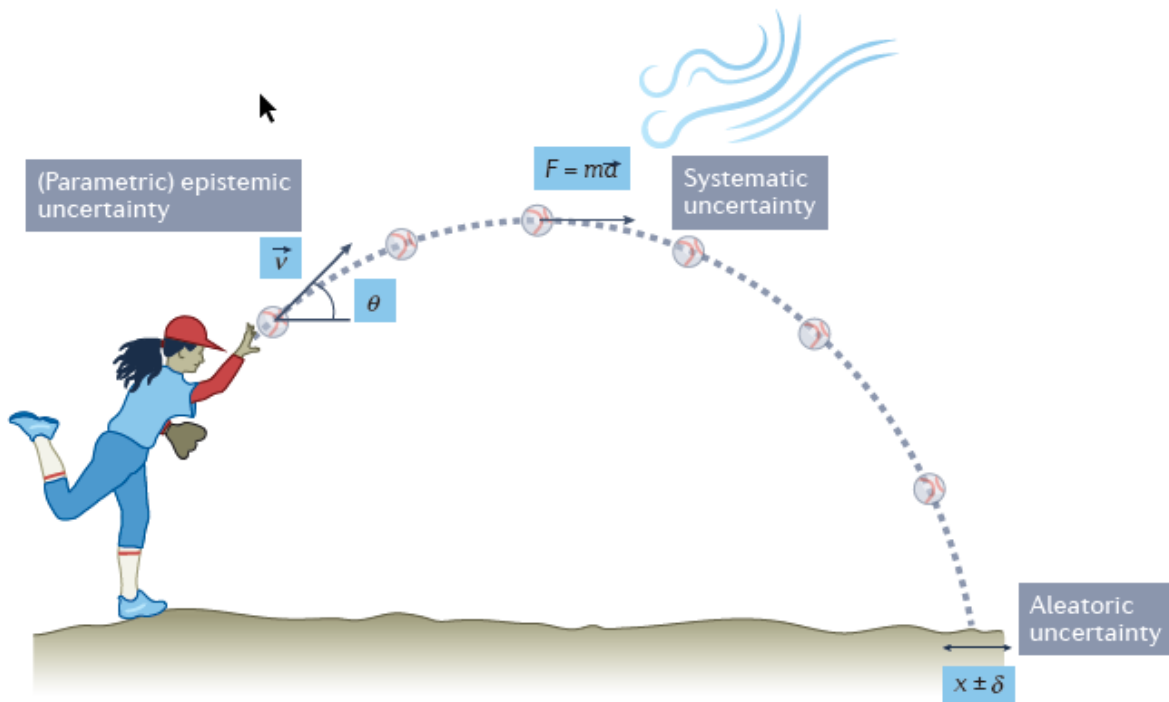


An introduction to simulation-based inference

51st SLAC Summer Institute

August 16, 2023

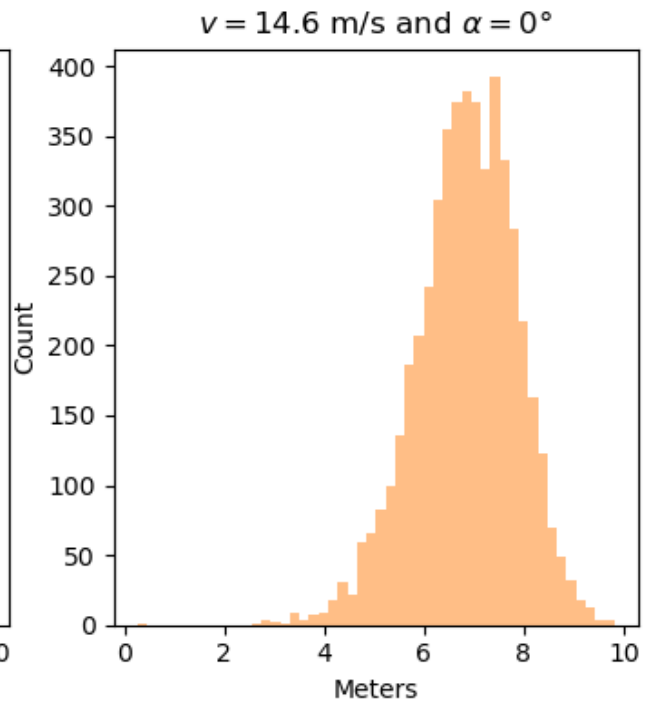
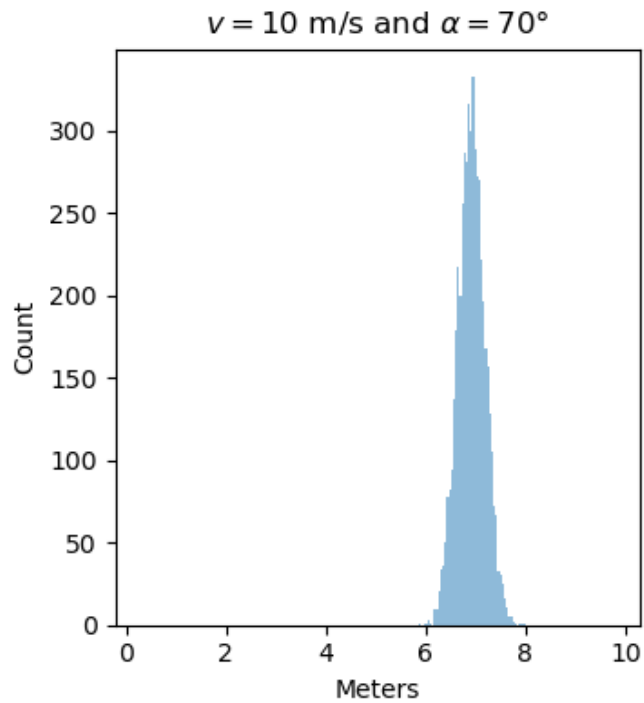
Gilles Louppe
g.louppe@uliege.be



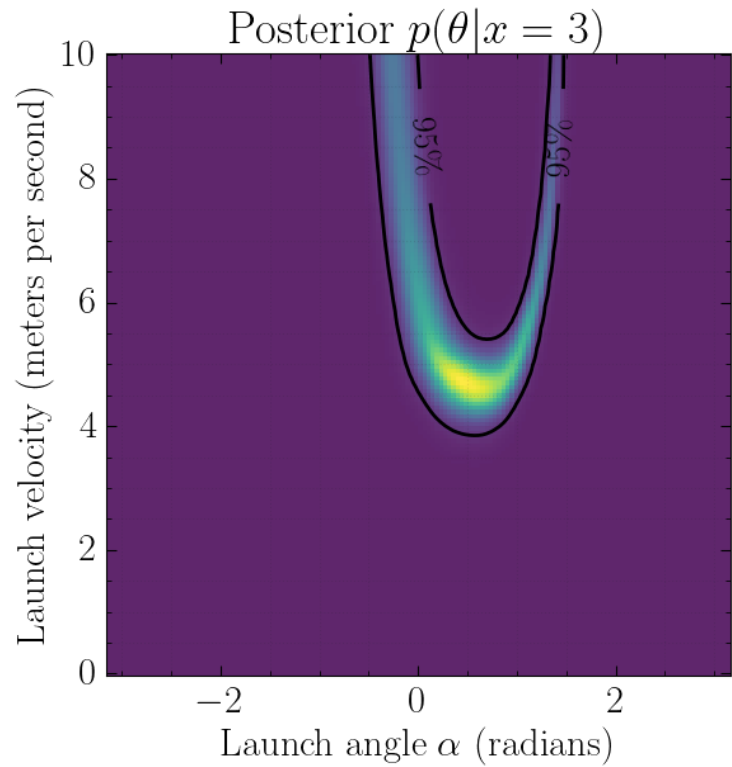
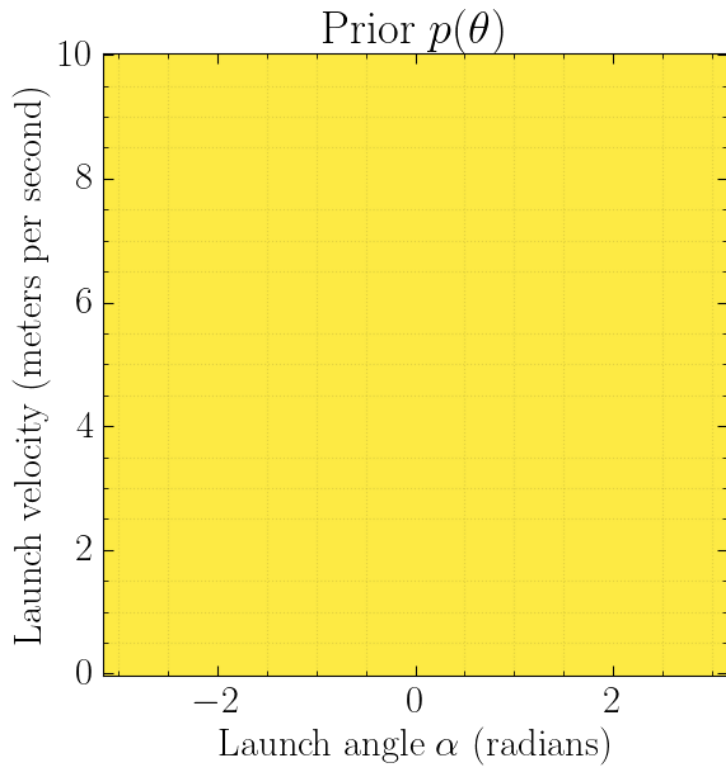
$$v_x = v \cos(\alpha), \quad v_y = v \sin(\alpha),$$

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

```
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()
```



What parameter values θ are the most plausible?



