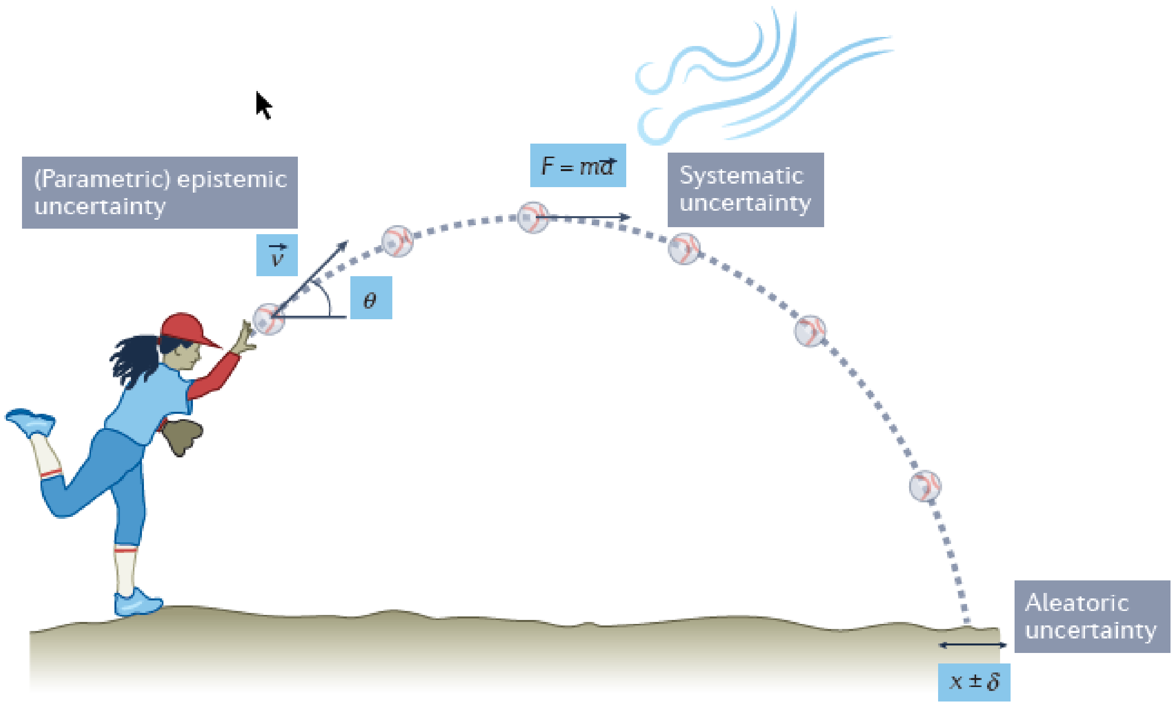
Simulation-based inference for the physical sciences

Grenoble Artificial Intelligence for Physical Sciences
May 29, 2024

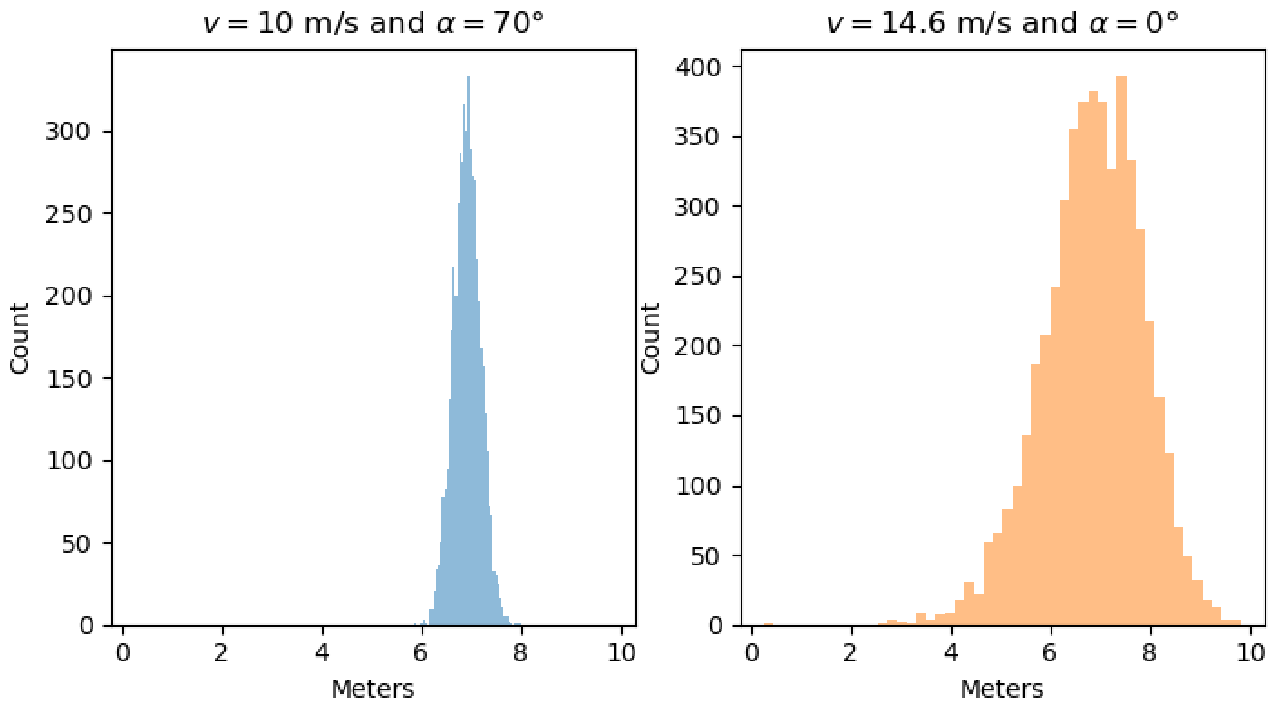
Gilles Louppe
g.louppe@uliege.be


(Parametric) epistemic uncertainty
$\vec{v}$
θ
$F = m\vec{a}$
Systematic uncertainty
Aleatoric uncertainty
$x \pm \delta$

