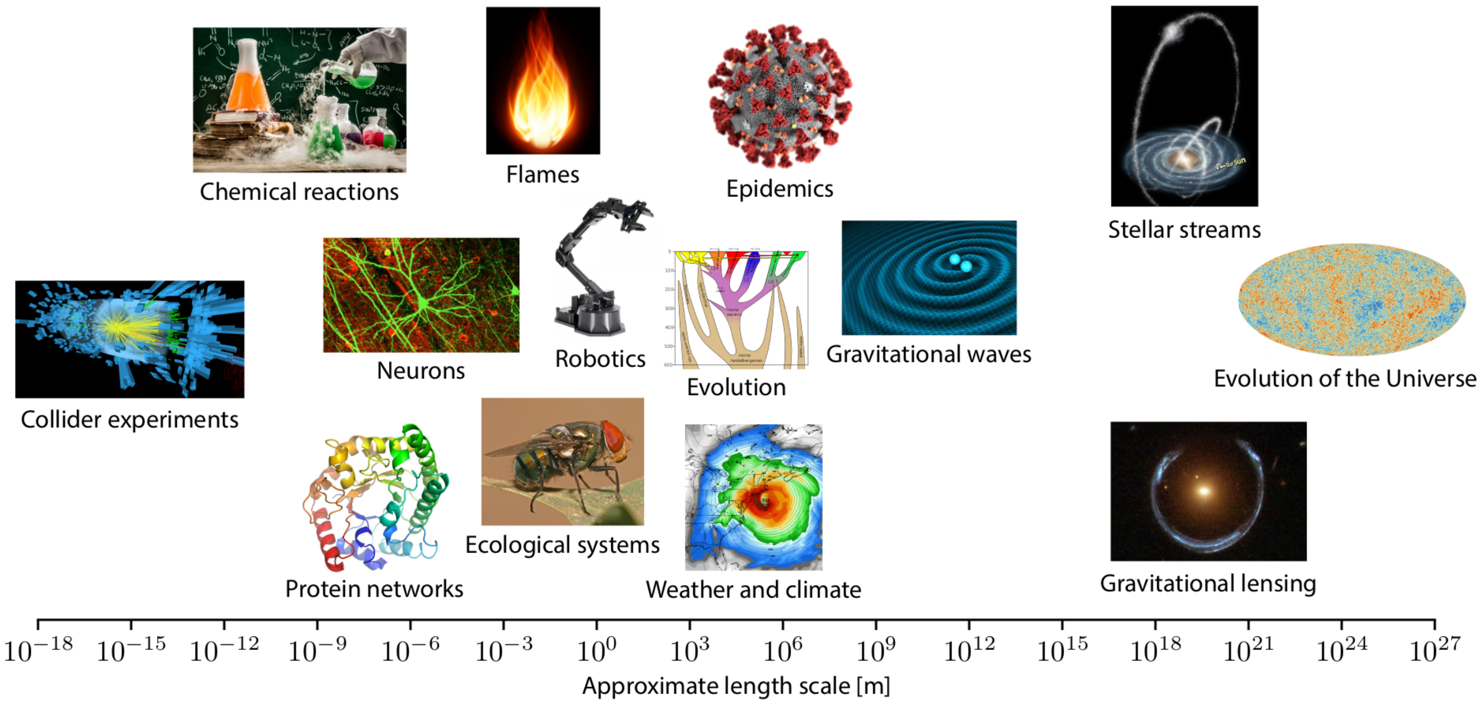
The frontier of simulation-based inference

February 12, AIMS Seminar Series, University of Oxford

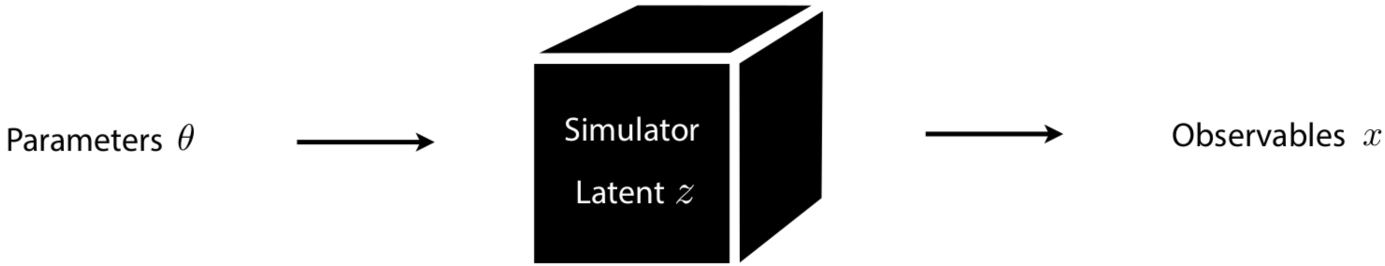
Gilles Louppe
g.louppe@uliege.be

This talk is inspired and adapted from previous talks given by my wonderful co-authors Kyle Cranmer and Johann Brehmer. Thanks to them!


Chemical reactions
Flames
Epidemics
Stellar streams
Collider experiments
Neurons
Robotics
Evolution
Gravitational waves
Evolution of the Universe
Protein networks
Ecological systems
Weather and climate
Gravitational lensing
10^{-18} 10^{-15} 10^{-12} 10^{-9} 10^{-6} 10^{-3} 10^{0} 10^{3} 10^{6} 10^{9} 10^{12} 10^{15} 10^{18} 10^{21} 10^{24} 10^{27}
Approximate length scale [m]

Simulation-based inference

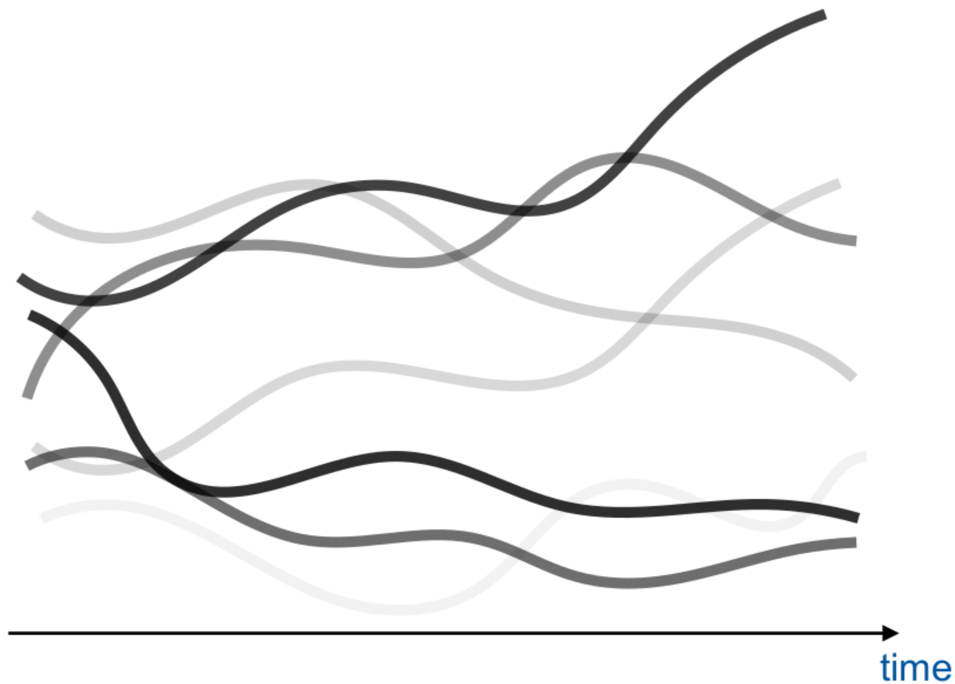


Prediction:

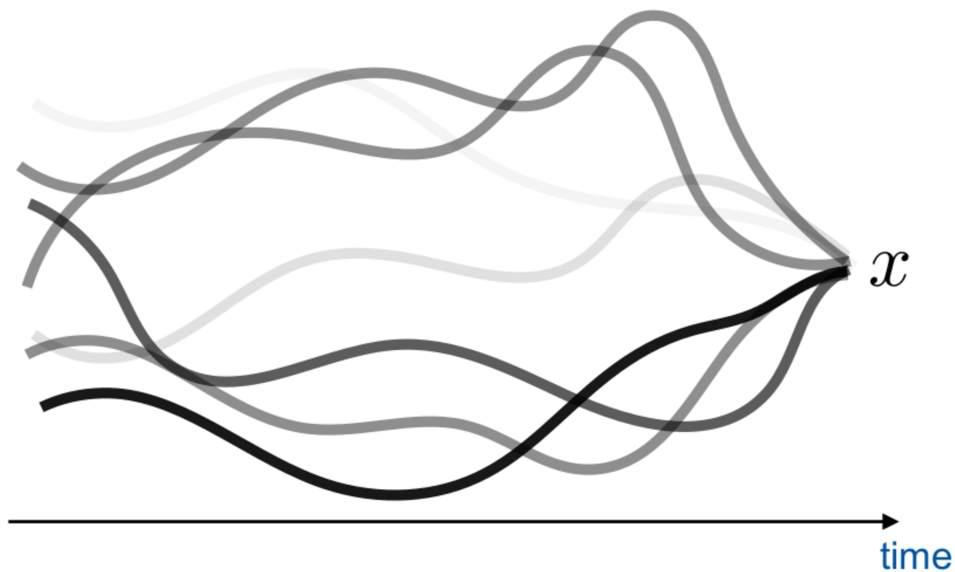
- Well-motivated mechanistic, causal model
- Simulator can generate samples $x \sim p(x|\theta)$

Inference:

- Interactions between low-level components lead to challenging inverse problems
- Likelihood $p(x|\theta) = \int \mathrm{d}z \; p(x, z|\theta)$ is intractable