Outline

1. Simulation-based inference
2. Algorithms
 - Neural ratio estimation
 - Neural posterior estimation
 - Neural score estimation
3. Diagnostics

Simulation-based inference

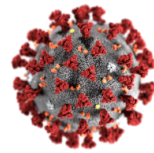
Scientific simulators



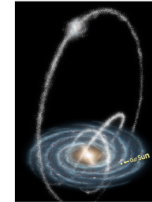
Chemical reactions



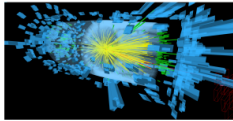
Flames



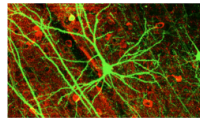
Epidemics



Stellar streams



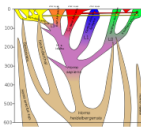
Collider experiments



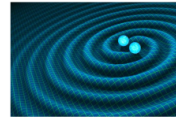
Neurons



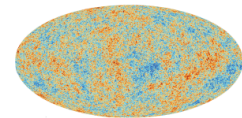
Robotics



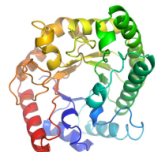
Evolution



Gravitational waves



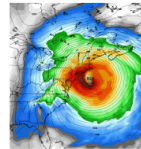
Evolution of the Universe



Protein networks



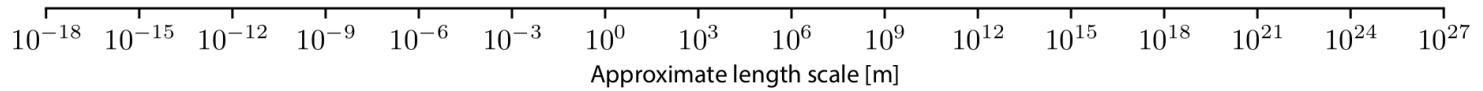
Ecological systems

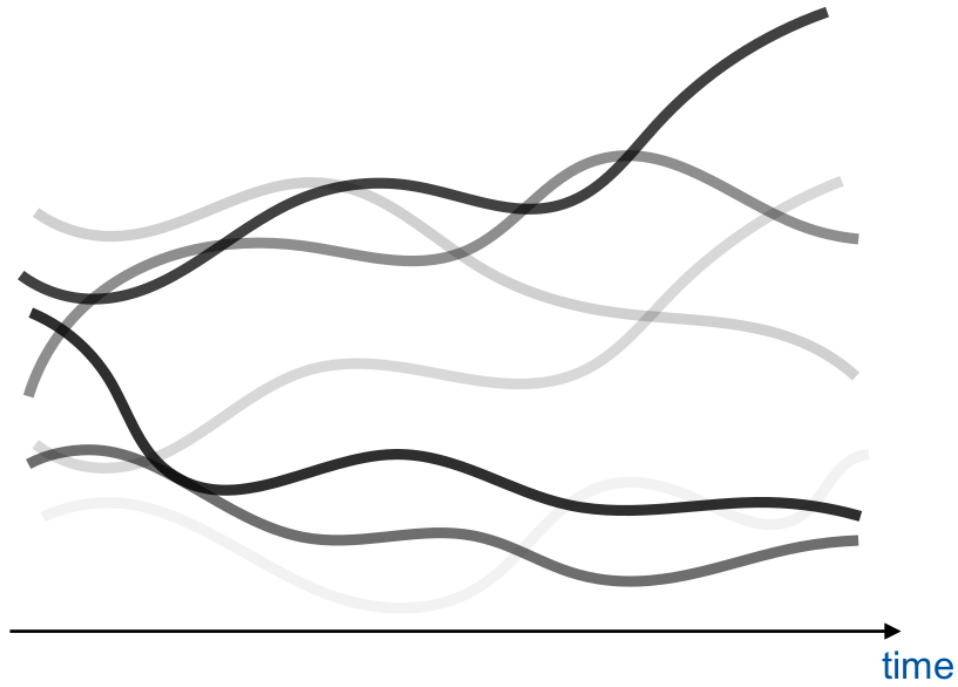


Weather and climate

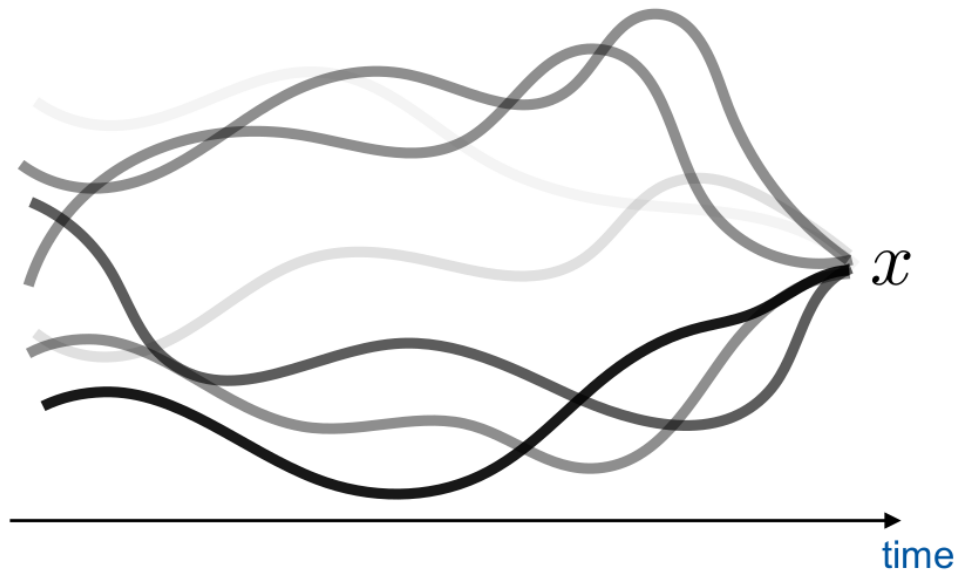


Gravitational lensing





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



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

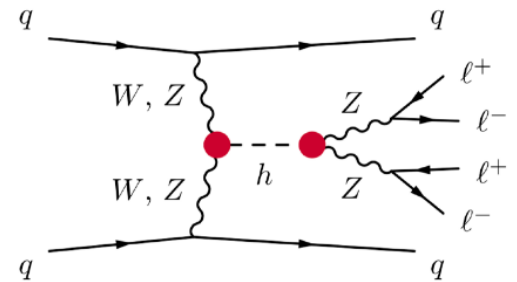
Latent variables

Parameters
of interest

Parton-level
momenta

Theory
parameters

$$z_p \longleftarrow \theta$$



Latent variables

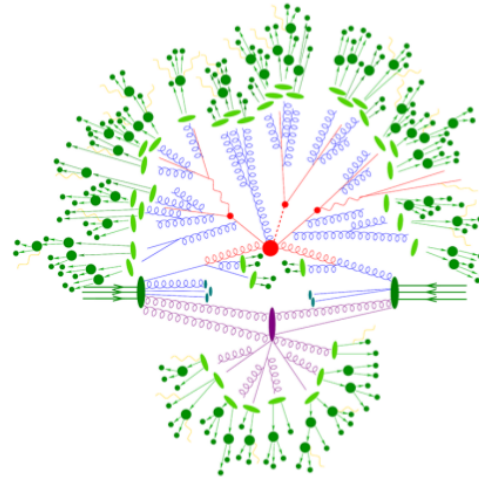
Parameters
of interest

Shower
splittings

Parton-level
momenta

Theory
parameters

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



Latent variables

Parameters of interest

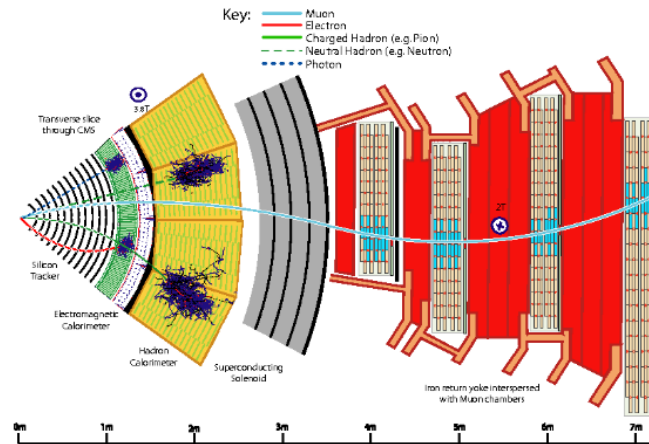
Detector interactions

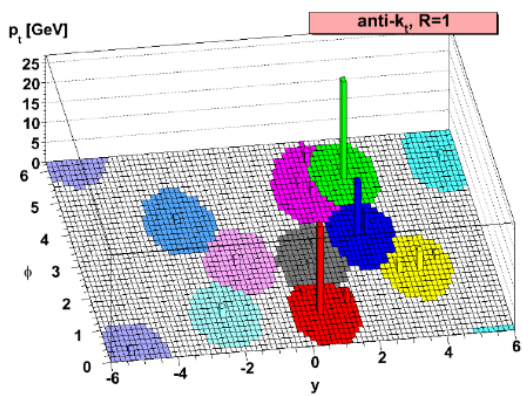
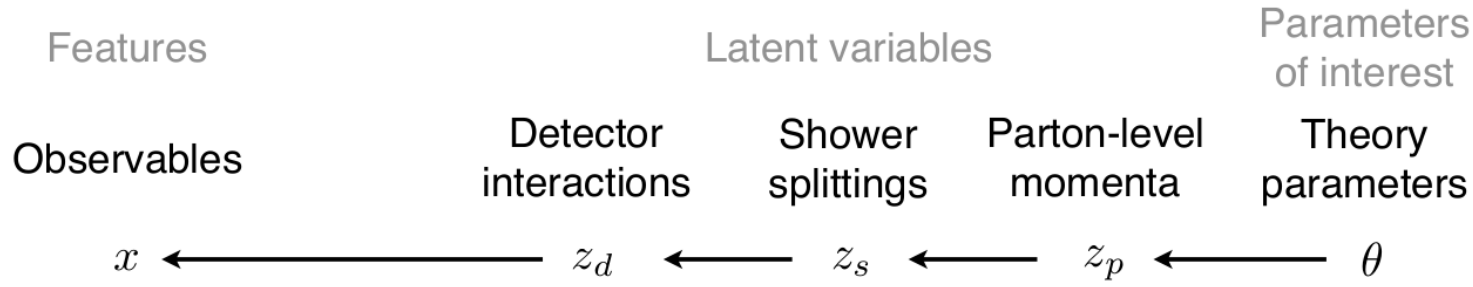
Shower splittings

