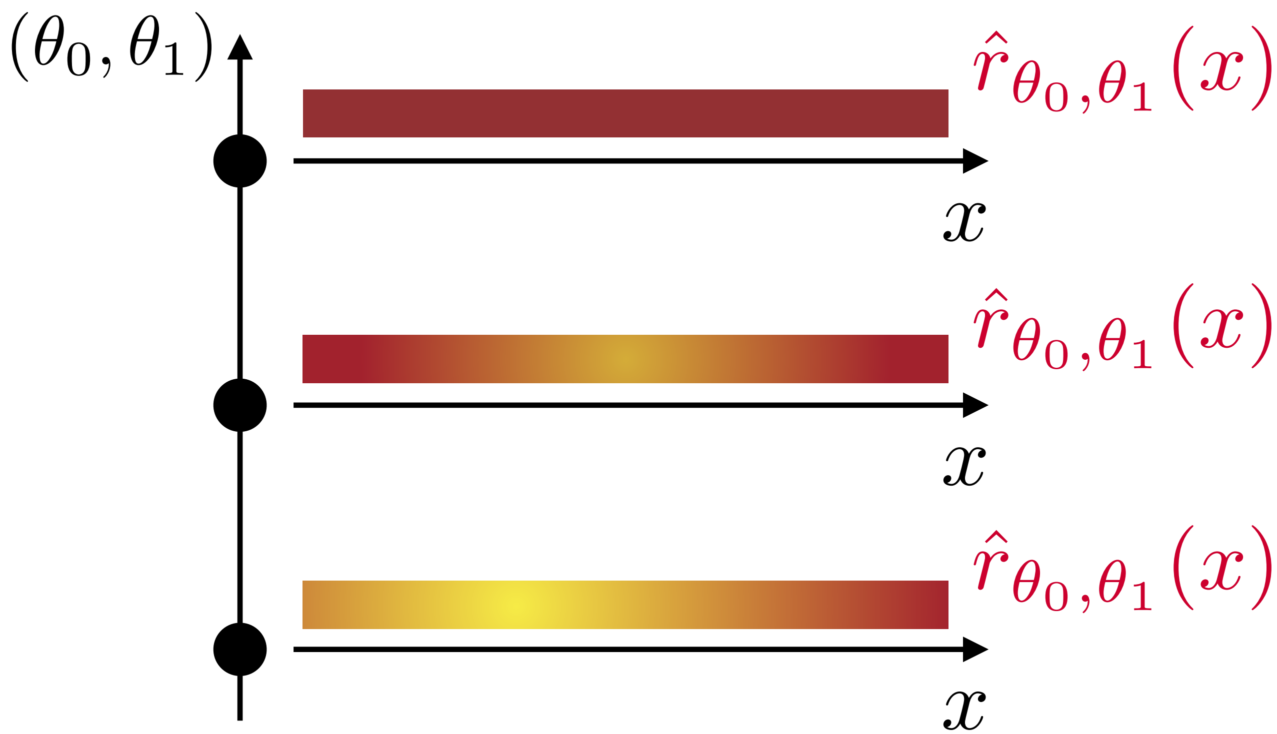
Two types of likelihood ratio estimators

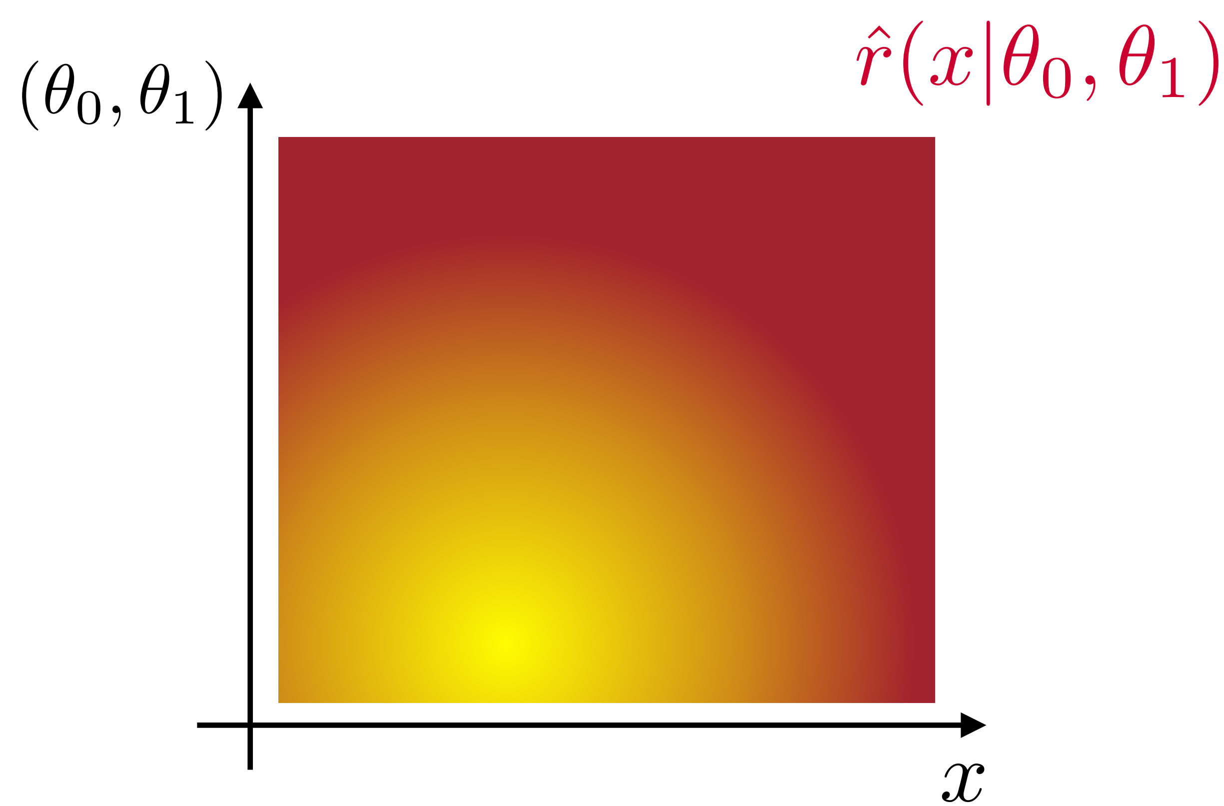
A) Point by point:

- first, define grid of parameter points $\{(\theta_0, \theta_1)\}$
- for each combination (θ_0, θ_1) , create separate estimator $\hat{r}_{\theta_0,\theta_1}(x)$
- final results can be interpolated between grid points

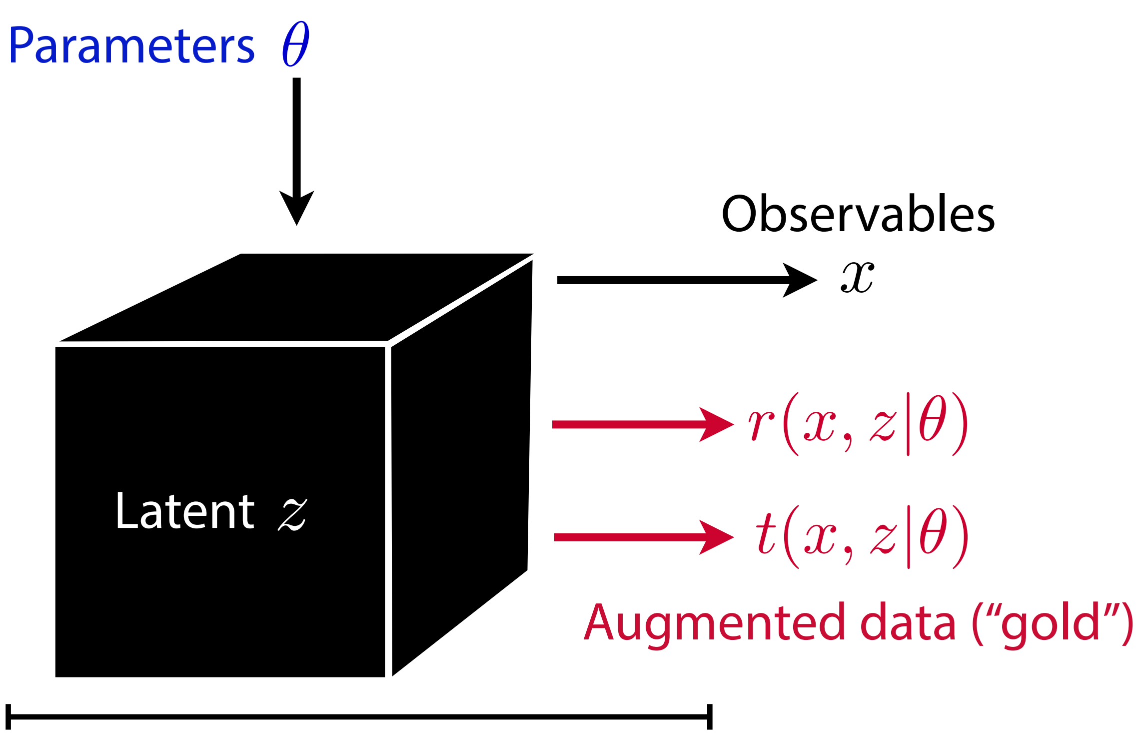
B) Parameterized: [P. Baldi et al. 1506.02169]

- create one estimator $\hat{r}(x|\theta_0, \theta_1)$ that is a function of θ_0 and θ_1
- no further interpolation necessary
- "borrows information" from close points





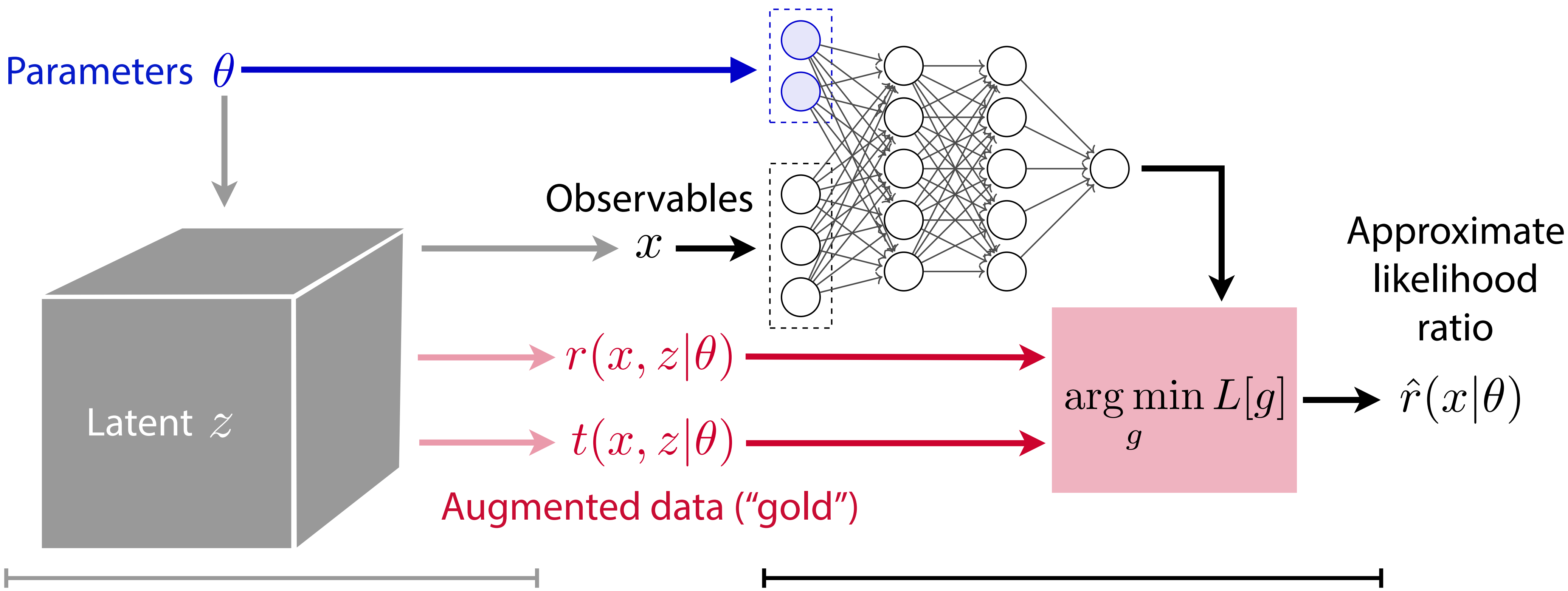
How do we get there?



1. Simulation

“Mining gold”: Extract additional information from simulator

How do we get there?