$$\theta, z, x \sim p(\theta, z, x)$$



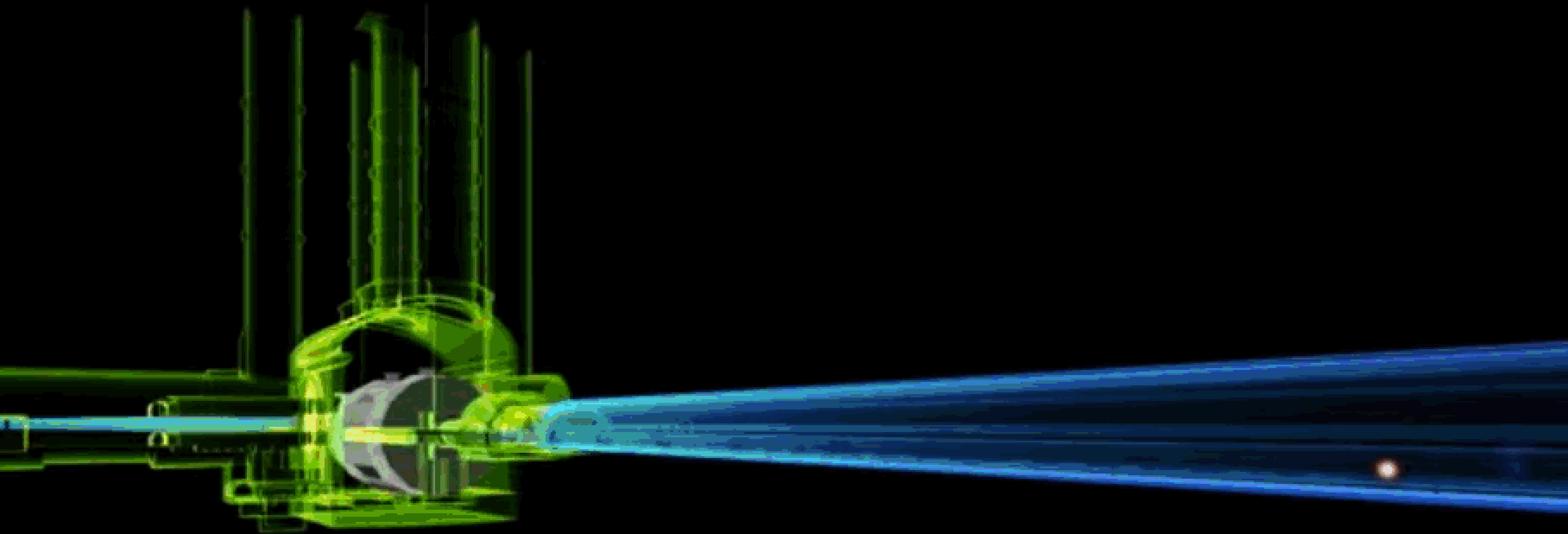
$$\theta, z \sim p(\theta, z|x)$$

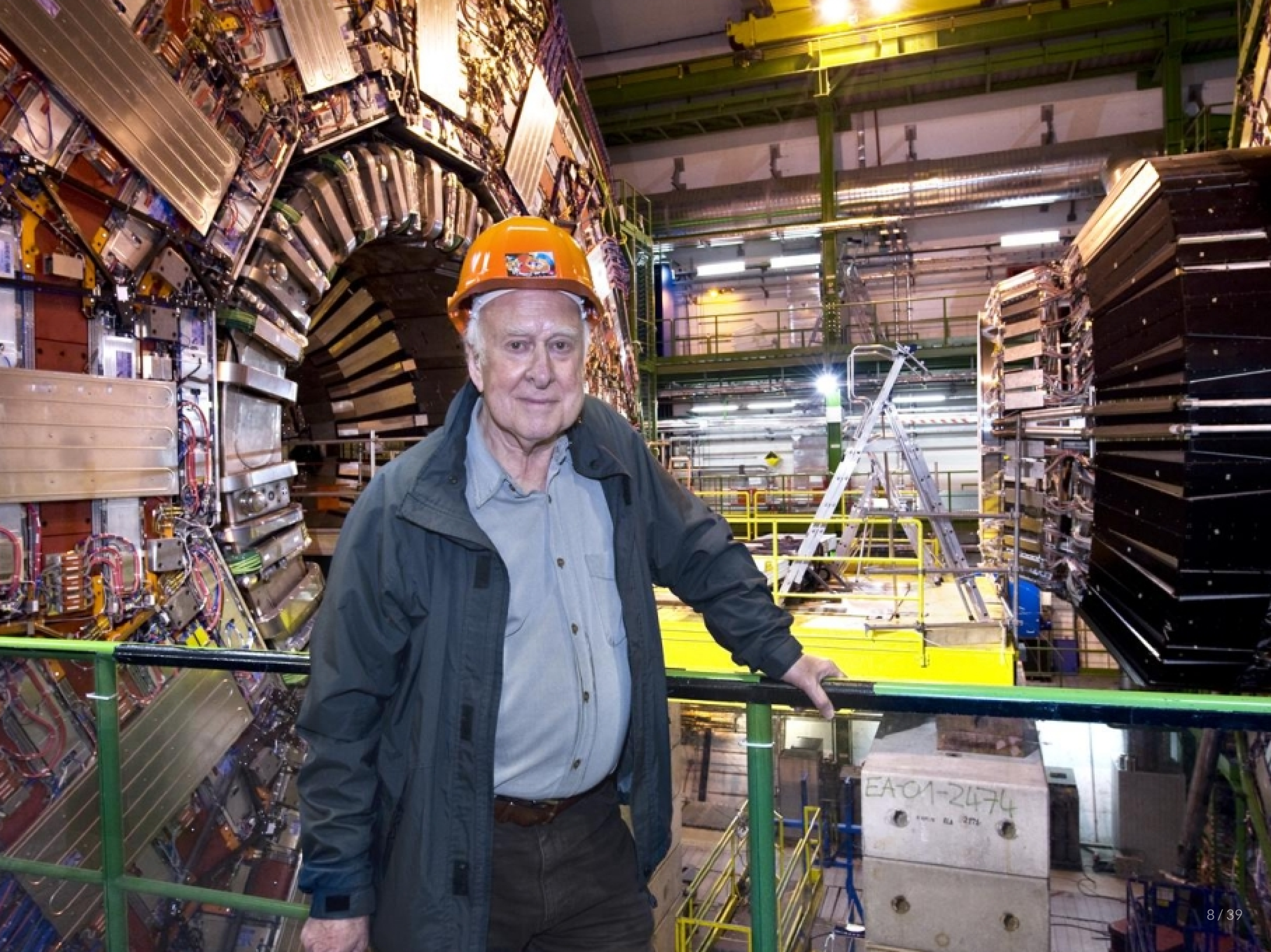
The Bean machine is a metaphore of simulation-based science:

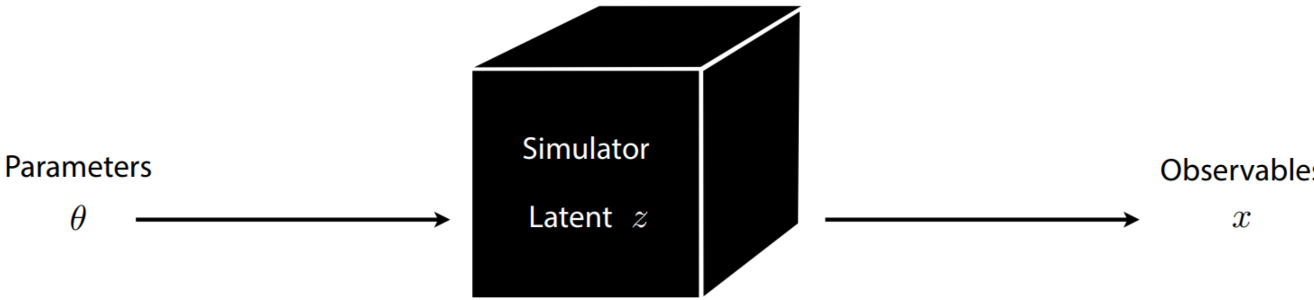
Bean machine	$\rightarrow$	Computer simulation
Parameters θ	$\rightarrow$	Model parameters θ
Buckets x	$\rightarrow$	Observables x
Random paths z	$\rightarrow$	Latent variables z (stochastic execution traces through simulator)

The case of particle physics

$$
\begin{gathered}
\mathcal{L}_{SM} = -\frac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \frac{1}{4}g_s^2 f^{abc}f^{ade}g^b_\mu g^c_\nu g^d_\mu g^e_\nu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
M^2 W^+_\mu W^-_\mu - \frac{1}{2}\partial_\nu Z^0_\mu\partial_\nu Z^0_\mu - \frac{1}{2c_w^2}M^2 Z^0_\mu Z^0_\mu - \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - igc_w(\partial_\nu Z^0_\mu(W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - Z^0_\nu(W^+_\mu\partial_\nu W^-_\mu - W^-_\mu\partial_\nu W^+_\mu) + Z^0_\mu(W^+_\nu\partial_\nu W^-_\mu - W^-_\nu\partial_\nu W^+_\mu)) - \\
igs_w(\partial_\nu A_\mu(W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu(W^+_\mu\partial_\nu W^-_\mu - W^-_\mu\partial_\nu W^+_\mu) + A_\mu(W^+_\nu\partial_\nu W^-_\mu - \\
W^-_\nu\partial_\nu W^+_\mu)) - \frac{1}{2} g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \frac{1}{2} g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2c_w^2(Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - \\
Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2(A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w(A_\mu Z^0_\nu(W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu) - \frac{1}{2}\partial_\mu H\partial_\mu H - 2M^2\alpha_h H^2 - \partial_\mu\phi^+\partial_\mu\phi^- - \frac{1}{2}\partial_\mu\phi^0\partial_\mu\phi^0 - \\
\beta_h\left(\frac{2M^2}{g^2} + \frac{2M}{g}H + \frac{1}{2}(H^2+\phi^0\phi^0+2\phi^+\phi^-)\right) + \frac{2M^4}{g^2}\alpha_h - \\
g\alpha_h M\left(H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-\right) - \\
\frac{1}{8} g^2\alpha_h\left(H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2H^2\right) - \\
gMW^+_\mu W^-_\mu H - \frac{1}{2} g\frac{M}{c_w^2}Z^0_\mu Z^0_\mu H - \\
\frac{1}{2} ig\left(W^+_\mu(\phi^0\partial_\mu\phi^- - \phi^-\partial_\mu\phi^0) - W^-_\mu(\phi^0\partial_\mu\phi^+ - \phi^+\partial_\mu\phi^0)\right) + \\
\frac{1}{2} g\left(W^+_\mu(H\partial_\mu\phi^- - \phi^-\partial_\mu H) + W^-_\mu(H\partial_\mu\phi^+ - \phi^+\partial_\mu H)\right) + \frac{1}{2} g\frac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) + \\
M\left(\frac{1}{c_w}Z^0_\mu\partial_\mu\phi^0 + W^+_\mu\partial_\mu\phi^- + W^-_\mu\partial_\mu\phi^+\right) - ig\frac{s_w^2}{c_w}MZ^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + igs_wMA_\mu(W^+_\mu\phi^- - \\
W^-_\mu\phi^+) - ig\frac{1-2c_w^2}{2c_w}Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + igs_wA_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) - \\
\frac{1}{4} g^2W^+_\mu W^-_\mu\left(H^2+(\phi^0)^2+2\phi^+\phi^-\right) - \frac{1}{8} g^2\frac{1}{c_w^2}Z^0_\mu Z^0_\mu\left(H^2+(\phi^0)^2+2(2s_w^2-1)^2\phi^+\phi^-\right) - \\
\frac{1}{2} g^2\frac{s_w^2}{c_w}Z^0_\mu\phi^0(W^+_\mu\phi^- + W^-_\mu\phi^+) - \frac{1}{2} ig^2\frac{s_w^2}{c_w}Z^0_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) + \frac{1}{2} g^2 s_w A_\mu\phi^0(W^+_\mu\phi^- + \\
W^-_\mu\phi^+) + \frac{1}{2} ig^2s_wA_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) - g^2\frac{s_w}{c_w}(2c_w^2-1)Z^0_\mu A_\mu\phi^+\phi^- - \\
g^2s_w^2A_\mu A_\mu\phi^+\phi^- + \frac{1}{2} ig_s\lambda^a_{ij}(\bar q^\sigma_i\gamma^\mu q^\sigma_j)g^a_\mu - \bar e^\lambda(\gamma\partial + m^\lambda_e)e^\lambda - \bar\nu^\lambda(\gamma\partial+m^\lambda_\nu)\nu^\lambda - \bar u^\lambda_j(\gamma\partial + \\
m^\lambda_u)u^\lambda_j - \bar d^\lambda_j(\gamma\partial + m^\lambda_d)d^\lambda_j + igs_wA_\mu\left(-(\bar e^\lambda\gamma^\mu e^\lambda) + \frac{2}{3}(\bar u^\lambda_j\gamma^\mu u^\lambda_j) - \frac{1}{3}(\bar d^\lambda_j\gamma^\mu d^\lambda_j)\right) + \\
\frac{ig}{4c_w}Z^0_\mu\{(\bar\nu^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar e^\lambda\gamma^\mu(4s_w^2-1-\gamma^5)e^\lambda) + (\bar d^\lambda_j\gamma^\mu(\tfrac{4}{3} s_w^2 - 1 - \gamma^5)d^\lambda_j) + \\
(\bar u^\lambda_j\gamma^\mu(1-\tfrac{8}{3} s_w^2+\gamma^5)u^\lambda_j)\} + \frac{ig}{2\sqrt{2}}W^+_\mu\left((\bar\nu^\lambda\gamma^\mu(1+\gamma^5)U^{lep}{}_{\lambda\kappa}e^\kappa) + (\bar u^\lambda_j\gamma^\mu(1+\gamma^5)C_{\lambda\kappa}d^\kappa_j)\right) + \\
\frac{ig}{2\sqrt{2}}W^-_\mu\left((\bar e^\kappa U^{lep\dagger}{}_{\kappa\lambda}\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar d^\kappa_j C^\dagger_{\kappa\lambda}\gamma^\mu(1+\gamma^5)u^\lambda_j)\right) + \\
\frac{ig}{2M\sqrt{2}}\phi^+\left(-m^\kappa_e(\bar\nu^\lambda U^{lep}{}_{\lambda\kappa}(1-\gamma^5)e^\kappa) + m^\lambda_\nu(\bar\nu^\lambda U^{lep}{}_{\lambda\kappa}(1+\gamma^5)e^\kappa\right) + \\
\frac{ig}{2M\sqrt{2}}\phi^-\left(m^\lambda_e(\bar e^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1+\gamma^5)\nu^\kappa) - m^\kappa_\nu(\bar e^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1-\gamma^5)\nu^\kappa\right) - \frac{g}{2}\frac{m^\lambda_\nu}{M}H(\bar\nu^\lambda\nu^\lambda) - \\
\frac{g}{2}\frac{m^\lambda_e}{M}H(\bar e^\lambda e^\lambda) + \frac{ig}{2}\frac{m^\lambda_\nu}{M}\phi^0(\bar\nu^\lambda\gamma^5\nu^\lambda) - \frac{ig}{2}\frac{m^\lambda_e}{M}\phi^0(\bar e^\lambda\gamma^5 e^\lambda) - \frac{1}{4}\bar\nu_\lambda\, M^R_{\lambda\kappa}\,(1-\gamma_5)\hat\nu_\kappa - \\
\frac{1}{4}\,\overline{\bar\nu_\lambda\, M^R_{\lambda\kappa}\,(1-\gamma_5)\hat\nu_\kappa} + \frac{ig}{2M\sqrt{2}}\phi^+\left(-m^\kappa_d(\bar u^\lambda_j C_{\lambda\kappa}(1-\gamma^5)d^\kappa_j) + m^\lambda_u(\bar u^\lambda_j C_{\lambda\kappa}(1+\gamma^5)d^\kappa_j\right) + \\
\frac{ig}{2M\sqrt{2}}\phi^-\left(m^\lambda_d(\bar d^\lambda_j C^\dagger_{\lambda\kappa}(1+\gamma^5)u^\kappa_j) - m^\kappa_u(\bar d^\lambda_j C^\dagger_{\lambda\kappa}(1-\gamma^5)u^\kappa_j\right) - \frac{g}{2}\frac{m^\lambda_u}{M}H(\bar u^\lambda_j u^\lambda_j) - \\
\frac{g}{2}\frac{m^\lambda_d}{M}H(\bar d^\lambda_j d^\lambda_j) + \frac{ig}{2}\frac{m^\lambda_u}{M}\phi^0(\bar u^\lambda_j\gamma^5u^\lambda_j) - \frac{ig}{2}\frac{m^\lambda_d}{M}\phi^0(\bar d^\lambda_j\gamma^5 d^\lambda_j) + \bar G^a\partial^2G^a + g_sf^{abc}\partial_\mu\bar G^aG^bg^c_\mu + \\
\bar X^+(\partial^2-M^2)X^+ + \bar X^-(\partial^2-M^2)X^- + \bar X^0(\partial^2-\frac{M^2}{c_w^2})X^0 + \bar Y\partial^2Y + igc_wW^+_\mu(\partial_\mu\bar X^0X^- - \\
\partial_\mu\bar X^+X^0) + igs_wW^+_\mu(\partial_\mu\bar YX^- - \partial_\mu\bar X^+Y) + igc_wW^-_\mu(\partial_\mu\bar X^-X^0 - \\
\partial_\mu\bar X^0X^+) + igs_wW^-_\mu(\partial_\mu\bar X^-Y - \partial_\mu\bar YX^+) + igc_wZ^0_\mu(\partial_\mu\bar X^+X^+ - \\
\partial_\mu\bar X^-X^-) + igs_wA_\mu(\partial_\mu\bar X^+X^+ - \\
\partial_\mu\bar X^-X^-) - \frac{1}{2} gM\left(\bar X^+X^+H + \bar X^-X^-H + \frac{1}{c_w^2}\bar X^0X^0H\right) + \frac{1-2c_w^2}{2c_w}igM\left(\bar X^+X^0\phi^+ - \bar X^-X^0\phi^-\right) + \\
\frac{1}{2c_w}igM\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + igMs_w\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + \\
\frac{1}{2} igM\left(\bar X^+X^+\phi^0 - \bar X^-X^-\phi^0\right) .
\end{gathered}
$$