$$v_x = v\cos(\alpha), \quad v_y = v\sin(\alpha),$$

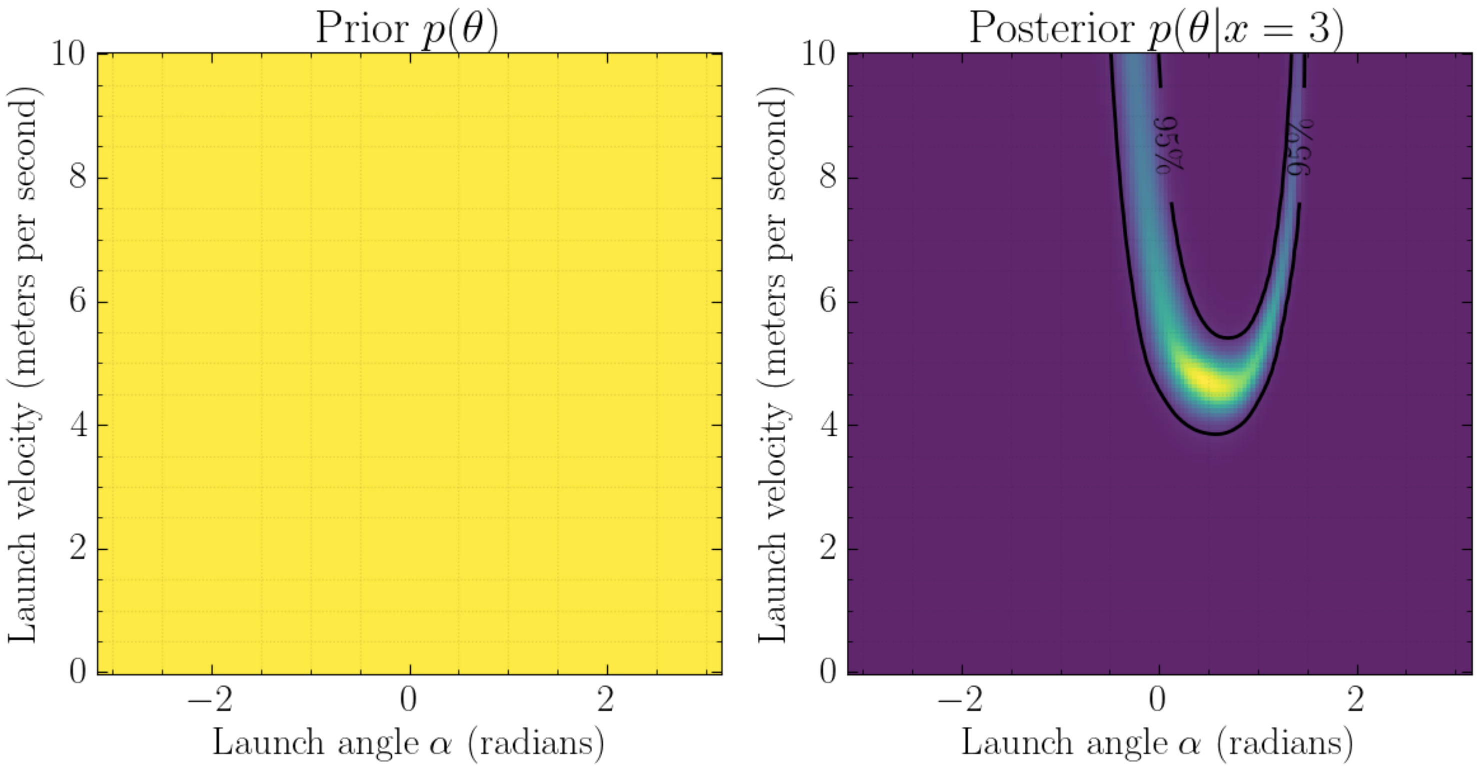
$$\frac{dx}{dt} = v_x\,, \quad \frac{dy}{dt} = v_y, \frac{dv_y}{dt} = -G.$$

```python
def simulate(v, alpha, dt=0.001):
    v_x = v * np.cos(alpha)  # x velocity m/s
    v_y = v * np.sin(alpha)  # y velocity m/s
    y = 1.1 + 0.3 * random.normal()
    x = 0.0

    while y > 0: # simulate until ball hits floor
        v_y += dt * -G  # acceleration due to gravity
        x += dt * v_x
        y += dt * v_y

    return x + 0.25 * random.normal()
```



$v = 10$ m/s and $\alpha = 70°$
Count
300
250
200
150
100
50
0
0
2
4
6
8
10
Meters
$v = 14.6$ m/s and $\alpha = 0°$
400
350
300
250
200
150
100
50
0
Count
0
2
4
6
8
10
Meters

What parameter values θ are the most plausible?


Prior $p(\theta)$
Posterior $p(\theta|x = 3)$
10
8
6
4
2
0
−2
0
2
Launch velocity (meters per second)
Launch angle α (radians)
95%
95%

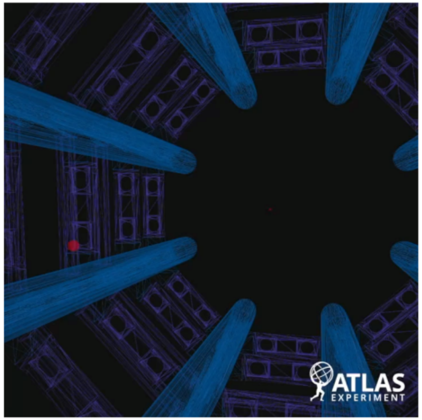
Simulation-based inference

Simulators as generative models

A simulator prescribes a generative model that can be used to simulate data $\mathbf{x}$.

Collider data

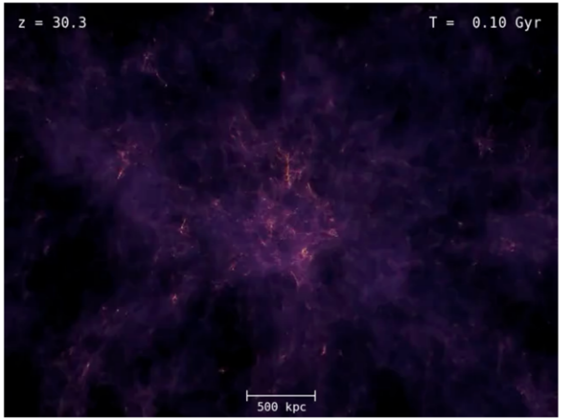
$\text{particles} \sim p(\text{particles})$



[C. Cesarotti with ATLAS]

Cosmology data

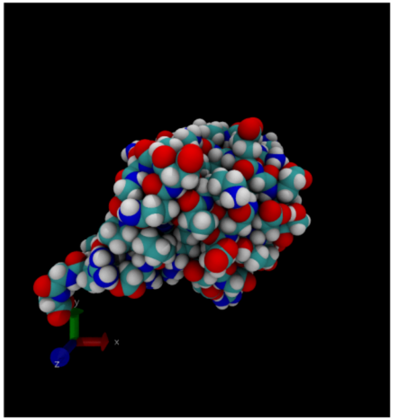
$\text{particles} \sim p(\text{particles})$



