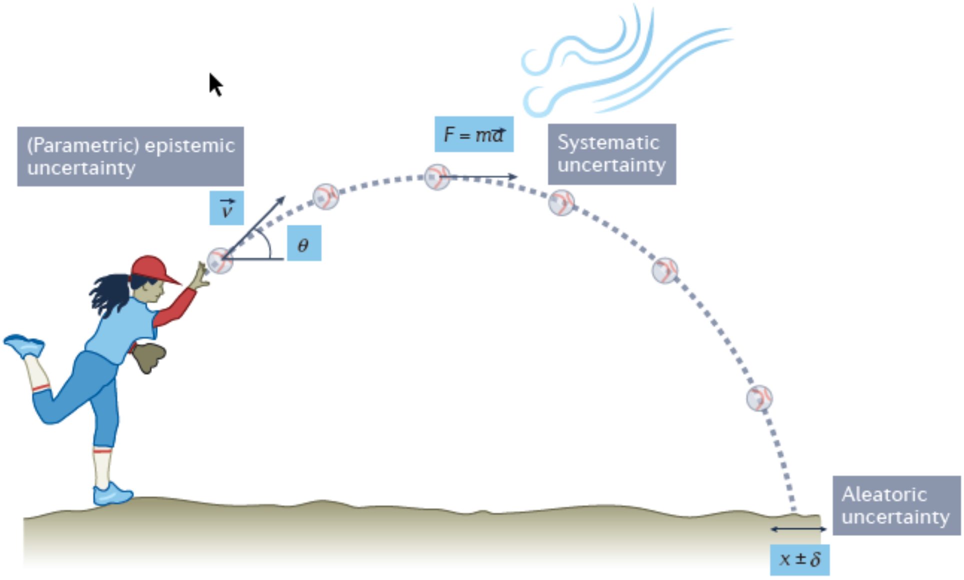
An introduction to simulation-based inference

AI and the uncertainty challenge in fundamental physics
November 28, 2023

Gilles Louppe
g.louppe@uliege.be
@glouppe

1 / 36


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

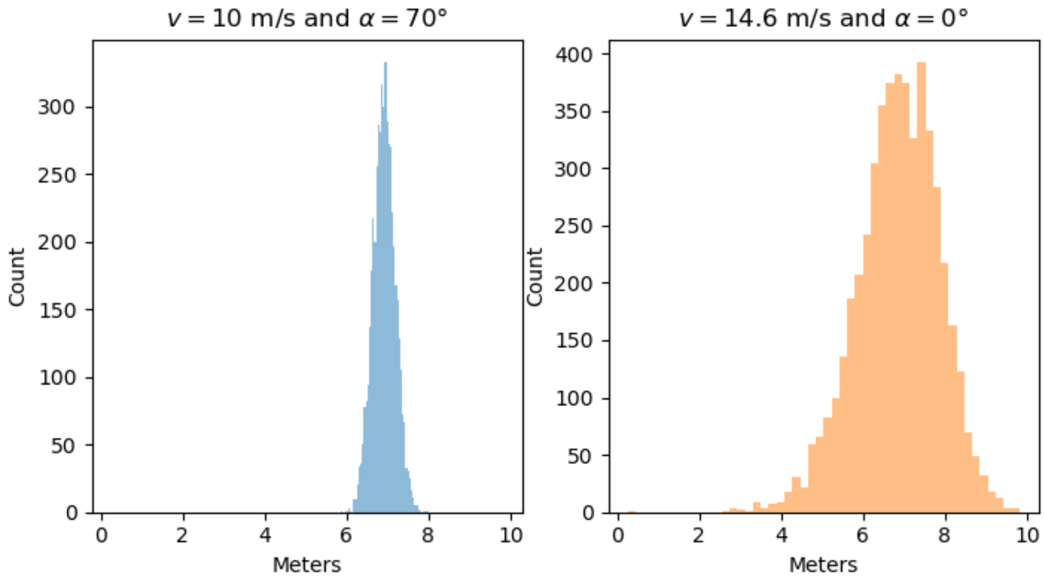
$$v_x = v\cos(\alpha), \;\; v_y = v\sin(\alpha),$$

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

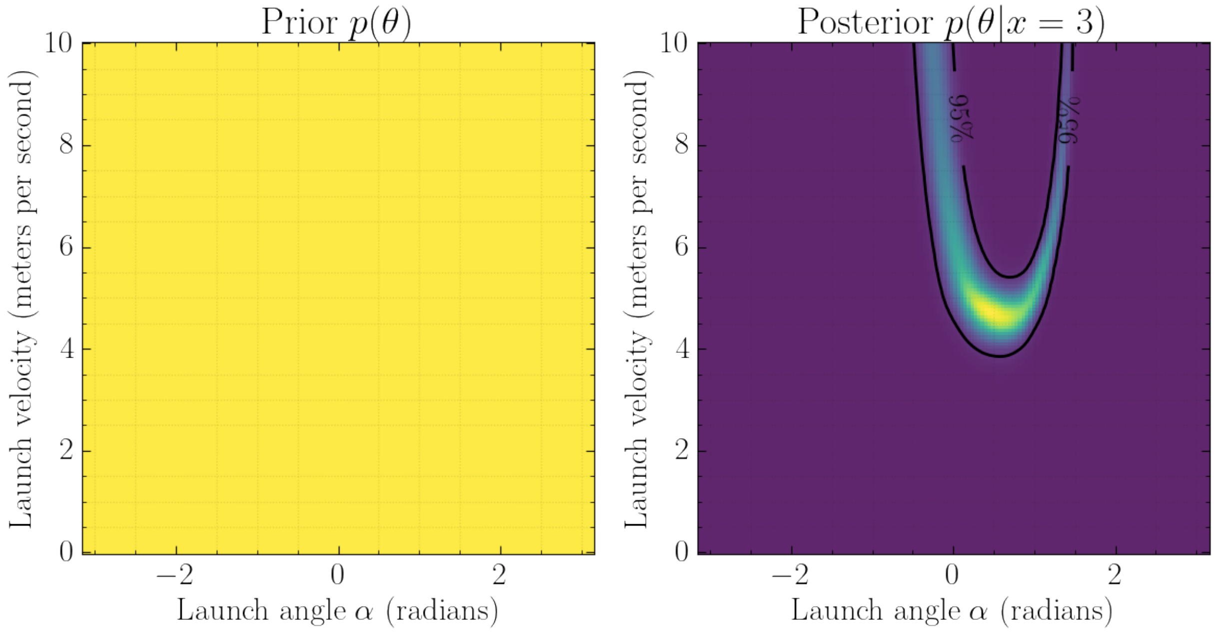
```python
def simulate(v, alpha, dt=0.001):
    v_x = v * np.cos(alpha)  # x velocity m/s
    v_y = v * np.sin(alpha)  # y velocity m/s
    y = 1.1 + 0.3 * random.normal()
    x = 0.0

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

What parameter values θ are the most plausible?