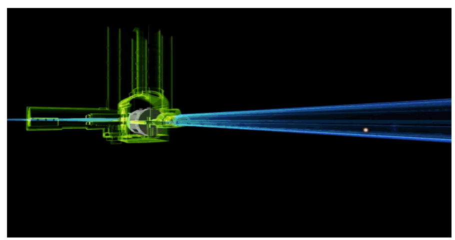



EA-01-2474

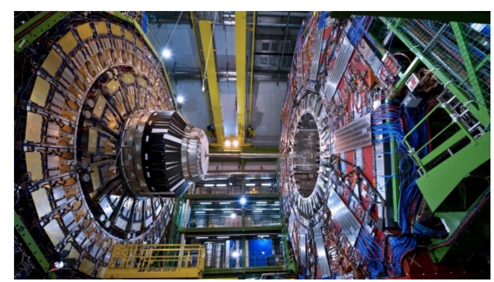


SM with parameters θ

Simulated observables x



Real observations x_{obs}



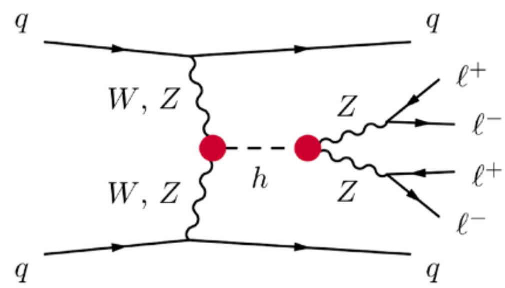
Latent variables

Parameters of interest

Parton-level momenta

Theory parameters

$$z_p \longleftarrow \theta$$

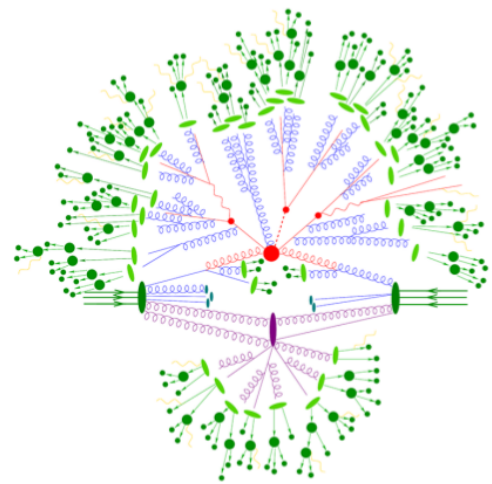


Latent variables

Parameters of interest

Shower splittings | Parton-level momenta | Theory parameters

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

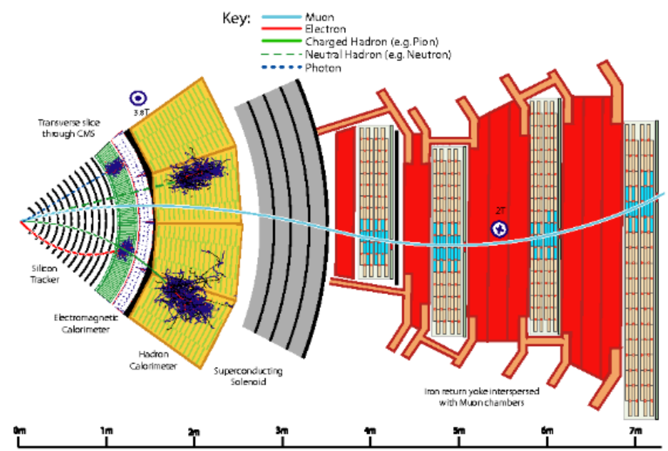


Latent variables

Parameters of interest

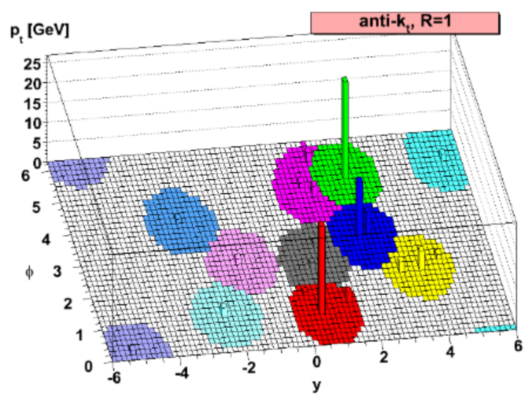
Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
z_d	z_s	z_p	θ

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



Features	Latent variables			Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
x	z_d	z_s	z_p	θ

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



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

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

Inference

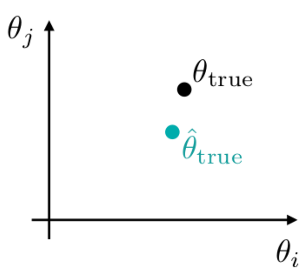
Problem statement(s)

Start with

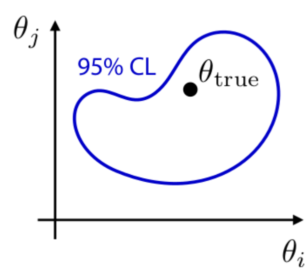
- a simulator that lets you generate N samples $x_i \sim p(x_i|\theta_i)$ (for parameters θ_i of our choice),
- observed data $x_{\text{obs}} \sim p(x_{\text{obs}}|\theta_{\text{true}})$,
- a prior $p(\theta)$.

Then,

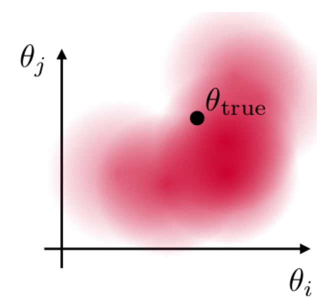
a) estimate θ_{true} (e.g., MLE)



b) construct confidence sets



c) estimate the posterior $p(\theta|x_{\text{obs}})$ (or sample from it)