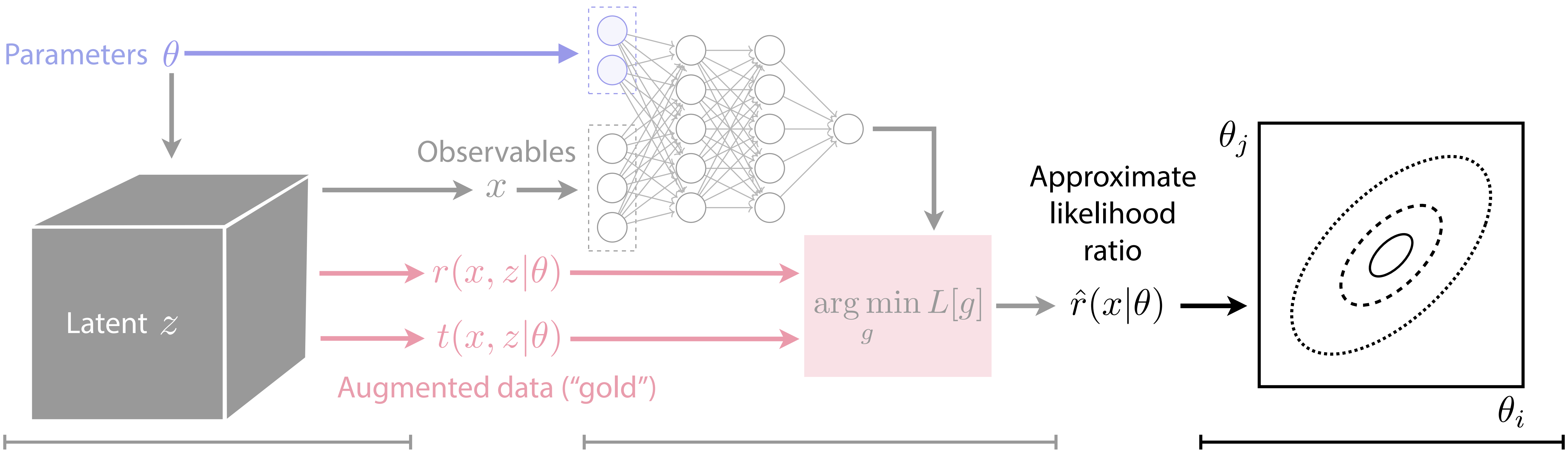
1. Simulation

“Mining gold”: Extract additional information from simulator

2. Machine Learning

Use this information to train parameterized estimator for likelihood ratio

How do we get there?



1. Simulation

"Mining gold": Extract additional information from simulator

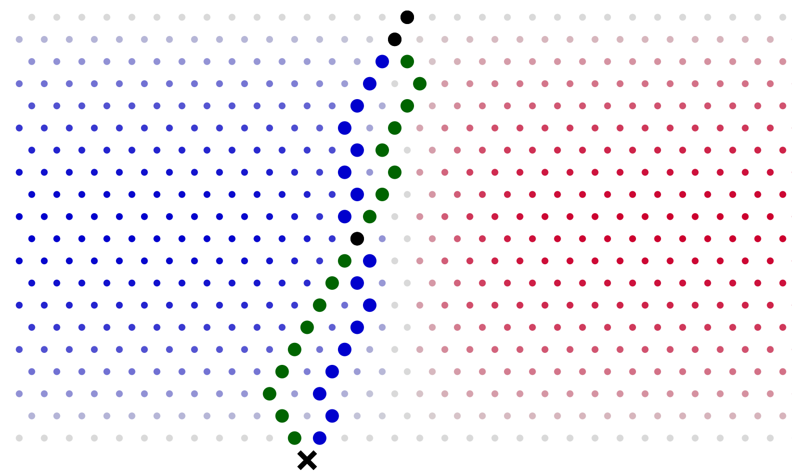
2. Machine Learning

Use this information to train parameterized estimator for likelihood ratio

3. Inference

Limit setting with standard hypothesis tests

Mining gold from the Galton board



- Remember: the likelihood

$$p(x|\theta) = \int \mathrm{d}z \; p(x, z|\theta)$$

and its ratios are intractable because of the integral over all possible paths z

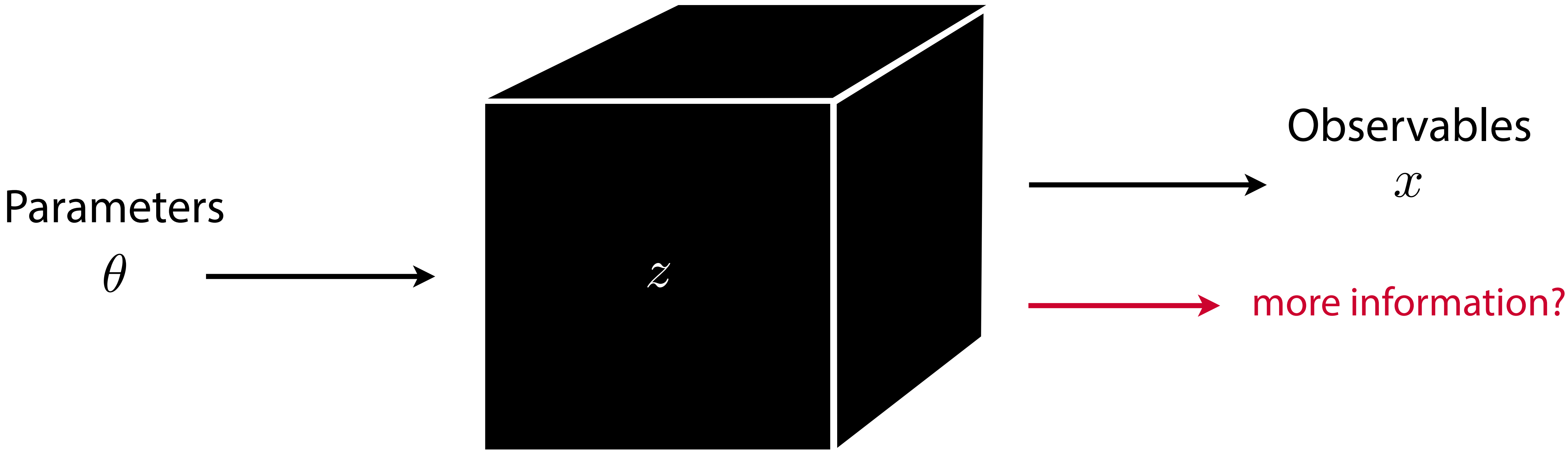
- But: we can calculate the probability of each individual path

$$p(x, z|\theta) = \prod_{\text{nails } i} p_i(x, z|\theta)$$

and the “joint likelihood ratio” conditional on a particular path

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z|\theta_0)}{p(x, z|\theta_1)}$$

Mining gold from LHC simulators



Mining gold from LHC simulators

Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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

$$\boxed{p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p} \, p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

Intractable integrals

Mining gold from LHC simulators

Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \, p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad \boxed{p(z_p|\theta)}$$

This parton-level likelihood can be evaluated!

Mining gold from LHC simulators

Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \, p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad \boxed{p(z_p|\theta)}$$

This parton-level likelihood can be evaluated!

$\Rightarrow$ We can calculate the joint likelihood ratio conditional on the specific evolution of an event:

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)} = \frac{p(x|z_d)}{p(x|z_d)} \frac{p(z_d|z_s)}{p(z_d|z_s)} \frac{p(z_s|z_p)}{p(z_s|z_p)} \boxed{\frac{p(z_p|\theta_0)}{p(z_p|\theta_1)}}$$

The value of gold

We have joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

We want likelihood ratio

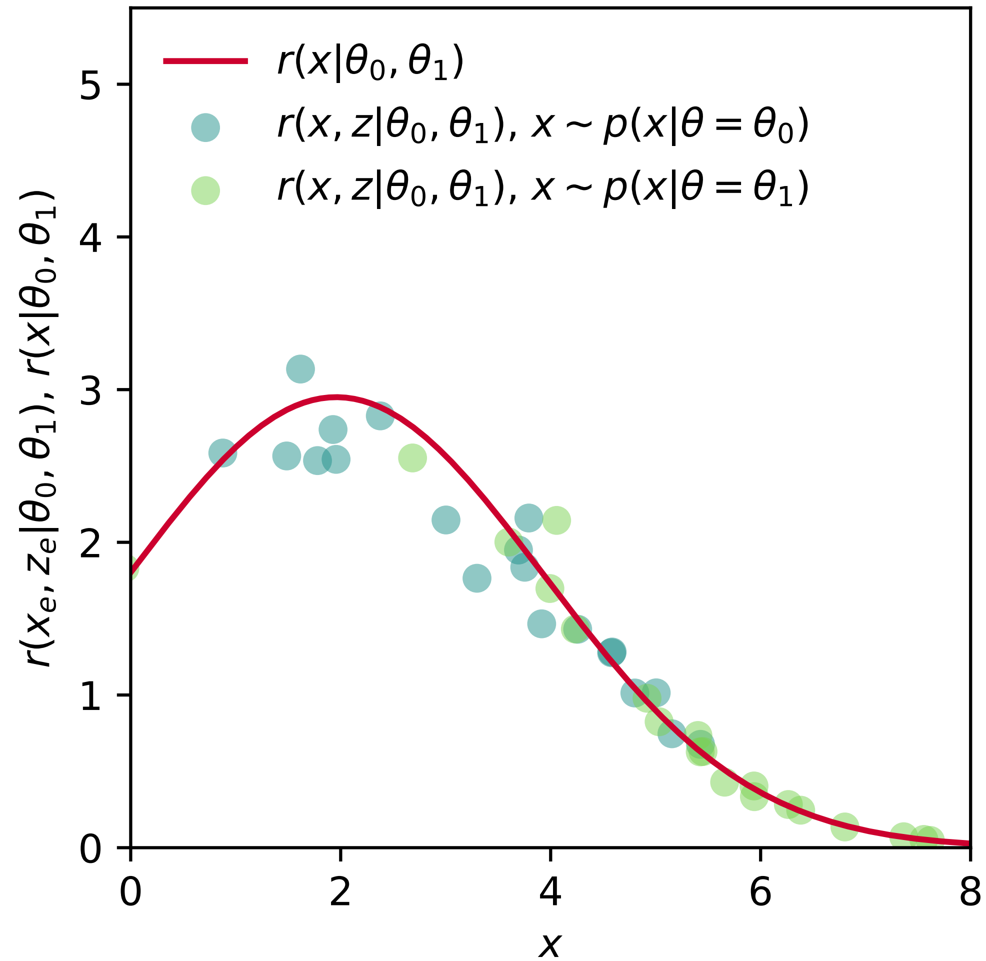
$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

The value of gold

We have joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

$r(x, z|\theta_0, \theta_1)$ are scattered around $r(x|\theta_0, \theta_1)$



We want likelihood ratio

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

The value of gold

We have joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

With $r(x, z|\theta_0, \theta_1)$,
we define the functional

$$L_r[\hat{r}(x|\theta_0, \theta_1)] = \int \mathrm{d}x \int \mathrm{d}z \; p(x, z|\theta_1) \left[\left(\hat{r}(x|\theta_0, \theta_1) - r(x, z|\theta_0, \theta_1) \right)^2 \right].$$

One can show it is minimized by

$$r(x|\theta_0, \theta_1) = \underset{\hat{r}(x|\theta_0, \theta_1)}{\arg\min} \; L_r[\hat{r}(x|\theta_0, \theta_1)] \; !$$

We want likelihood ratio

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

Enter machine learning

Need to minimize a functional:

Extremization

Functional

Variational family

$$r(x|\theta_0, \theta_1) = \arg\min_{\hat{r}(x|\theta_0,\theta_1)} L_r[\hat{r}(x|\theta_0, \theta_1)]$$

Enter machine learning

Need to minimize a functional:

Extremization

Functional

Variational family

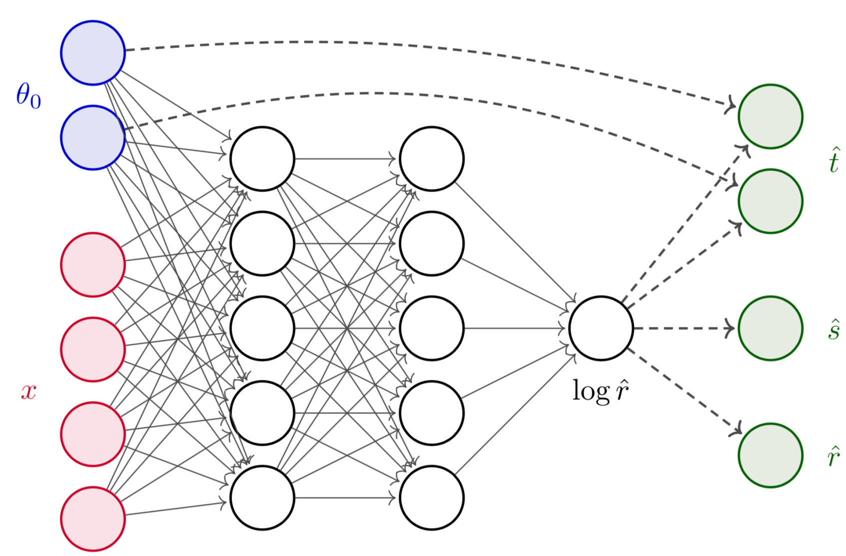
$$r(x|\theta_0,\theta_1) = \arg\min_{\hat{r}(x|\theta_0,\theta_1)} L_r[\hat{r}(x|\theta_0,\theta_1)]$$

This is exactly what machine learning does!