Parton-level momenta

Theory parameters

z_d ← z_s ← z_p ← θ





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

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

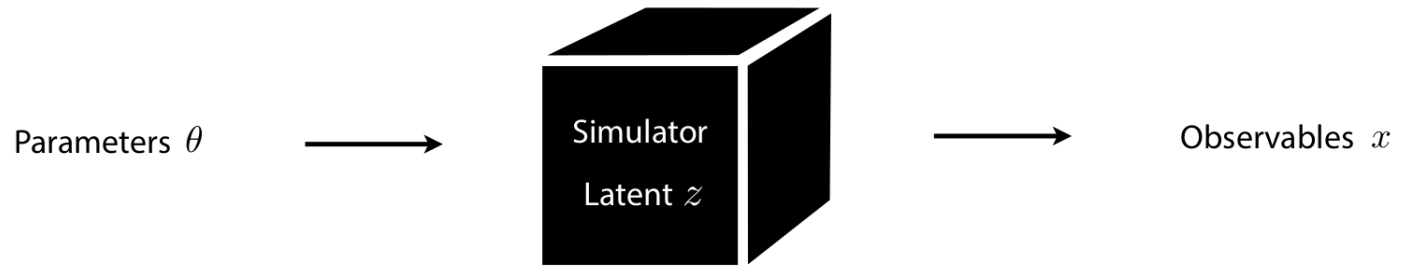
Bayesian inference

Start with

- a simulator that can generate N samples $\mathbf{x}_i \sim p(\mathbf{x}_i | \theta_i)$,
- a prior model $p(\theta)$,
- observed data $\mathbf{x}_{\text{obs}} \sim p(\mathbf{x}_{\text{obs}} | \theta_{\text{true}})$.

Then, estimate the posterior

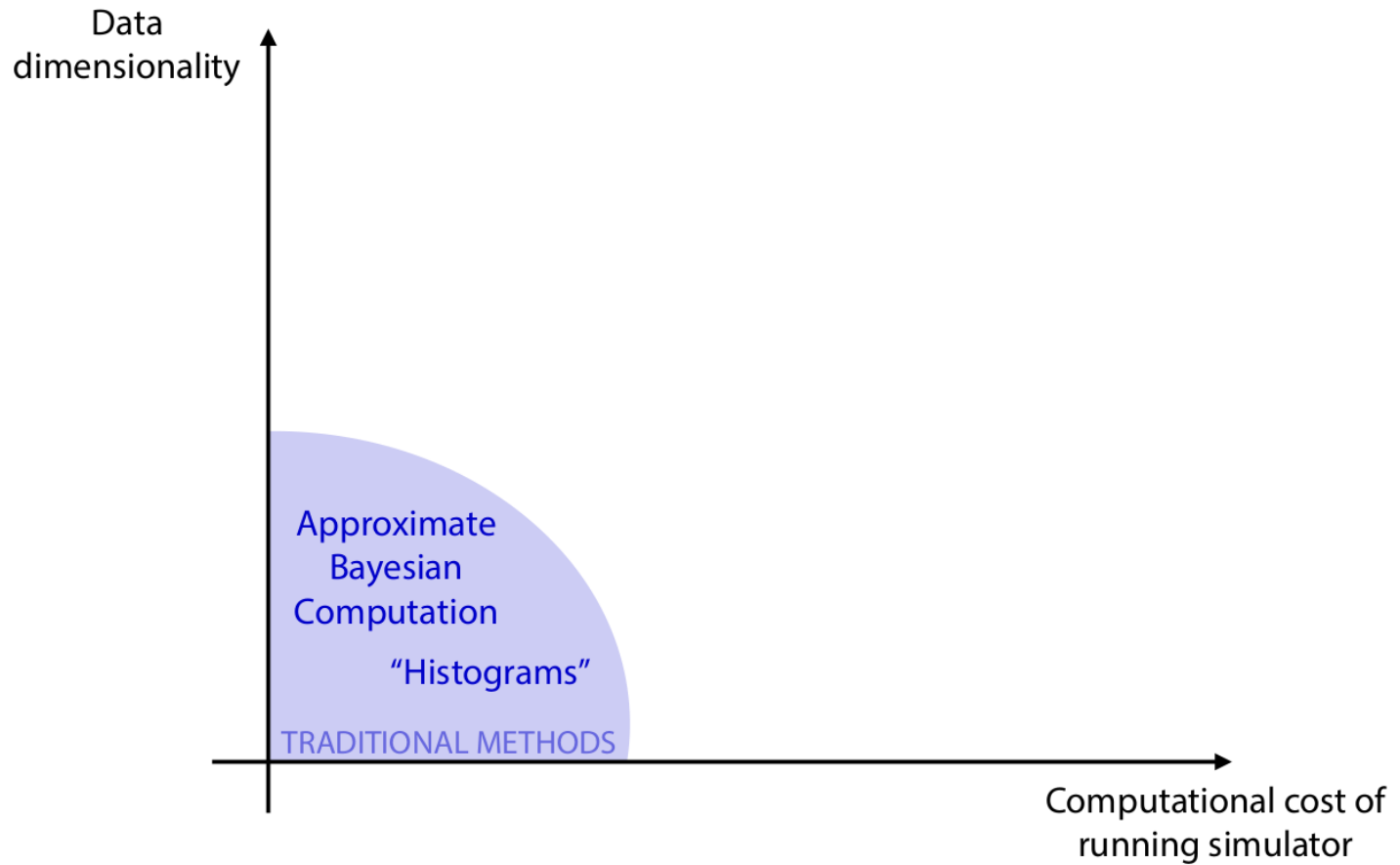
$$p(\theta | \mathbf{x}_{\text{obs}}) = \frac{p(\mathbf{x}_{\text{obs}} | \theta)p(\theta)}{p(\mathbf{x}_{\text{obs}})}.$$



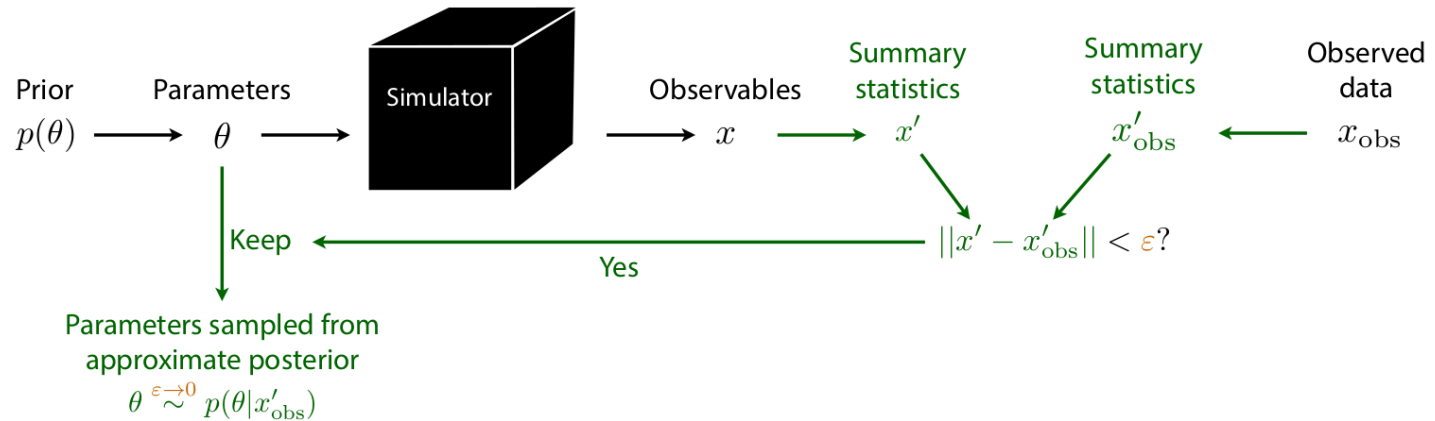
- Prediction:
- Well-motivated mechanistic, causal model
 - Simulator can generate samples $x \sim p(x|\theta)$

- Inference:
- Interactions between low-level components lead to challenging inverse problems
 - Likelihood $p(x|\theta) = \int dz p(x, z|\theta)$ is intractable