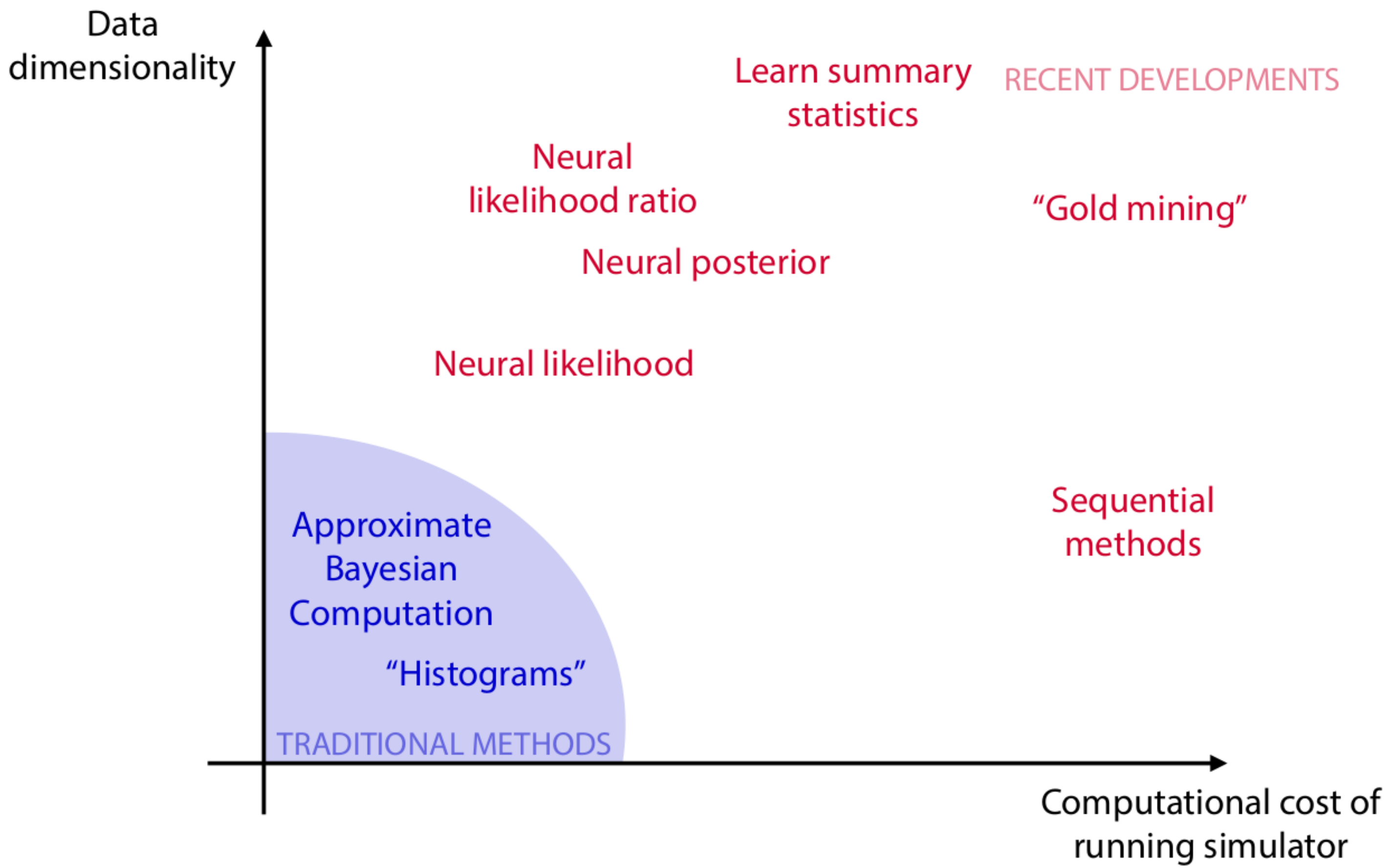
Ingredients

Statistical inference requires the computation of key ingredients, such as

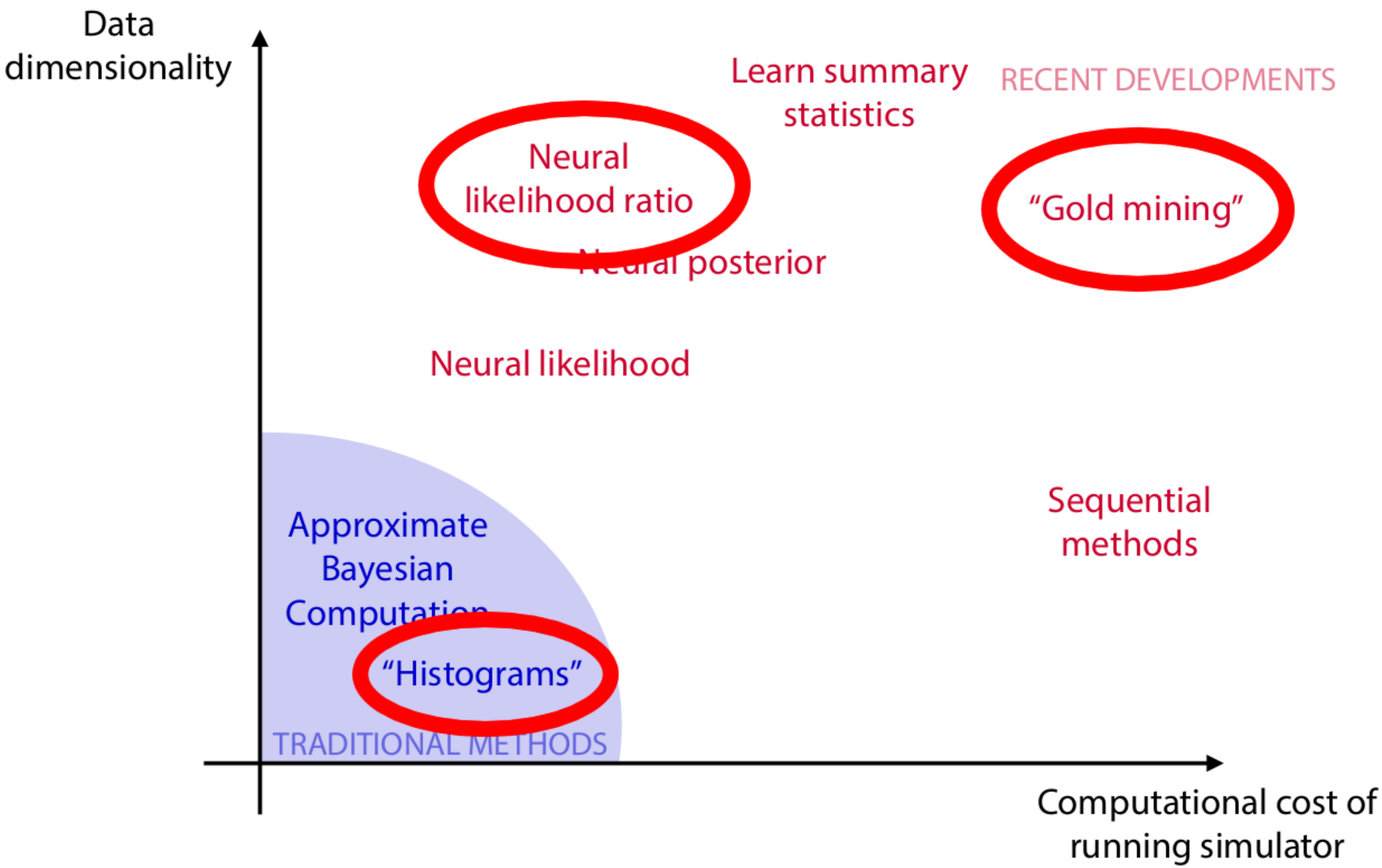
- the likelihood $p(x|\theta)$,
- the likelihood ratio $r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$,
- or the posterior $p(\theta|x)$,

but none are usually tractable in simulation-based science!

Inference algorithms



Inference algorithms

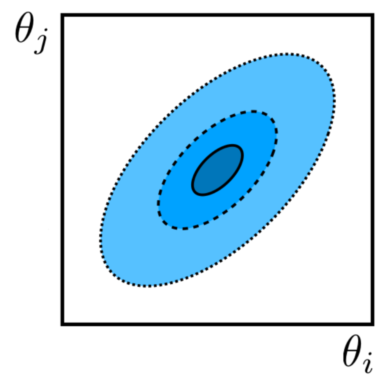


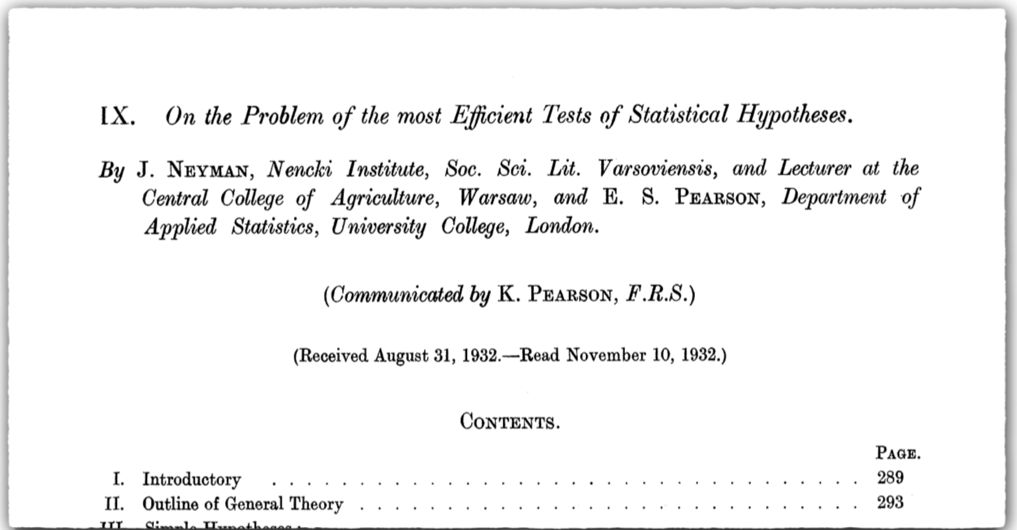
The frequentist (physicist's) way

The Neyman-Pearson lemma states that the likelihood ratio

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

is the most powerful test statistic to discriminate between a null hypothesis θ_0 and an alternative θ_1.





IX. *On the Problem of the most Efficient Tests of Statistical Hypotheses.*

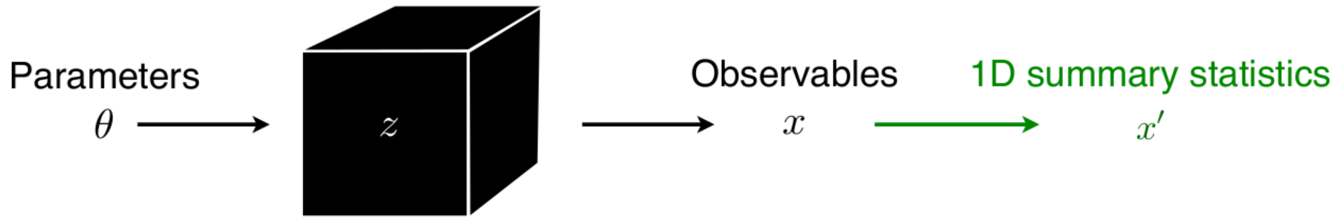
By J. Neyman, *Nencki Institute, Soc. Sci. Lit. Varsoviensis, and Lecturer at the Central College of Agriculture, Warsaw, and* E. S. Pearson, *Department of Applied Statistics, University College, London.*

(*Communicated by* K. Pearson, *F.R.S.*)

(Received August 31, 1932.—Read November 10, 1932.)

Contents.

		Page.
I.	Introductory	289
II.	Outline of General Theory	293

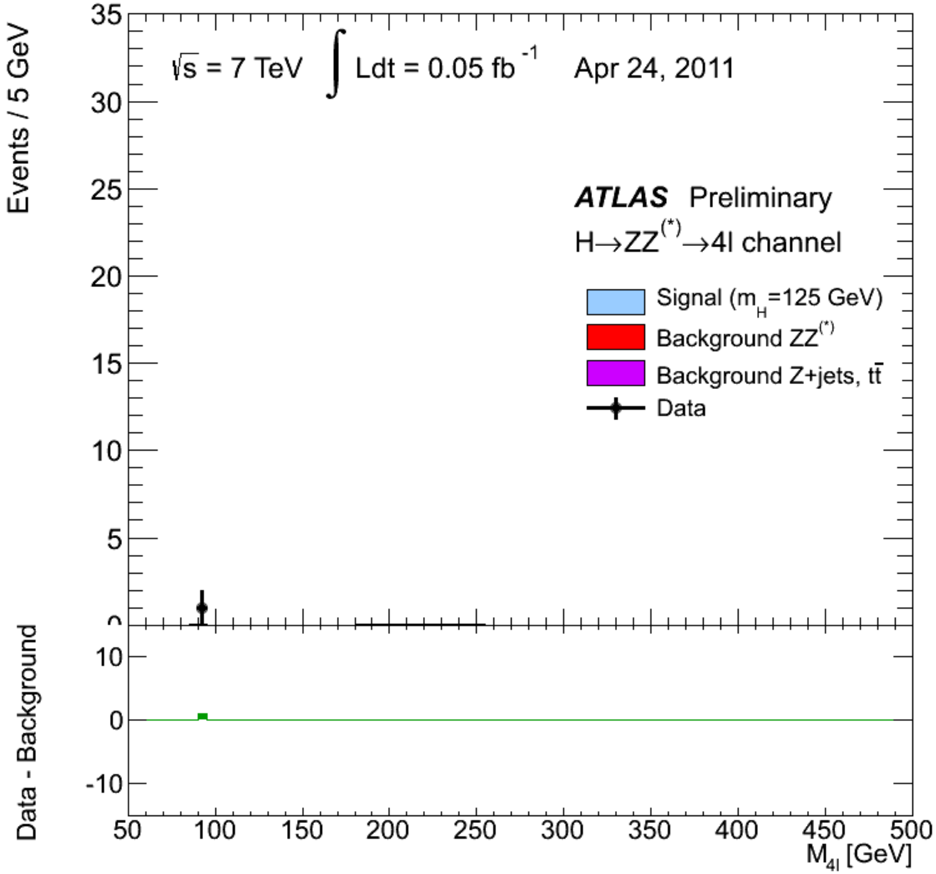


Define a projection function $s : \mathcal{X} \rightarrow \mathbb{R}$ mapping observables x to a summary statistic $x' = s(x)$.

Then, approximate the likelihood $p(x|\theta)$ with the surrogate $\hat{p}(x|\theta) = p(x'|\theta)$.

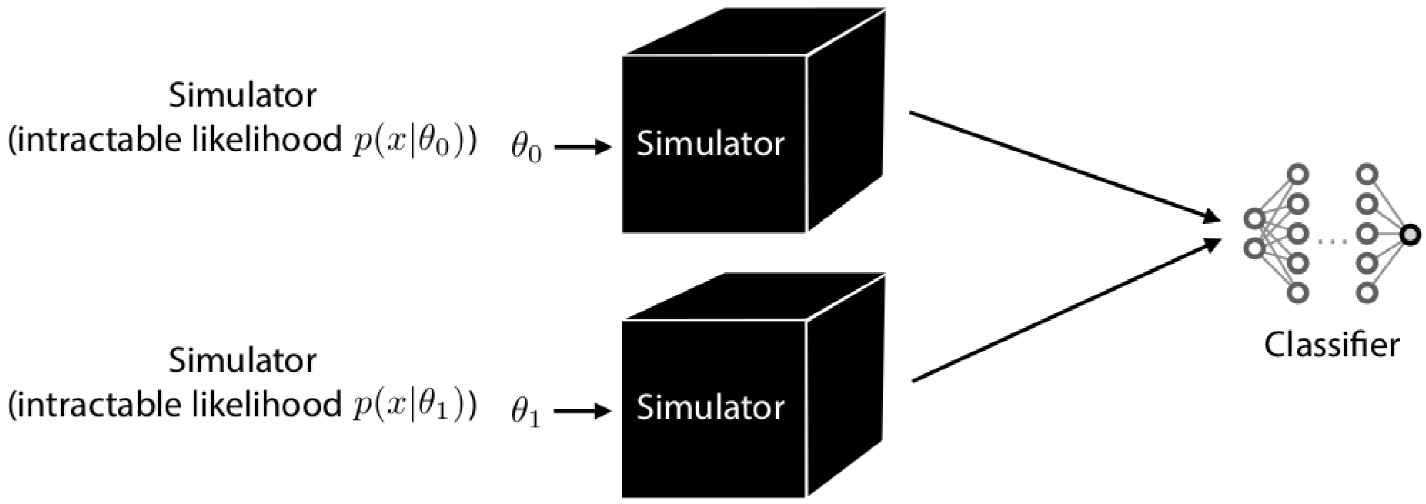
From this it comes

$$\frac{p(x|\theta_0)}{p(x|\theta_1)} \approx \frac{\hat{p}(x|\theta_0)}{\hat{p}(x|\theta_1)} = \hat{r}(x|\theta_0, \theta_1).$$



Discovery of the Higgs boson at 5-σ

The likelihood ratio trick



The solution $\hat{s}$ found after training approximates the optimal classifier

$$\hat{s}(x) \approx s^*(x) = \frac{p(x|\theta_1)}{p(x|\theta_0) + p(x|\theta_1)}.$$

Therefore,

$$r(x|\theta_0, \theta_1) \approx \hat{r}(x|\theta_0, \theta_1) = \frac{1 - \hat{s}(x)}{\hat{s}(x)}.$$

To avoid retraining a classifier $\hat{s}$ for every (θ_0, θ_1) pair, fix θ_1 to θ_{ref} and train a single parameterized classifier $\hat{s}(x|\theta_0, \theta_{\text{ref}})$ where θ_0 is also given as input.

