Constraining Effective Field Theories with Machine Learning

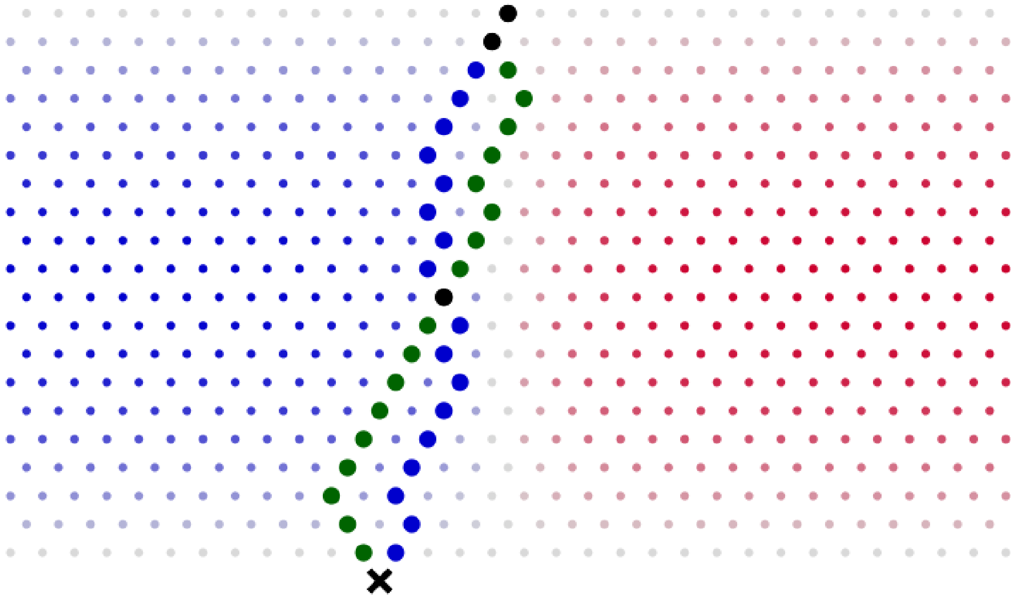
ATLAS ML workshop, October 15-17 2018

Gilles Louppe
g.louppe@uliege.be

with Johann Brehmer, Kyle Cranmer and Juan Pavez.

LIÈGE université

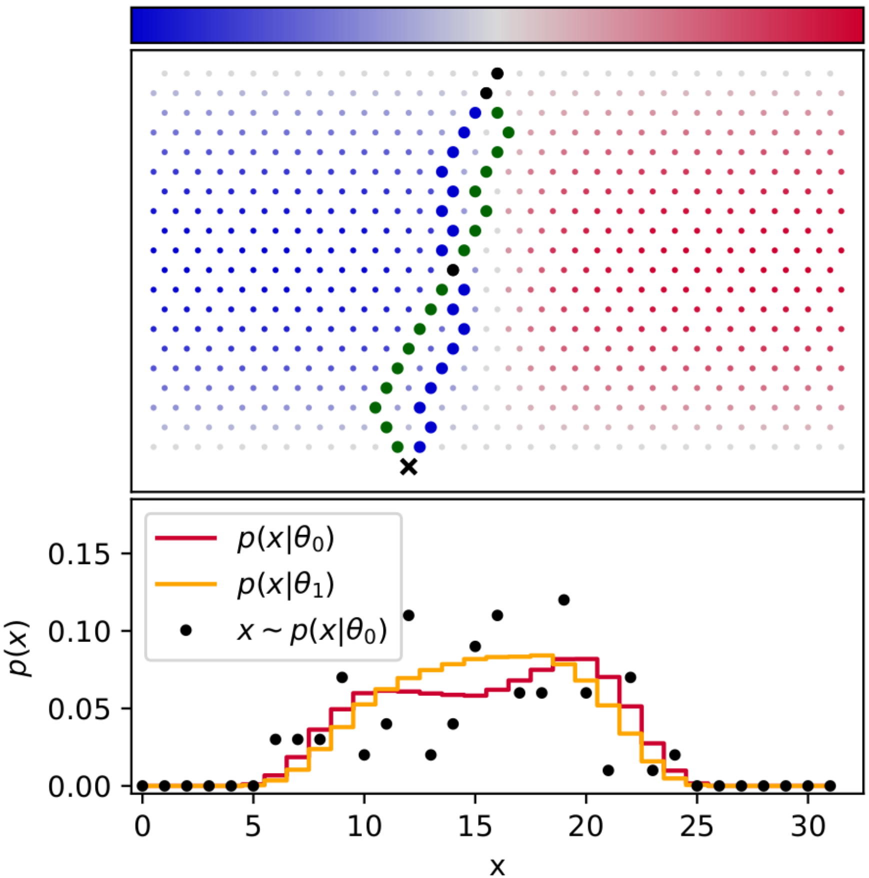
Refs: GaltonBoard.com



The probability of ending in bin x corresponds to the total probability of all the paths z from start to x.

$$p(x|\theta) = \int p(x, z|\theta) dz = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

What if we shift or remove some of the pins?



