Simulation-based inference and its applications in neuroscience

March 31, 2023

Gilles Louppe
g.louppe@uliege.be

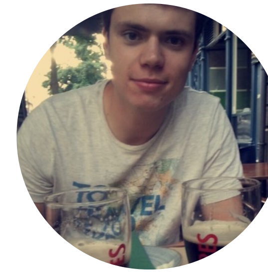
LIÈGE université


Kyle Cranmer


Johann Brehmer


Joeri Hermans


Antoine Wehenkel


Norman Marlier


Siddharth Mishra-Sharma

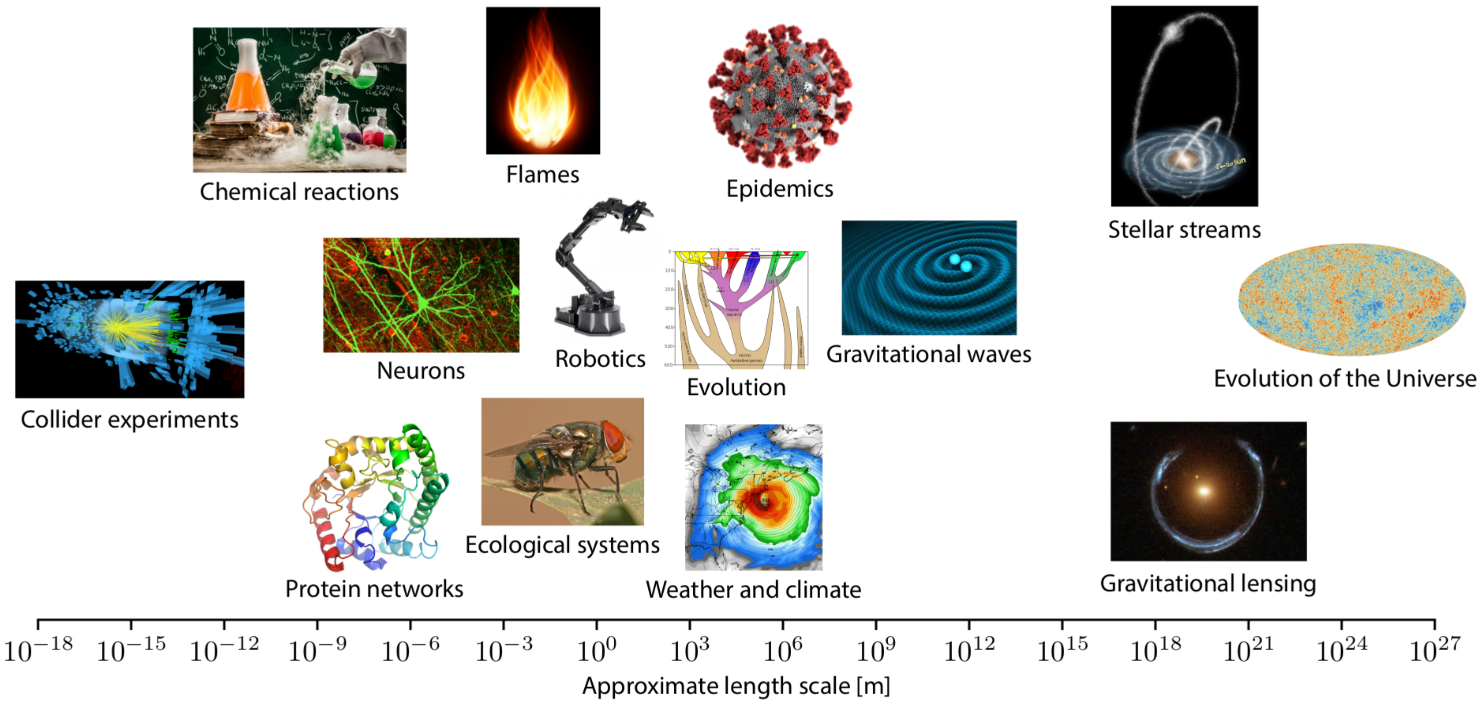

Christoph Weniger