Loss function

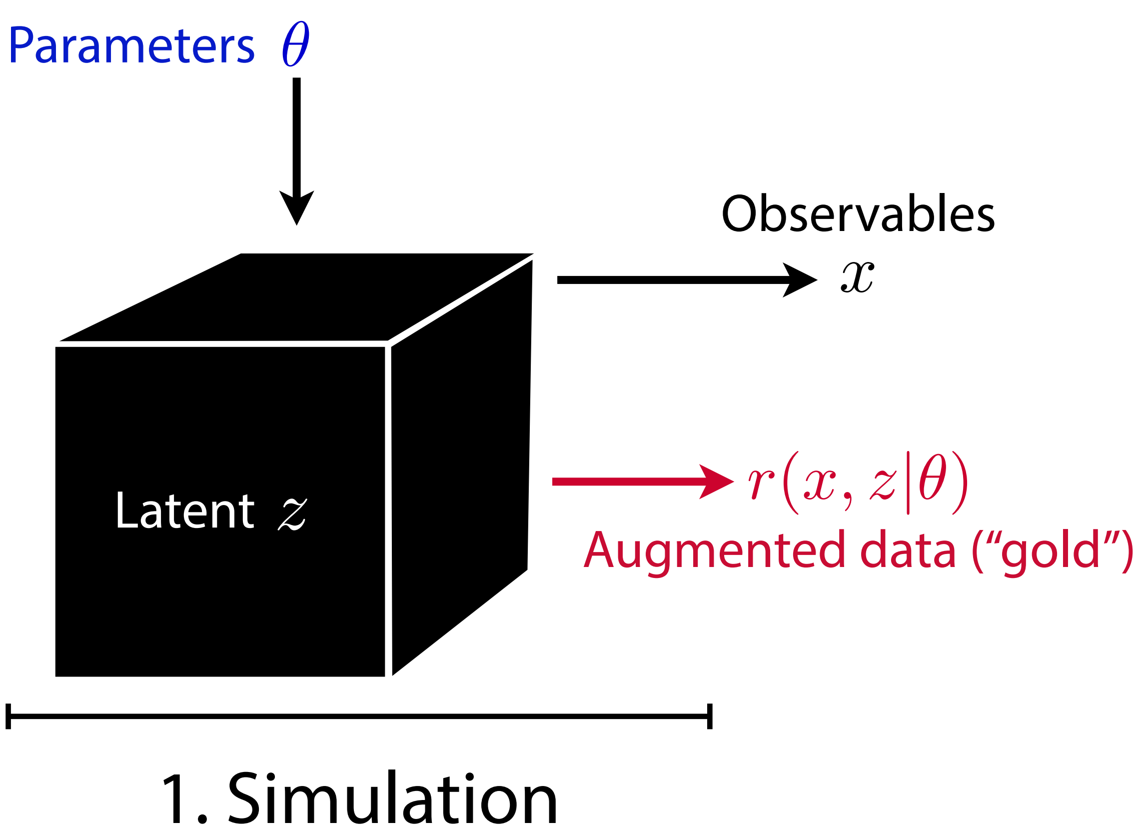
Stochastic gradient descent

Neural network

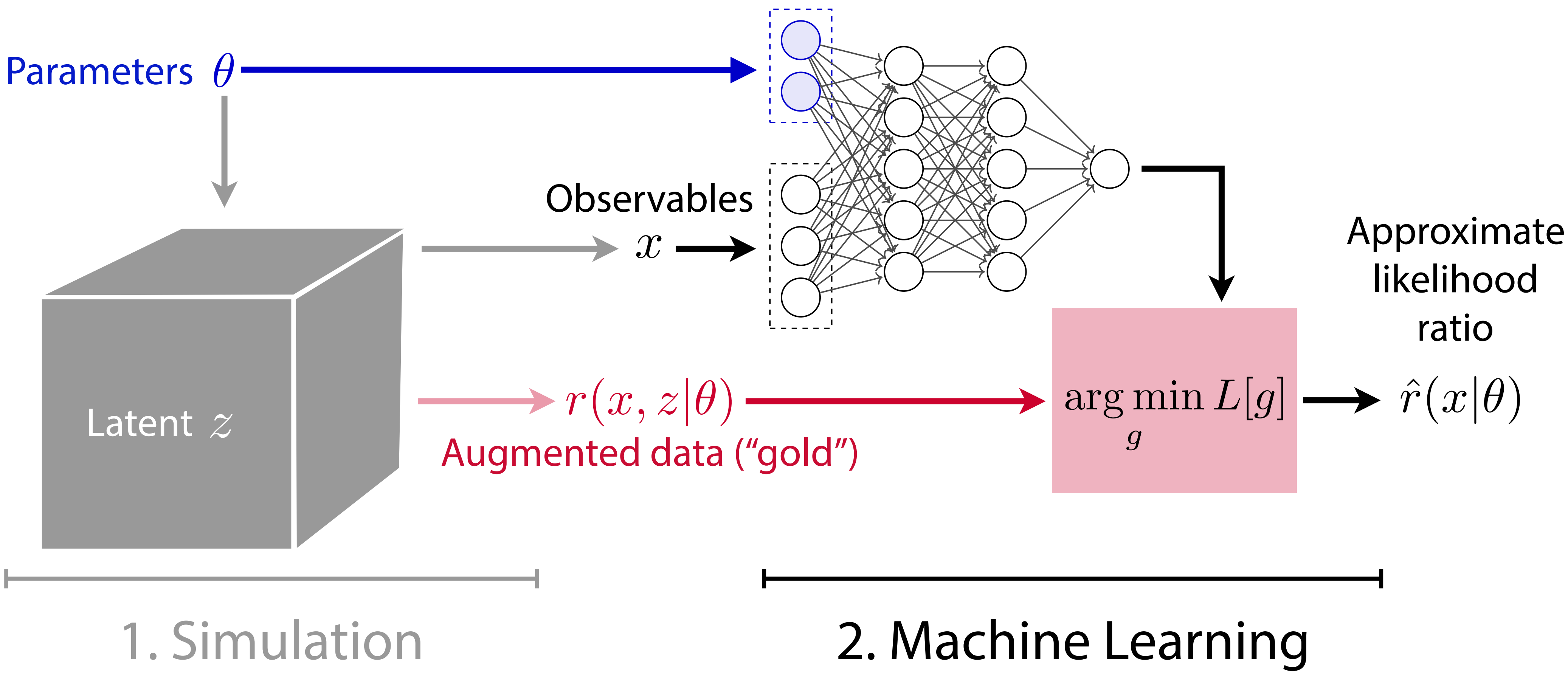


$\Rightarrow$ We implement $\hat{r}(x|\theta_0,\theta_1)$
as a neural network trained on the data available from the simulator

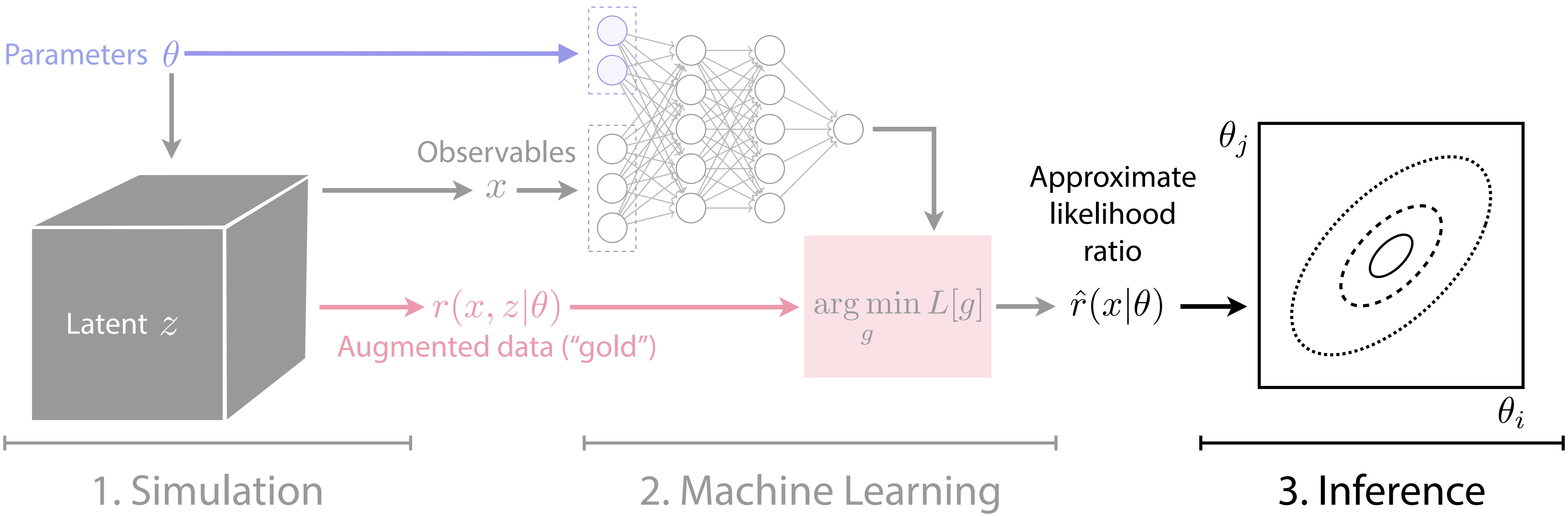
What we have so far



What we have so far

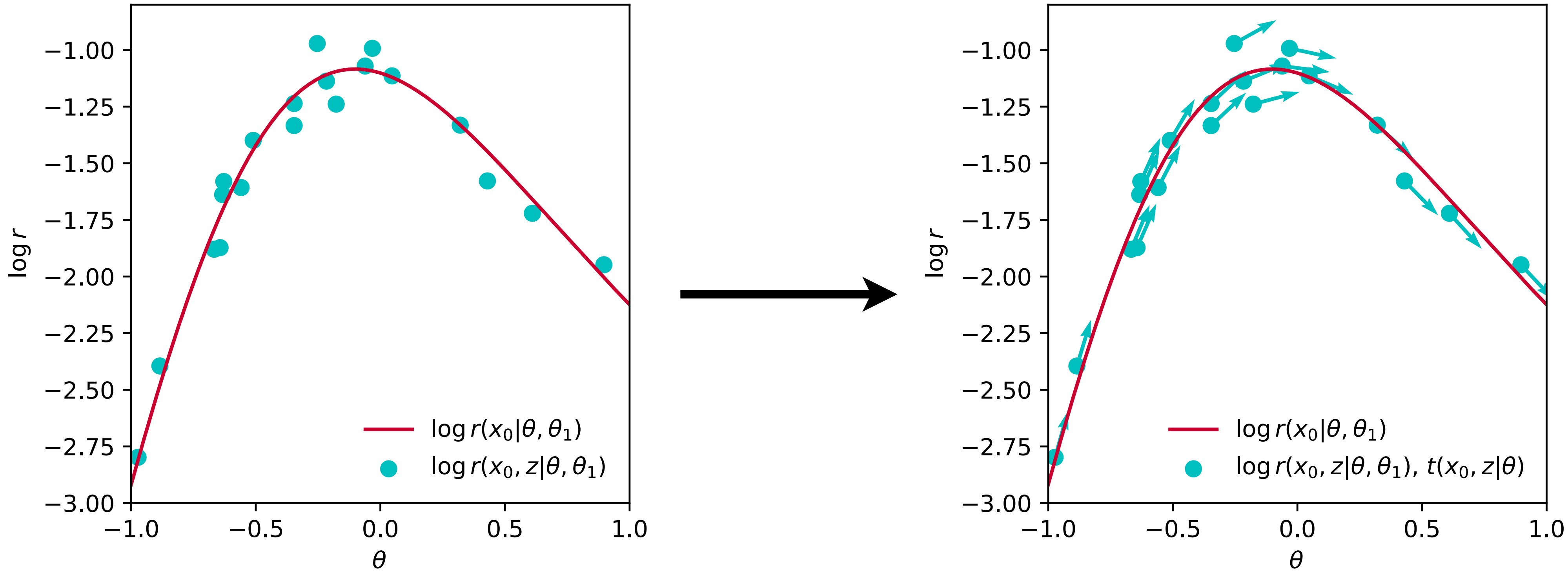


What we have so far



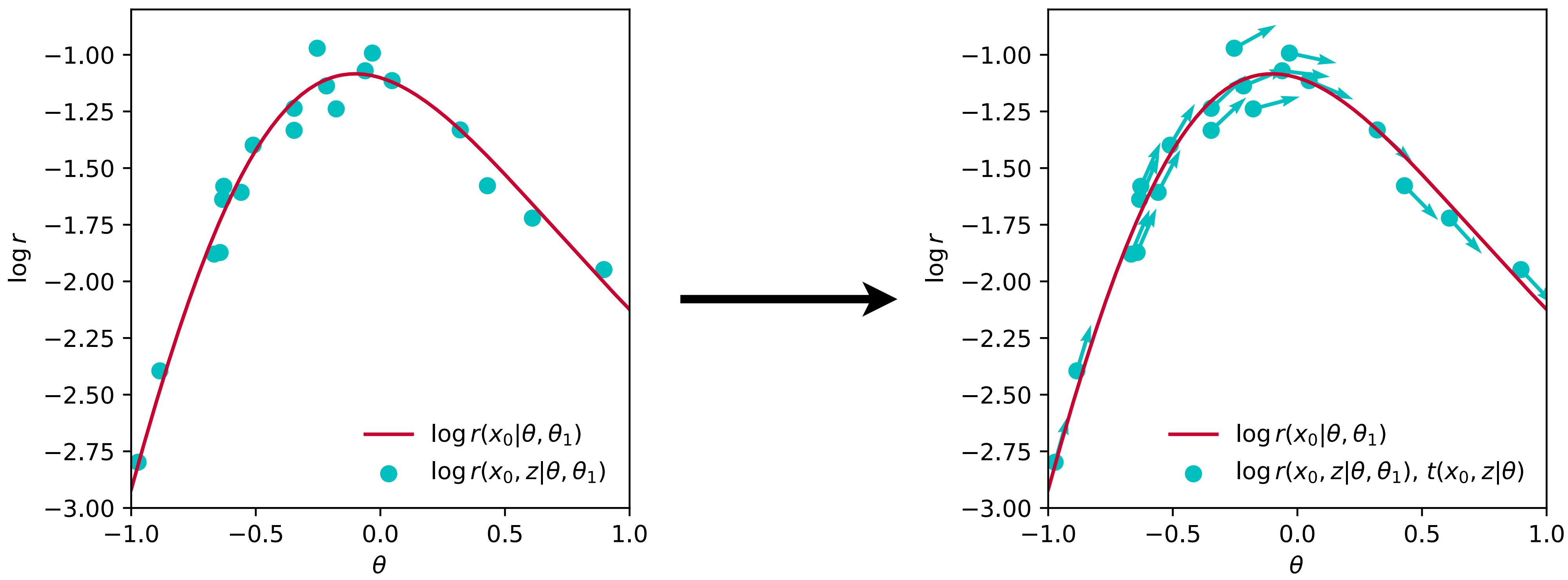
One more piece: the score

- Knowing derivative often helps fitting



One more piece: the score

- Knowing derivative often helps fitting



- In our case, the relevant quantity is the score

$$t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0} .$$

  - The score fully characterizes the likelihood function in the neighborhood of θ_0
  - The score itself is intractable. But…