The probability of ending in bin x still corresponds to the total probability of all the paths z from start to x:

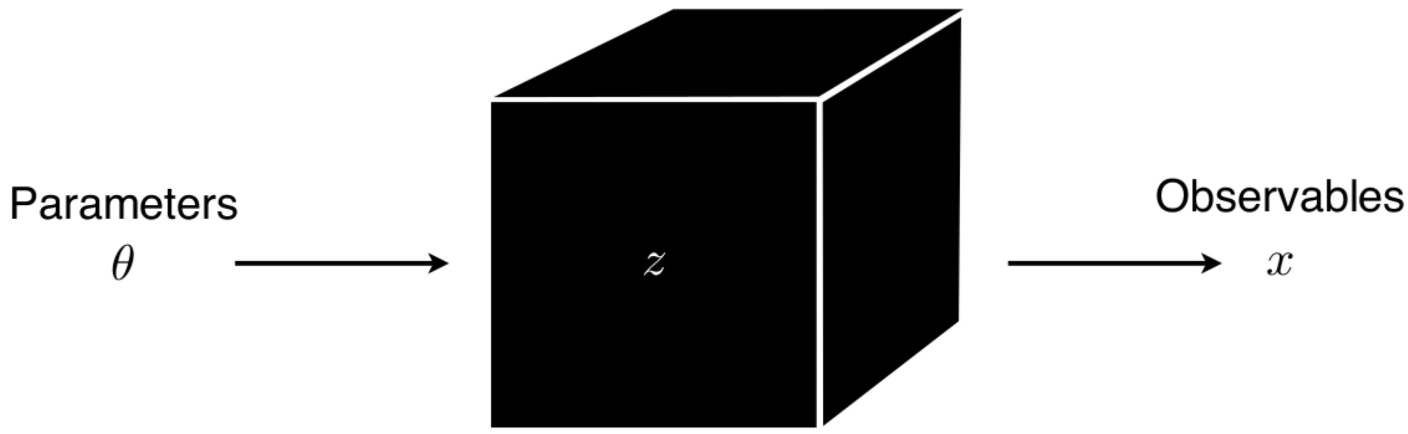
$$p(x|\theta) = \int p(x, z|\theta) dz$$

- But this integral can no longer be simplified analytically!
- As n grows larger, evaluating $p(x|\theta)$ becomes intractable since the number of paths grows combinatorially.
- Generating observations remains easy: drop balls.

The Galton board is a metaphore for the simulator-based scientific method:

- the Galton board device is the equivalent of the scientific simulator
- θ are parameters of interest
- z are stochastic execution traces through the simulator
- x are observables

Inference in this context requires likelihood-free algorithms.

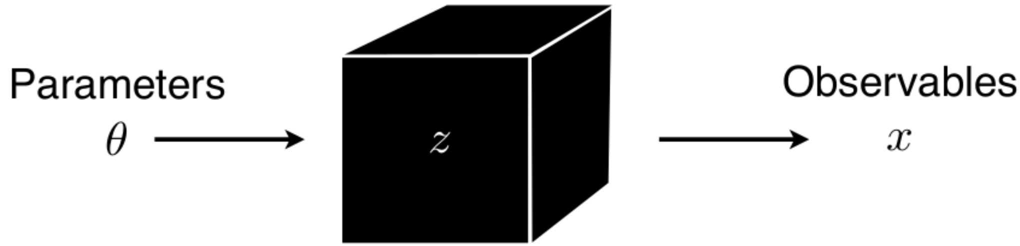


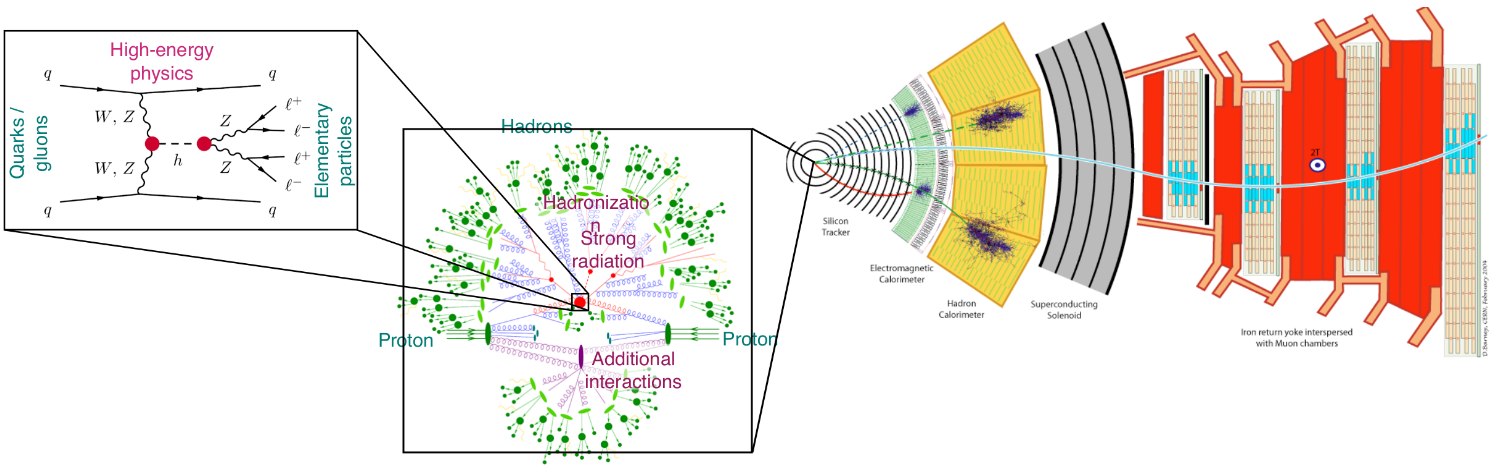
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