[Aquarius simulation]

Molecular dynamics

$\text{configurations} \sim p(\text{configurations})$



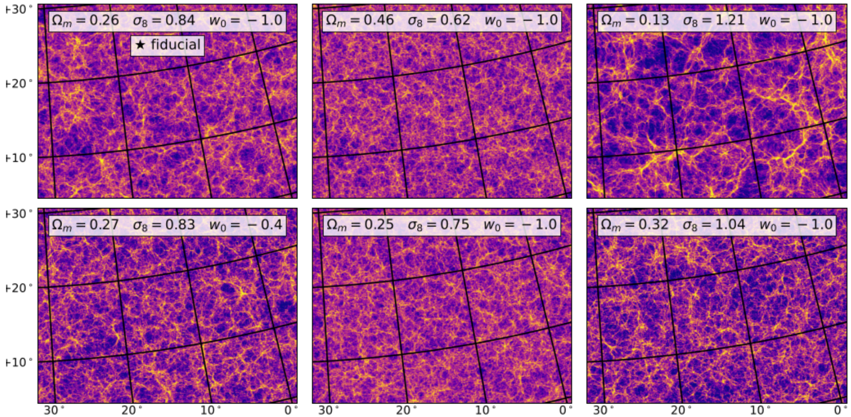
[E. Cances et al]

Conditional simulators

A conditional simulator prescribes a way to sample from the likelihood $p(\mathbf{x}|\theta)$, where θ is a set of conditioning variables or parameters.

Cosmology data

$$\text{map} \sim p(\text{map} \mid \{\Omega_m, \sigma_8, w_0\})$$



[Kacprzak et al 2022]

$$x \sim p(x; \mathscr{M})$$

Model

or

$$x \sim p(x \mid \theta)$$

Model parameters

What can we do with generative models?

Produce samples and make predictions

$$\mathbf{x} \sim p(\mathbf{x}|\theta)$$

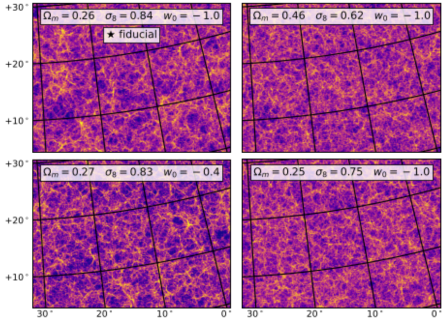
Evaluate densities

$$p(\mathbf{x}|\theta)$$

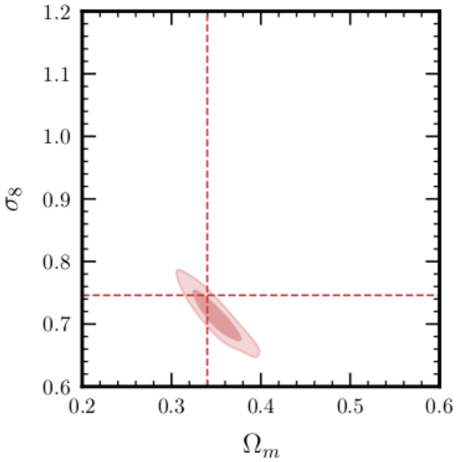
$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})}$$

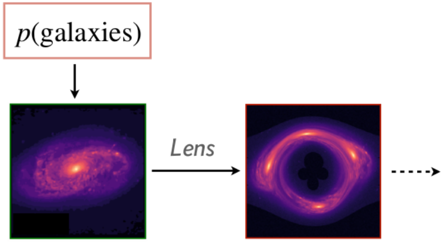
Encode complex priors

$$p(\mathbf{x})$$



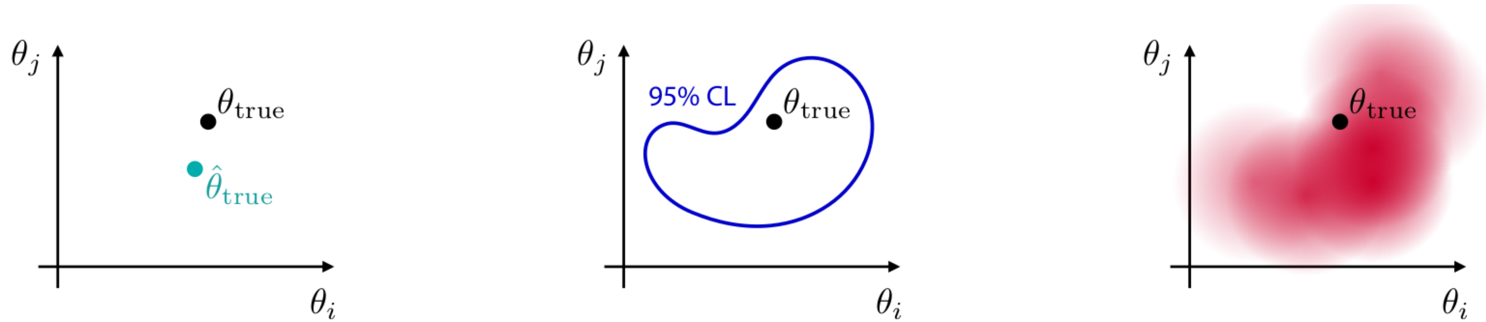
[Kacprzak et al 2022]





—
Credits: Siddharth Mishra-Sharma, 2023.