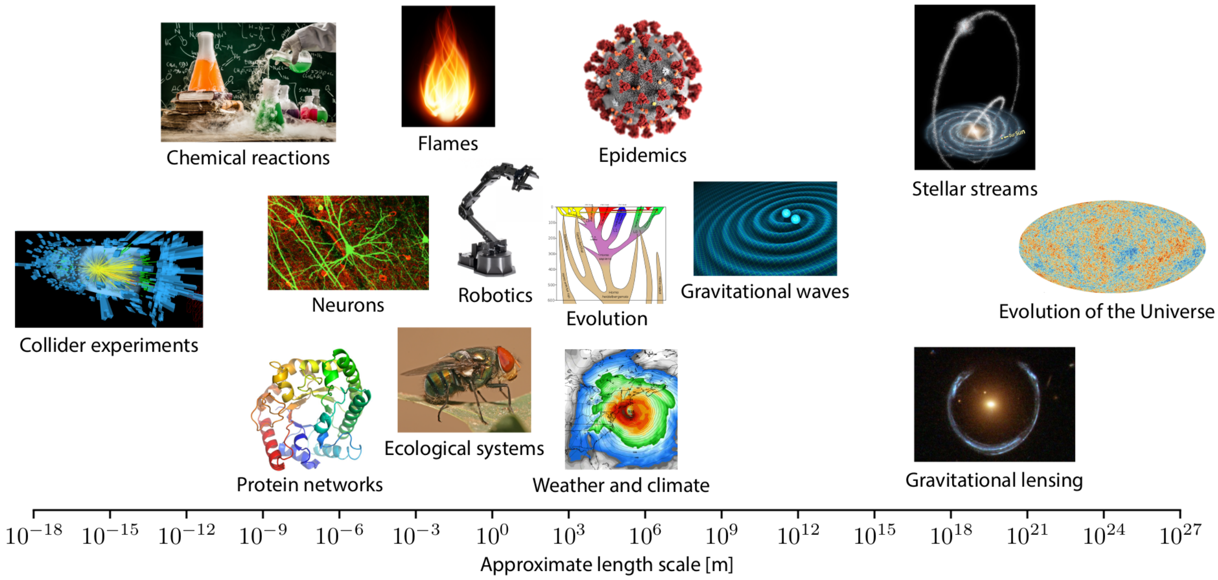
Prior $p(\theta)$
Launch velocity (meters per second)
Launch angle α (radians)
Posterior $p(\theta|x = 3)$
95%
95%
Launch velocity (meters per second)
Launch angle α (radians)

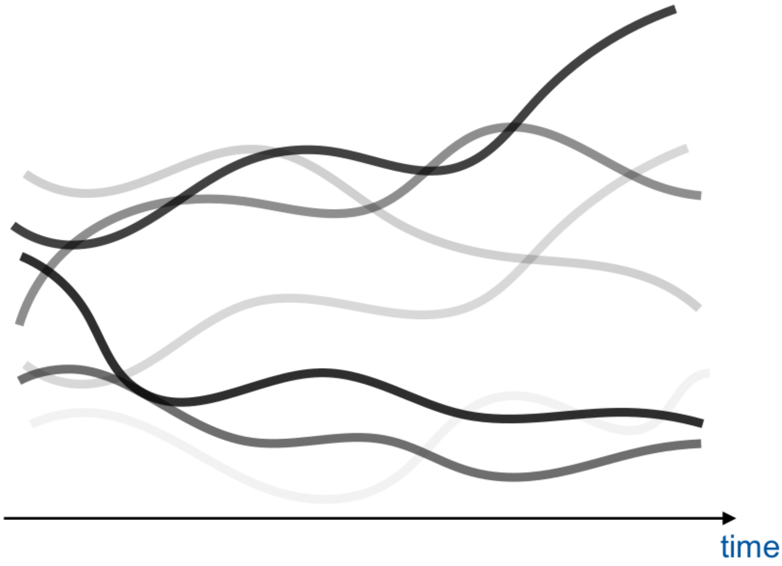
Outline

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

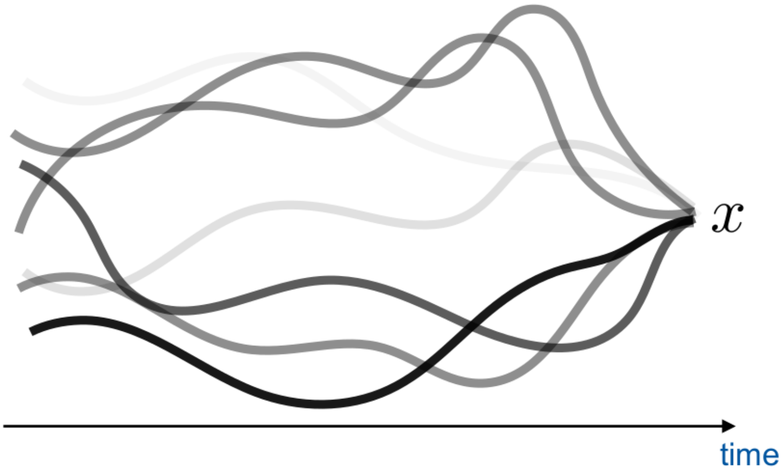
Simulation-based inference

Scientific simulators





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



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

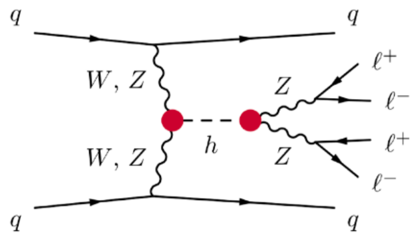
Latent variables

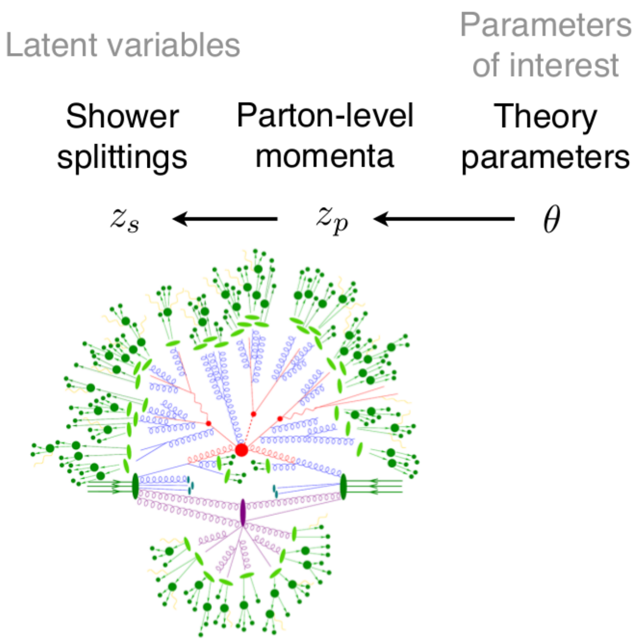
Parameters of interest

Parton-level momenta

Theory parameters

$z_p \longleftarrow \theta$

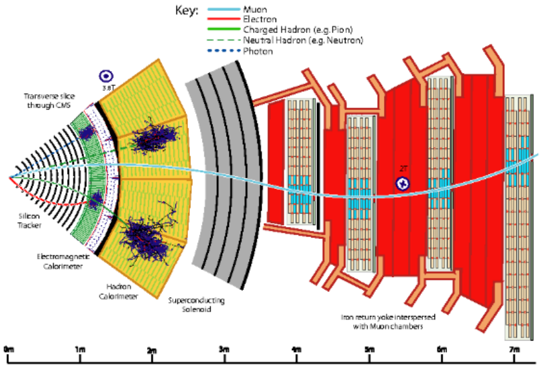



Latent variables
Parameters of interest
Shower splittings
Parton-level momenta
Theory parameters
z_s ← z_p ← θ

Latent variables

Parameters of interest

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

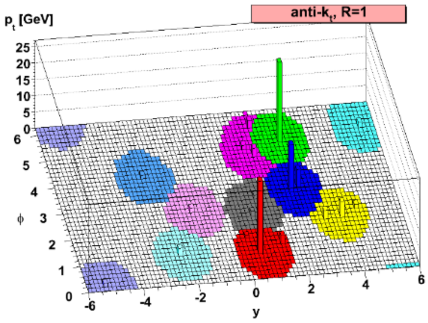


12 / 36

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]

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

Bayesian inference

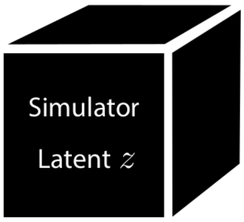
Start with

- a simulator that can generate N samples $x_i \sim p(x_i|\theta_i)$,
- a prior model $p(\theta)$,
- observed data x_{obs}.