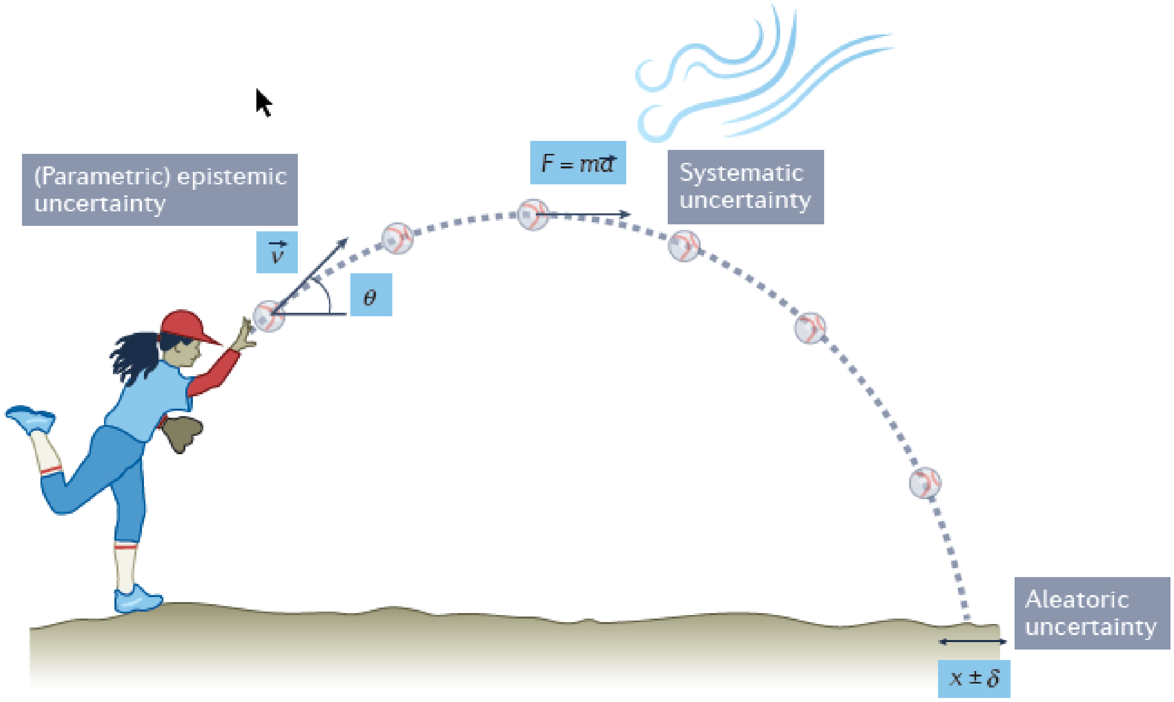

Arnaud Delaunoy


Malavika Vasist


Francois Rozet

xxx update with new faces: pedro, alex


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]

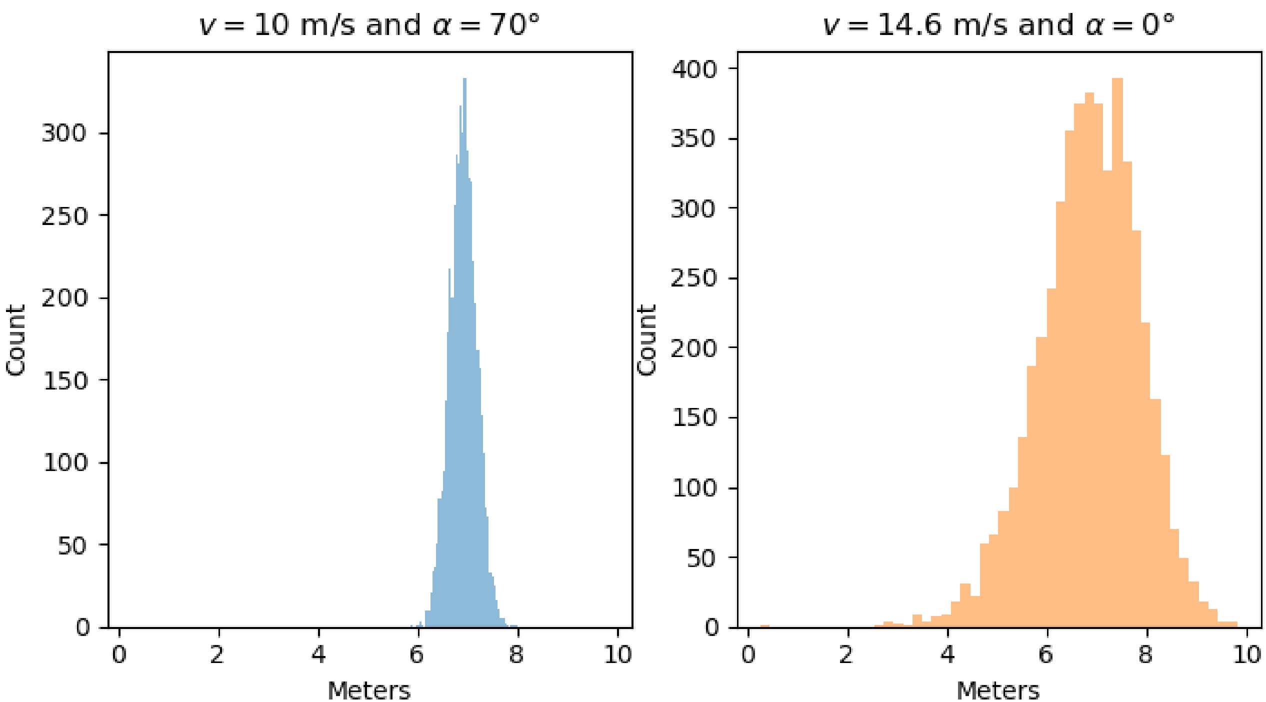

(Parametric) epistemic uncertainty
$\vec{v}$
θ
$F = m\vec{a}$
Systematic uncertainty
Aleatoric uncertainty
$x \pm \delta$