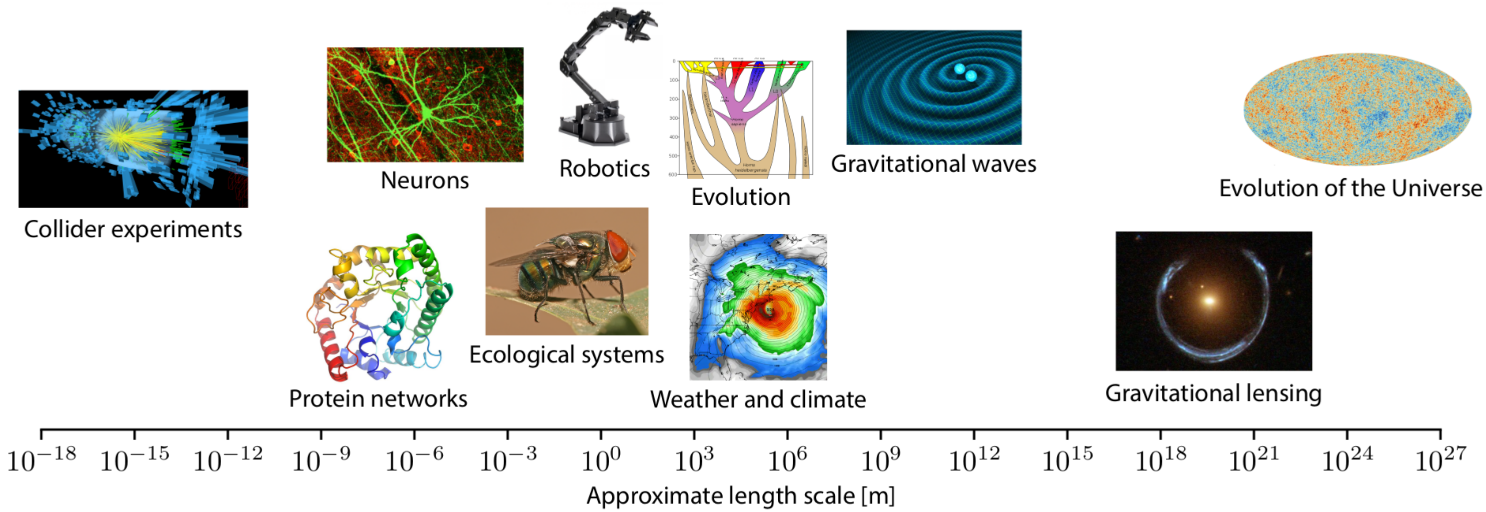
Inference



- Frequentist inference: find the parameters $\hat{\theta}$ that maximizes the likelihood $p(\mathbf{x}|\theta)$ or build a confidence interval thereof.
- Bayesian inference: compute the posterior distribution

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})}$$

of the parameters θ given the data $\mathbf{x}$.



Examples of inference problems across the physical sciences

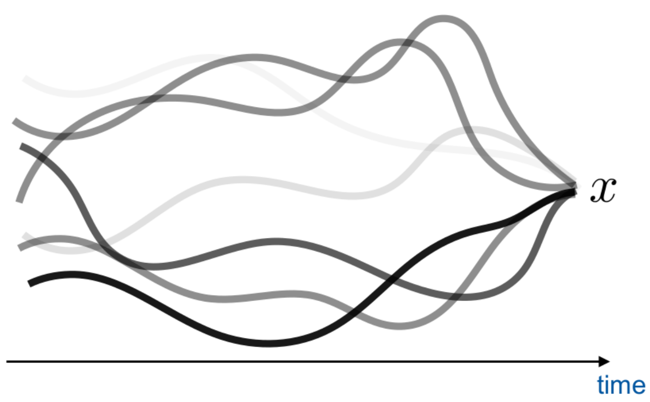
- Discovering new particles in high-energy physics
- Data assimilation in weather forecasting
- Estimating gravitational wave parameters
- Retrieving atmospheric properties of exoplanets
- Constraining cosmological models from galaxy surveys

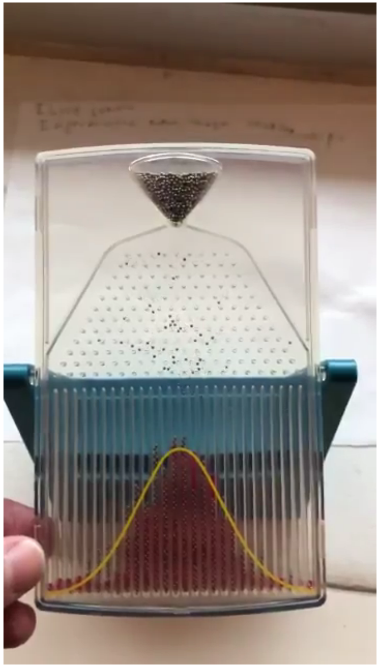
Intractable likelihoods

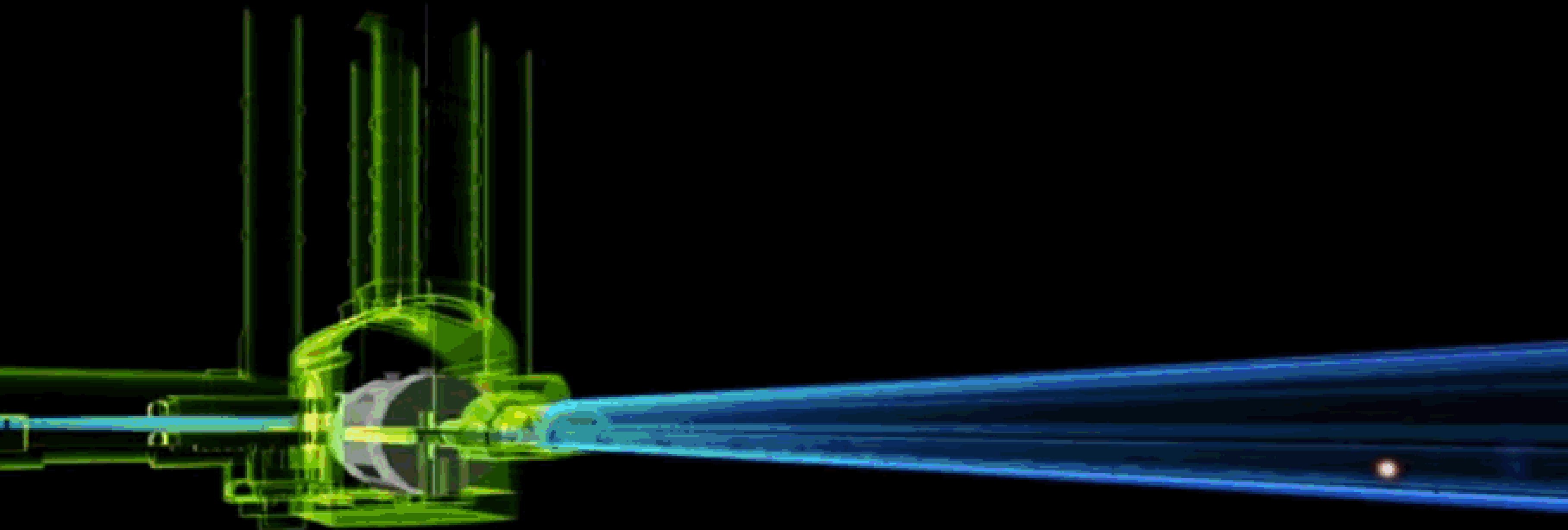
The (modeled) data generating process may involve additional latent variables $\mathbf{z}$ that are not observed, leading to likelihoods

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

In this case, evaluating the likelihood becomes intractable.