Then, estimate the posterior

$$p(\theta|x_{\text{obs}}) = \frac{p(x_{\text{obs}}|\theta)p(\theta)}{p(x_{\text{obs}})}.$$

Parameters θ →



→ Observables x

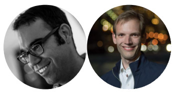
Prediction:

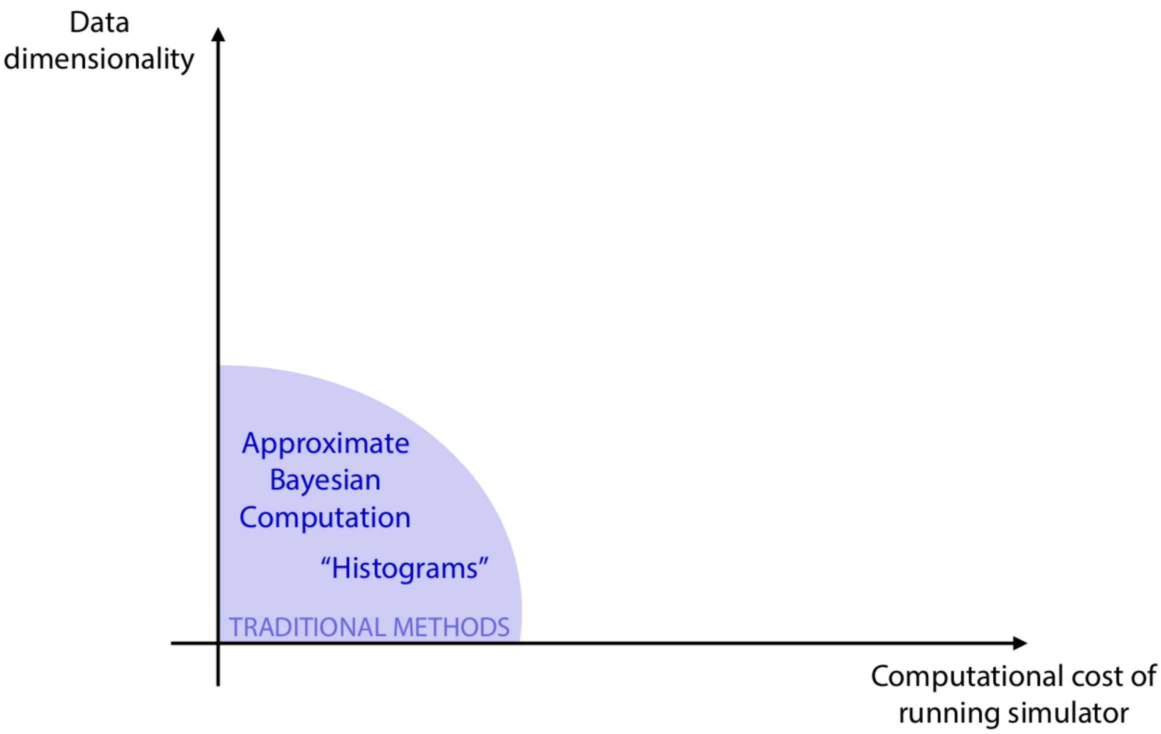
- Well-motivated mechanistic, causal model
- Simulator can generate samples $x \sim p(x|\theta)$

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

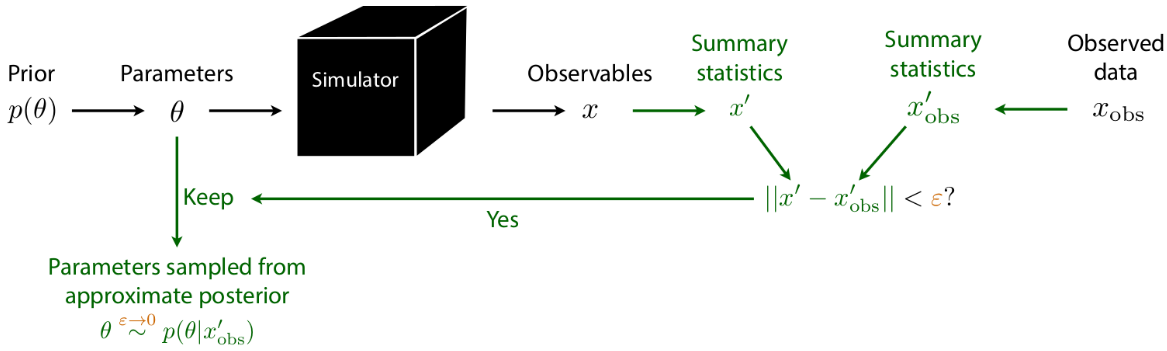
Algorithms





—
Credits: Cranmer, Brehmer and Louppe, 2020.