Learning the score

Similar to the joint likelihood ratio,
we can calculate the joint score

$$t(x, z|\theta_0) \equiv \nabla_\theta \log p(x, z_d, z_s, z|\theta)\Big|_{\theta_0}$$

We want score

$$t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0}$$

Learning the score

Similar to the joint likelihood ratio,
we can calculate the joint score

$$t(x,z|\theta_0) \equiv \nabla_\theta \log p(x, z_d, z_s, z|\theta)\Big|_{\theta_0}$$

We want score

$$t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0}$$

Given $t(x,z|\theta_0)$,
we define the functional

$$L_t[\hat{t}(x|\theta_0)] = \int \mathrm{d}x \int \mathrm{d}z \; p(x,z|\theta_0) \; \left[\Big(\hat{t}(x|\theta_0) - t(x,z|\theta_0) \Big)^2 \right].$$

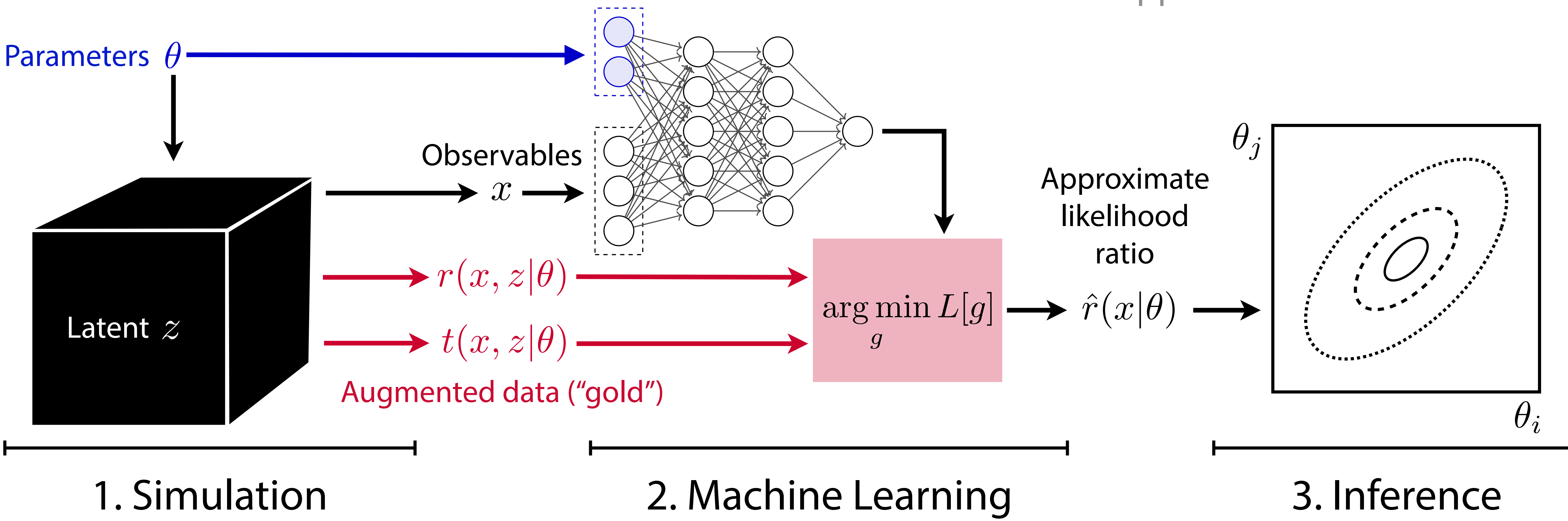
One can show it is minimized by

$$t(x|\theta_0) = \underset{\hat{t}(x|\theta_0)}{\arg\min} \, L_t[\hat{t}(x|\theta_0)] \, .$$

Again, we implement this minimization
through machine learning

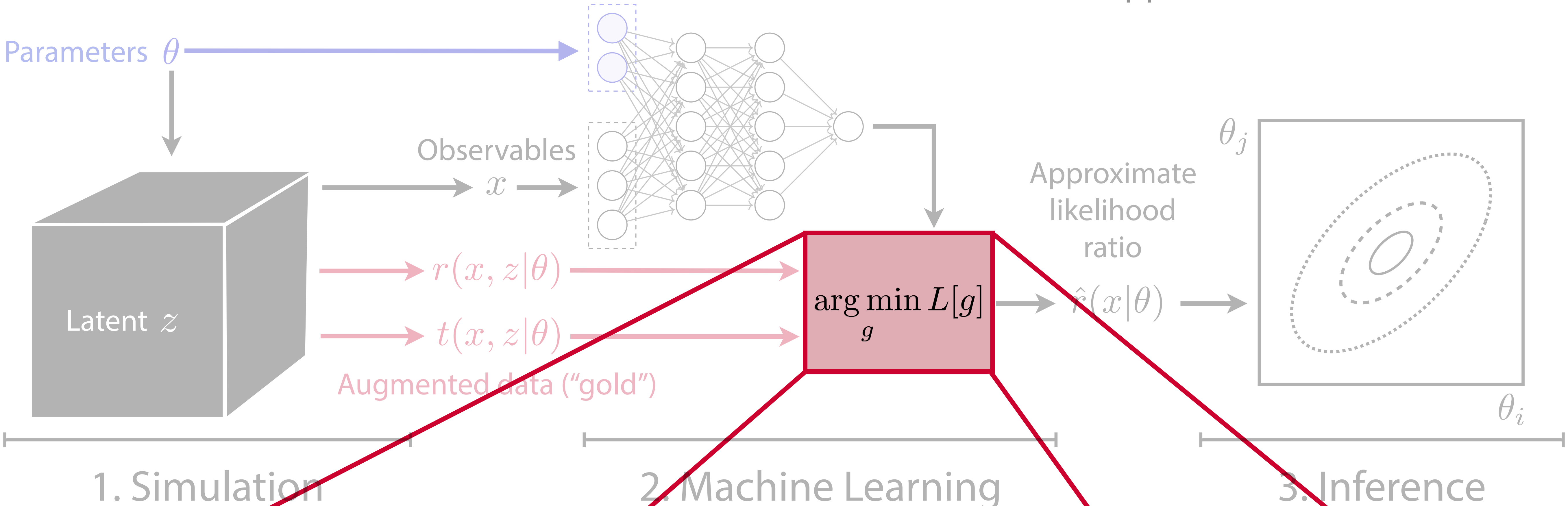
Putting the pieces together: RASCAL[1]

[1] **R**atio **a**nd **sc**ore **a**pproximate **l**ikelihood ratio



Putting the pieces together: RASCAL[1]

[1] Ratio and score approximate likelihood ratio



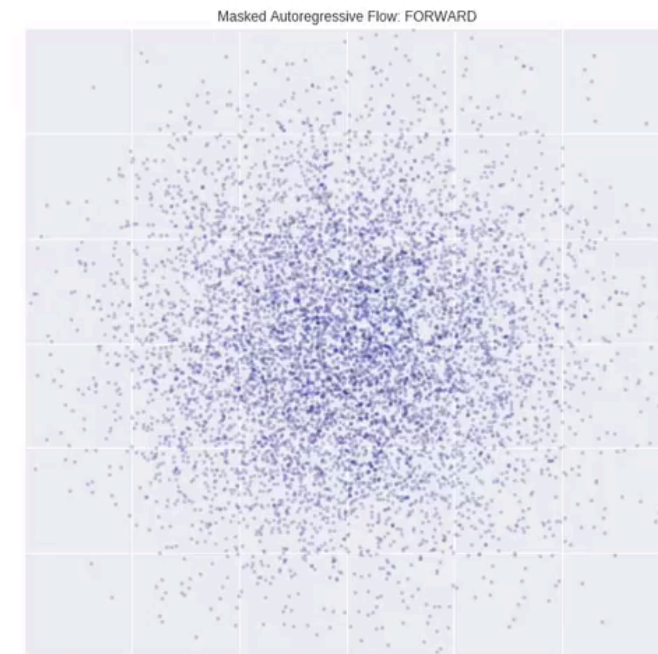
Combining ratio and score information:

- If ratio estimator is parameterized and differentiable, we can calculate the estimated score $\hat{t}(x|\theta_0) = \nabla_{\theta_0} \log \hat{r}(x|\theta_0, \theta_1)$
- Combined loss function:

$$L_{\text{RASCAL}}[\hat{r}(x|\theta_0, \theta_1)] = L_r[\hat{r}(x|\theta_0, \theta_1)] + \alpha\, L_t[\nabla_{\theta_0} \log \hat{r}(x|\theta_0, \theta_1)]$$

Alternatives and extensions

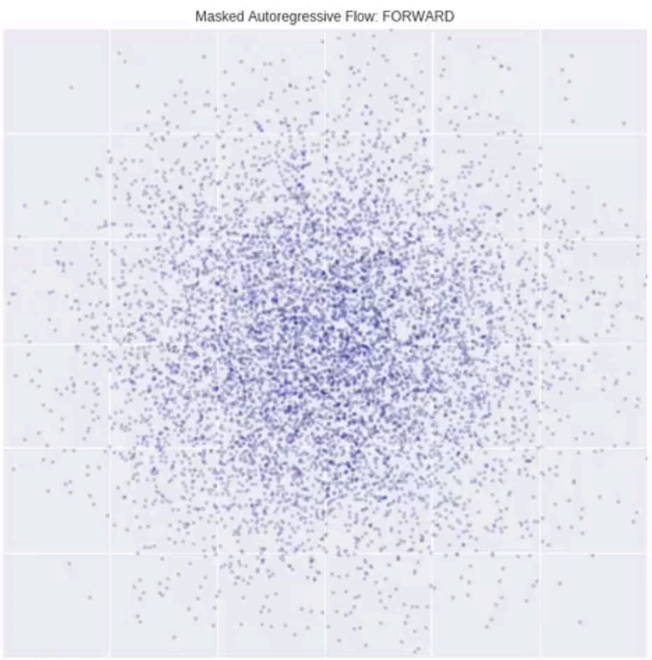
- More than one way to the likelihood (ratio)!
  - SCANDAL: combine with neural density estimators, e.g. Masked Autoregressive Flows [G. Papamakarios, T. Pavlakou, I. Murray 1705.07057]
  - SALLY / SALLINO: use estimated score as "optimal observable"



[Source: Alex Mordvintsev]