$$p(\mathbf{z}_p|\theta)$$

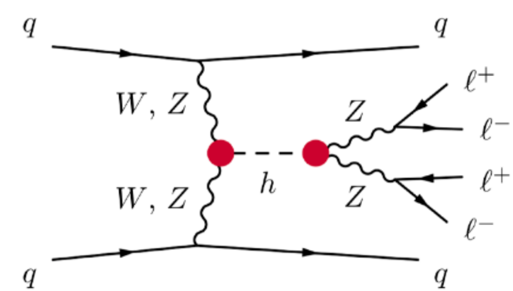
Latent variables

Parameters of interest

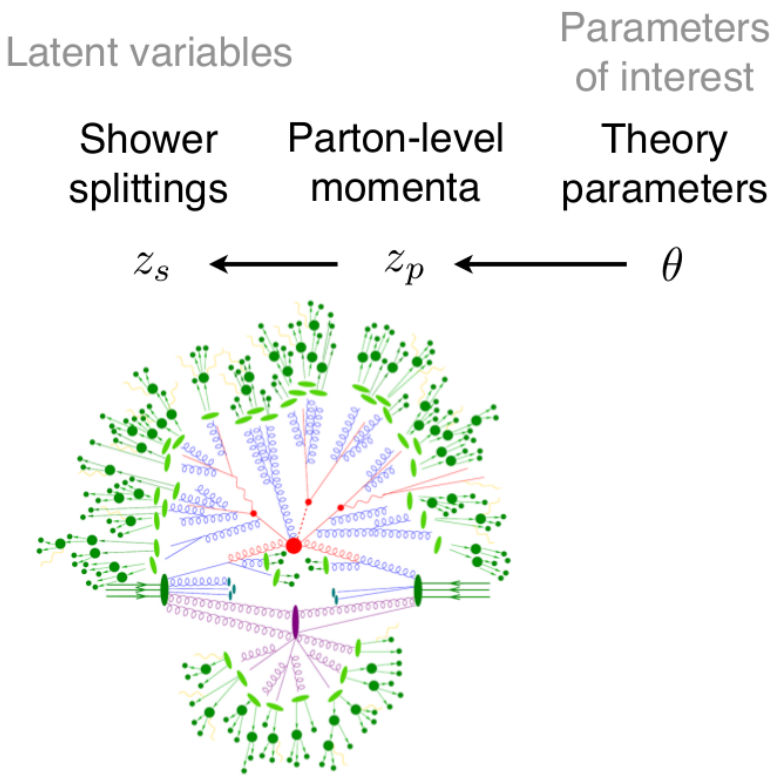
Parton-level momenta

Theory parameters

$$z_p \longleftarrow \theta$$

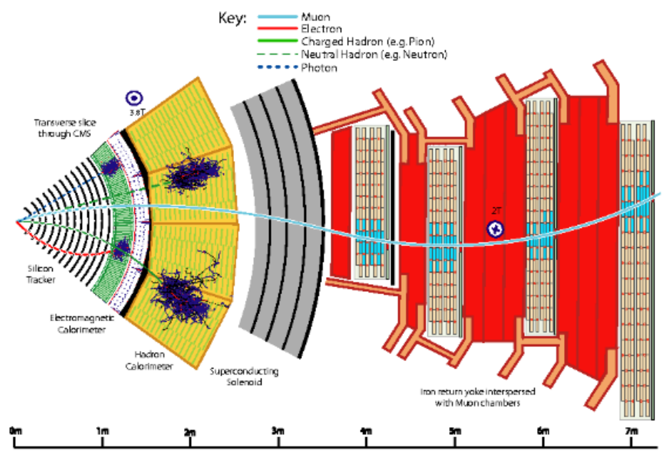


$$p(\mathbf{z}_s|\theta) = \int p(\mathbf{z}_p|\theta)p(\mathbf{z}_s|\mathbf{z}_p)d\mathbf{z}_p$$



$$p(\mathbf{z}_d|\theta) = \iint p(\mathbf{z}_p|\theta)p(\mathbf{z}_s|\mathbf{z}_p)p(\mathbf{z}_d|\mathbf{z}_s)d\mathbf{z}_p d\mathbf{z}_s$$

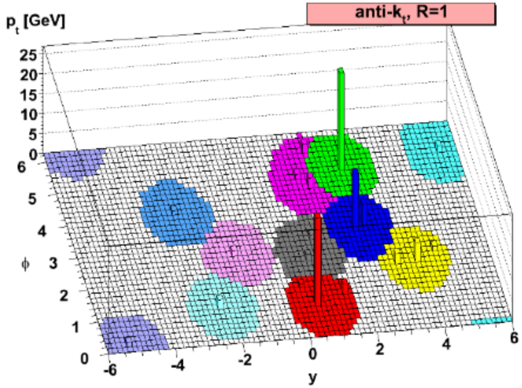
Latent variables			Parameters of interest
Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
z_d ←	z_s ←	z_p ←	θ



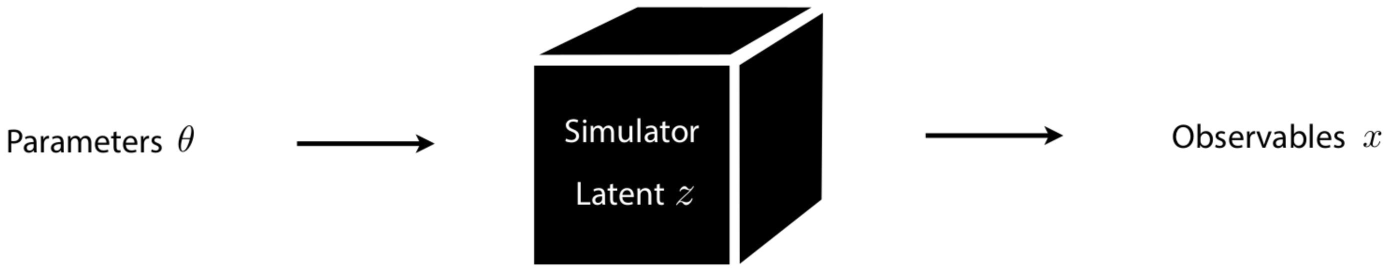
$$p(\mathbf{x}|\theta) = \underbrace{\iiint}_{\text{yikes!}} p(\mathbf{z}_p|\theta)p(\mathbf{z}_s|\mathbf{z}_p)p(\mathbf{z}_d|\mathbf{z}_s)p(\mathbf{x}|\mathbf{z}_d)d\mathbf{z}_p d\mathbf{z}_s d\mathbf{z}_d$$