Inference:

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



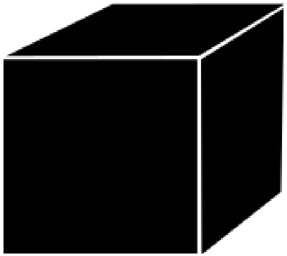


The Galton board of particle physics

Likelihood-free inference methods

Treat the simulator as a black box

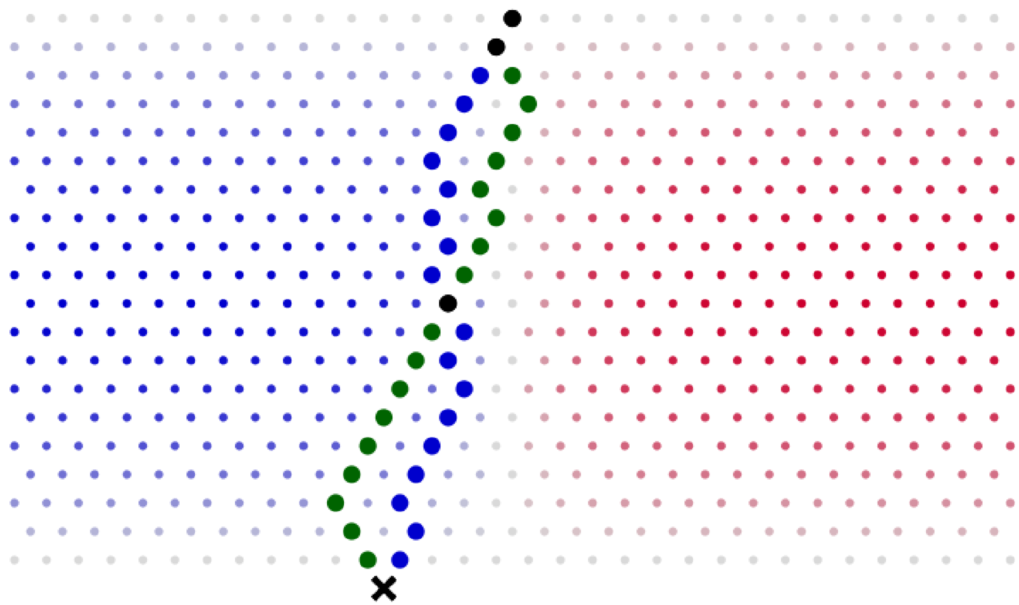
- Histograms of observables, Approximate Bayesian computation.
  - *Rely on summary statistics.*
- Machine learning algorithms
  - *Density estimation, CARL, autoregressive models, normalizing flows, etc.*



Use latent structure

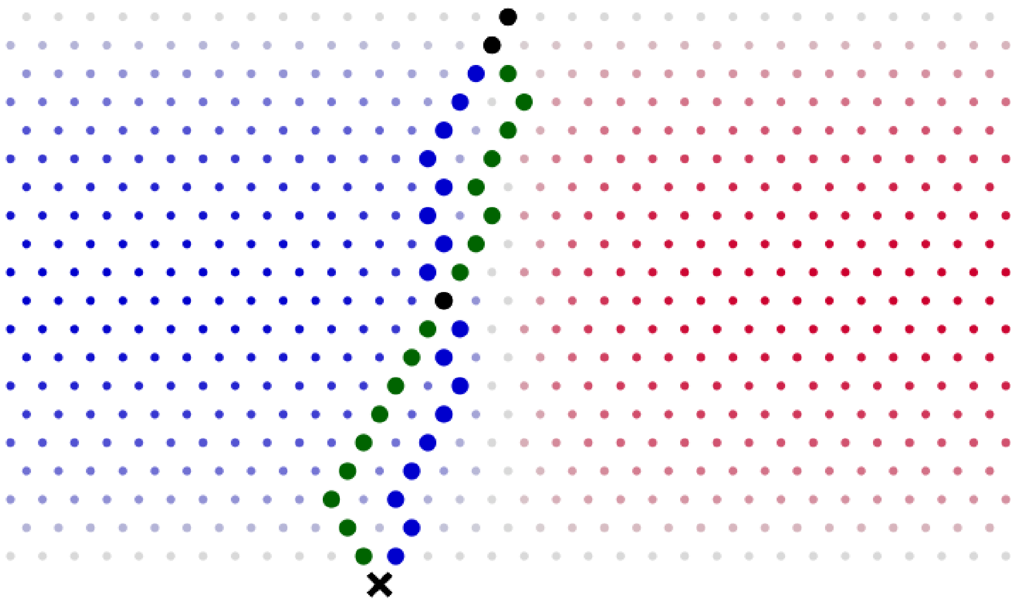
- Matrix Element Method, Optimal Observables, Shower deconstruction, Event Deconstruction.
  - *Neglect or approximate shower + detector, explicitly calculate z integral.*
- *new* Mining gold from the simulator.
  - *Leverage matrix-element information + Machine Learning.*

Mining gold from simulators



$p(x|\theta)$ is usually intractable.