Algorithms

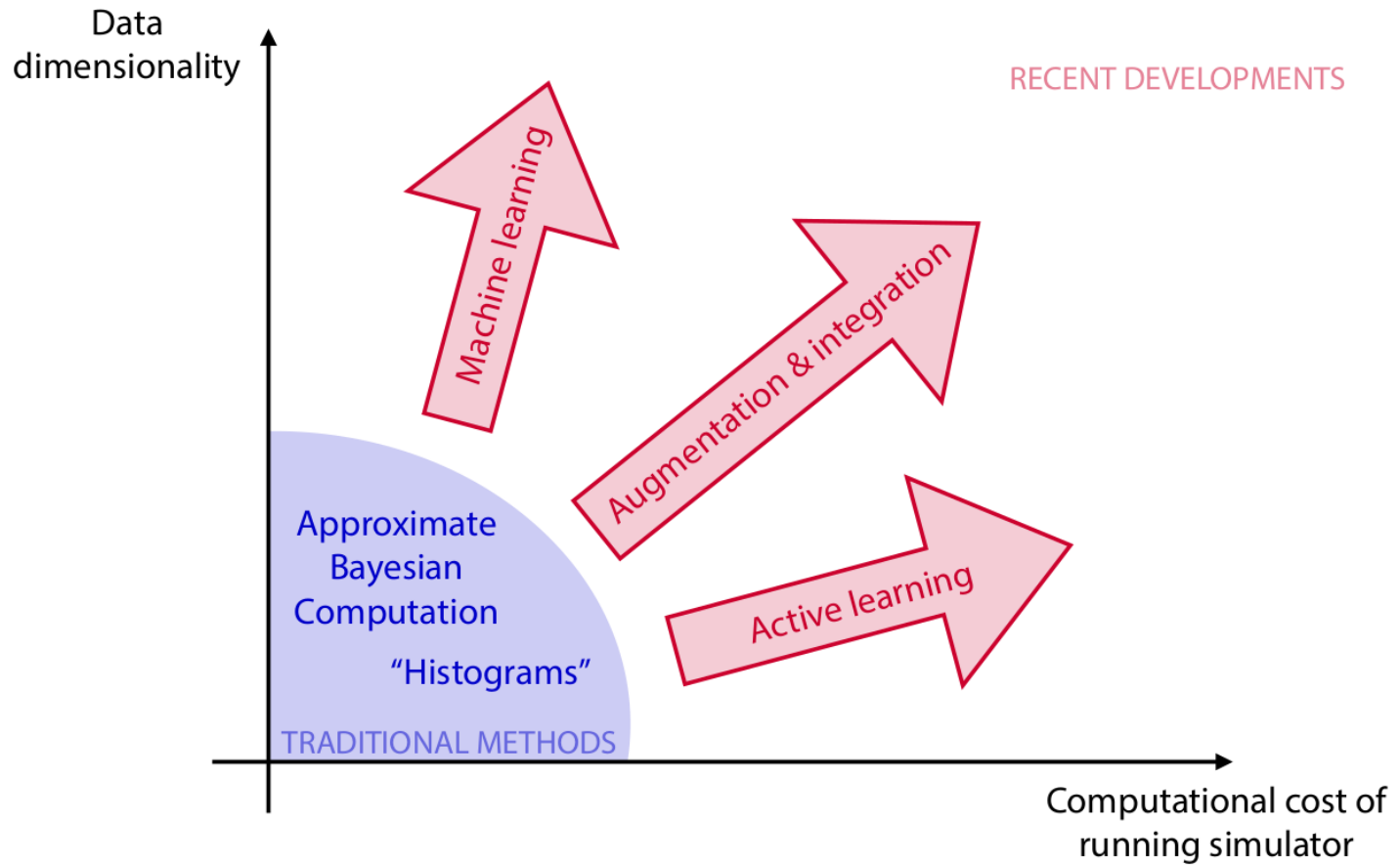


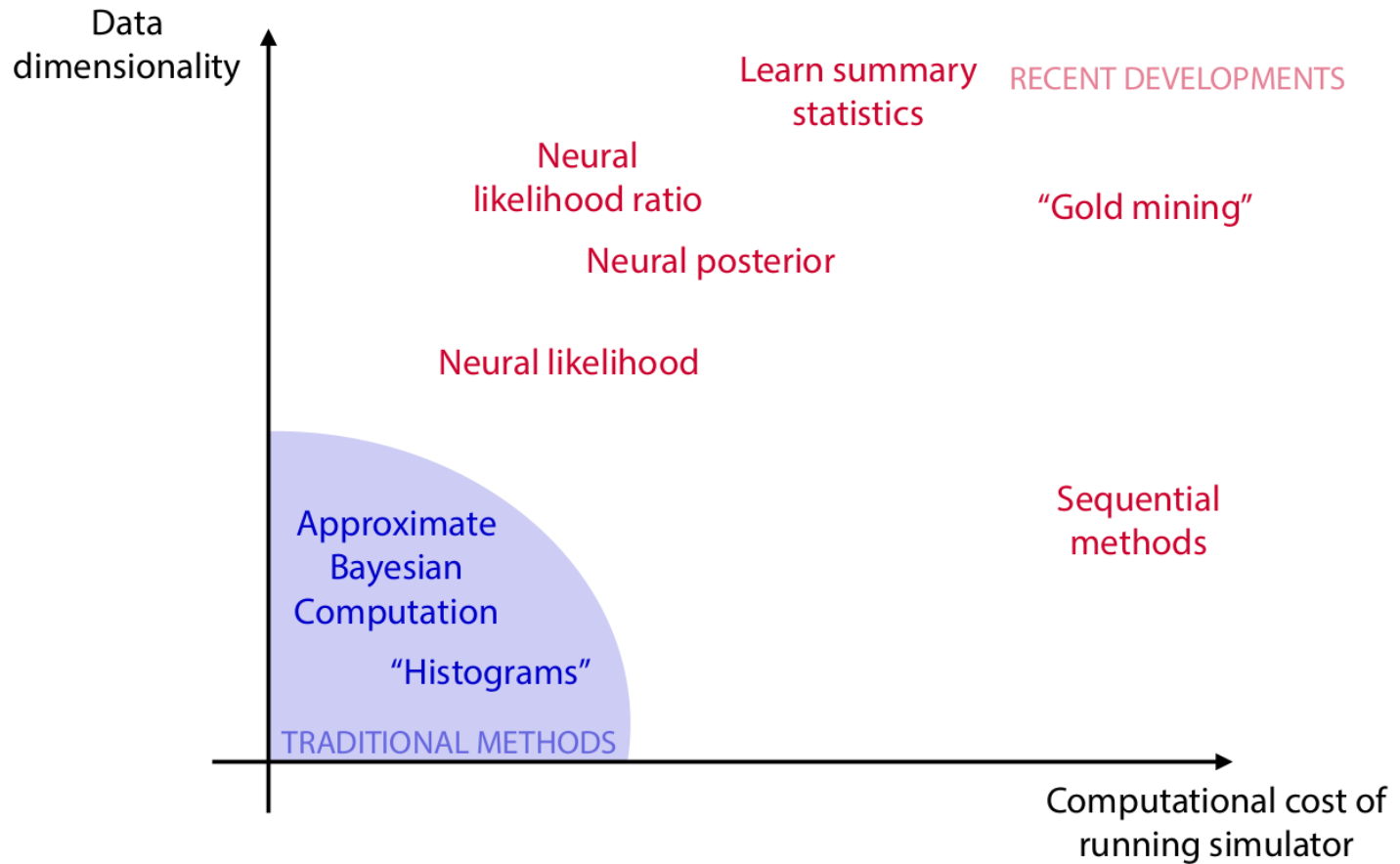
Approximate Bayesian Computation (ABC)



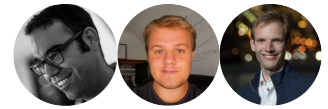
Issues:

- How to choose x' ? ϵ ? $\| \cdot \|$?
- No tractable posterior.
- Need to run new simulations for new data or new prior.

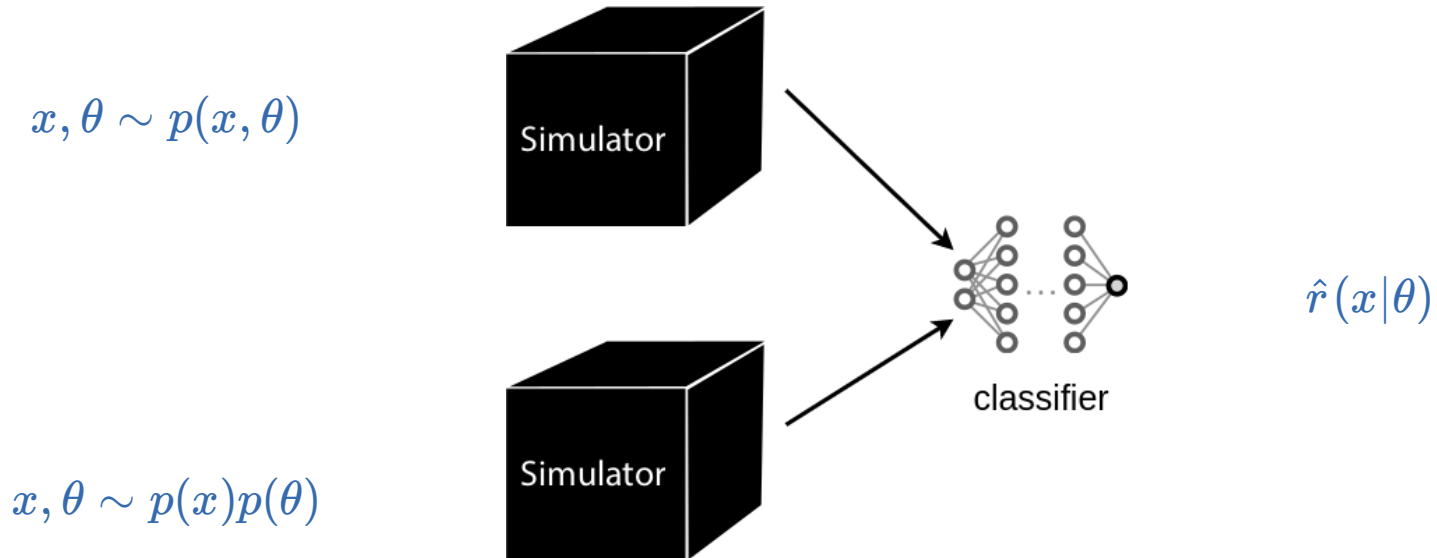


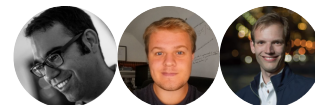


Neural ratio estimation



The likelihood-to-evidence $r(x|\theta) = \frac{p(x|\theta)}{p(x)} = \frac{p(x,\theta)}{p(x)p(\theta)}$ ratio can be learned, even if neither the likelihood nor the evidence can be evaluated:



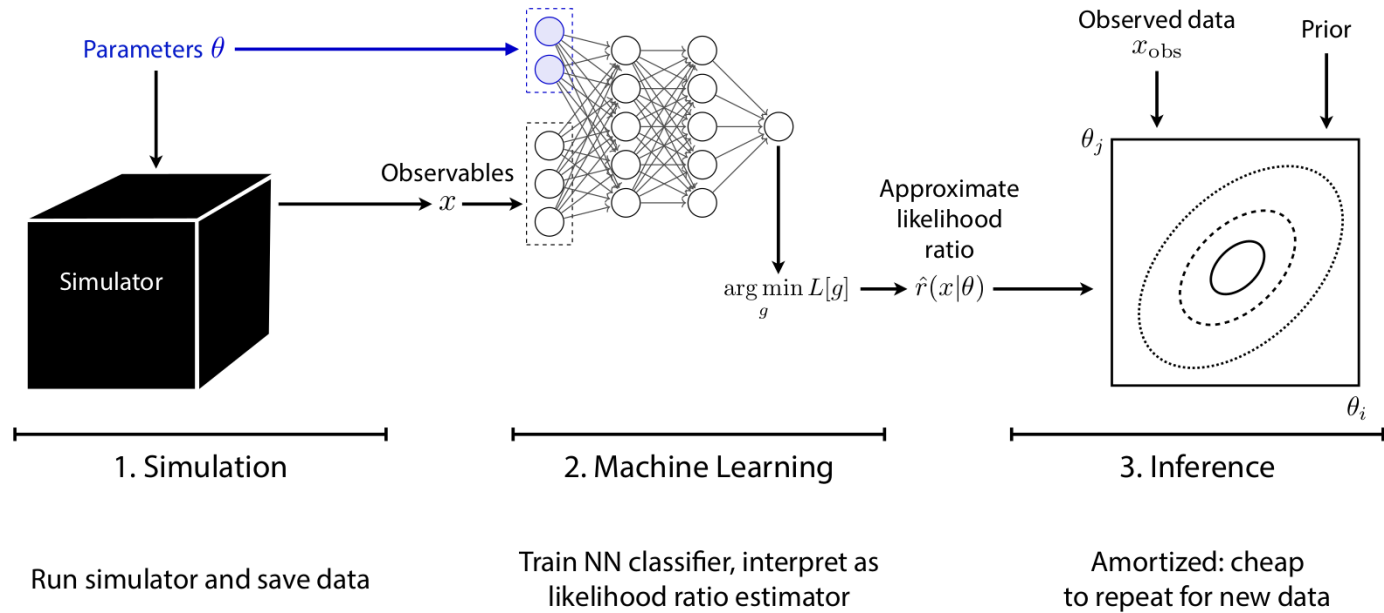
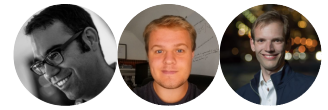


The solution d found after training approximates the optimal classifier

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

Therefore,

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



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

Constraining dark matter with stellar streams



Palomar 5 (Pal5) stream

Pal5 was discovered in 2001 as the first thin stream formed from a globular cluster. Its current orbit takes it far over the galactic center.

Globular clusters

These hives typically hold 100,000 stars or fewer and give rise to long, thin streams.

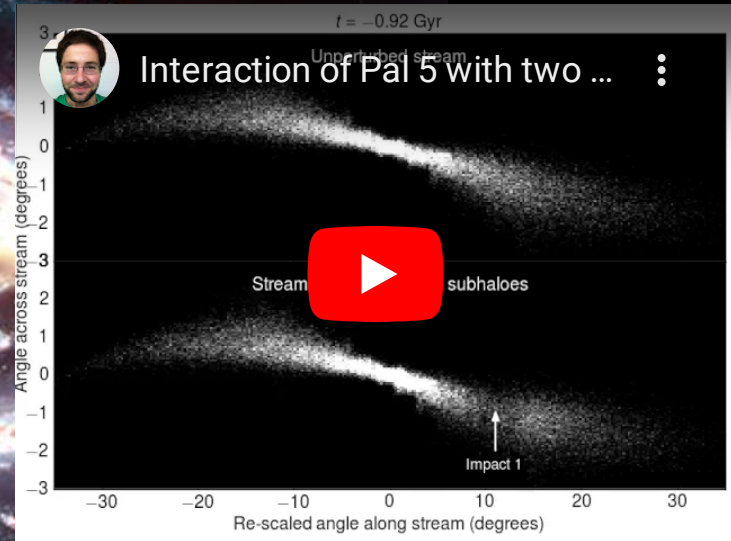
Gap

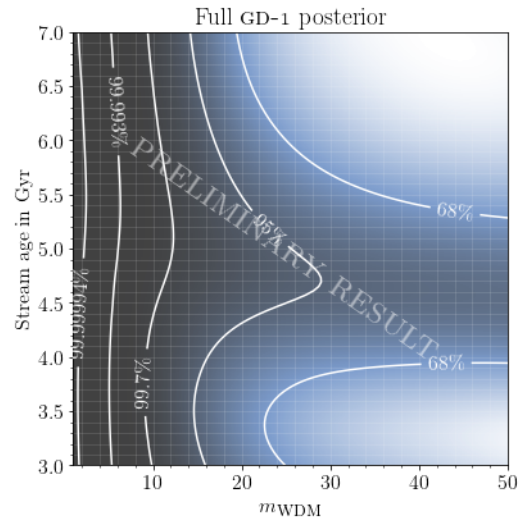
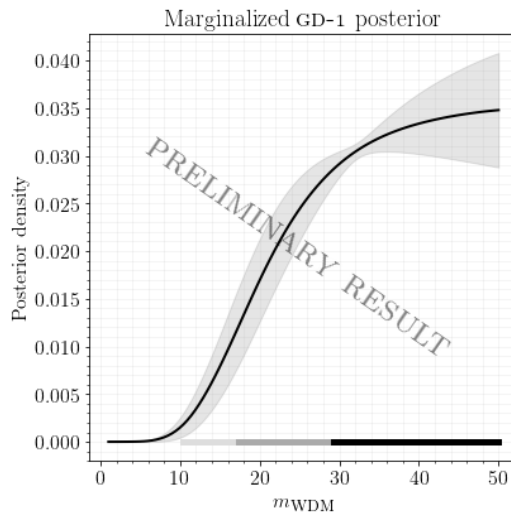
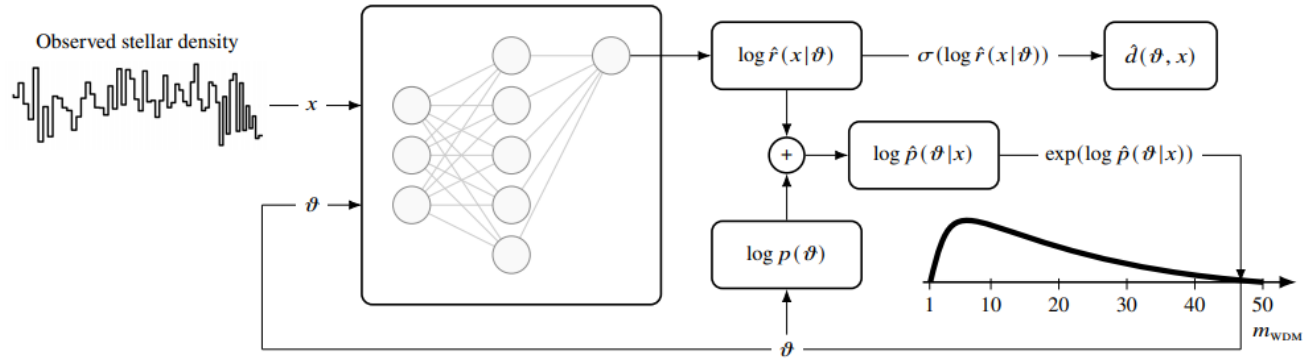
Sun

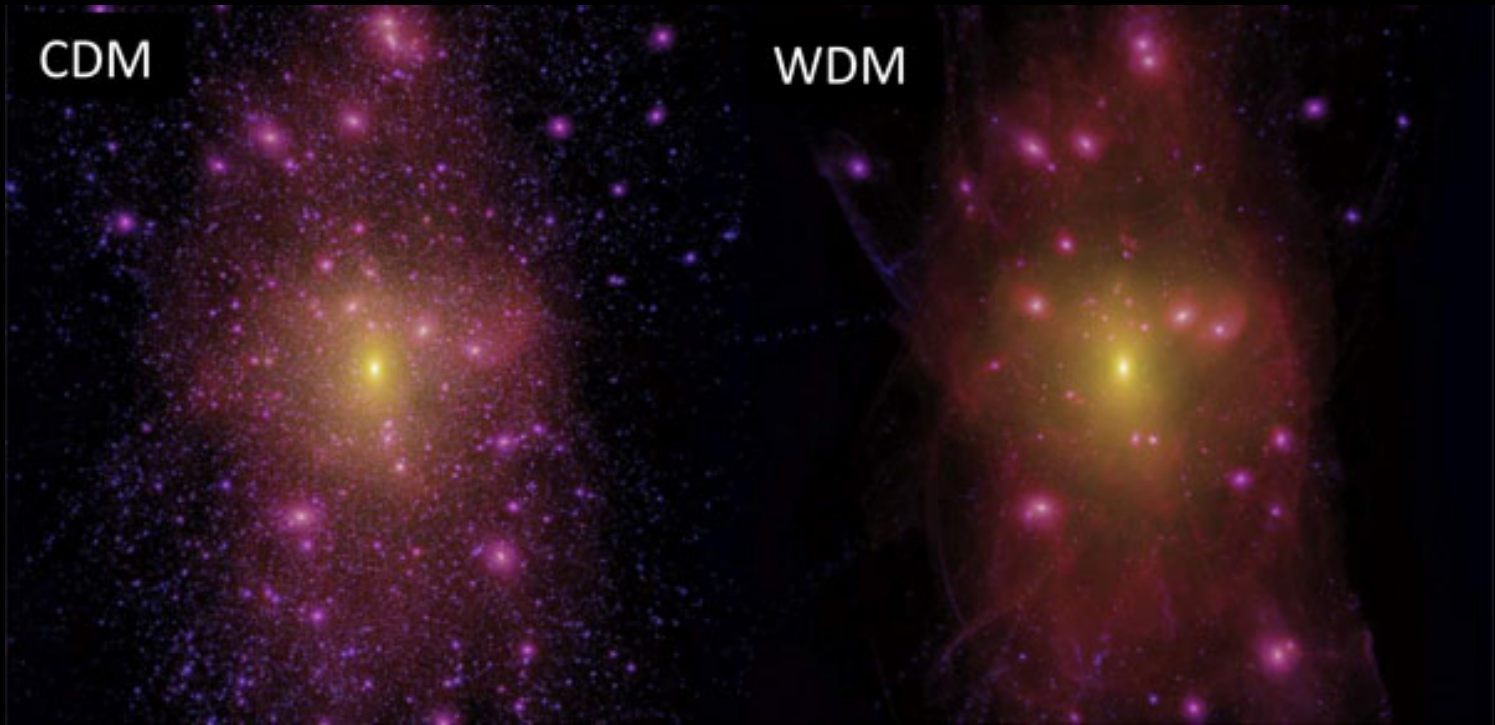
GD1 stream

Discovered in 2006, GD1 is the longest known thin stream, stretching across more than half the northern sky. It contains a gap that could be the scar of a dark matter collision 500 million years ago.