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

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



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



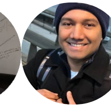
Prediction:

- 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

Statistical inference becomes challenging when the likelihood $p(\mathbf{x}|\theta)$ is implicit or intractable. **Simulation-based inference algorithms are needed.**

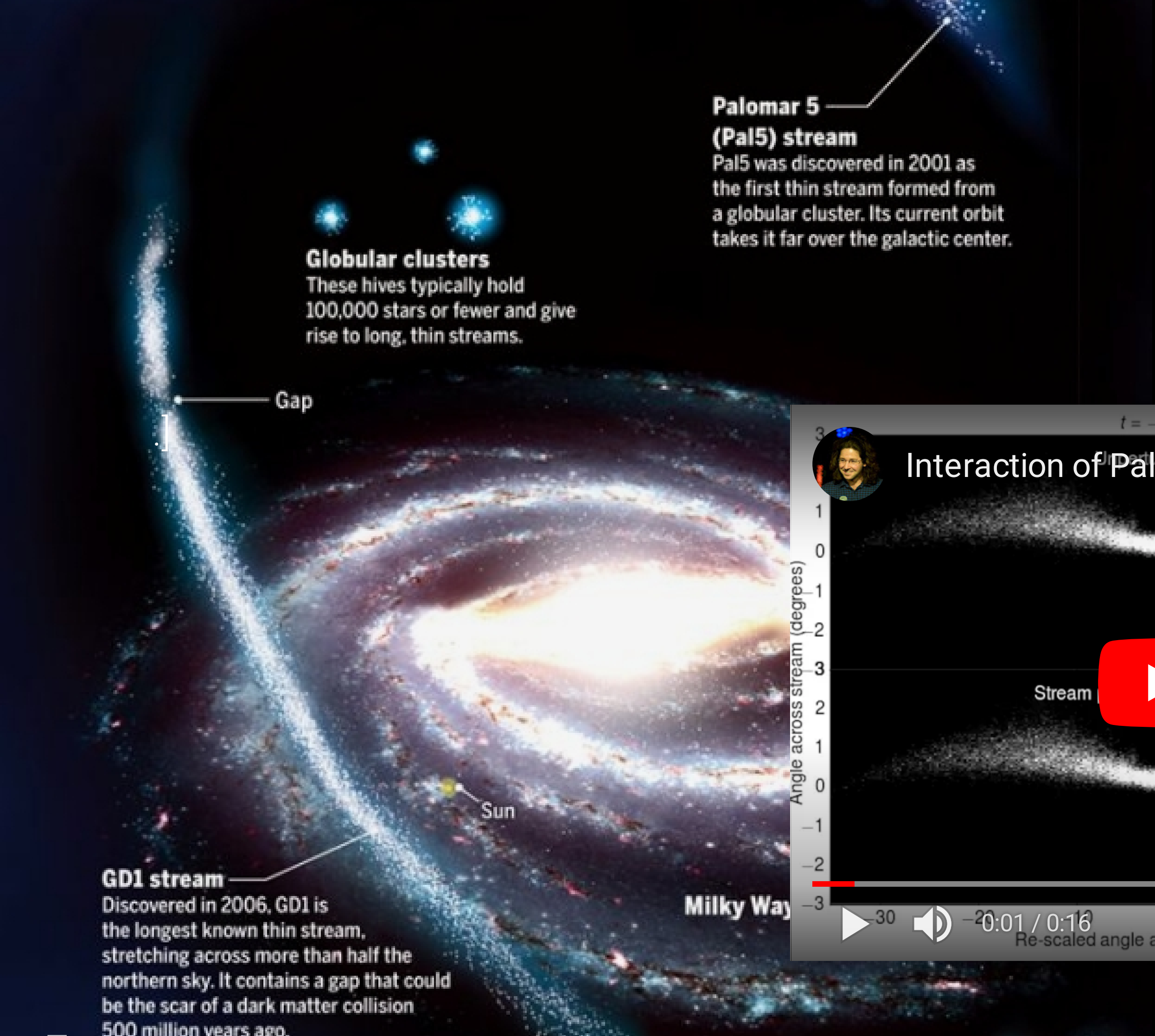



A case study

Hermans et al, "Constraining dark matter with stellar streams", 2021.

Can we constrain the nature of dark matter from cosmological observations?

Constraining dark matter with stellar streams



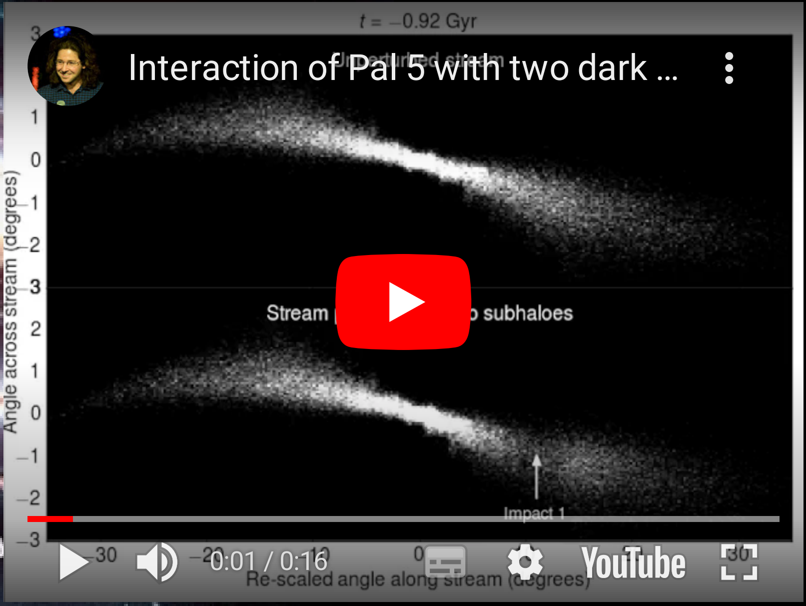


Image credits: C. Bickel/Science; D. Erkal.