What about $p(x, z|\theta)$?



$$p(x, z|\theta) = p(z_1|\theta)p(z_2|z_1, \theta) \dots p(z_T|z_{<T}, \theta)p(x|z_{\leq T}, \theta)$$
$$= p(z_1|\theta)p(z_2|\theta) \dots p(z_T|\theta)p(x|z_T)$$
$$= p(x|z_T) \prod_t \theta^{z_t} (1-\theta)^{1-z_t}$$

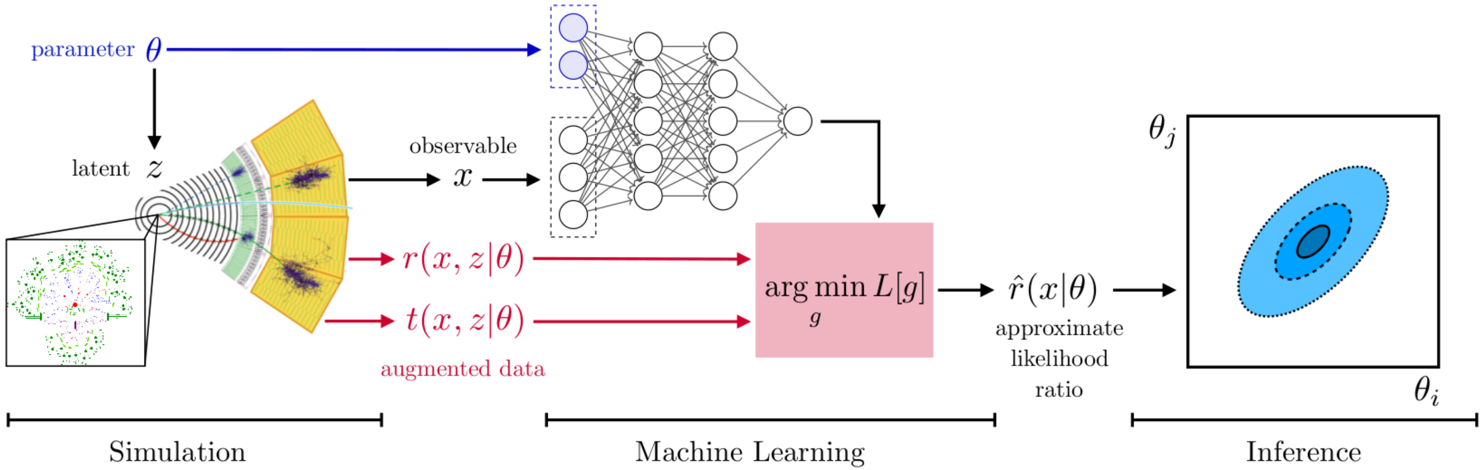
This can be computed as the ball falls down the board!

As the trajectory $z = z_1, ..., z_T$ and the observable x are emitted, it is often possible:

- to calculate the joint likelihood $p(x, z|\theta)$;
- to calculate the joint likelihood ratio $r(x, z|\theta_0, \theta_1)$;
- to calculate the joint score $t(x, z|\theta_0) = \nabla_\theta \log p(x, z|\theta)\big|_{\theta_0}$.

We call this process mining gold from your simulator!

Rascal



Refs: Brehmer et al, 2018 (arXiv:1805.12244)