Milky Way

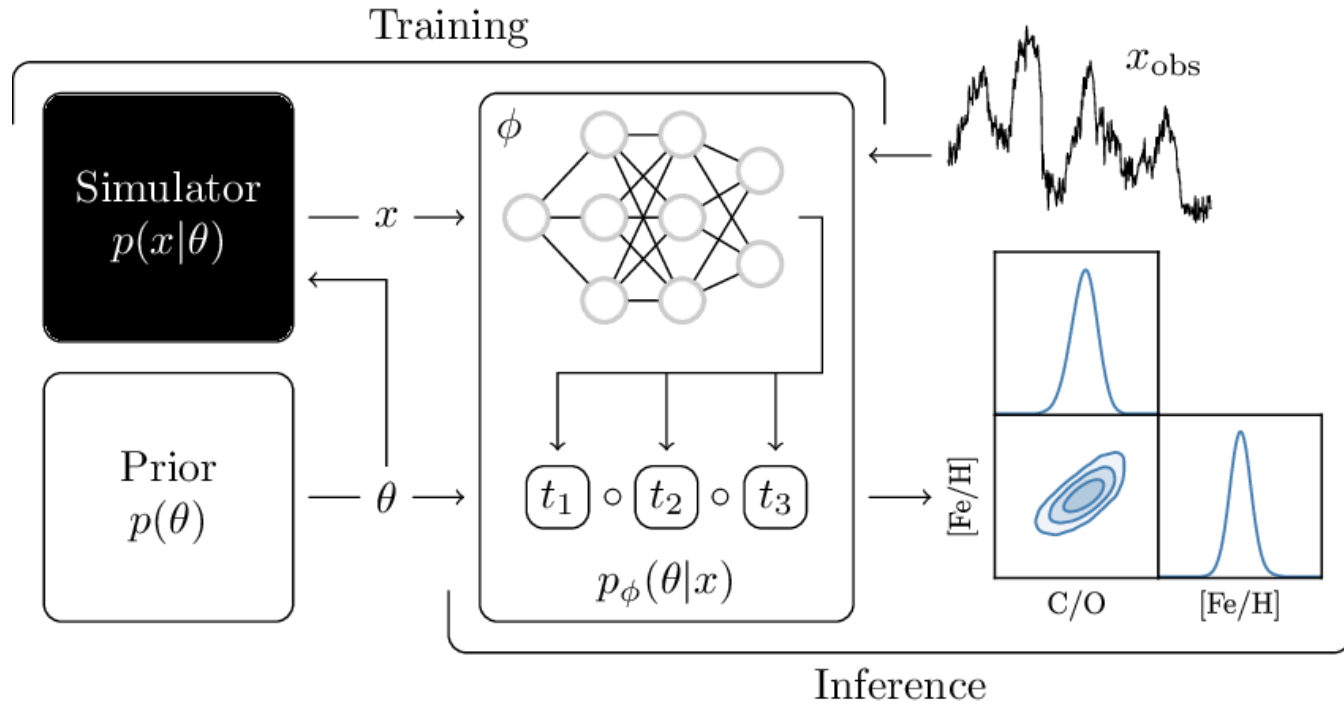






Preliminary results for GD-1 suggest a **preference for CDM over WDM.**

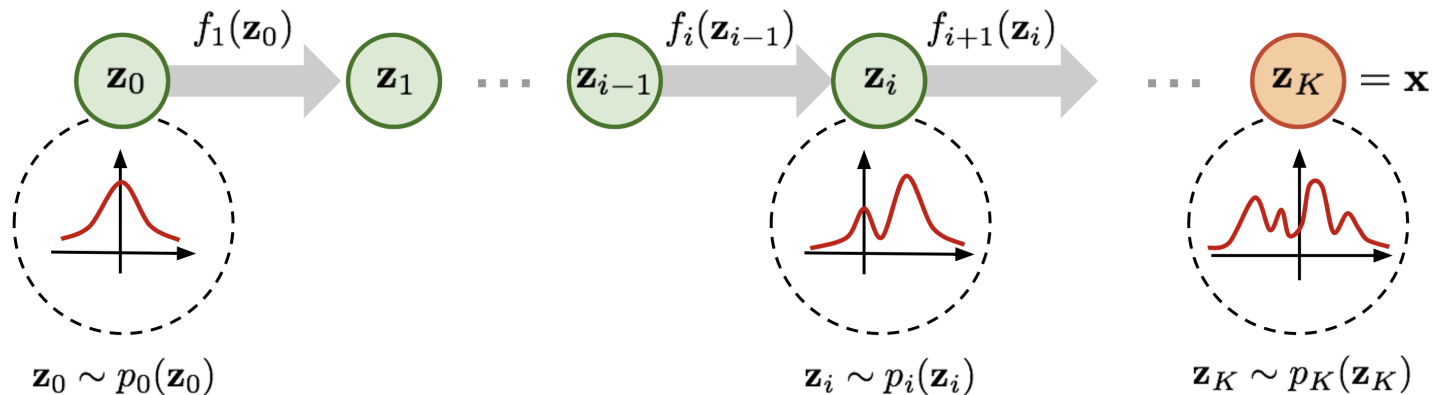
Neural Posterior Estimation



$$\min_{q_\phi} \mathbb{E}_{p(x)} [\text{KL}(p(\theta|x) || q_\phi(\theta|x))]$$

Normalizing flows

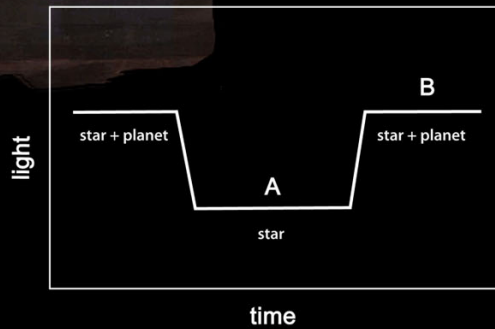
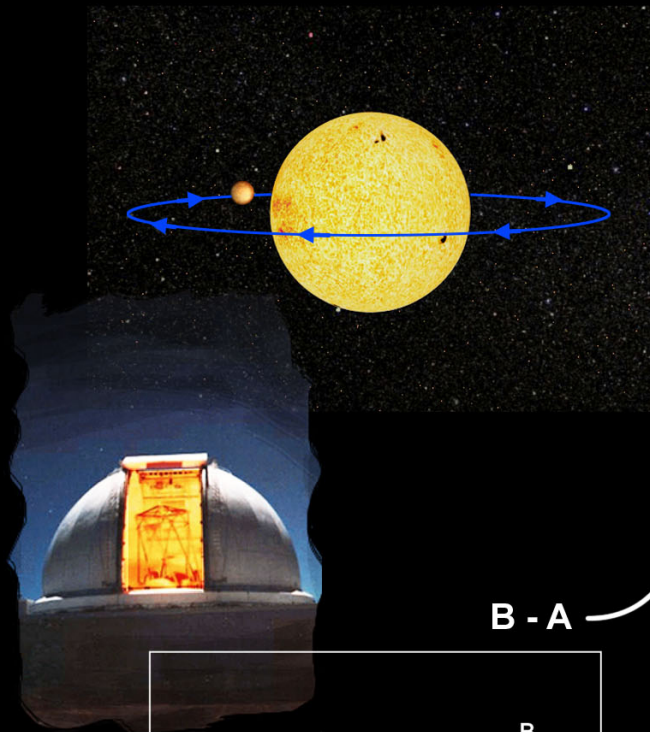
A normalizing flow is a sequence of invertible transformations f_k that map a simple distribution p_0 to a more complex distribution p_K :



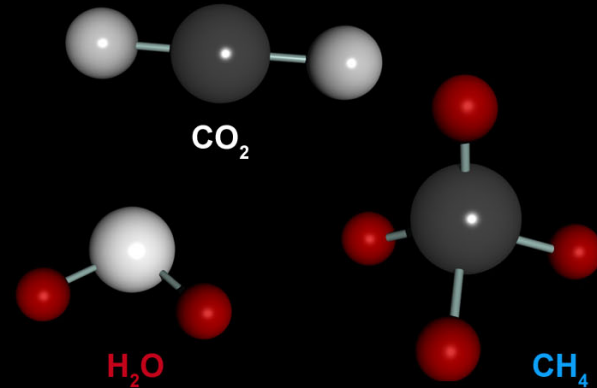
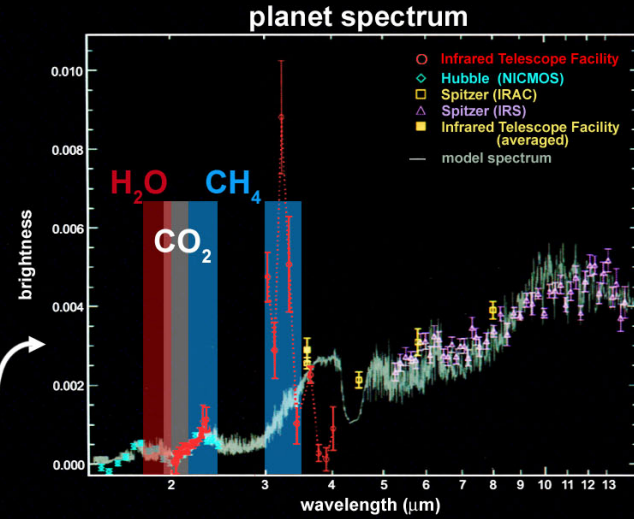
By the change of variables formula, the log-likelihood of a sample \mathbf{x} is given by