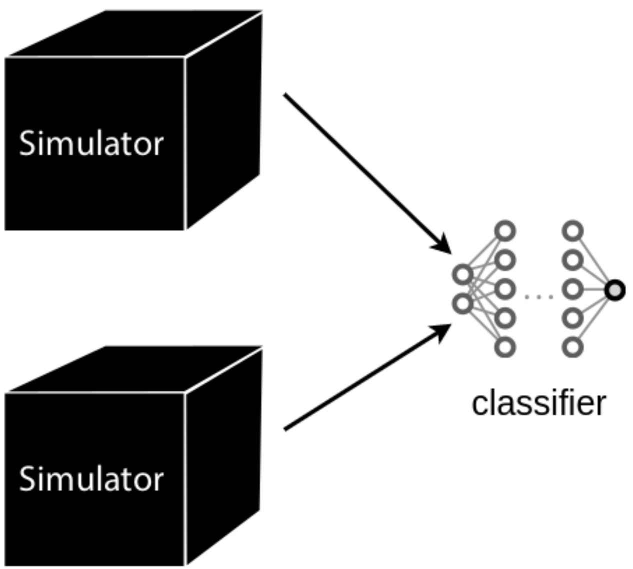
$$p(m_{\mathrm{WDM}}, t_{\mathrm{age}} | \mathrm{GD1}) = \frac{p(\mathrm{GD1} | m_{\mathrm{WDM}}, t_{\mathrm{age}}) p(m_{\mathrm{WDM}}, t_{\mathrm{age}})}{p(\text{GD-1})}$$

Neural ratio estimation (NRE)

The likelihood-to-evidence $r(\mathbf{x}|\theta) = \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x})} = \frac{p(\theta,\mathbf{x})}{p(\theta)p(\mathbf{x})}$ ratio can estimated from a binary classifier $d(\theta, \mathbf{x})$, even if neither the likelihood nor the evidence can be evaluated.

$$\theta, \mathbf{x} \sim p(\theta, \mathbf{x})$$

$$\theta, \mathbf{x} \sim p(\theta)p(\mathbf{x})$$



$$\hat{r}(\mathbf{x}|\theta)$$

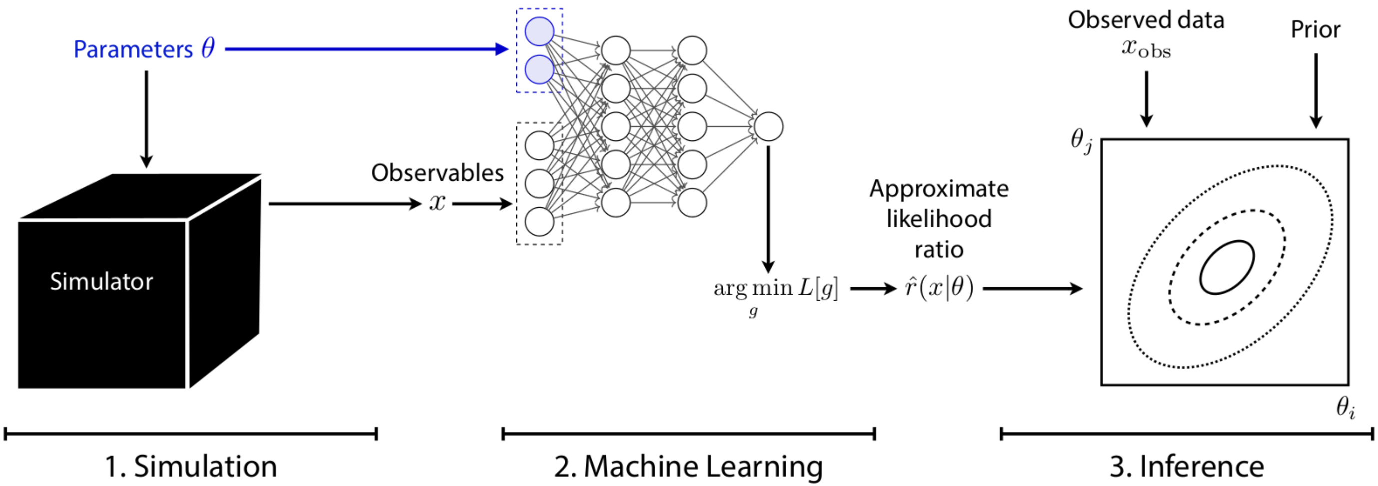
—
Credits: Cranmer et al, 2015; Hermans et al, 2020.

The solution d found after training approximates the optimal classifier

$$d(\theta, \mathbf{x}) \approx d^*(\theta, \mathbf{x}) = \frac{p(\theta, \mathbf{x})}{p(\theta, \mathbf{x}) + p(\theta)p(\mathbf{x})}.$$

Therefore,

$$r(\mathbf{x}|\theta) = \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x})} = \frac{p(\theta, \mathbf{x})}{p(\theta)p(\mathbf{x})} \approx \frac{d(\theta, \mathbf{x})}{1 - d(\theta, \mathbf{x})} = \hat{r}(\mathbf{x}|\theta).$$



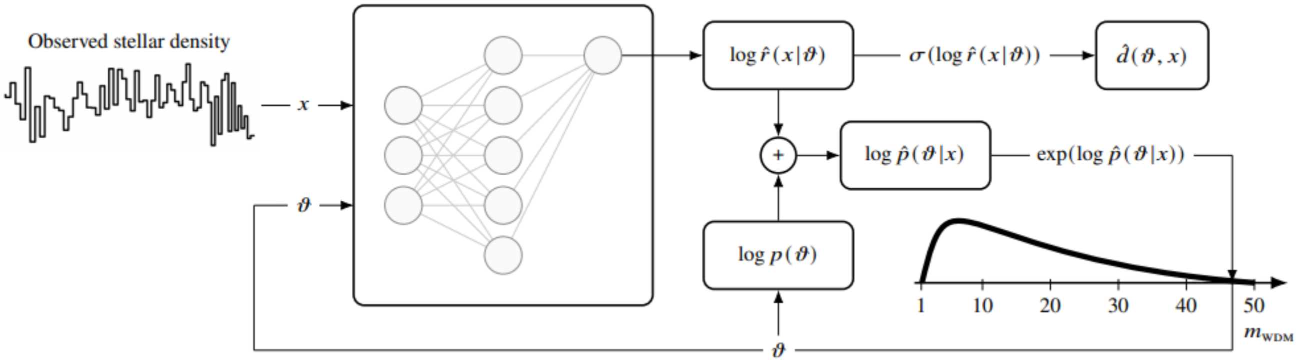
Run simulator and save data

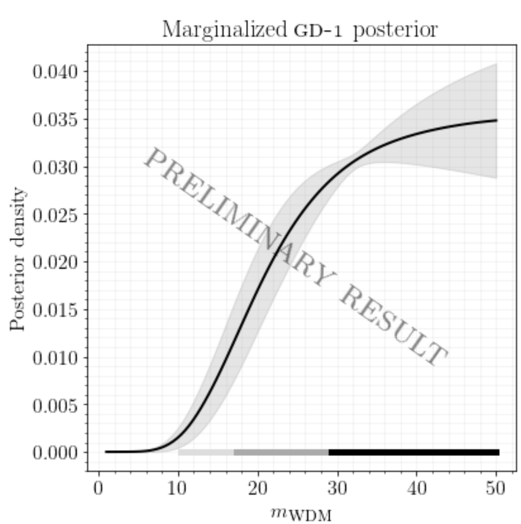
Train NN classifier, interpret as likelihood ratio estimator