Constraining Effective Field Theories, effectively

LHC processes

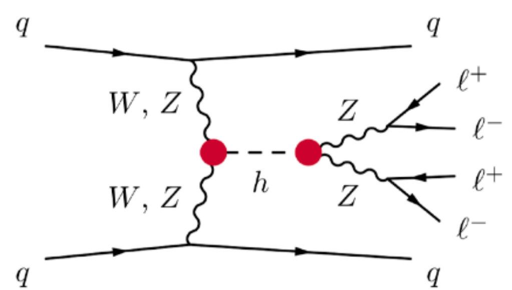
Latent variables

Parameters of interest

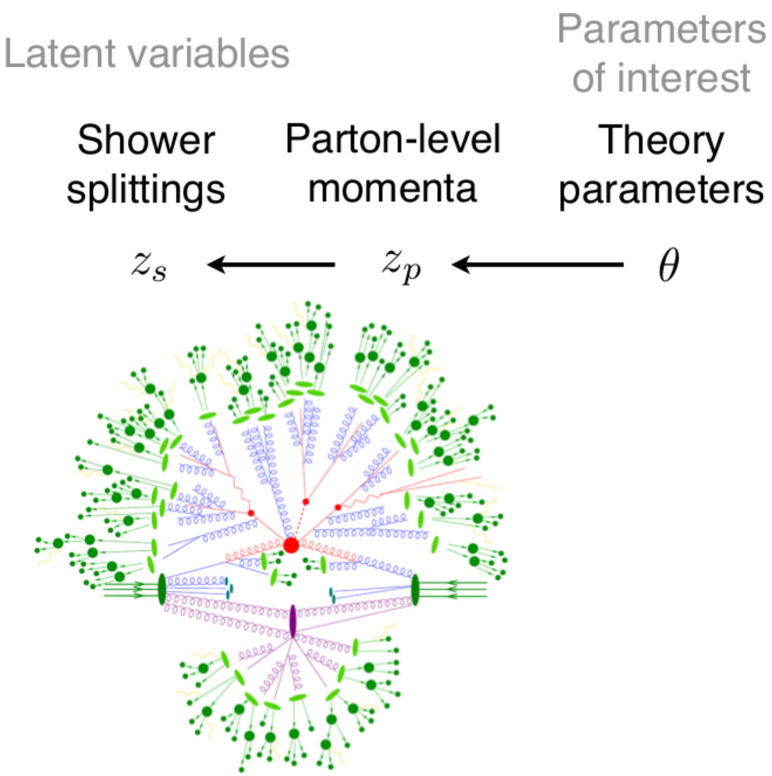
Parton-level momenta

Theory parameters

$z_p \longleftarrow \theta$



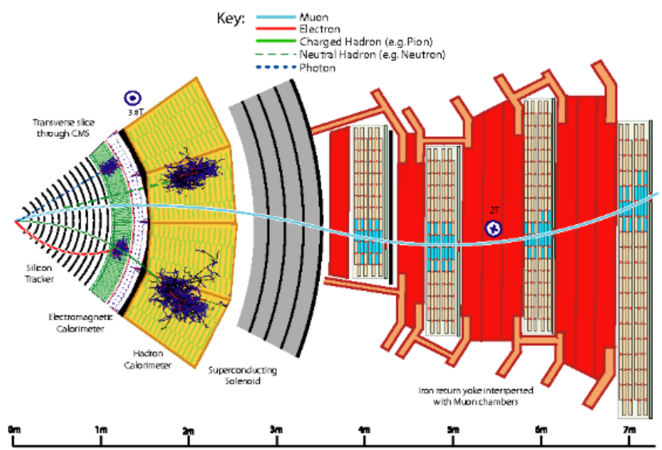
LHC processes



LHC processes

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

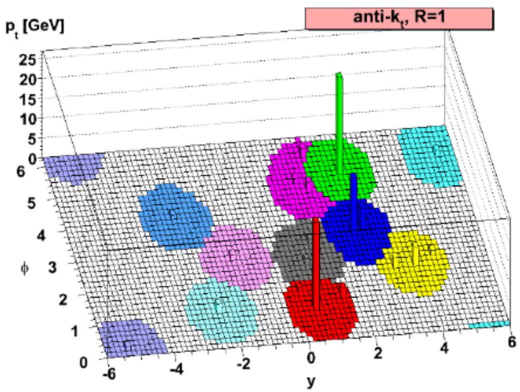
$$z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$



LHC processes

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

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



[Image source: M. Cacciari, G. Salam, G. Soyez 0802.1189]

$$p(x|\theta) = \underbrace{\iiint}_{\text{intractable}} p(z_p|\theta)p(z_s|z_p)p(z_d|z_s)p(x|z_d)dz_p dz_s dz_d$$

Key insights:

- The distribution of parton-level momenta

$$p(z_p|\theta) = \frac{1}{\sigma(\theta)} \frac{d\sigma(\theta)}{dz_p},$$

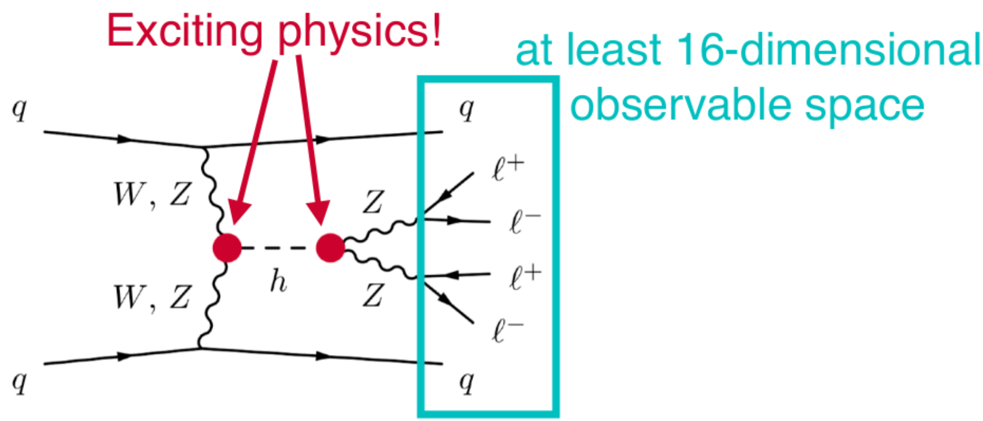
  where $\sigma(\theta)$ and $\frac{d\sigma(\theta)}{dz_p}$ are the total and differential cross sections, is tractable.