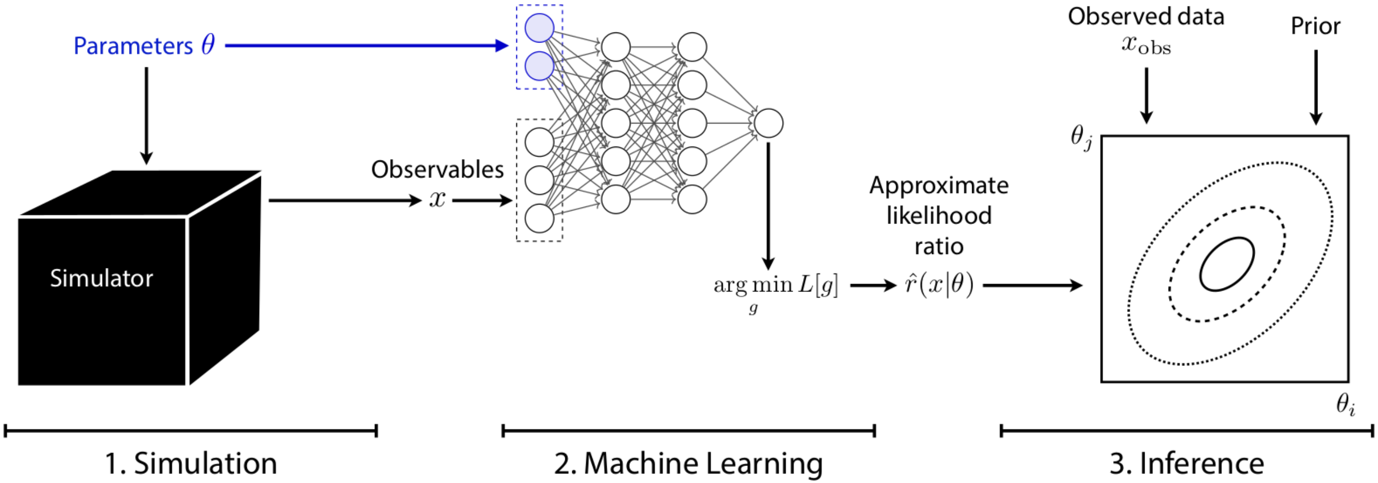
Therefore, we have

$$\hat{r}(x|\theta_0, \theta_{\text{ref}}) = \frac{1 - \hat{s}(x|\theta_0, \theta_{\text{ref}})}{\hat{s}(x|\theta_0, \theta_{\text{ref}})}$$

such that for any (θ_0, θ_1),

$$r(x|\theta_0, \theta_1) \approx \frac{\hat{r}(x|\theta_0, \theta_{\text{ref}})}{\hat{r}(x|\theta_1, \theta_{\text{ref}})}.$$

Inference



Run simulator and save data

Train NN classifier, interpret as likelihood ratio estimator

Amortized: cheap to repeat for new data

Amortizing Bayes

The Bayes rule can be rewritten as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = r(x|\theta)p(\theta) \approx \hat{r}(x|\theta)p(\theta),$$

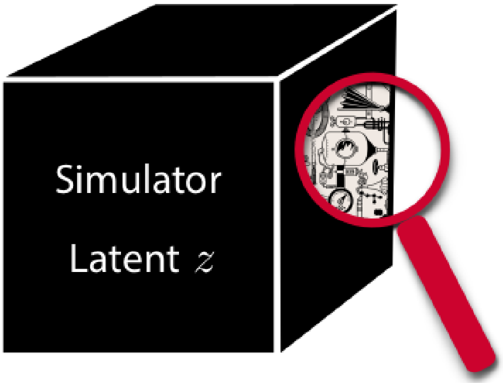
where $r(x|\theta) = \frac{p(x|\theta)}{p(x)}$ is the likelihood-to-evidence ratio.

As before, it can be approximated e.g. from a neural network classifier, but trained to distinguish $x, \theta \sim p(x, \theta)$ from $x, \theta \sim p(x)p(\theta)$, hence resulting in

$$\hat{r}(x|\theta) \approx \frac{p(x, \theta)}{p(x)p(\theta)} = \frac{p(x|\theta)}{p(x)}.$$

This enables direct and amortized posterior evaluation.

Gold mining



We cannot compute $p(x|\theta) = \int p(x,z|\theta)\mathrm{d}z$.

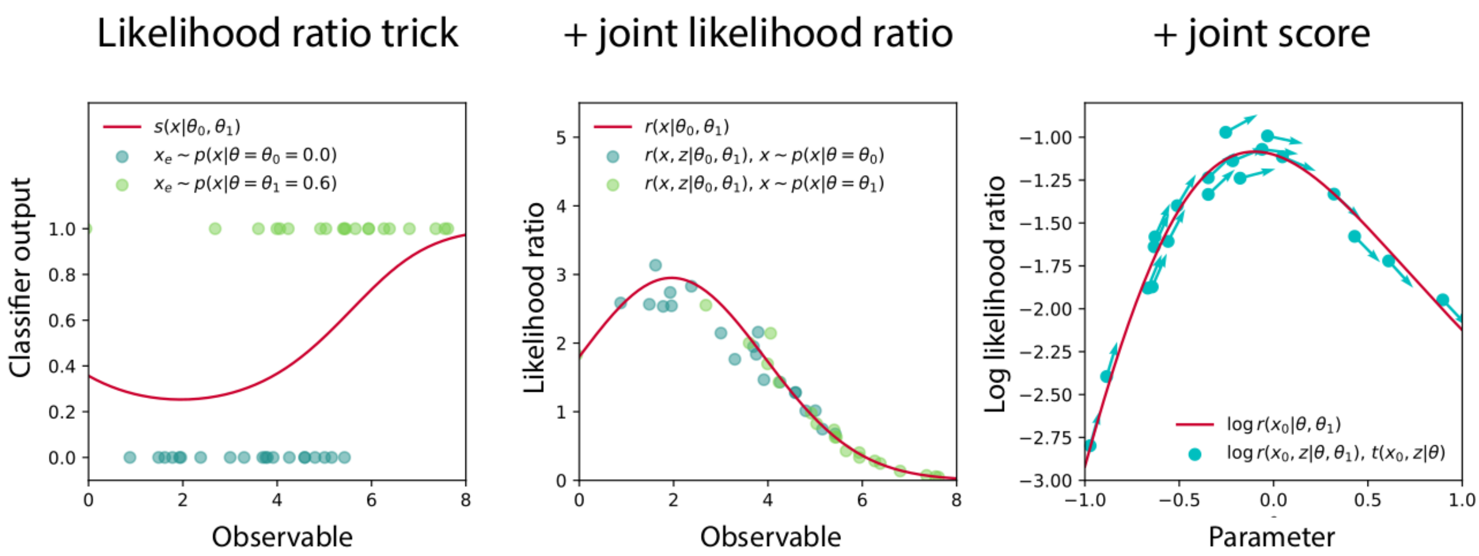
However, using techniques from probabilistic programming we can often extract

- the joint likelihood ratio $r(x,z|\theta) = \frac{p(x,z|\theta)}{p_{\mathrm{ref}}(x,z)}$
- the joint score $t(x,z|\theta) = \nabla_\theta \log p(x,z|\theta)$.

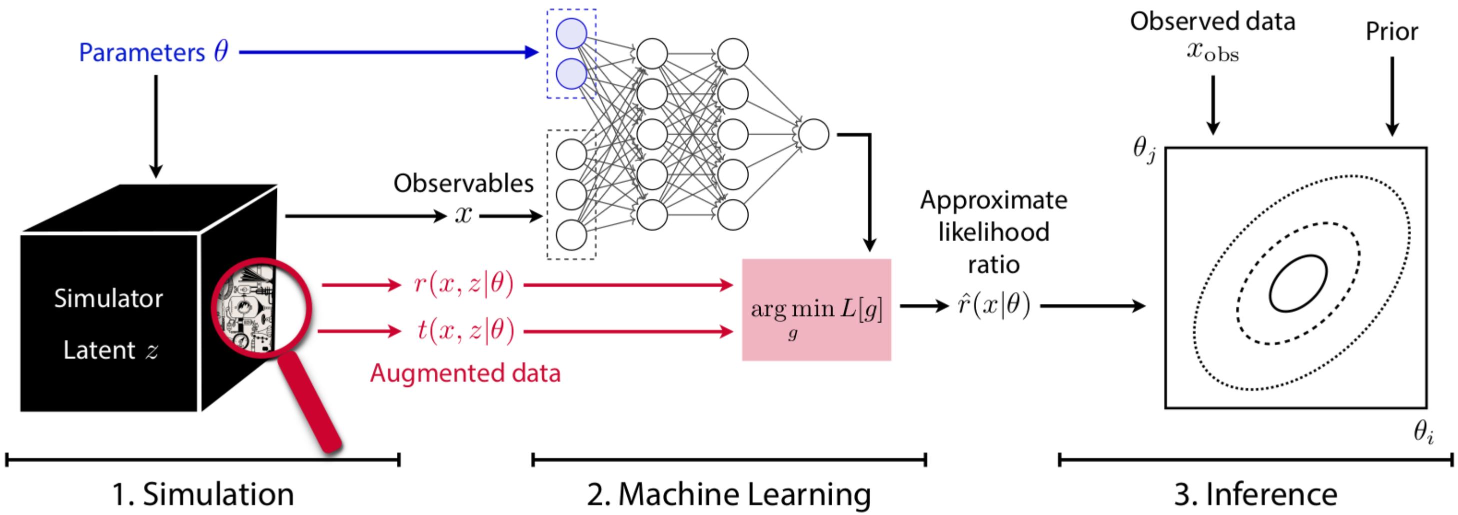
This is interesting because

- the joint likelihood ratio is an unbiased estimator of the likelihood ratio,
- the joint score provides unbiased gradient information

$\Rightarrow$ use them as labels in supervised NN training!



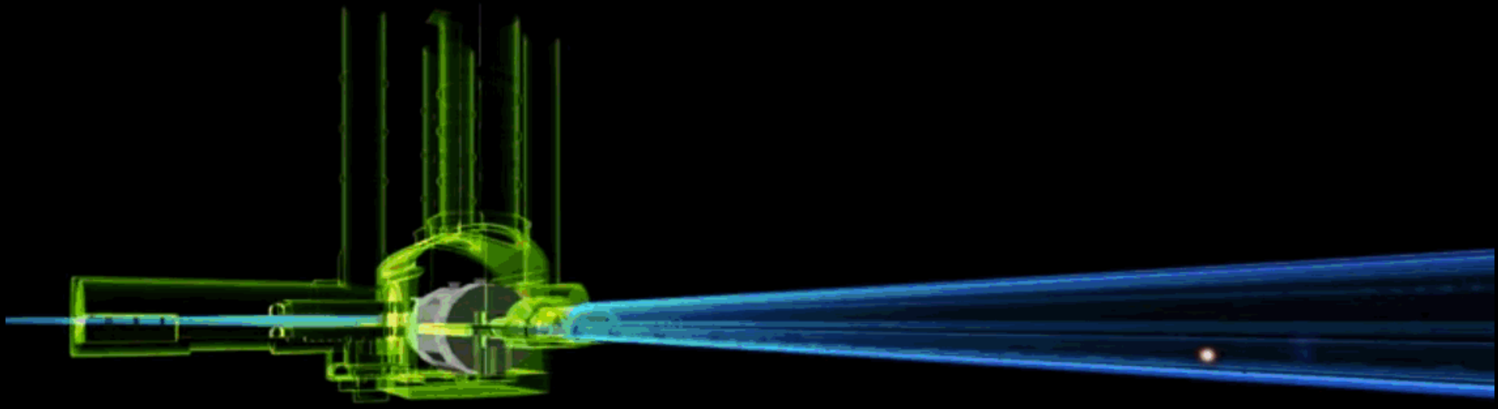
RASCAL



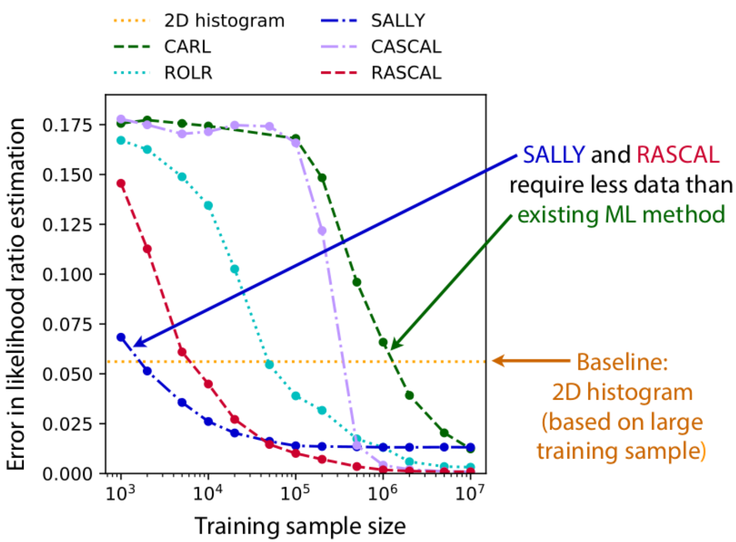
Extract joint likelihood ratio
and joint score from simulator

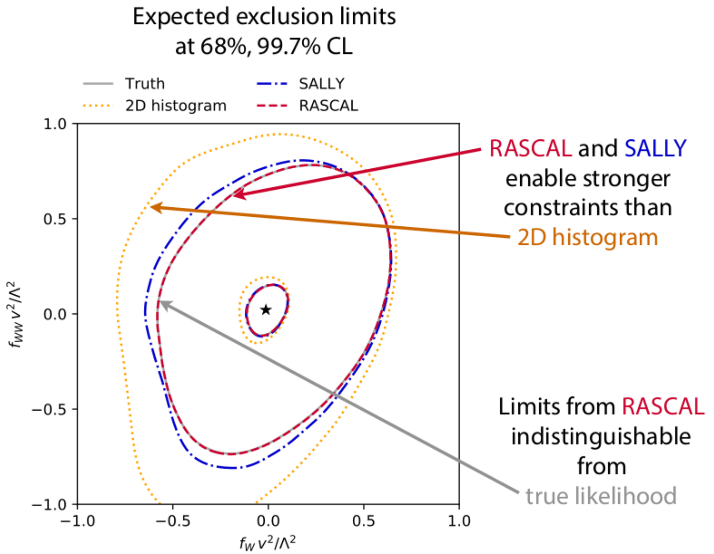
Augment training data &
use as labels in new loss functions
⇒ improve training efficiency

Showtime!