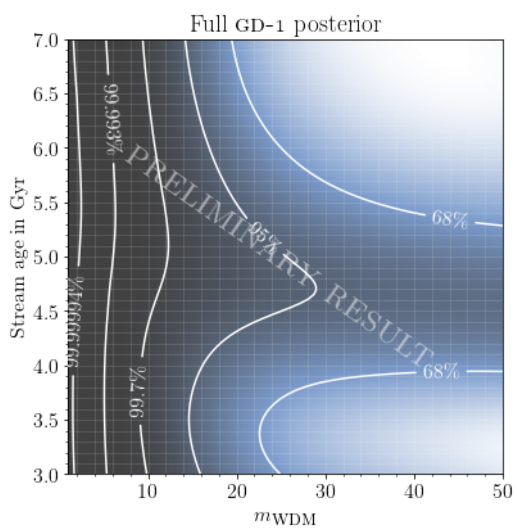
Amortized: cheap to repeat for new data

$$p(\theta|\mathbf{x}) \approx \hat{r}(\mathbf{x}|\theta)p(\theta)$$

NRE for stellar streams







—

Credits: Hermans et al, 2021.

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

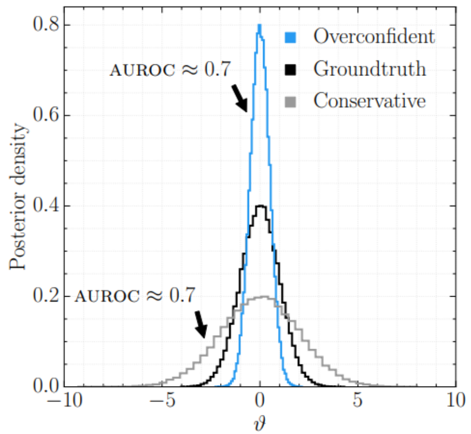
Wait a minute Gilles...
I can't claim that in a paper!
Your neural network must be wrong!

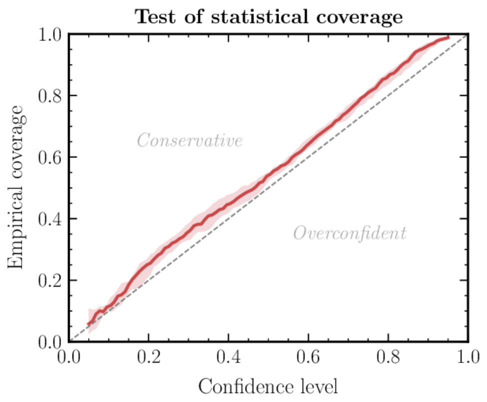
Expected coverage

$$\text{EC}(\hat{p}, \alpha) = \mathbb{E}_{p(\theta,\mathbf{x})}[\theta \in \Theta_{\hat{p}(\theta|\mathbf{x})}(\alpha)]$$

If the expected coverage is close to the nominal coverage probability α, then the approximate posterior $\hat{p}$ is calibrated.

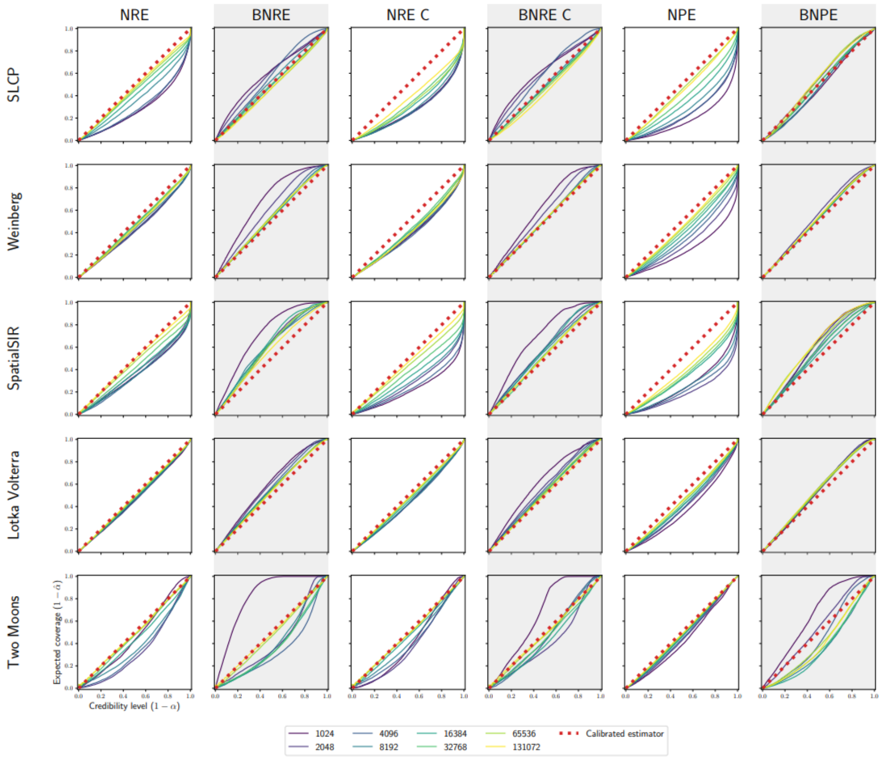
- If $\text{EC} < \alpha$, then the posterior is underdispersed and overconfident.
- If $\text{EC} > \alpha$, then the posterior is overdispersed and conservative.





—
Credits: Hermans et al, 2021; Siddharth Mishra-Sharma, 2021.

Balancing inference for conservative posteriors



Conservative posteriors can be obtained by enforcing d to be balanced, i.e. such that $\mathbb{E}_{p(\theta,\mathbf{x})}\left[d(\theta,\mathbf{x})\right] = \mathbb{E}_{p(\theta)p(\mathbf{x})}\left[1 - d(\theta,\mathbf{x})\right]$.

—

Credits: Delaunoy et al, 2022; Delaunoy et al, 2023.

Summary

Simulation-based inference is a major evolution in the statistical capabilities for science, as it enables the analysis of complex models and data without simplifying assumptions.

Obstacles remain to be overcome, such as the curse of dimensionality, the need for large amounts of data, or the necessary robustness of the inference network.