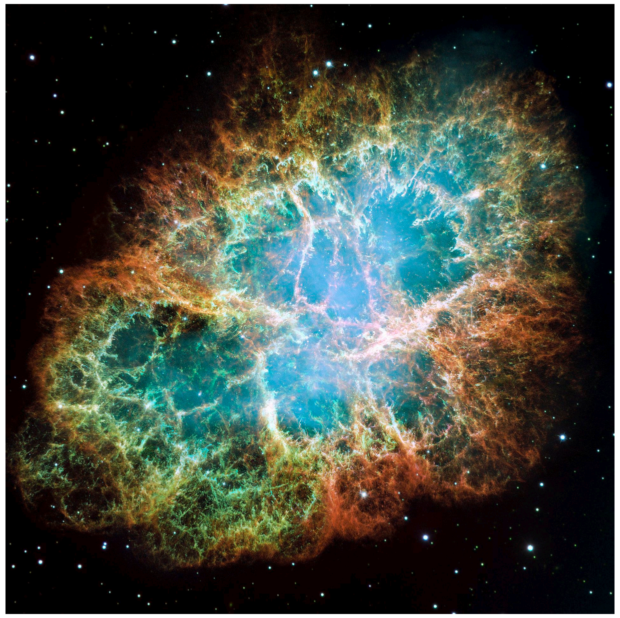
Alternatives and extensions

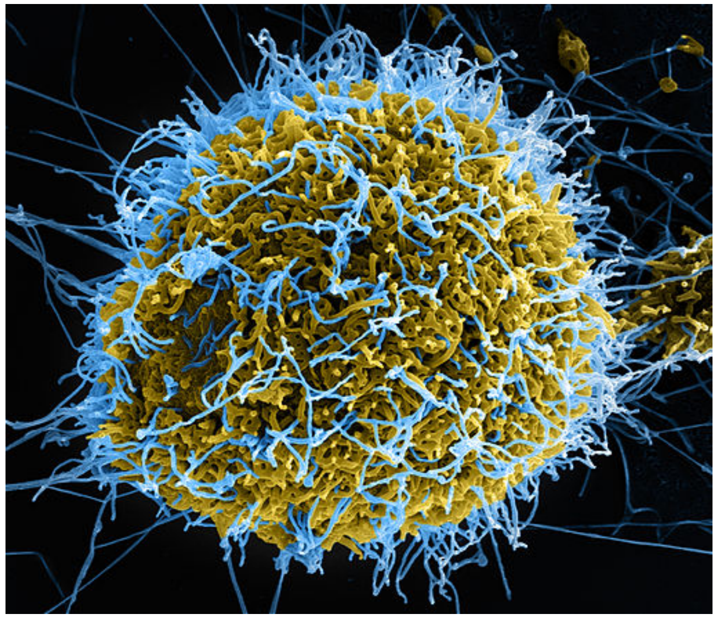
- More than one way to the likelihood (ratio)!
  - SCANDAL: combine with neural density estimators, e.g. Masked Autoregressive Flows [G. Papamakarios, T. Pavlakou, I. Murray 1705.07057]
  - SALLY / SALLINO: use estimated score as "optimal observable"



[Source: Alex Mordvintsev]

- Nuisance parameters / systematic uncertainties
- More general than particle physics
  - Currently being adapted to cosmology and epidemiology
  - Bayesian inference





[Sources: NASA, NIAID]

Comparison with established methods

- Histograms of observables, Approximate Bayesian Computation
  Rely on summary statistics
- Machine learning techniques
  Density networks, CARL, autoregressive models, normalizing flows, …

- Matrix Element Method, Optimal Observables, Shower Deconstruction
  Neglect or approximate shower + detector, explicitly calculate z integral

New!

- Mining gold from the simulator
  Leverage matrix-element information + machine learning

	Histograms, ABC	NDE	MEM, OO	RASCAL etc
High-dimensional observables		✓	✓	✓
Realistic shower, detector sim.	✓	✓		✓
Asymptotically exact		✓		✓
Uses matrix element information			✓	✓
Evaluation	fast	fast	expensive	fast

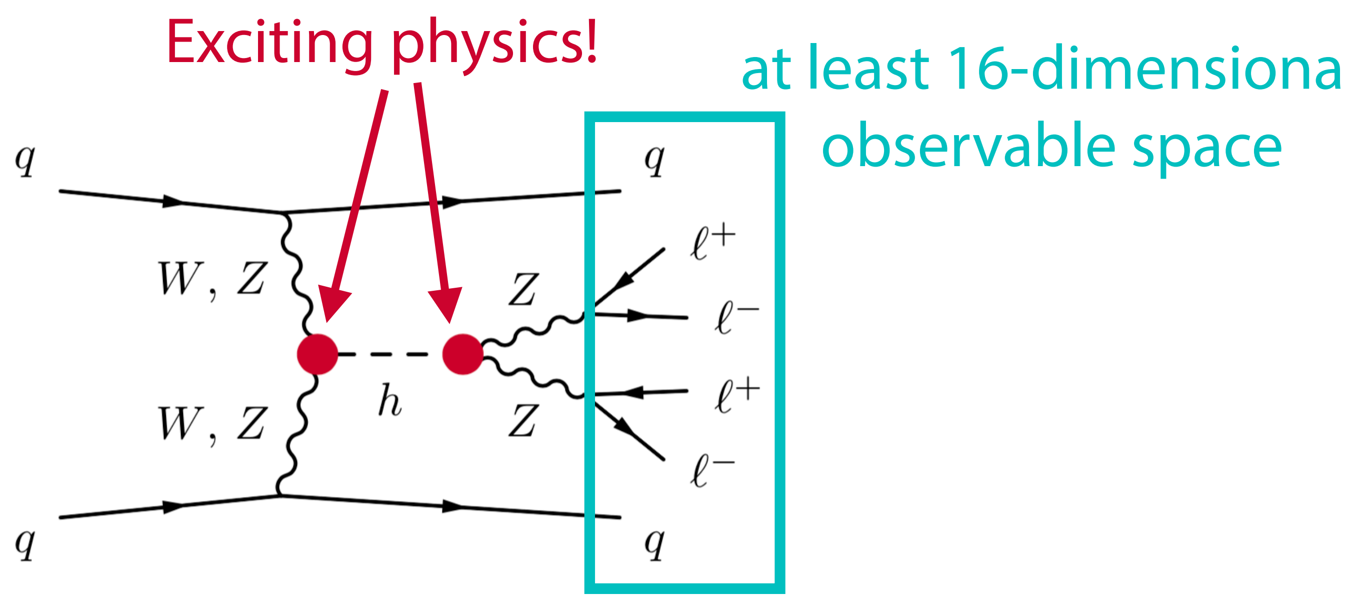

DEGUSSA
FEINGOLD
999.9
1000

EFT example

[JB, K. Cranmer, G. Louppe, J. Pavez 1805.00013, 1805.00020, 1805.12244;
with M. Stoye 1808.00973]

Proof of concept

- Higgs production in weak boson fusion:

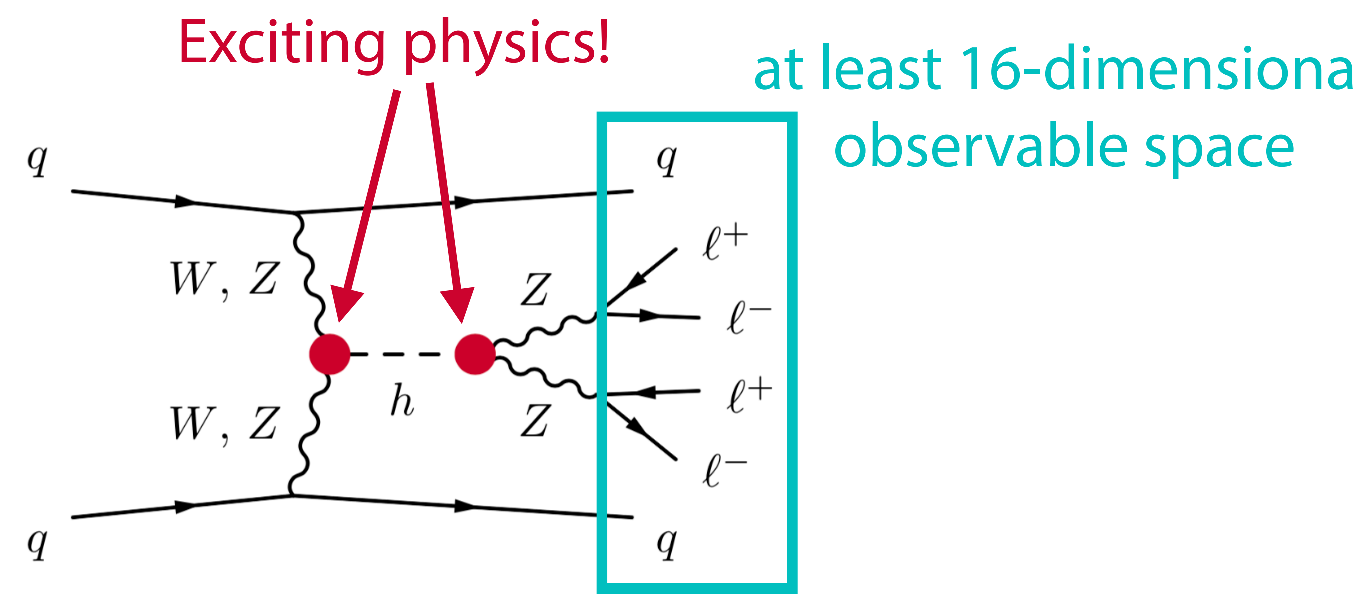


- Goal: constraints on two EFT parameters

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \boxed{\frac{f_W}{\Lambda^2}} \underbrace{\frac{\mathrm{i}g}{2} \, (D^\mu \phi)^\dagger \, \sigma^a \, D^\nu \phi \, W^a_{\mu\nu}}_{\mathcal{O}_W} - \boxed{\frac{f_{WW}}{\Lambda^2}} \underbrace{\frac{g^2}{4} \, (\phi^\dagger \phi) \, W^a_{\mu\nu} \, W^{\mu\nu\, a}}_{\mathcal{O}_{WW}}$$

Proof of concept

- Higgs production in weak boson fusion:



- Goal: constraints on two EFT parameters

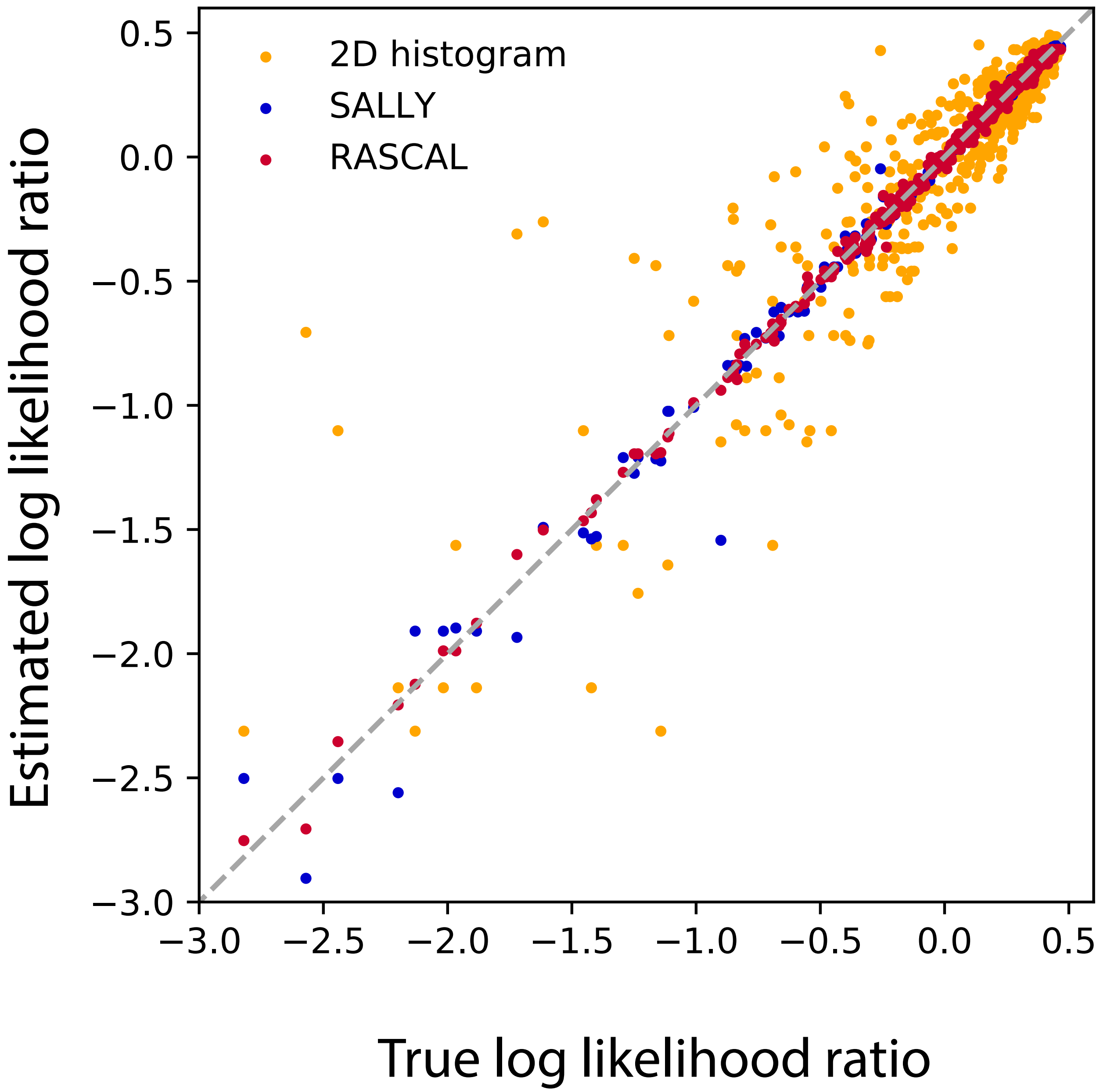
$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \boxed{\frac{f_W}{\Lambda^2}} \underbrace{\frac{\mathrm{i}g}{2} \, (D^\mu \phi)^\dagger \, \sigma^a \, D^\nu \phi \, W^a_{\mu\nu}}_{\mathcal{O}_W} - \boxed{\frac{f_{WW}}{\Lambda^2}} \underbrace{\frac{g^2}{4} \, (\phi^\dagger \phi) \, W^a_{\mu\nu} \, W^{\mu\nu\, a}}_{\mathcal{O}_{WW}}$$