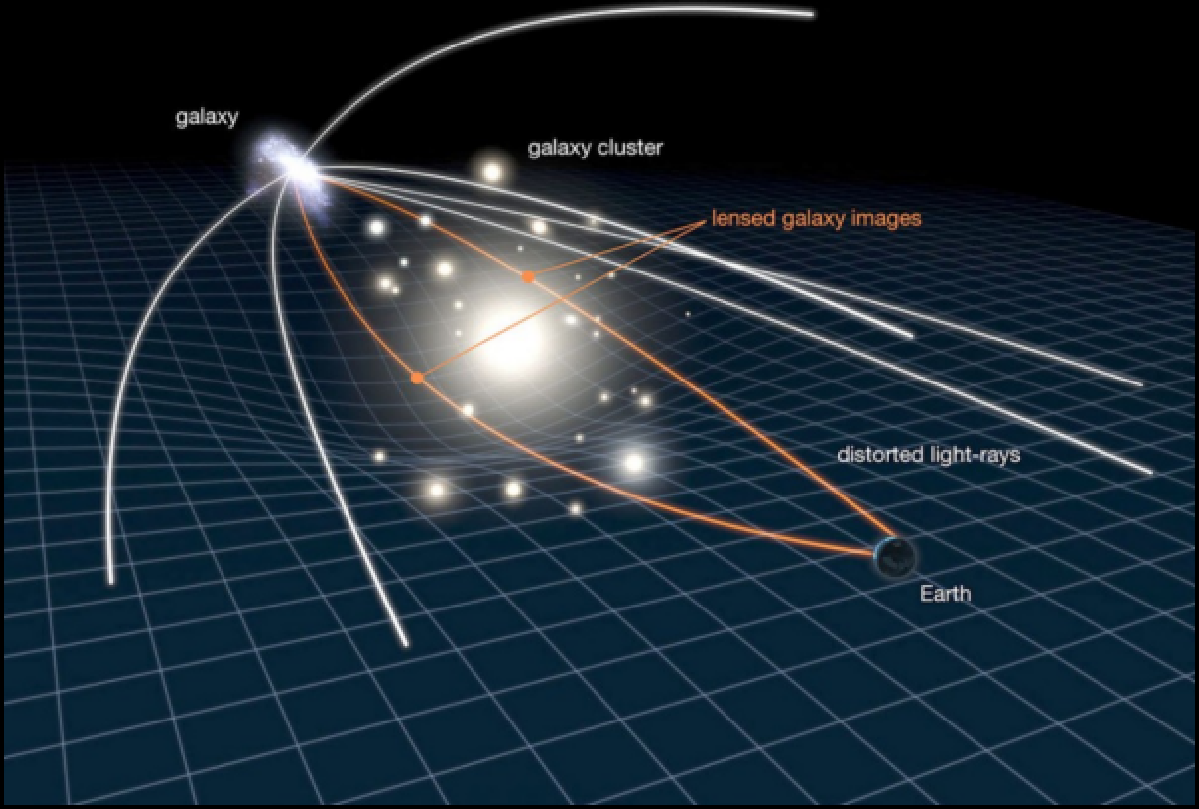
Case 1: Hunting new physics at particle colliders



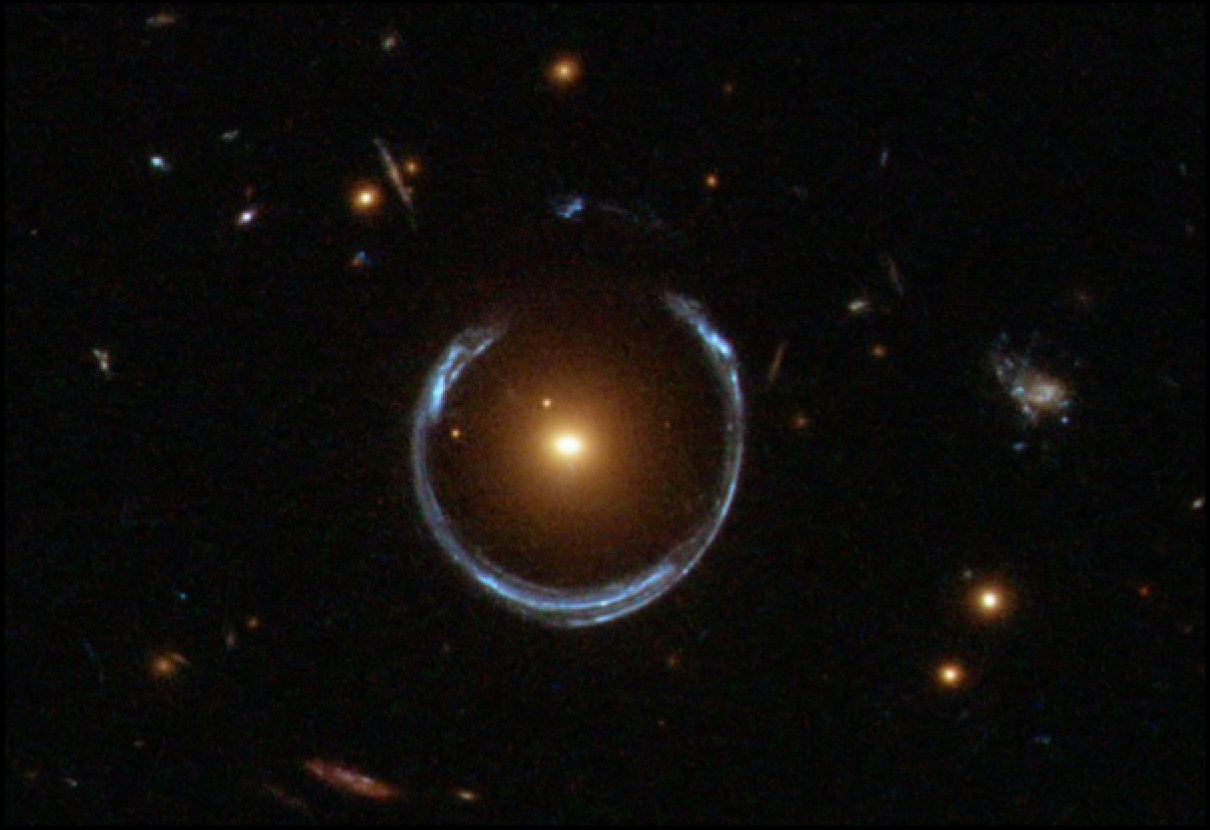


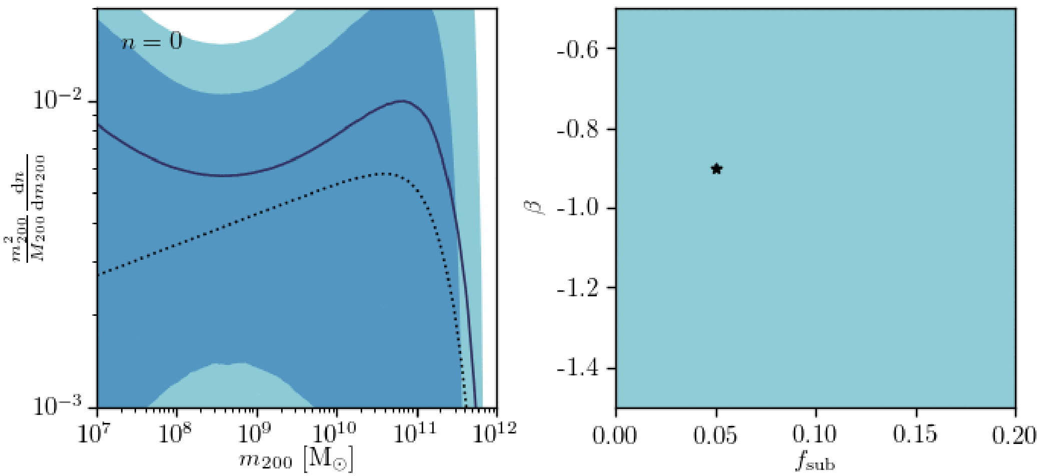
With enough training data, the ML algorithms get the likelihood function right.

Using more information from the simulator improves sample efficiency substantially.

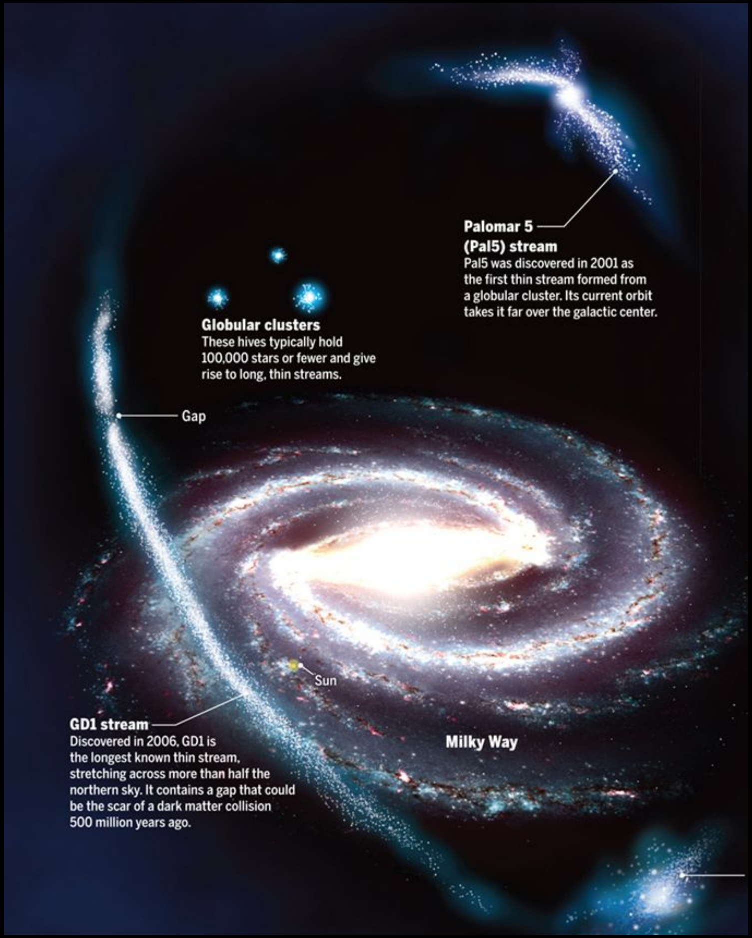


Case 2: Dark matter substructure from gravitational lensing





$n = 0$
10^{-2}
10^{-3}
$\frac{m_{200}^2}{M_{200}} \frac{dn}{dm_{200}}$
10^7 10^8 10^9 10^{10} 10^{11} 10^{12}
m_{200} [$M_\odot$]
-0.6
-0.8
-1.0
-1.2
-1.4
β
0.00 0.05 0.10 0.15 0.20
$f_{\rm sub}$



Case 3: Constraining dark matter with stellar streams

—
Image credits: C. Bickel/Science.