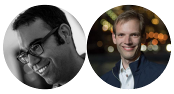
Approximate Bayesian Computation (ABC)

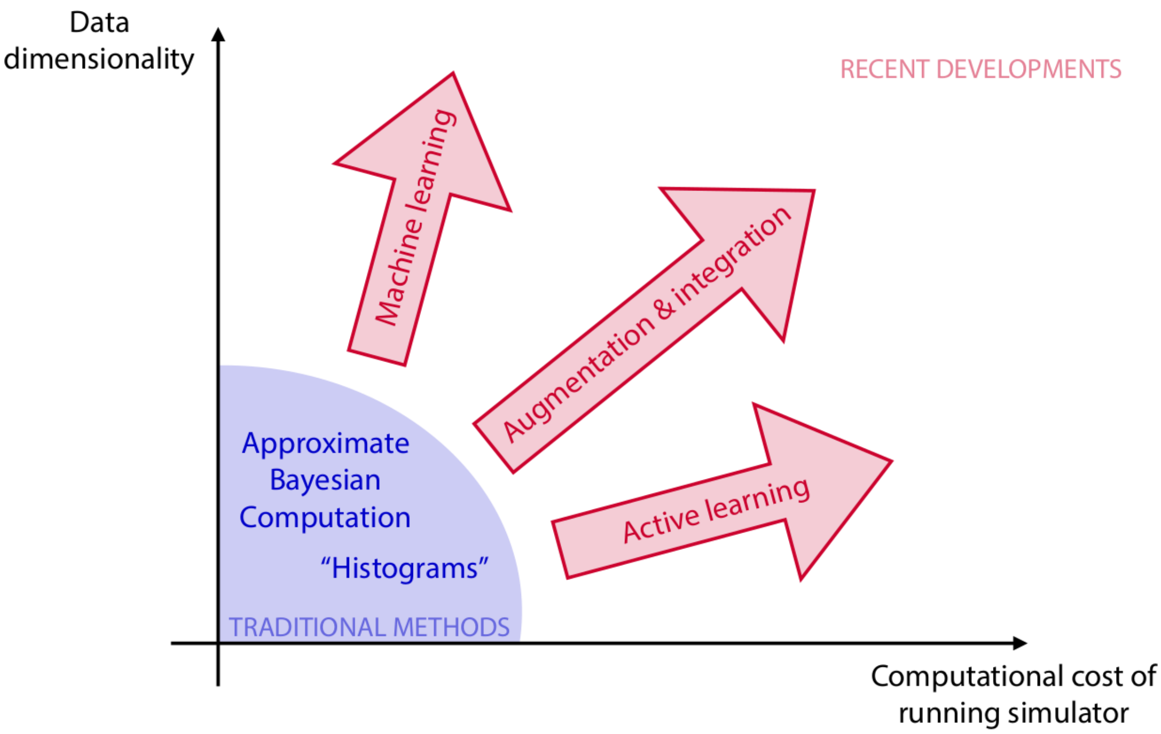


Issues:

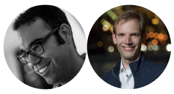
- How to choose x'? ϵ? $||\cdot||$?
- No tractable posterior.
- Need to run new simulations for new data or new prior.

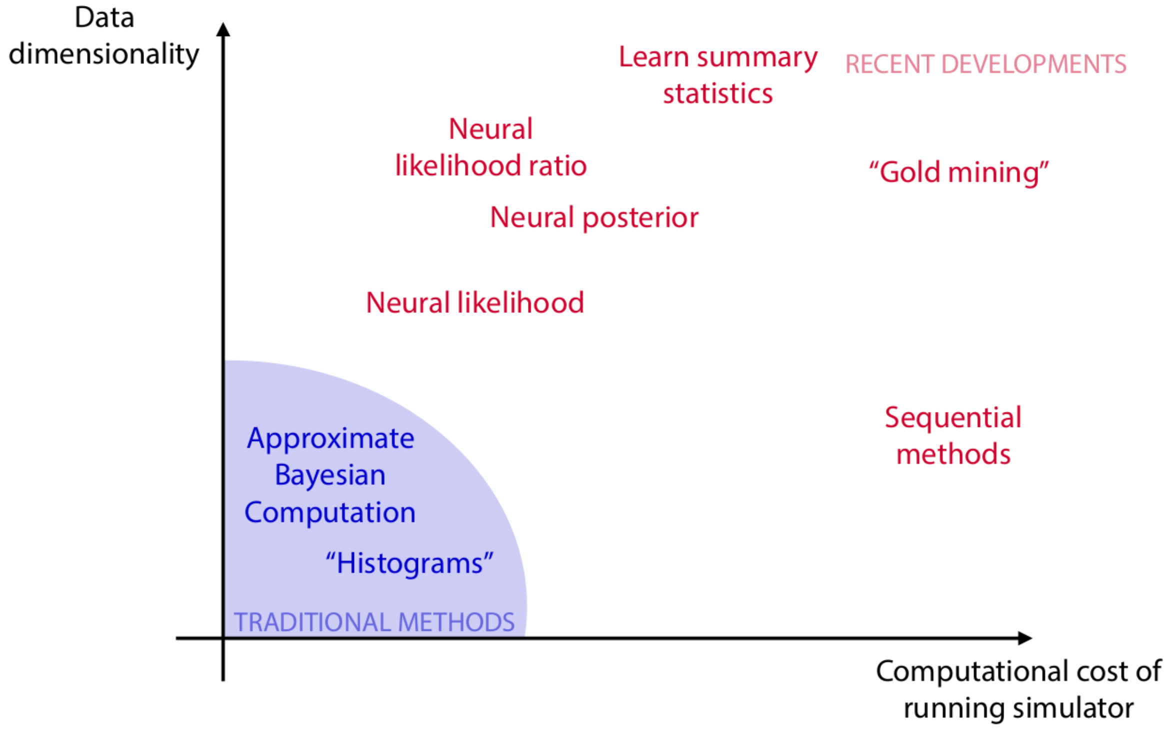
—
Credits: Johann Brehmer.





—
Credits: Cranmer, Brehmer and Louppe, 2020.





—

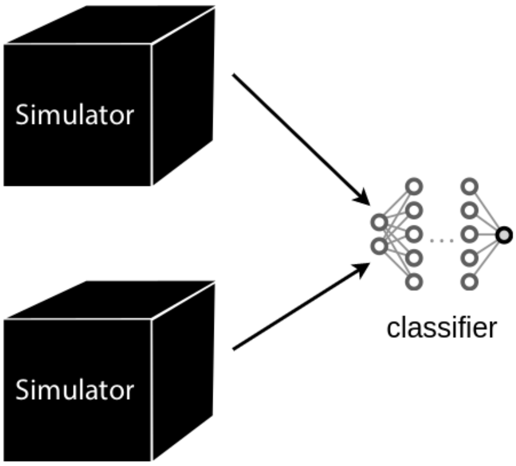
Credits: Cranmer, Brehmer and Louppe, 2020.