- Two setups:
  - Simplified setup in which we can compare to true likelihood
  - "Realistic" simulation with approximate detector effects

- Simulation: MadGraph [J. Alwall et al. 1405.0301] + MadMax
  [K. Cranmer, T. Plehn hep-ph/0605268; T. Plehn, P. Schichtel, D. Wiegand 1311.2591]

Precise likelihood ratio estimates

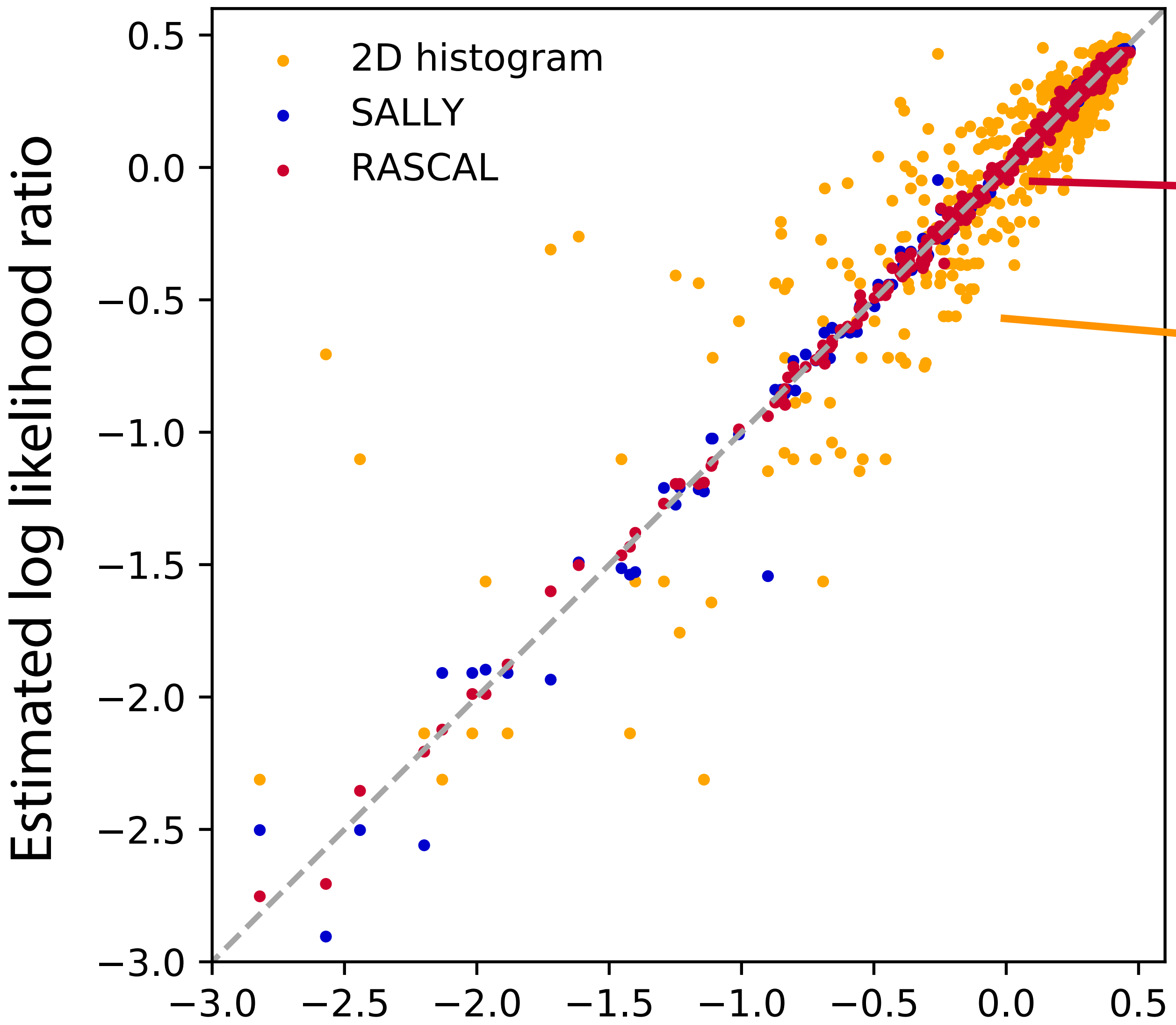
Each point = 1 event



(usually don't have this)

Precise likelihood ratio estimates

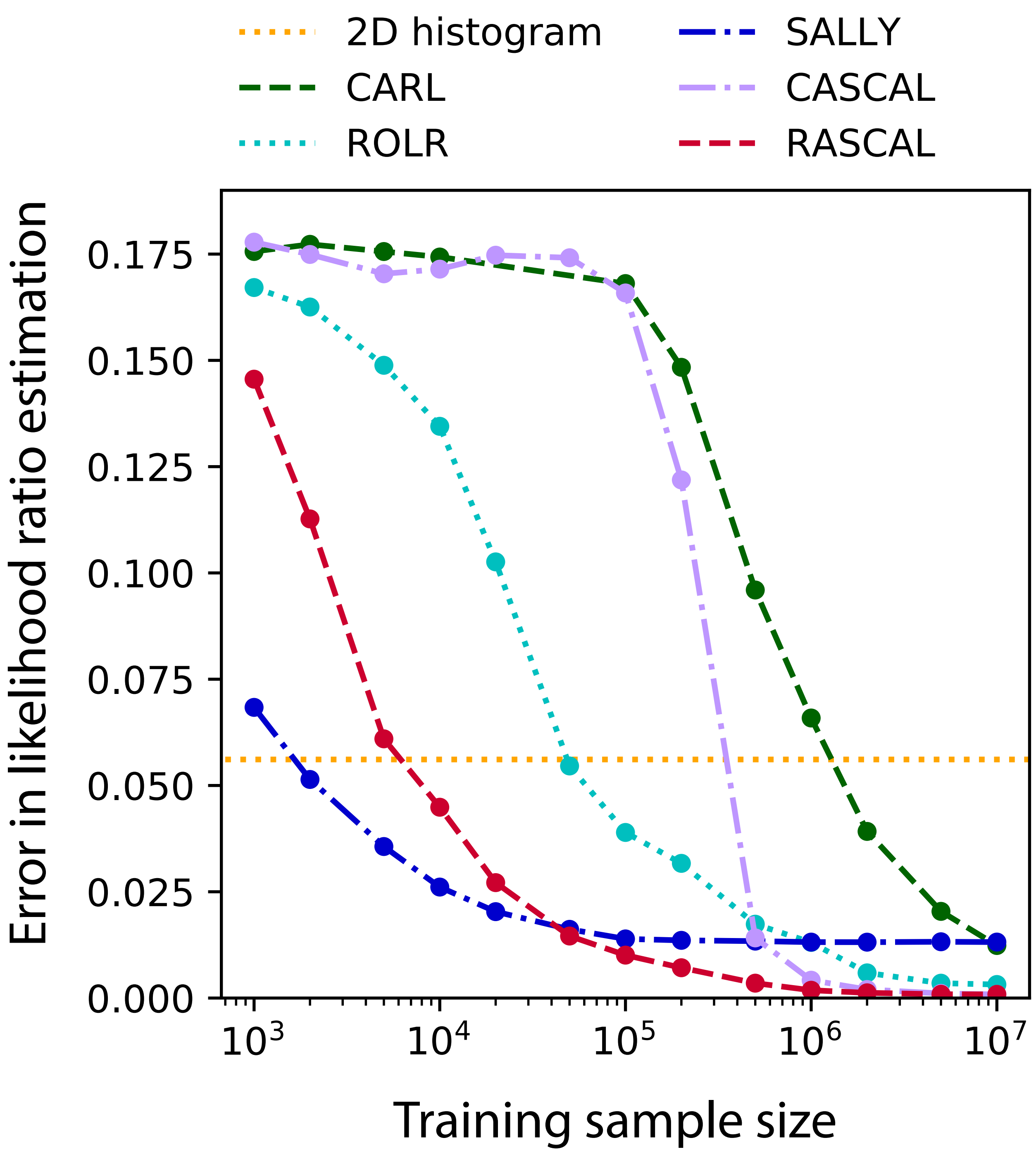
Each point = 1 event



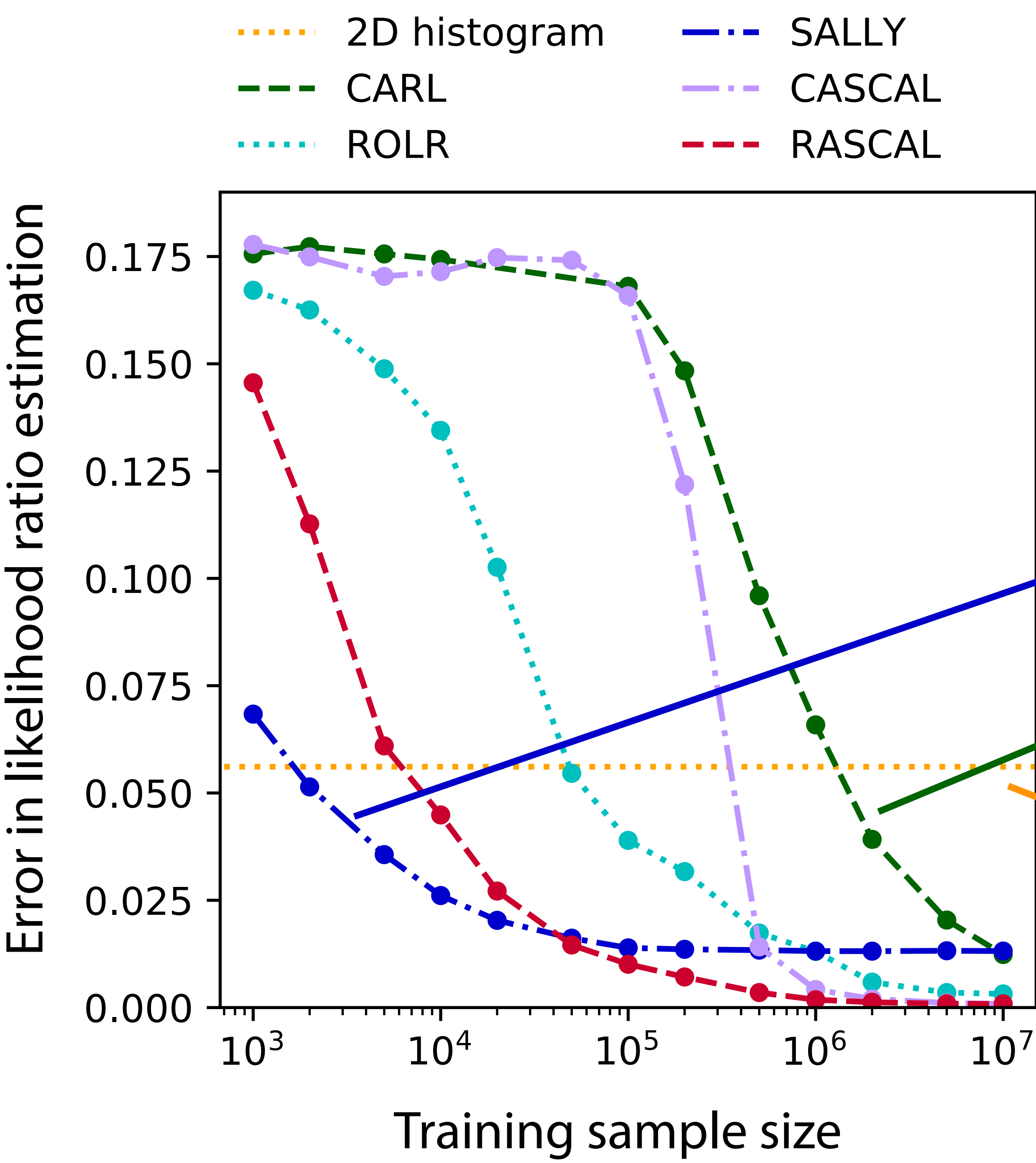
True log likelihood ratio
(usually don’t have this)