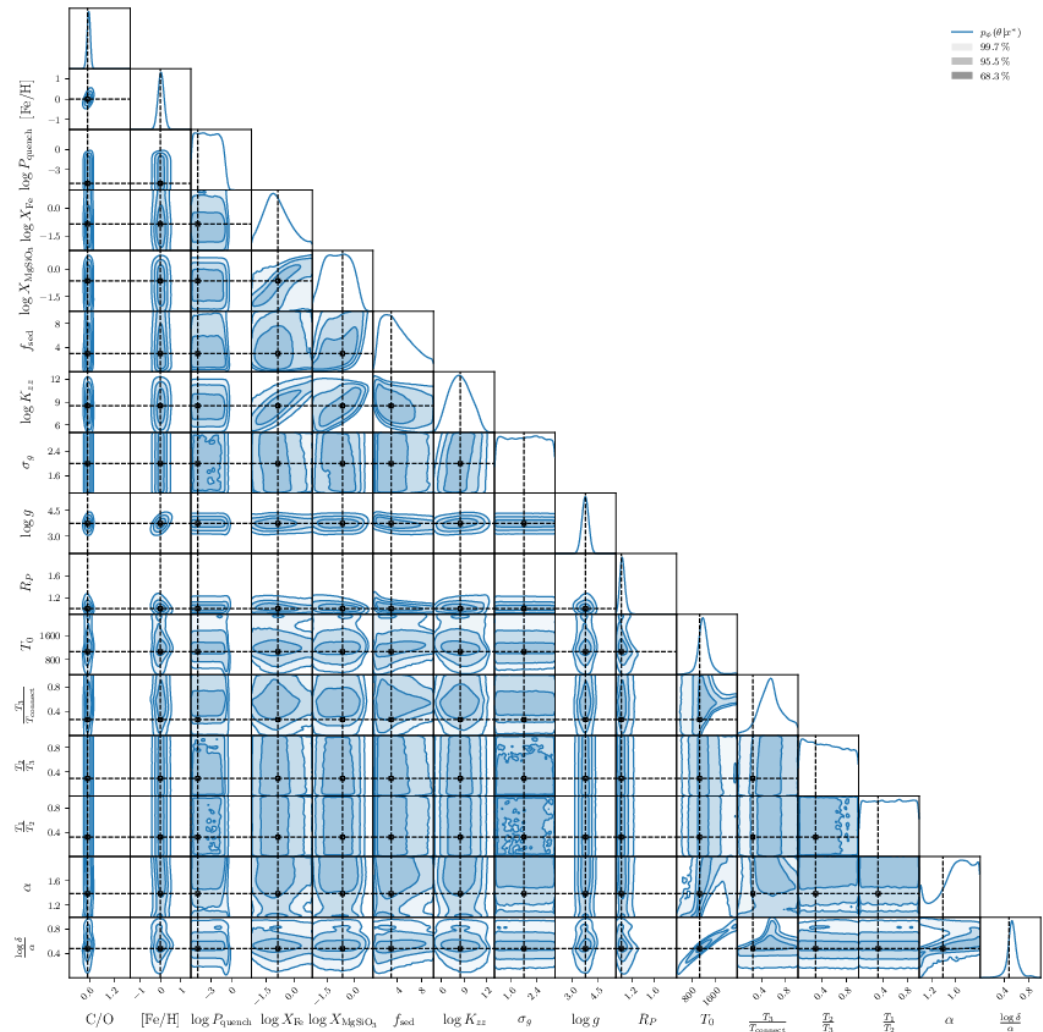
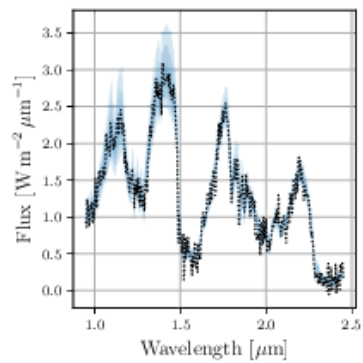
$$\log p(\mathbf{x}) = \log p(\mathbf{z}_0) - \sum_{k=1}^K \log |\det J_{f_k}(\mathbf{z}_{k-1})|.$$

Exoplanet atmosphere characterization



B - A

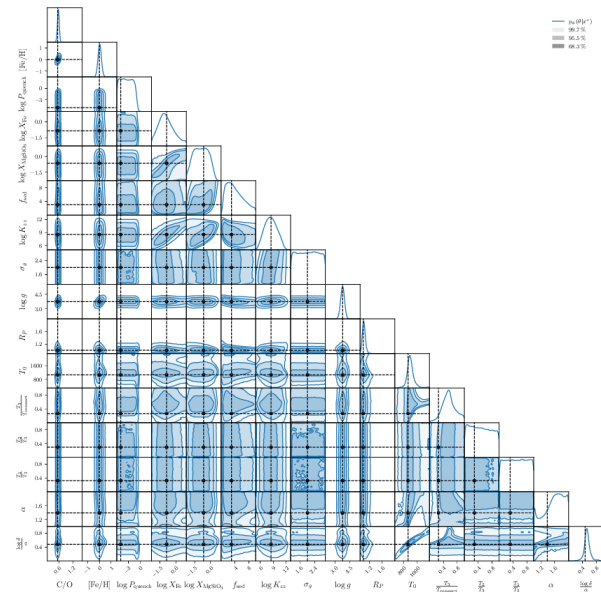




Diagnostics

$$\hat{p}(\theta|x) = \text{sbi}(p(x|\theta), p(\theta), x)$$

We must make sure our approximate simulation-based inference algorithms can (at least) actually realize faithful inferences on the (expected) observations.



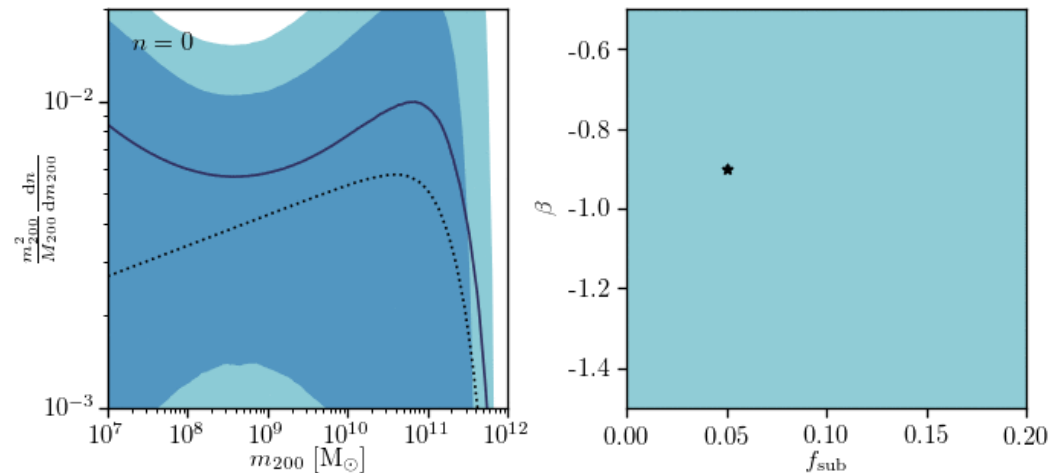
How do we know this is good enough?



Mode convergence

The maximum a posteriori estimate converges towards the nominal value θ^* for an increasing number of independent and identically distributed observables $x_i \sim p(x|\theta^*)$:

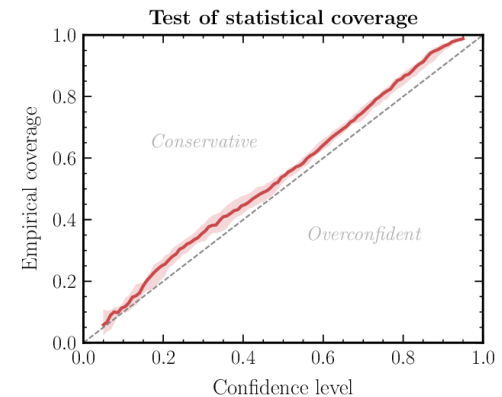
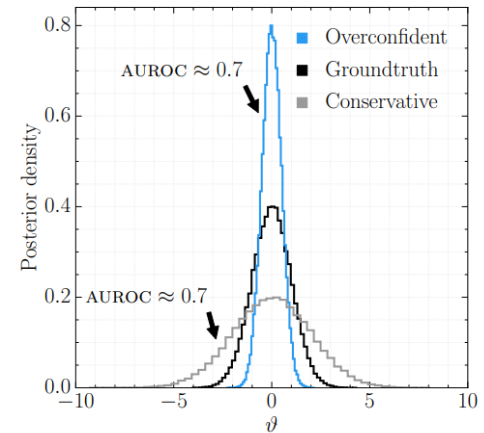
$$\begin{aligned} & \lim_{N \rightarrow \infty} \arg \max_{\theta} p(\theta | \{x_i\}_{i=1}^N) \\ &= \lim_{N \rightarrow \infty} \arg \max_{\theta} p(\theta) \prod_{x_i} r(x_i | \theta) = \theta^* \end{aligned}$$