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:

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

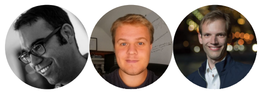
$x, \theta \sim p(x)p(\theta)$



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

—

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

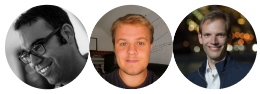


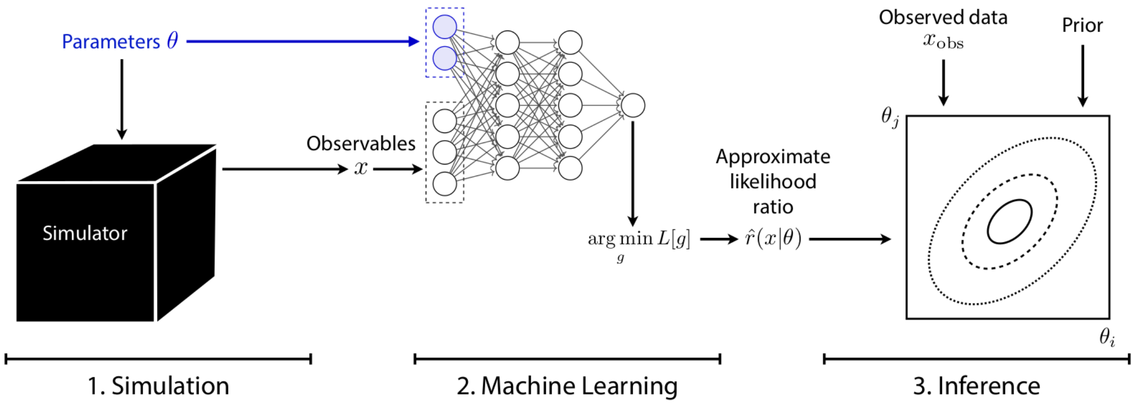
The solution d found after training approximates the optimal classifier

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

Therefore,

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





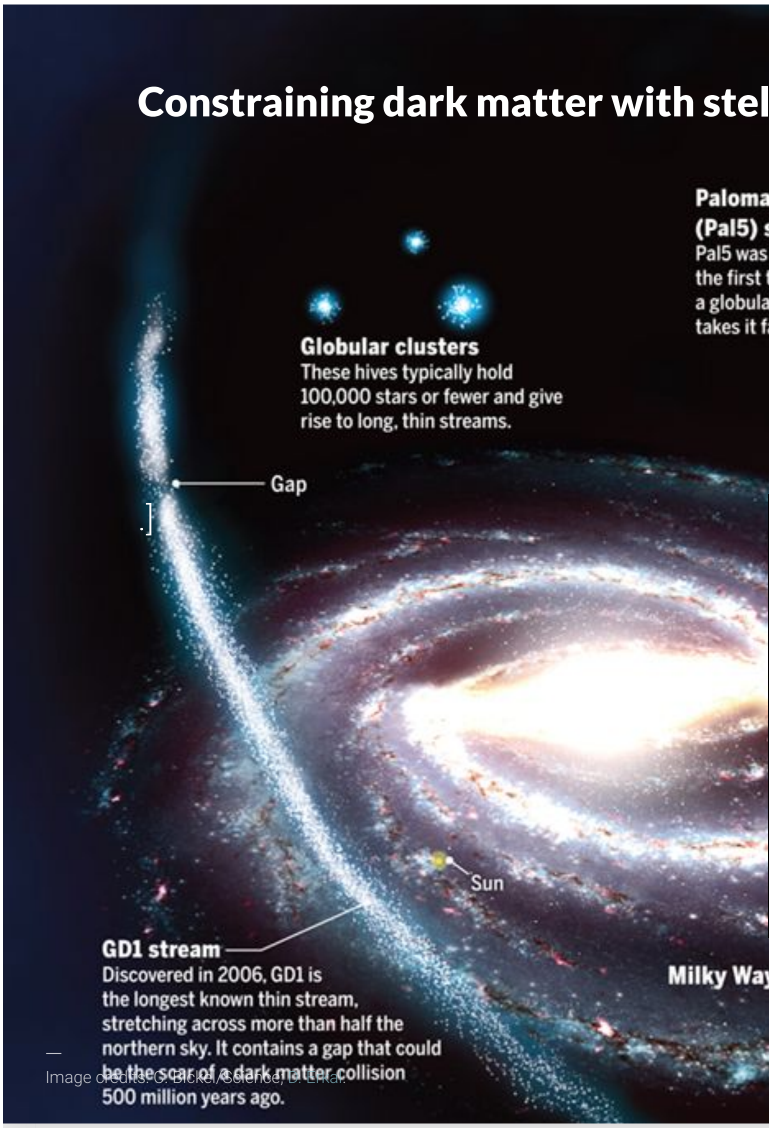
Run simulator and save data

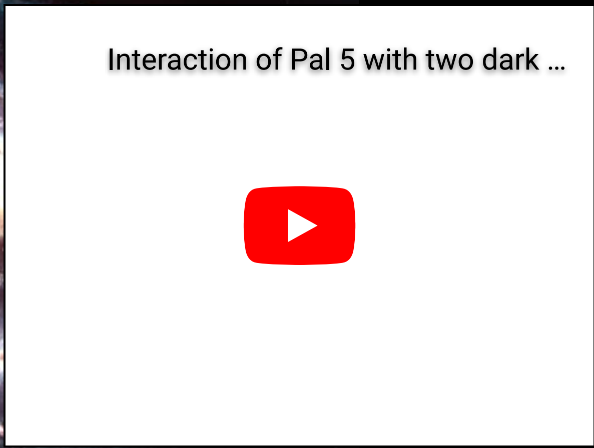
Train NN classifier, interpret as likelihood ratio estimator