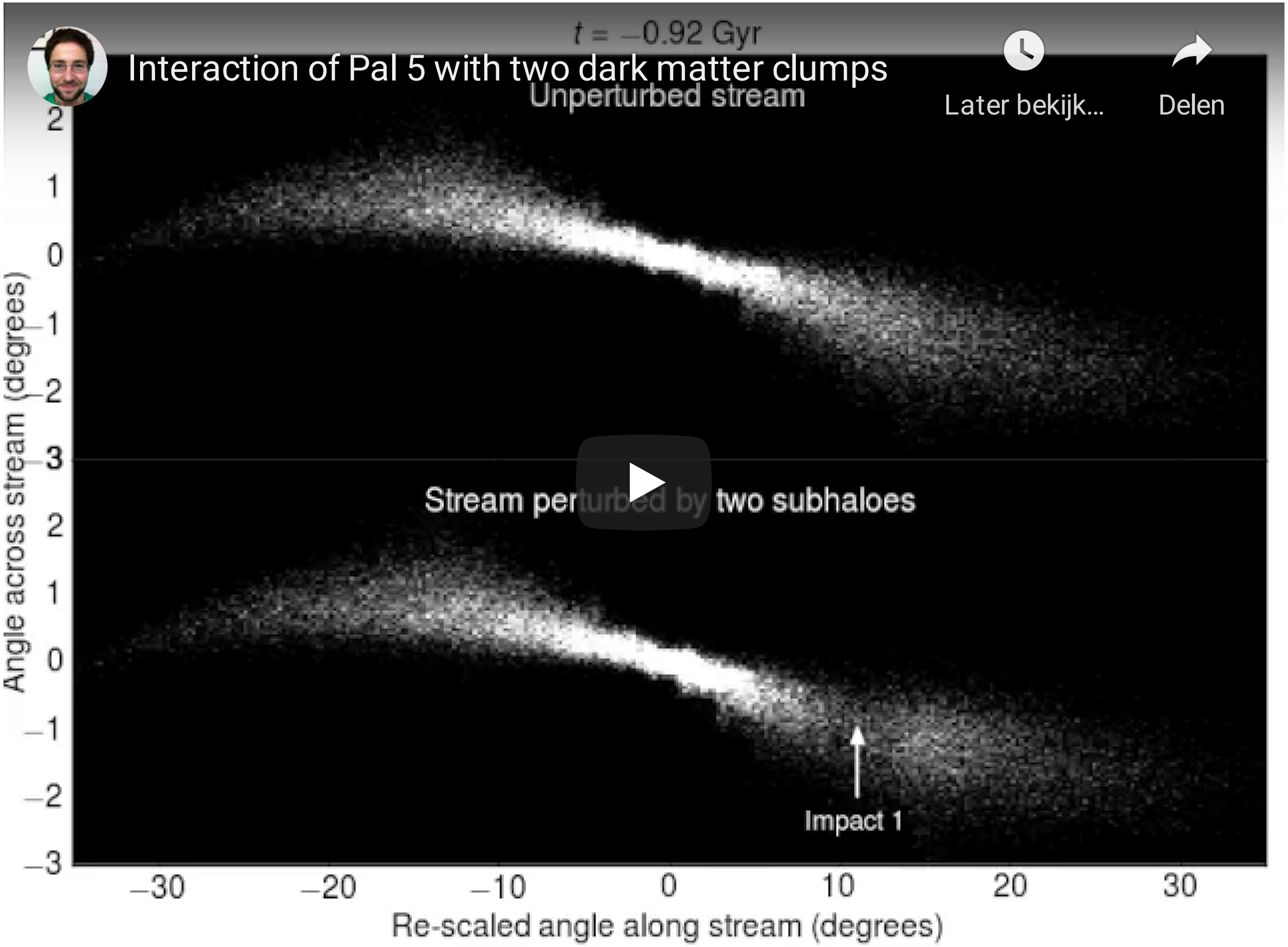


Image credits: D. Erkal.

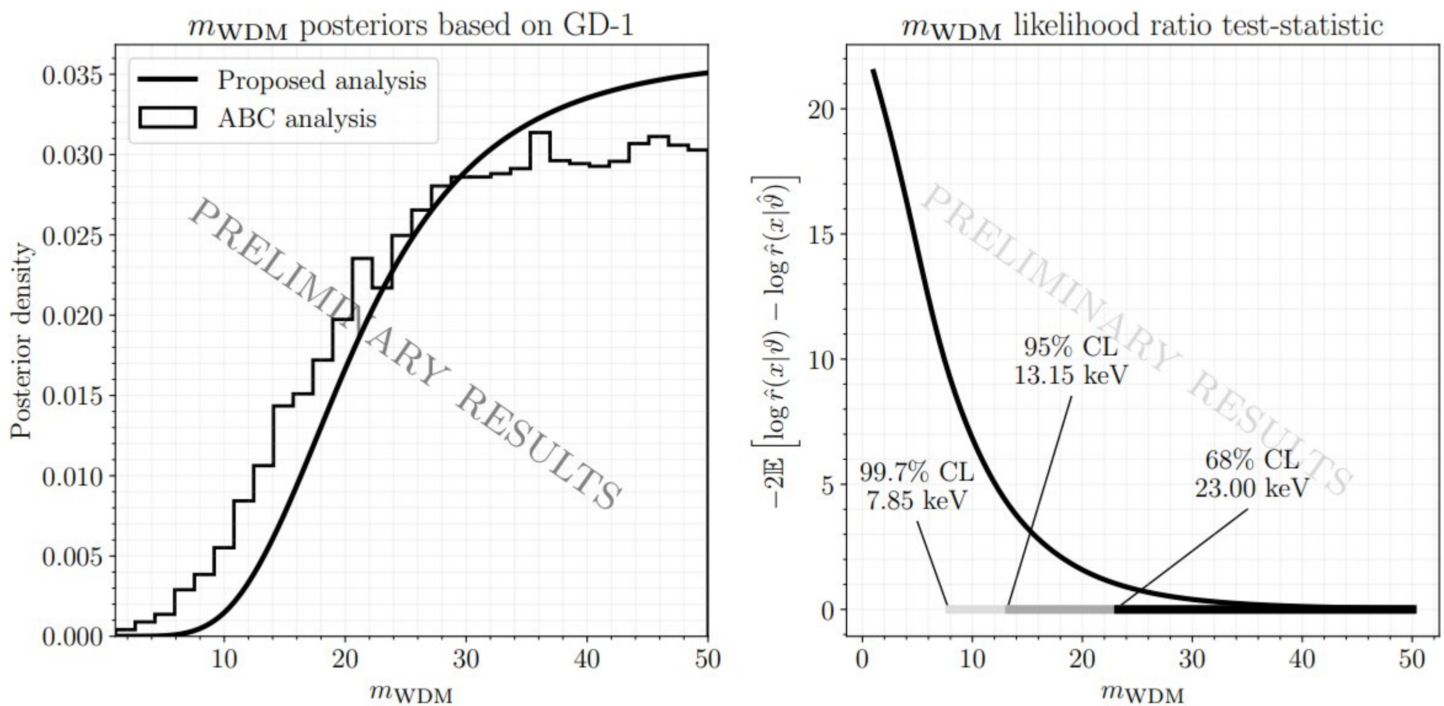
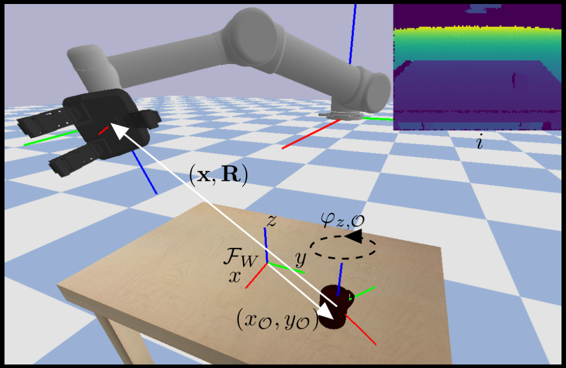
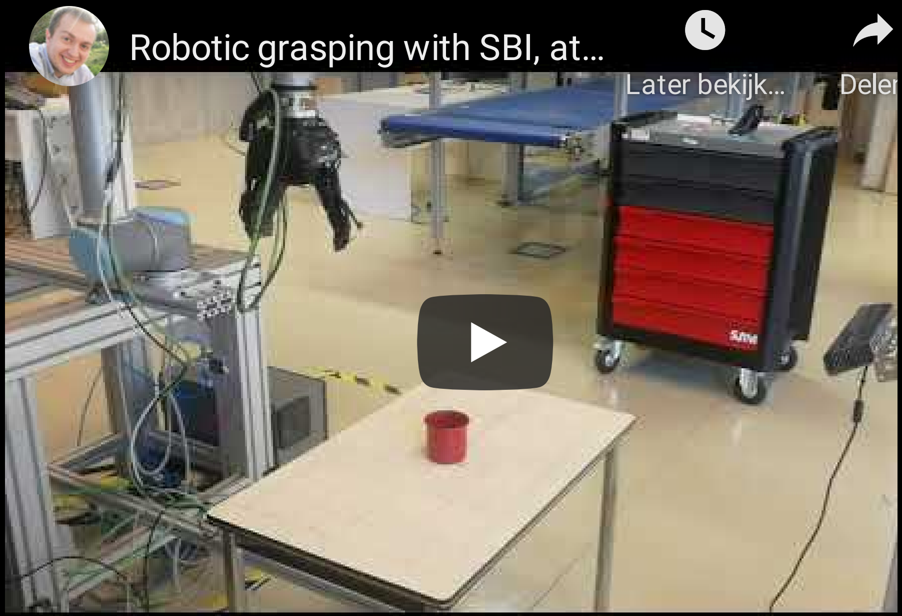


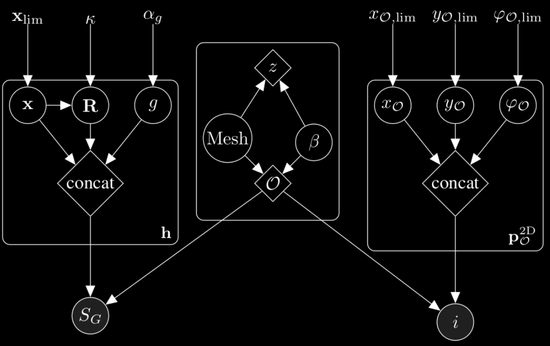
Figure 6. Age-marginalized results based on the observed stellar density variations of GD-1. *The results shown here illustrate the power of the proposed methodology, but should be considered as preliminary, since e.g. baryonic effects are not yet fully included in the simulation model. (Left)* Direct comparison of the reference ABC and the proposed analysis. Both posteriors indicate a preference for CDM over WDM within the assumed simulation model. We find that the proposed method is able to put stronger constraints on $m_{\rm WDM}$. *(Right)* Likelihood ratio test-statistic used to derive the lower limit confidence intervals. </>

Preliminary results for GD-1 suggest a preference for CDM over WDM.