- Downstream processes $p(z_s|z_p)$, $p(z_d|z_s)$ and $p(x|z_d)$ do not depend on θ.

$\Rightarrow$ This implies that both $r(x, z|\theta_0, \theta_1)$ and $t(x, z|\theta_0)$ can be mined. E.g.,

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

Proof of concept



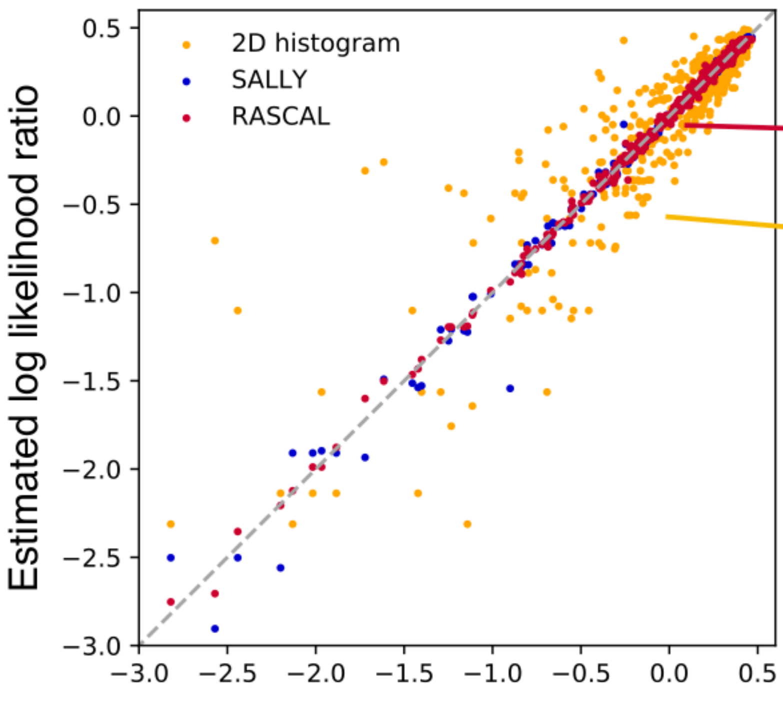
Higgs production in weak boson fusion

Goal: Constraints on two theory parameters:

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

Precise likelihood ratio estimates

Each point = 1 event

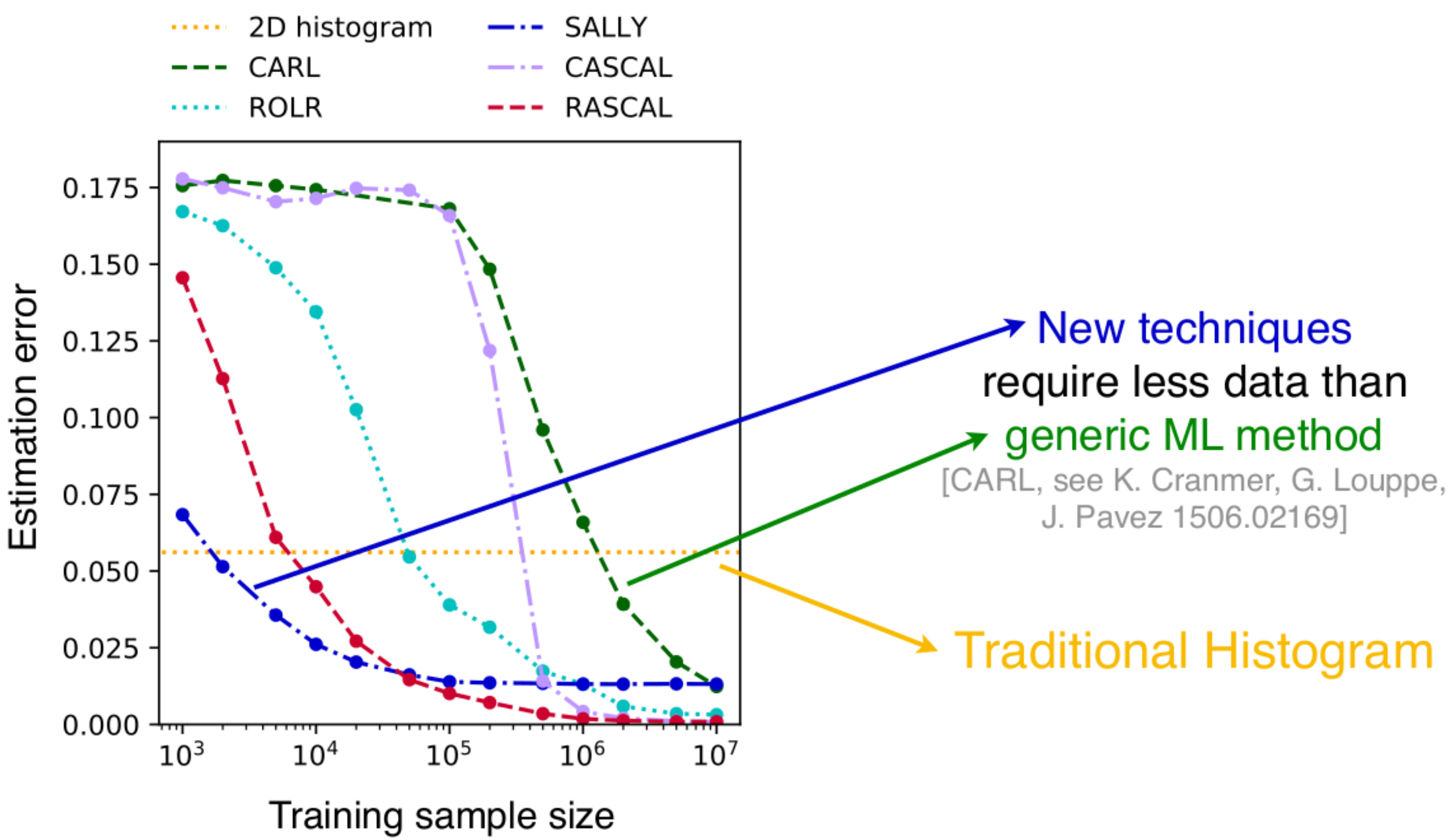


True log likelihood ratio
(usually don't have this)