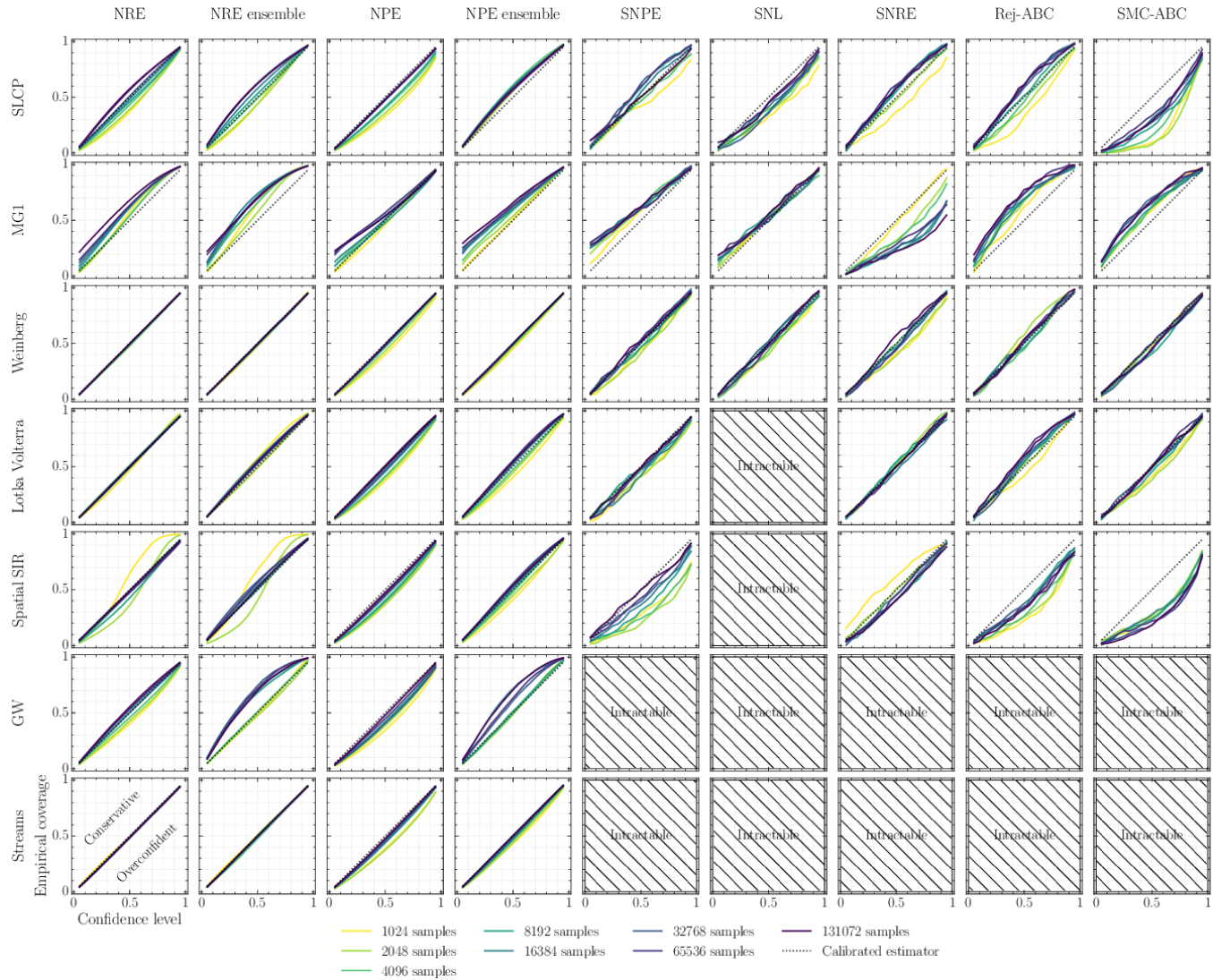




Coverage diagnostic

- For $x, \theta \sim p(x, \theta)$, compute the $1 - \alpha$ credible interval based on $\hat{p}(\theta|x)$.
- If the fraction of samples for which θ is contained within the interval is larger than the nominal coverage probability $1 - \alpha$, then the approximate posterior $\hat{p}(\theta|x)$ has coverage.





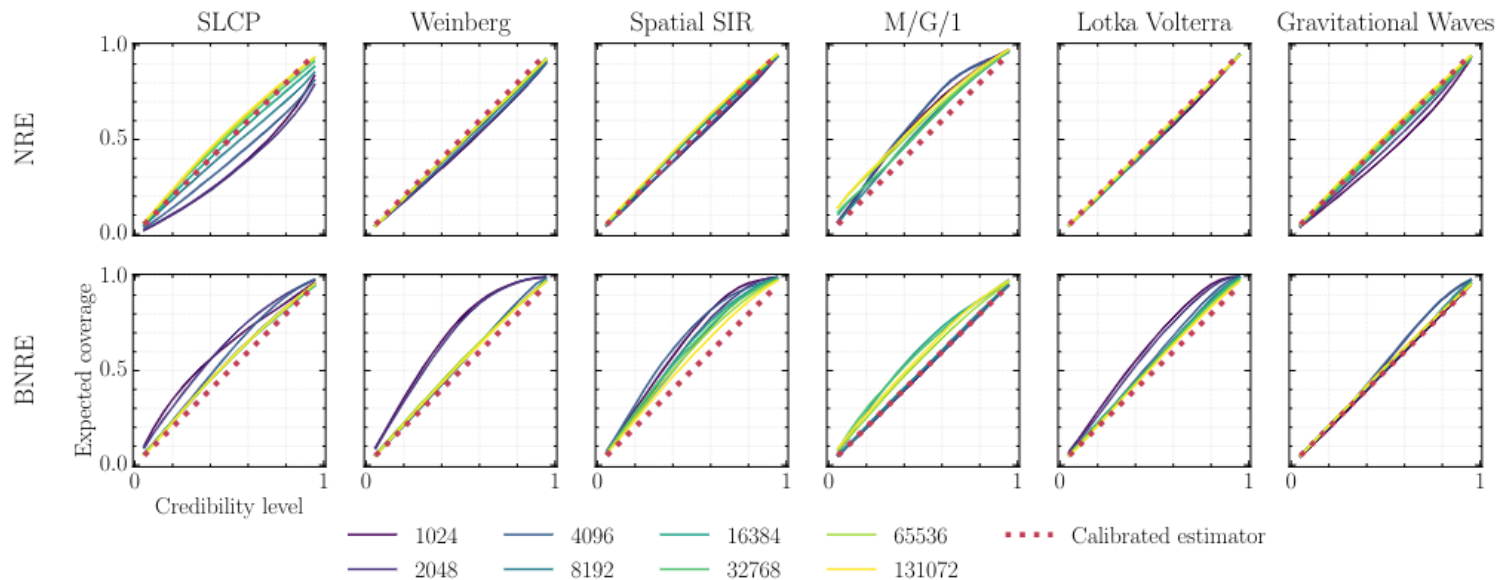
What if diagnostics fail?

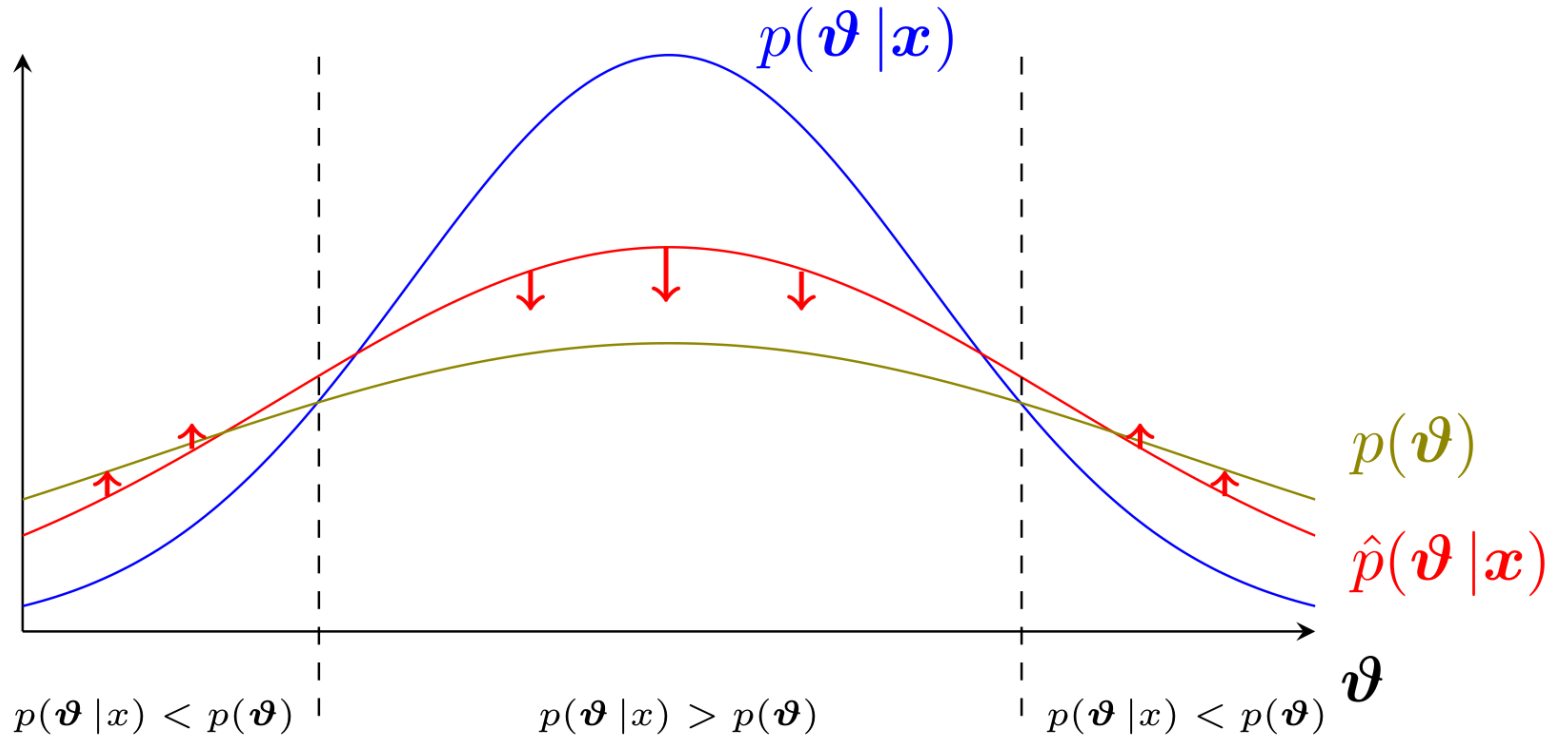
Balanced NRE



Enforce neural ratio estimation to be **conservative** by using binary classifiers \hat{d} that are balanced, i.e. such that

$$\mathbb{E}_{p(\theta, x)} \left[\hat{d}(\theta, x) \right] = \mathbb{E}_{p(\theta)p(x)} \left[1 - \hat{d}(\theta, x) \right].$$





Summary

Advances in deep learning have enabled new approaches to statistical inference.

This is major evolution in the statistical capabilities for science, as it enables the analysis of complex models and data without simplifying assumptions.

Inference remains approximate and requires careful validation.

Obstacles remain to be overcome, such as the curse of dimensionality and the need for large amounts of data.