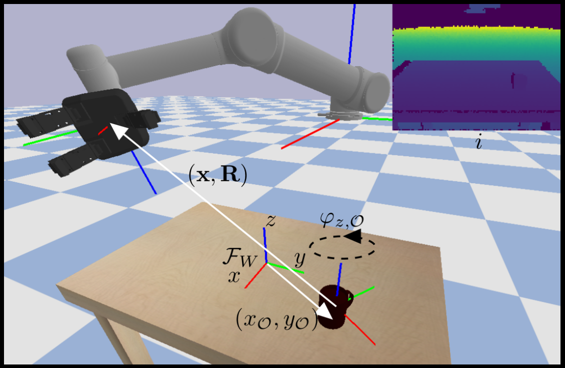
Case 4: Robotic grasping

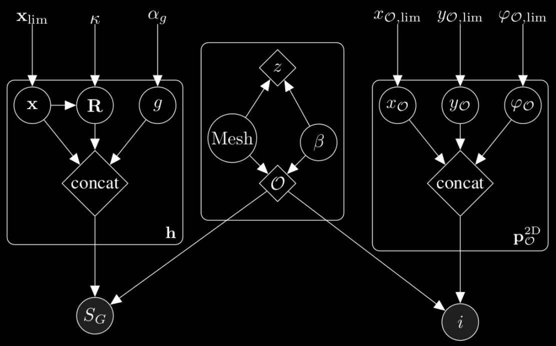


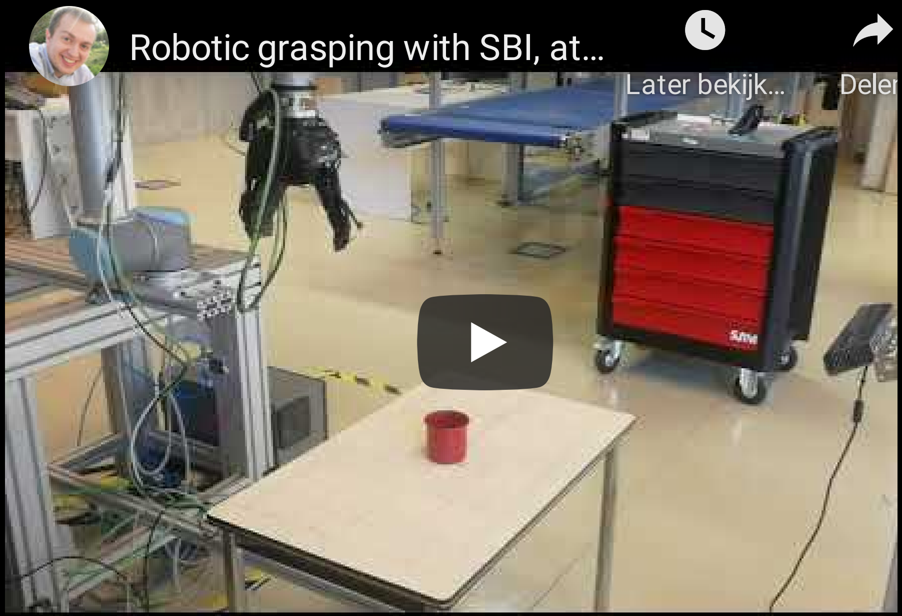




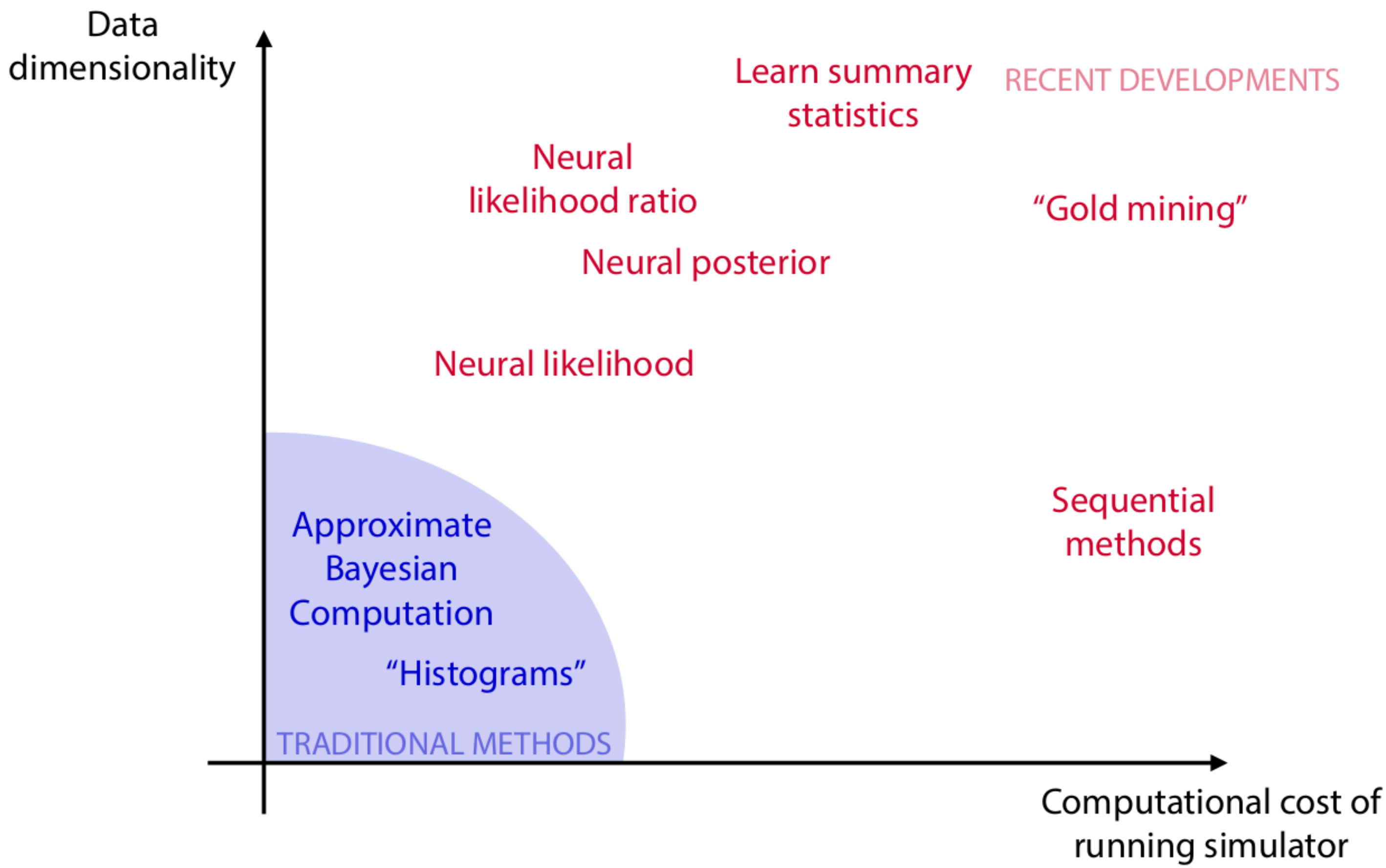
Case 5: Inference in hierarchial m

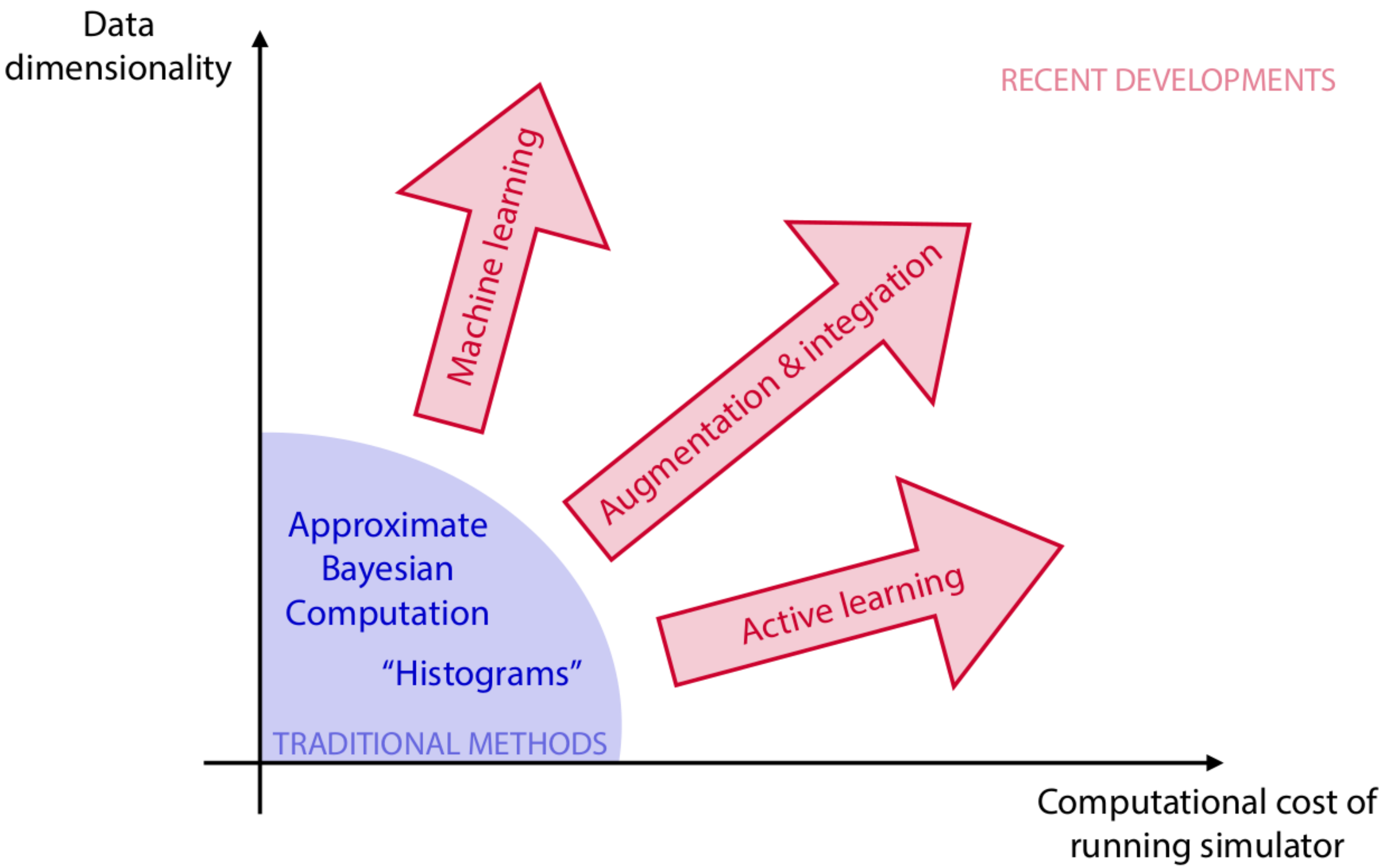






The frontier


Data dimensionality
Learn summary statistics
RECENT DEVELOPMENTS
Neural likelihood ratio
"Gold mining"
Neural posterior
Neural likelihood
Sequential methods
Approximate Bayesian Computation
"Histograms"
TRADITIONAL METHODS
Computational cost of running simulator


Data
dimensionality
RECENT DEVELOPMENTS
Machine learning
Augmentation & integration
Active learning
Approximate
Bayesian
Computation
"Histograms"
TRADITIONAL METHODS
Computational cost of
running simulator

COLLOQUIUM PAPER

The frontier of simulation-based inference

Kyle Cranmer[a,b,1], Johann Brehmer[a,b], and Gilles Louppe[c]

[a]Center for Cosmology and Particle Physics, New York University, New York, NY 10003; [b]Center for Data Science, New York University, New York, NY 10011; and [c]Montefiore Institute, University of Liège, B-4000 Liège, Belgium

Edited by Jitendra Malik, University of California, Berkeley, CA, and approved April 10, 2020 (received for review November 4, 2019)

Many domains of science have developed complex simulations to describe phenomena of interest. While these simulations provide high-fidelity models, they are poorly suited for inference and lead to challenging inverse problems. We review the rapidly developing field of simulation-based inference and identify the forces giving additional momentum to the field. Finally, we describe how the frontier is expanding so that a broad audience can appreciate the profound influence these developments may have on science.

statistical inference | implicit models | likelihood-free inference | approximate Bayesian computation | neural density estimation

Mechanistic models can be used to predict how systems will behave in a variety of circumstances. These run the gamut of distance scales, with notable examples including particle physics, molecular dynamics, protein folding, population genetics, neuroscience, epidemiology, economics, ecology, climate science, astrophysics, and cosmology. The expressiveness of programming languages facilitates the development of complex, high-fidelity simulations and the power of modern computing provides the ability to generate synthetic data from them. Unfortunately, these simulators are poorly suited for statistical inference. The source of the challenge is that the probability density (or likelihood) for a given observation—an essential ingredient for both frequentist and Bayesian inference methods—is typically intractable. Such models are often referred to as implicit models and contrasted against prescribed models where the likelihood for an observation can be explicitly calculated (1). The problem setting of statistical inference under intractable likelihoods has been dubbed likelihood-free inference—although it is a bit of a misnomer as typically one attempts to estimate the intractable likelihood, so we feel the term simulation-based inference is more apt.

The intractability of the likelihood is an obstruction for scientific progress as statistical inference is a key component of the scientific method. In areas where this obstruction has appeared,

the simulator—is being recognized as a key idea to improve the sample efficiency of various inference methods. A third direction of research has stopped treating the simulator as a black box and focused on integrations that allow the inference engine to tap into the internal details of the simulator directly.

Amidst this ongoing revolution, the landscape of simulation-based inference is changing rapidly. In this review we aim to provide the reader with a high-level overview of the basic ideas behind both old and new inference techniques. Rather than discussing the algorithms in technical detail, we focus on the current frontiers of research and comment on some ongoing developments that we deem particularly exciting.

Simulation-Based Inference

Simulators. Statistical inference is performed within the context of a statistical model, and in simulation-based inference the simulator itself defines the statistical model. For the purpose of this paper, a simulator is a computer program that takes as input a vector of parameters θ, samples a series of internal states or latent variables $z_i \sim p_i(z_i|\theta, z_{<i})$, and finally produces a data vector $x \sim p(x|\theta, z)$ as output. Programs that involve random samplings and are interpreted as statistical models are known as probabilistic programs, and simulators are an example. Within this general formulation, real-life simulators can vary substantially:

- The parameters θ describe the underlying mechanistic model and thus affect the transition probabilities $p_i(z_i|\theta, z_{<i})$. Typically the mechanistic model is interpretable by a domain scientist and θ has relatively few components and a fixed dimensionality. Examples include coefficients found in the Hamiltonian of a physical system, the virulence and incubation rate of a pathogen, or fundamental constants of Nature.
- The latent variables z that appear in the data-generating process may directly or indirectly correspond to a physically meaningful state of a system, but typically this state is unobservable in practice. The structure of the latent space varies substantially between simulators. The latent variables may be continuous

STATISTICS

In summary

- Much of modern science is based on simulators making precise predictions, but in which inference is challenging.
- Machine learning enables powerful inference methods.
- They work in problems from the smallest to the largest scales.
- Further advances in machine learning will translate into scientific progress.

Thanks!







References

- Hermans, J., Banik, N., Weniger, C., Bertone, G., & Louppe, G. (2020). Towards constraining warm dark matter with stellar streams through neural simulation-based inference. arXiv preprint arXiv:2011.14923.
- Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. Proceedings of the National Academy of Sciences, 117(48), 30055-30062.
- Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., Cranmer, K. (2019). Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning. arXiv preprint arXiv 1909.02005.
- Hermans, J., Begy, V., & Louppe, G. (2019). Likelihood-free MCMC with Approximate Likelihood Ratios. arXiv preprint arXiv:1903.04057.
- 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.