RASCAL
much more precise than
traditional histogram

Less training data needed



Less training data needed

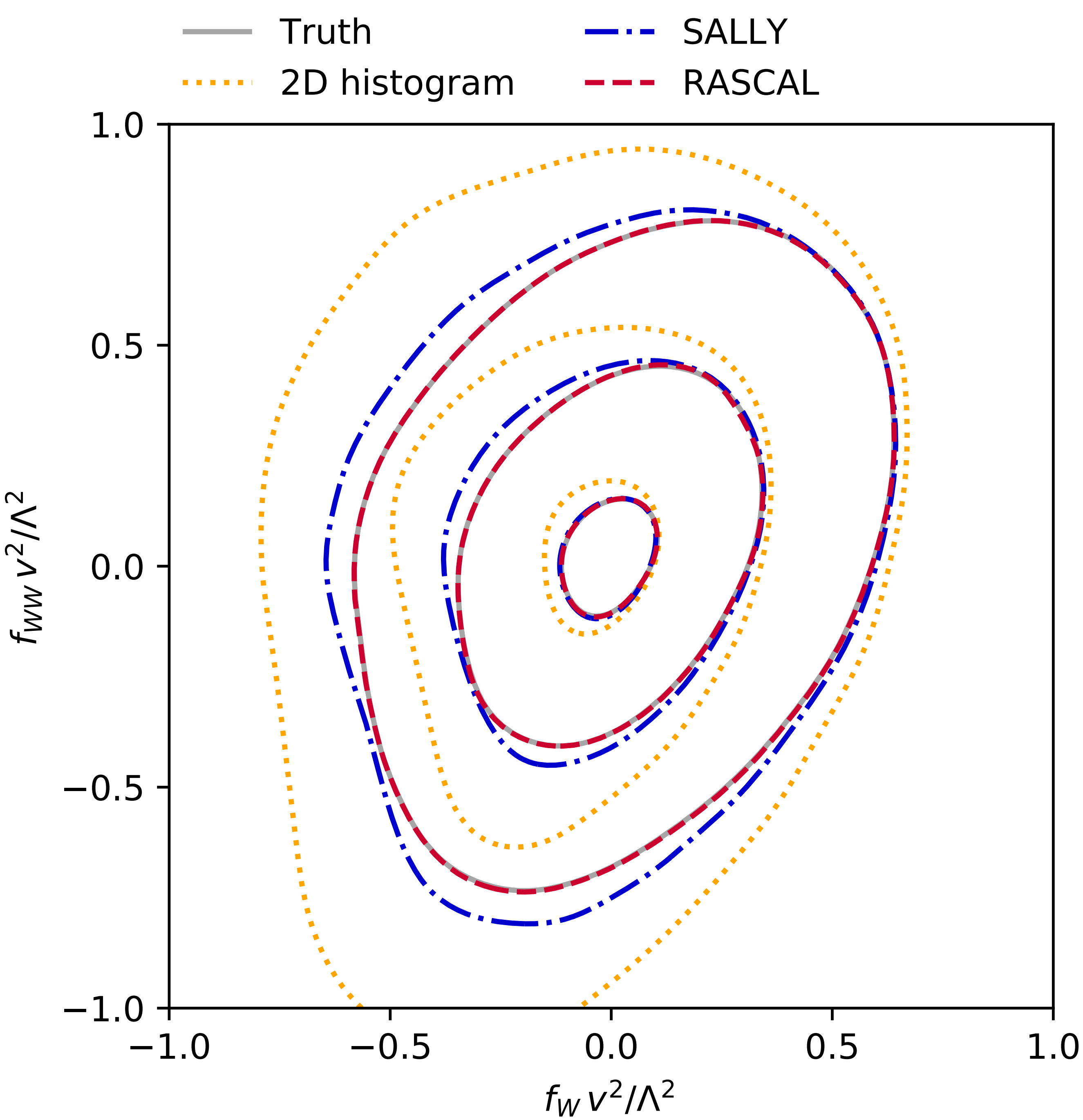


New techniques
require less data than
generic ML method
[CARL, see K. Cranmer, G. Louppe, J. Pavez 1506.02169]

Traditional histogram

Stronger bounds

Expected exclusion limits at 68%, 95%, 99.7% CL

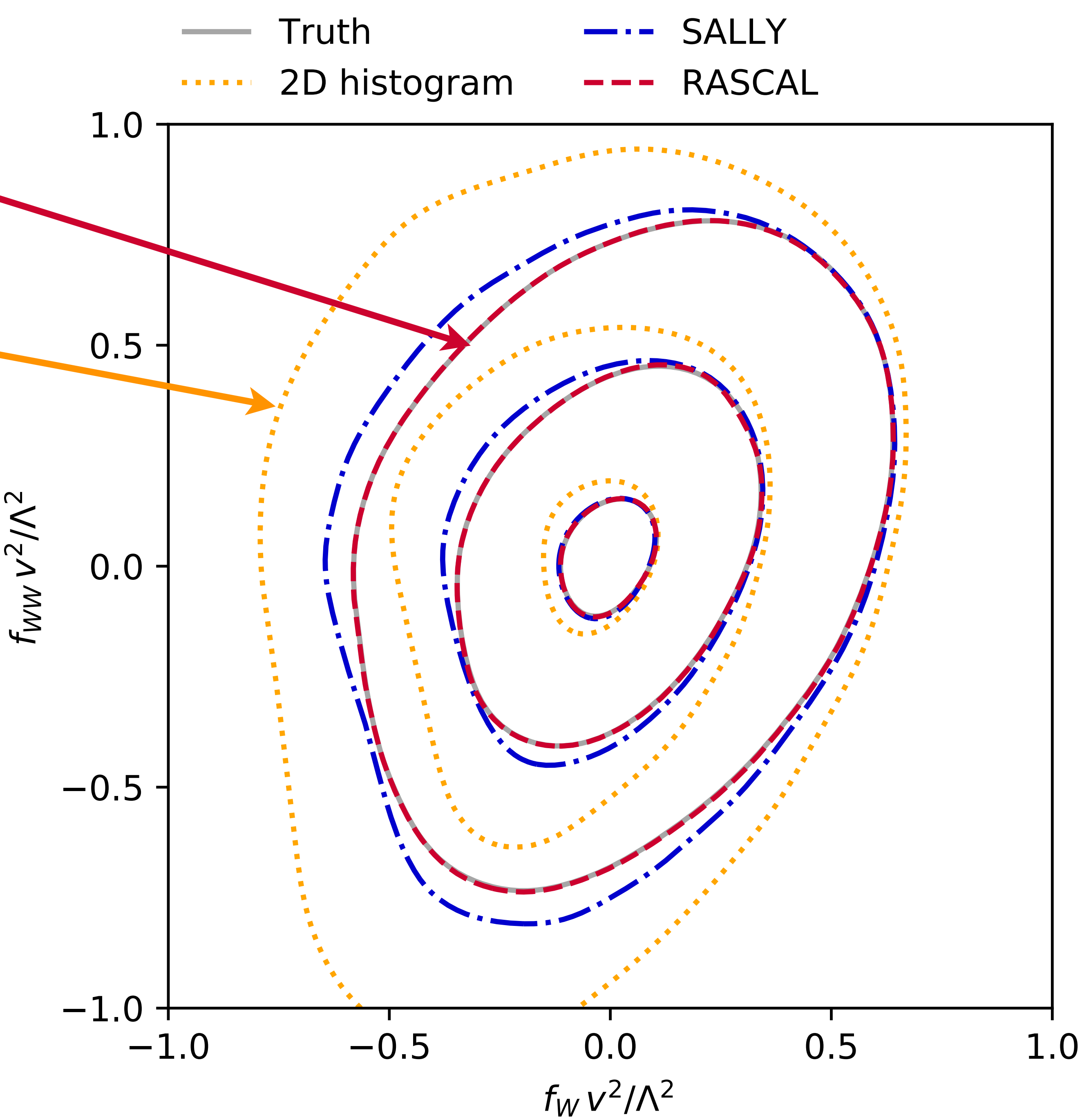


36 events, assuming SM

Stronger bounds

RASCAL
enables stronger
limits than
traditional histogram

Expected exclusion limits at 68%, 95%, 99.7% CL



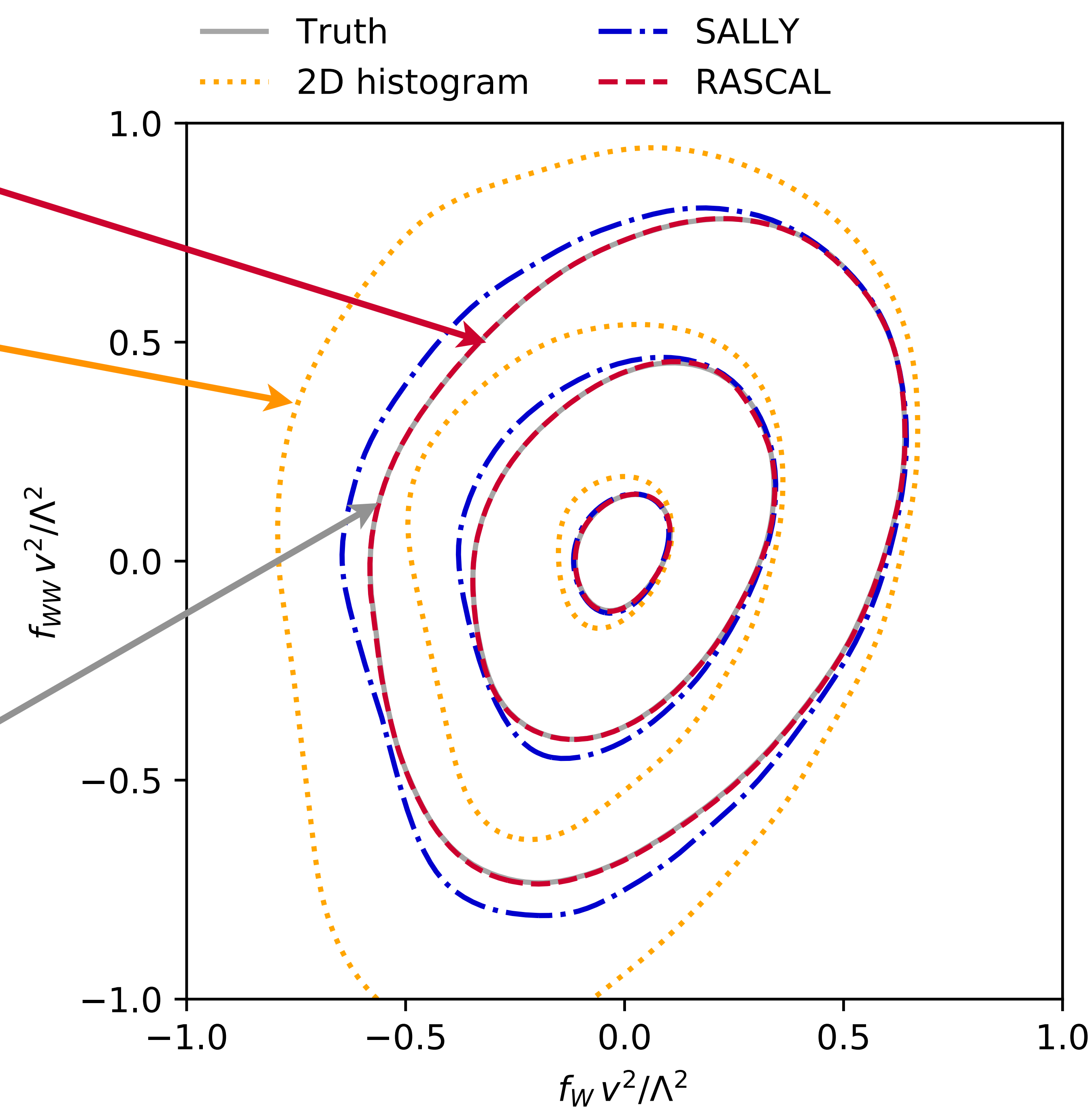
36 events, assuming SM

Stronger bounds

RASCAL
enables stronger
limits than
traditional histogram

Limits from RASCAL
virtually indistinguishable
from true likelihood
(usually we don't have that)

Expected exclusion limits at 68%, 95%, 99.7% CL



36 events, assuming SM

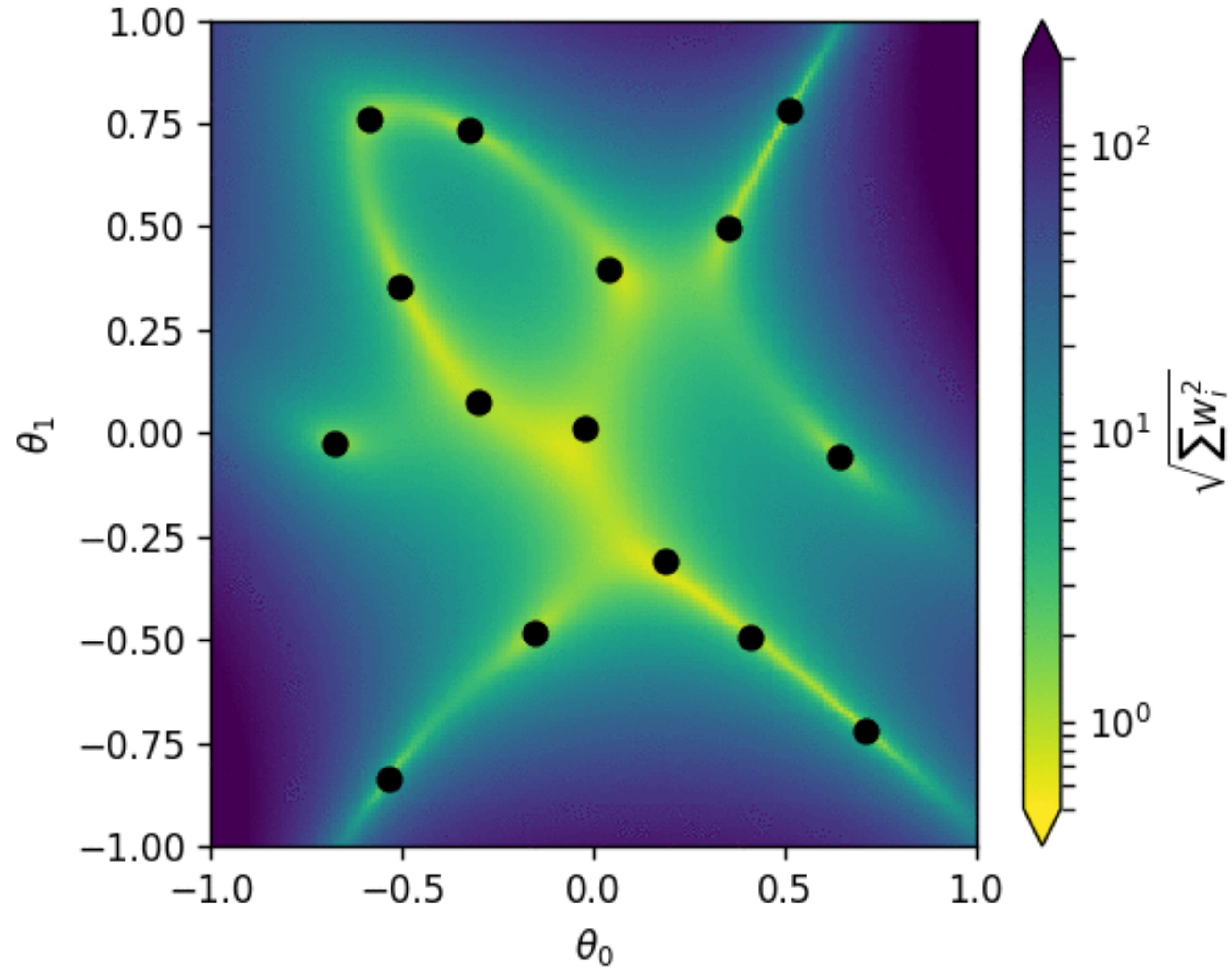
MadMiner

[JB, K. Cranmer, F. Kling in progress]

Can I use any of this?