Amortized: cheap to repeat for new data

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

Constraining dark matter with stellar streams

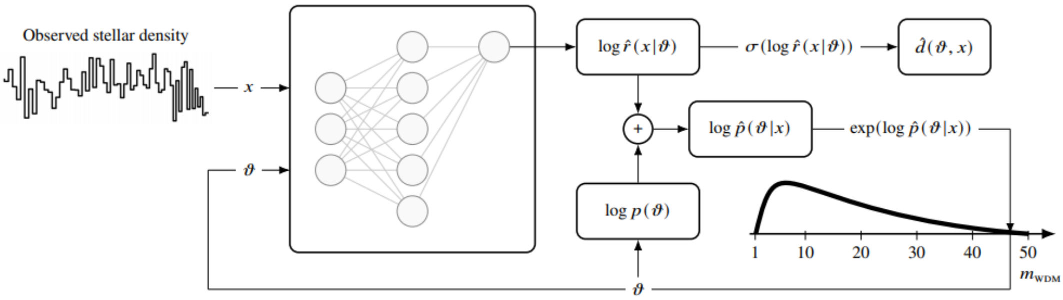


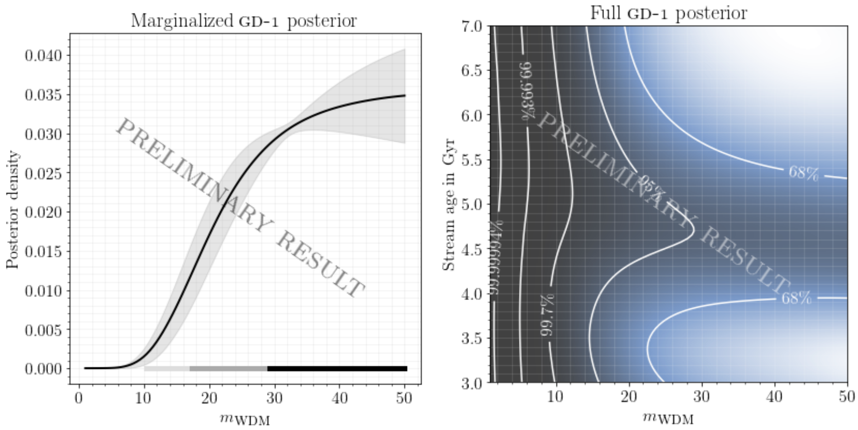


—

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

22 / 36



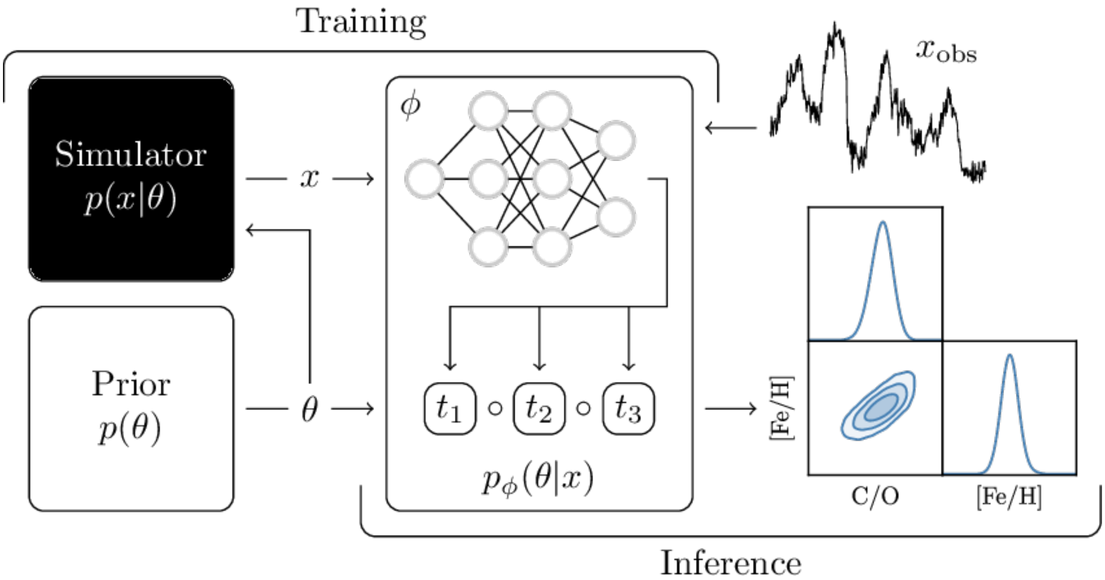


—

Credits: Hermans et al, 2021.

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

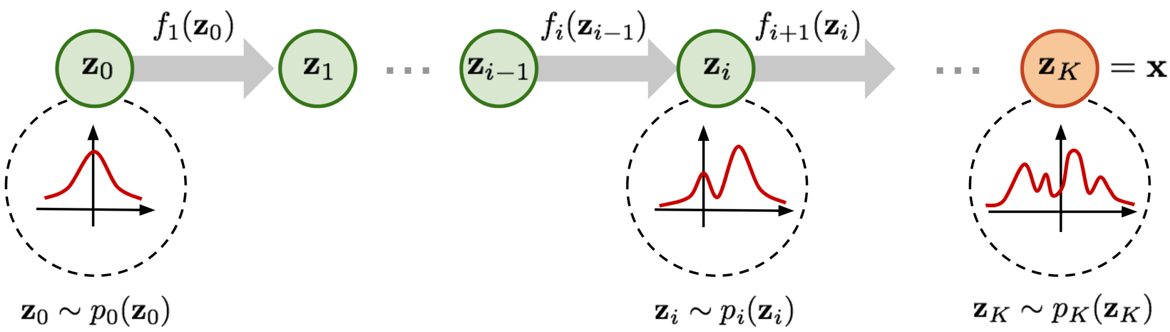
Neural Posterior Estimation



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

Normalizing flows

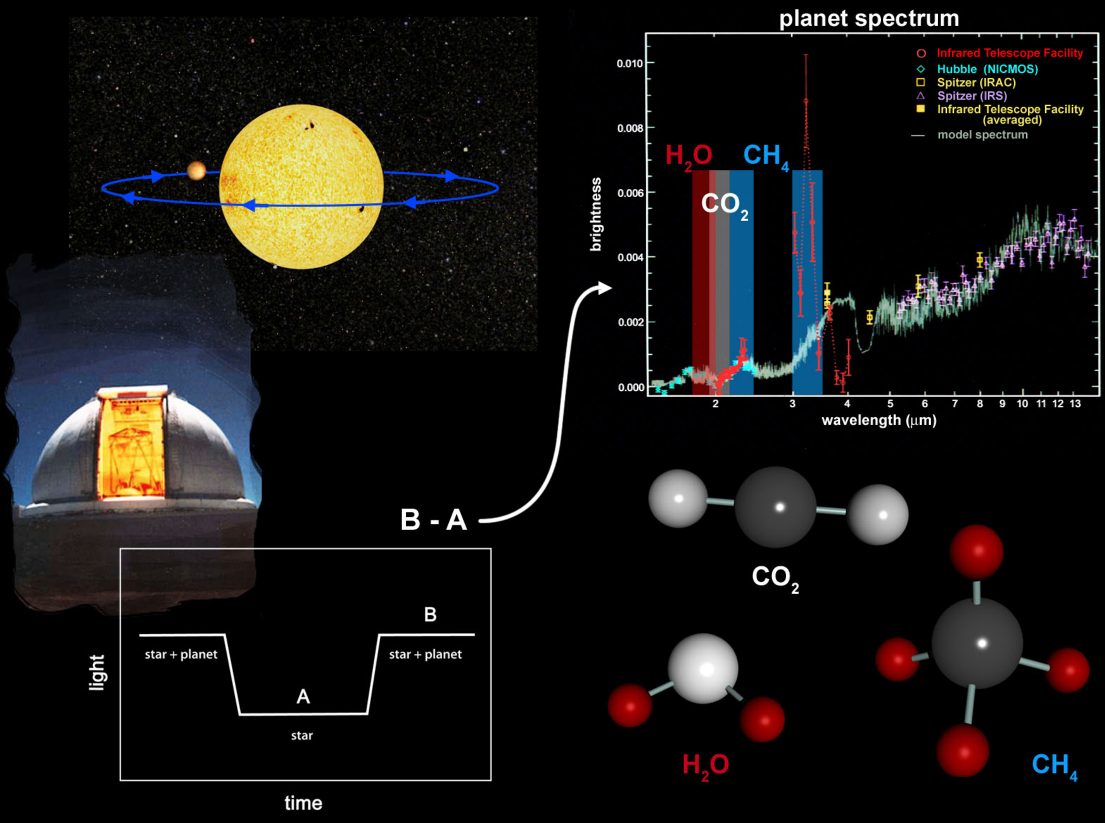
A normalizing flow is a sequence of invertible transformations f_k that map a simple distribution p_0 to a more complex distribution p_K:



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