RASCAL
much more precise than
traditional histogram

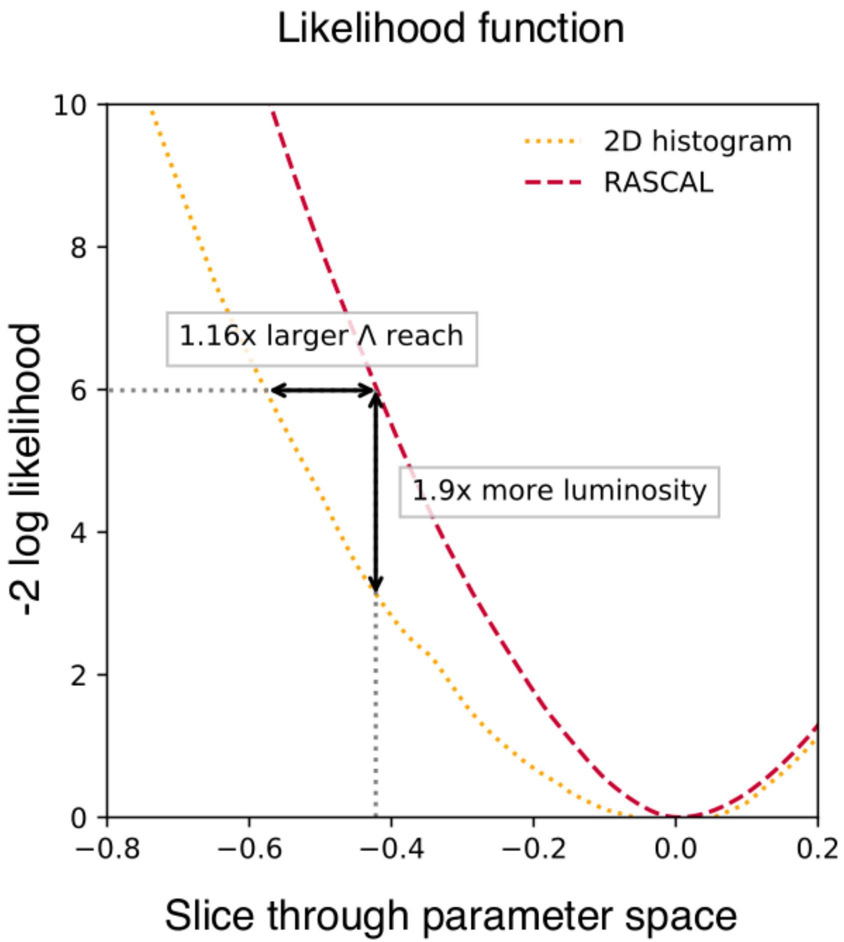
Credits: Johann Brehmer

Increased data efficiency



Credits: Johann Brehmer

Better sensitivity



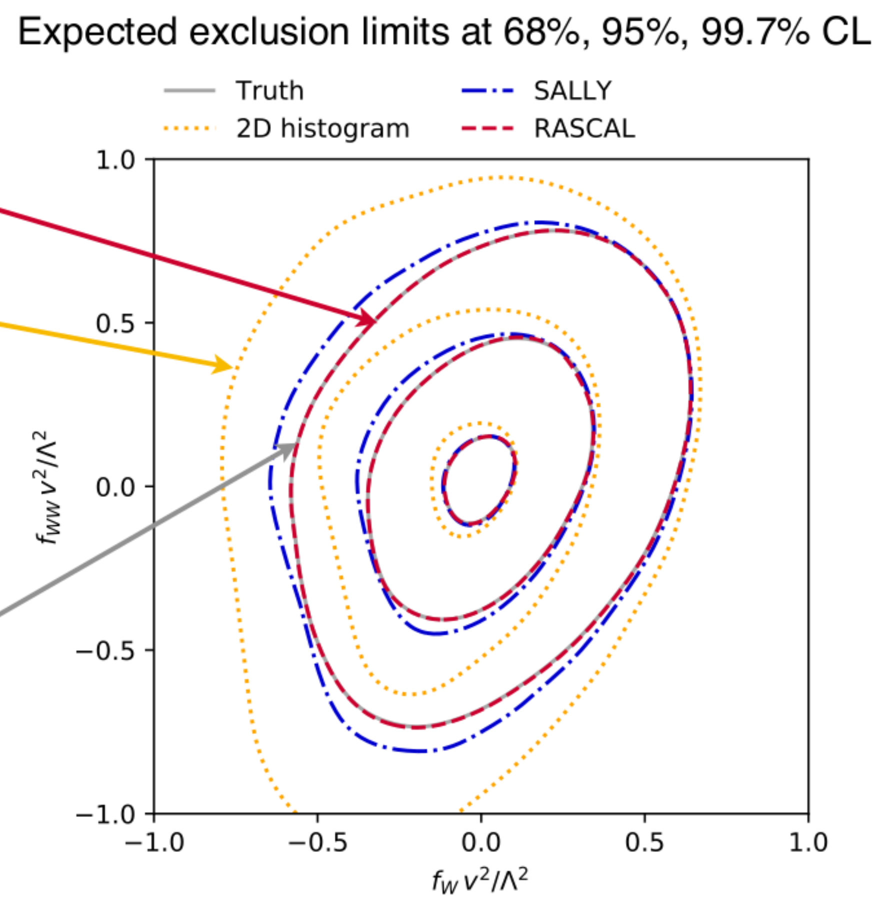
36 events, assuming SM

Credits: Johann Brehmer

Stronger bounds

RASCAL enables stronger limits than traditional histogram

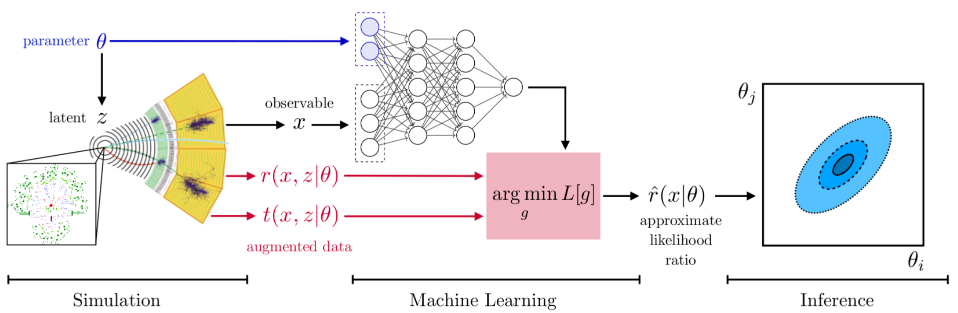
Limits from RASCAL virtually indistinguishable from true likelihood
(usually we don't have that)

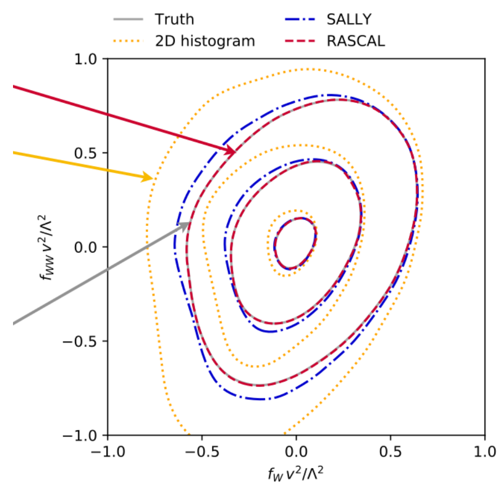


36 events, assuming SM

Summary

- Many LHC analysis (and much of modern science) are based on "likelihood-free" simulations.
- New inference algorithms:
  - Leverage more information from the simulator
  - Combine with the power of machine learning
- First application to LHC physics: stronger EFT constraints with less simulations.





Collaborators



Johann Brehmer, Kyle Cranmer and Juan Pavez

References

- Stoye, M., Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2018). Likelihood-free inference with an improved cross-entropy estimator. arXiv preprint arXiv:1808.00973.
- Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2018). Mining gold from implicit models to improve likelihood-free inference. arXiv preprint arXiv:1805.12244.
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. (2018). Constraining Effective Field Theories with Machine Learning. arXiv preprint arXiv:1805.00013.
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. (2018). A Guide to Constraining Effective Field Theories with Machine Learning. arXiv preprint arXiv:1805.00020.
- Cranmer, K., Pavez, J., & Louppe, G. (2015). Approximating likelihood ratios with calibrated discriminative classifiers. arXiv preprint arXiv:1506.02169.

The end.