Yes! To make that as painless as possible, we're working on the python package MadMiner:

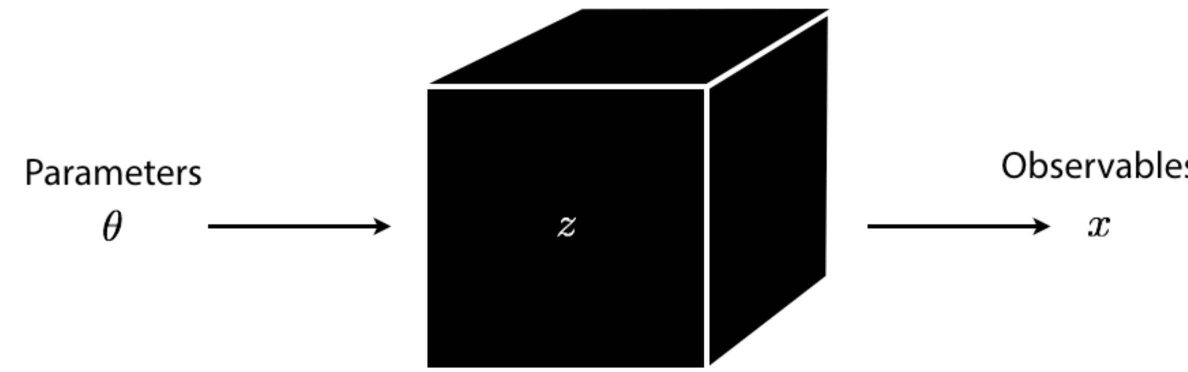
- "Mining gold" from MadGraph + Pythia + detector simulation
- Morphing: reconstruct full dependence on model parameters from few MC runs
- Likelihood ratio estimation with RASCAL and friends
- Calculate Fisher information (truth or reco level)



Come visit us at github.com/johannbrehmer/madminer!

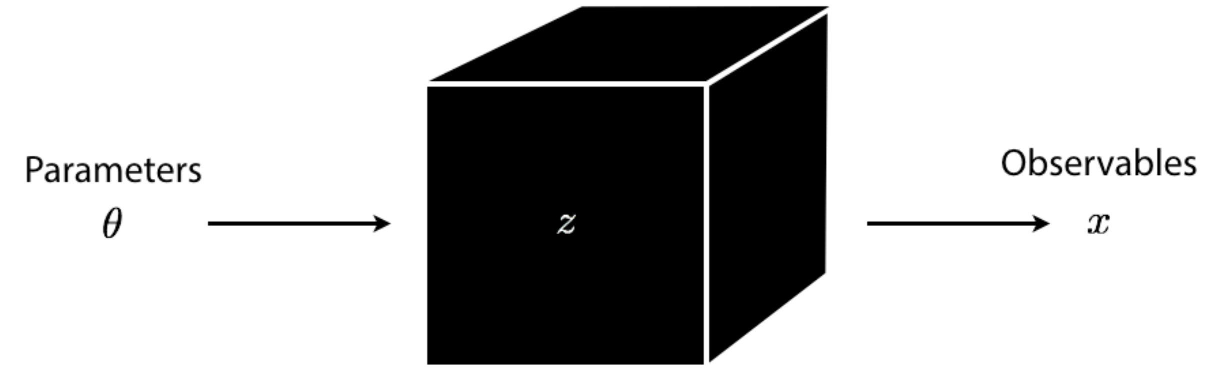
A new approach to simulator-based inference

- Many LHC analyses (and much of modern science) are based on simulations, “likelihood-free”

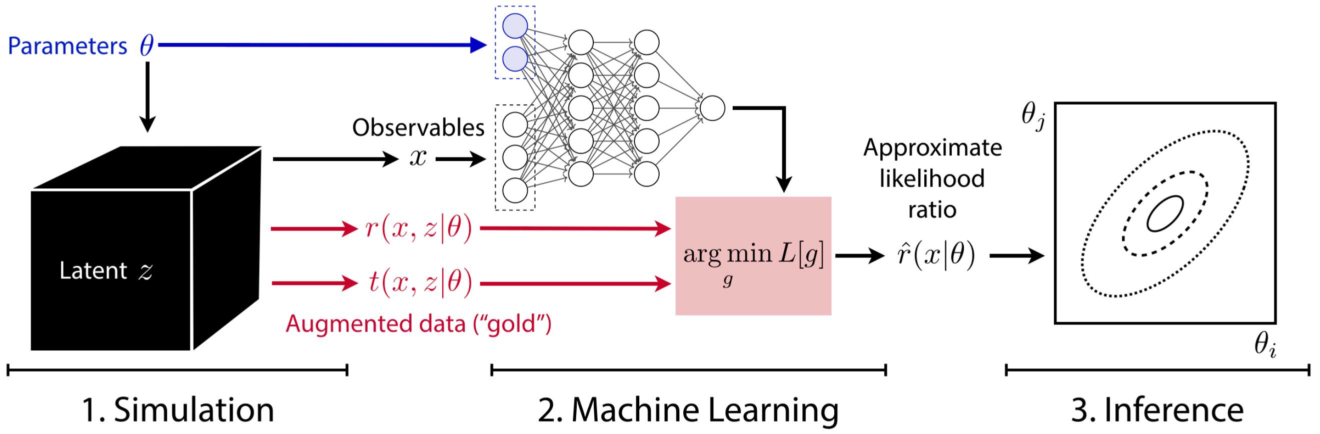


A new approach to simulator-based inference

- Many LHC analyses (and much of modern science) are based on simulations, “likelihood-free”

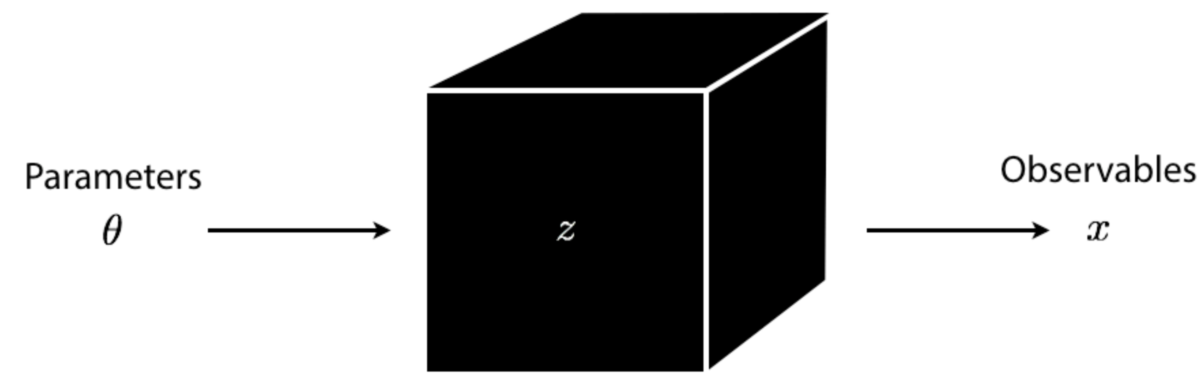


- New multivariate inference techniques: Leverage more information from simulator + power of machine learning

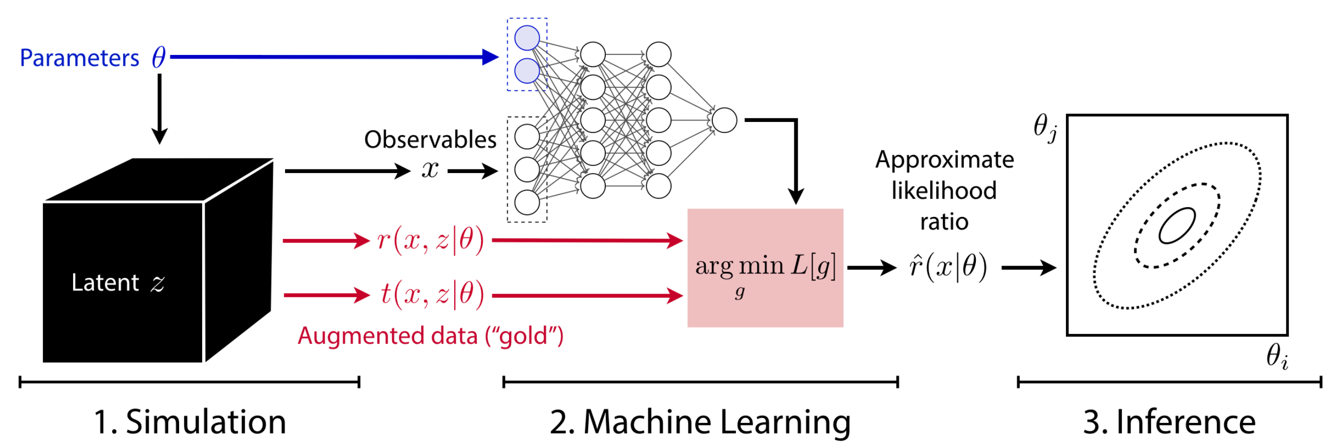


A new approach to simulator-based inference

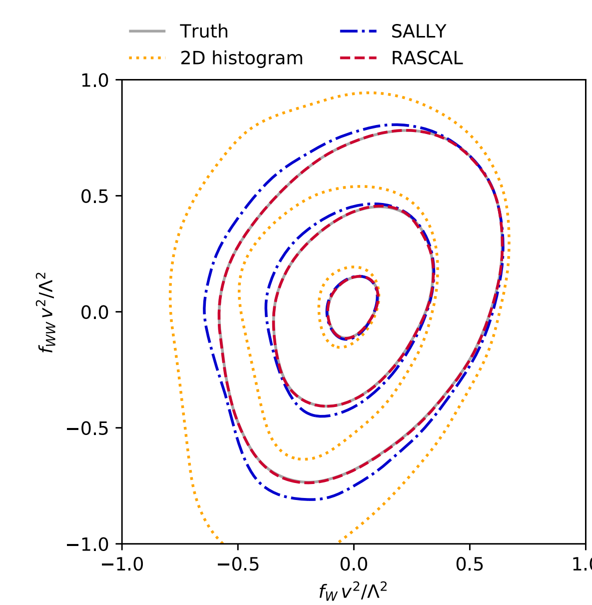
- Many LHC analyses (and much of modern science) are based on simulations, "likelihood-free"



- New multivariate inference techniques: Leverage more information from simulator + power of machine learning



- First application to Higgs physics: Stronger EFT constraints with less data



References


Kyle Cranmer


Gilles Louppe


Juan Pavez


Markus Stoye


Felix Kling


Tilman Plehn


Tim Tait

JB, KC, FK, TP:	Better Higgs Measurements Through Information Geometry	[1612.05261]
JB, FK, TP, TT:	Better Higgs-CP Measurements Through Information Geometry	[1712.02350]
JB, KC, GL, JP:	Constraining Effective Field Theories with Machine Learning	[1805.00013]
JB, KC, GL, JP:	A Guide to Constraining Effective Field Theories with Machine Learning	[1805.00020]
JB, GL, JP, KC:	Mining gold from implicit models to improve likelihood-free inference	[1805.12244]
MS, **JB**, GL, JP, KC:	Likelihood-free inference with an improved cross-entropy estimator	[1808.00973]
JB, KC, FK:	MadMiner	In preparation

Thanks to Kyle and Gilles for inspiring many of my slides!