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

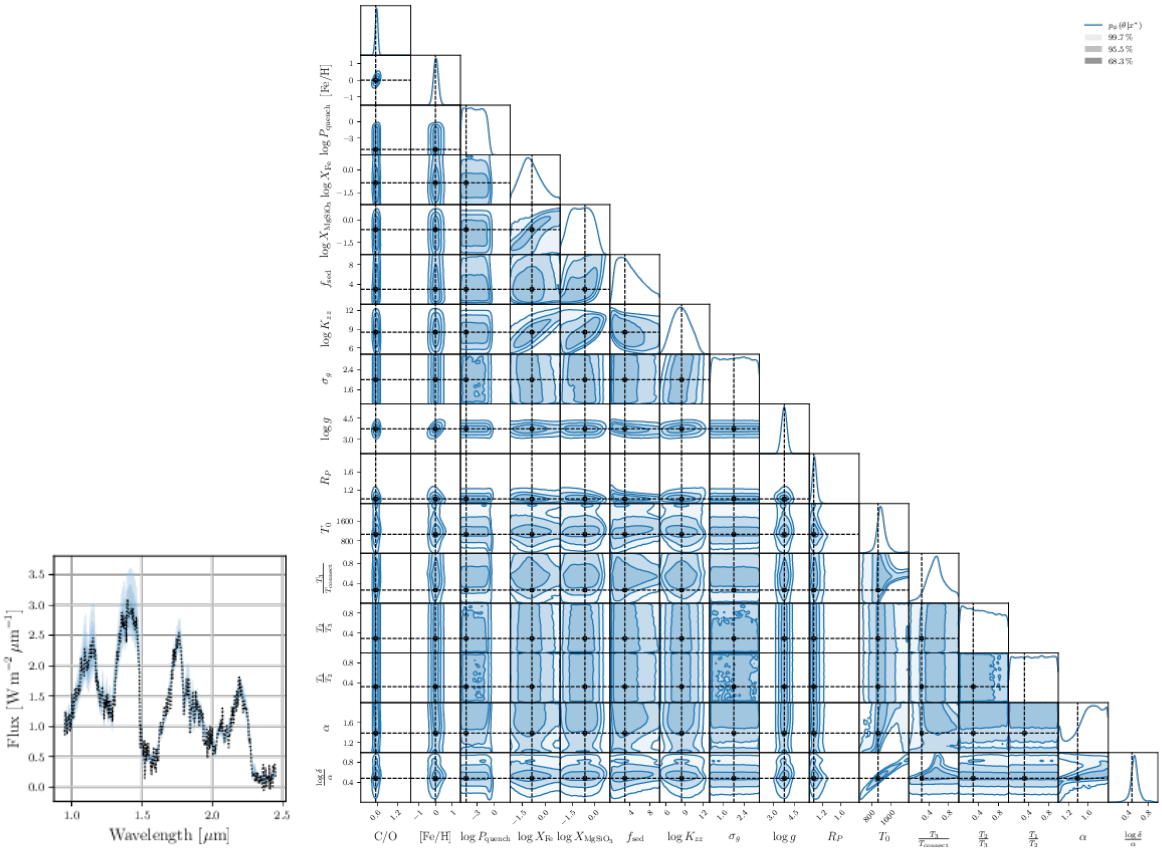
Exoplanet atmosphere characterization



—

Credits: NSA/JPL-Caltech, 2010.

27 / 36



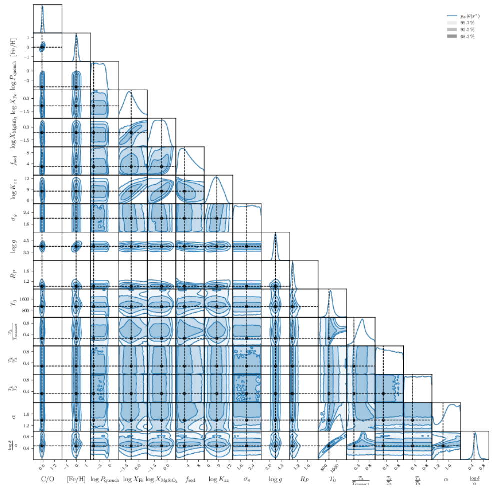
—
Credits: Vasist et al, 2023.

28 / 36

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.

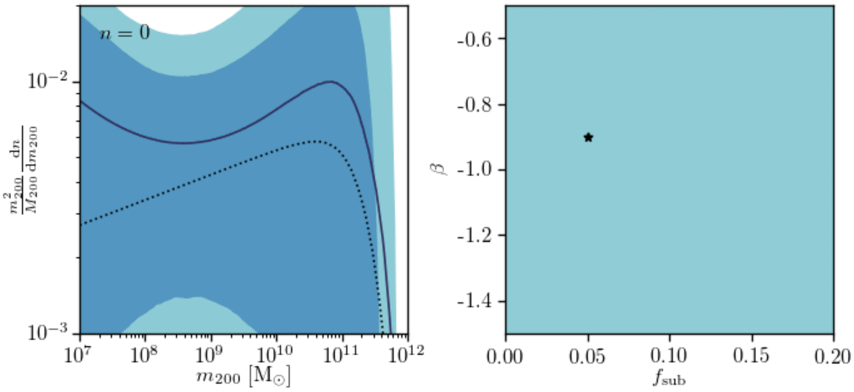


How certain are you of your uncertainties?

Mode convergence

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

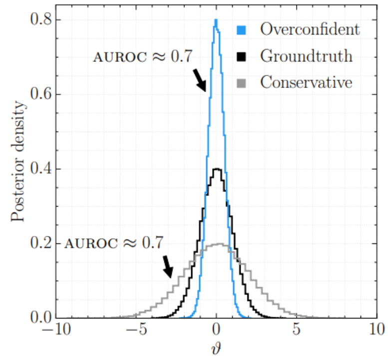
$$\lim_{N \to \infty} \arg\max_\theta p(\theta|\{x_i\}_{i=1}^N)$$
$$= \lim_{N \to \infty} \arg\max_\theta p(\theta) \prod_{x_i} r(x_i|\theta) = \theta^*$$

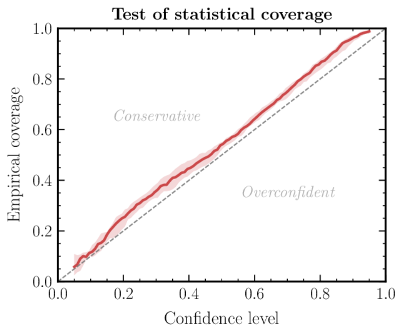


—
Credits: Brehmer et al, 2019.

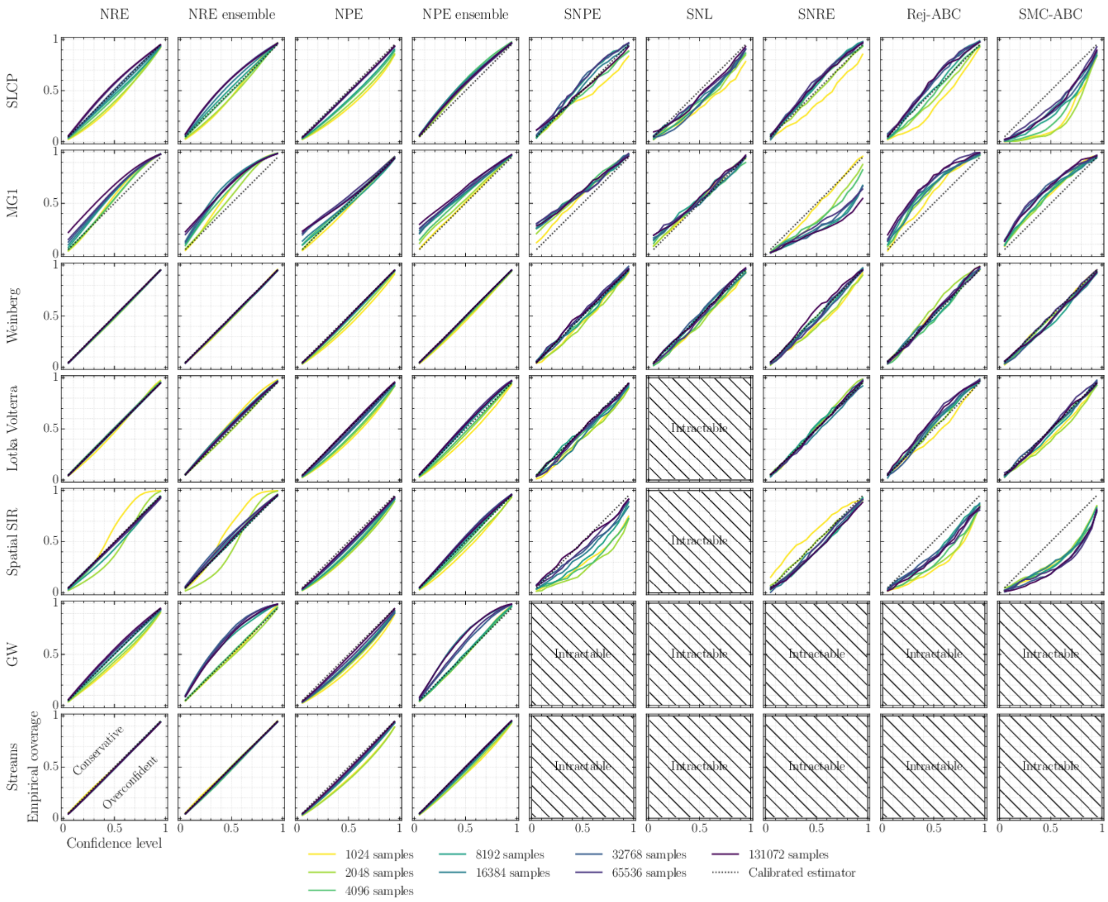
Coverage diagnostic

- 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.





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



—

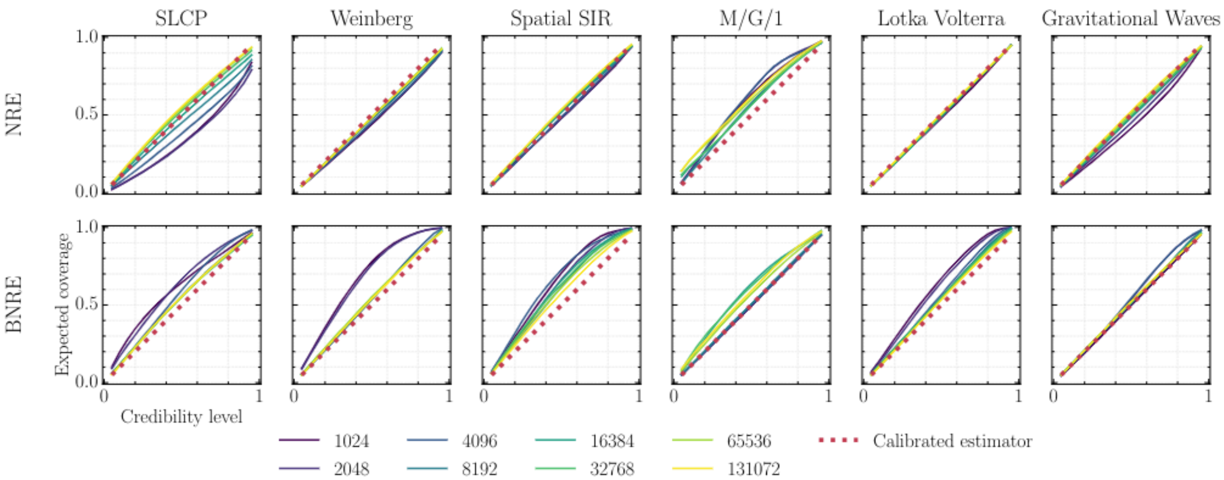
Credits: Hermans et al, 2021.

What if diagnostics fail?

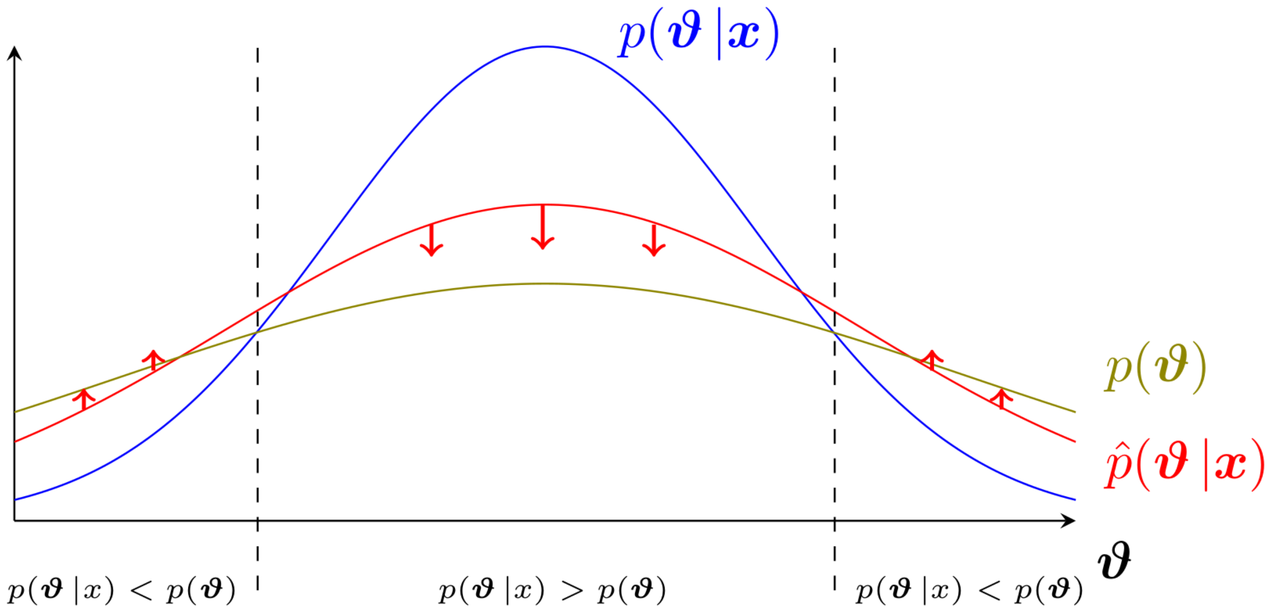
Balanced NRE

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

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



—
Credits: Delaunoy et al, 2022.



—
Credits: Delaunoy et al, 2022.

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.

The end.

36 / 36