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

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

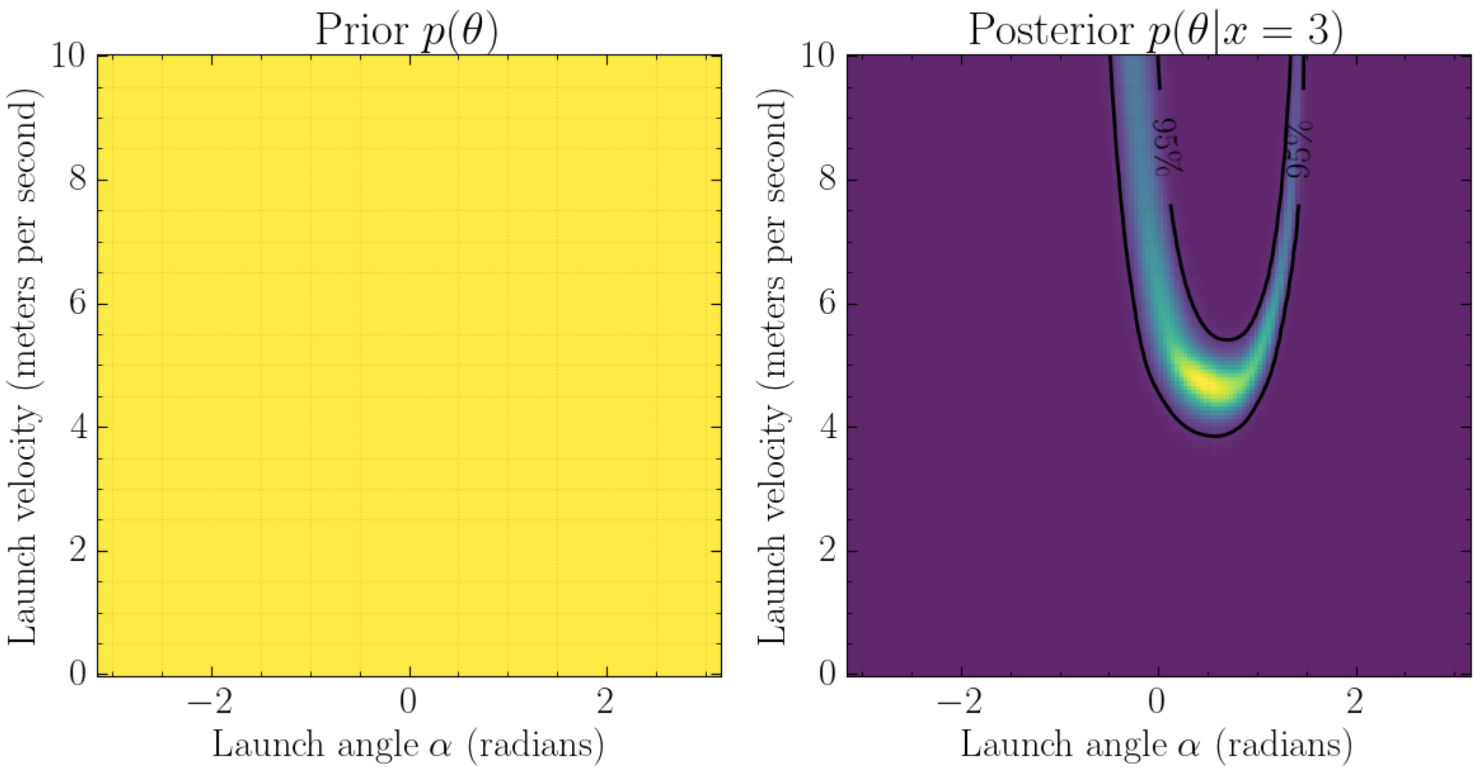
```python
def simulate(v, alpha, dt=0.001):
    v_x = v * np.cos(alpha)  # x velocity m/s
    v_y = v * np.sin(alpha)  # y velocity m/s
    y = 1.1 + 0.3 * random.normal()
    x = 0.0

    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()
```

The computer simulator defines the likelihood function $p(x|\theta)$ implicitly.



What parameter values θ are the most plausible?

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

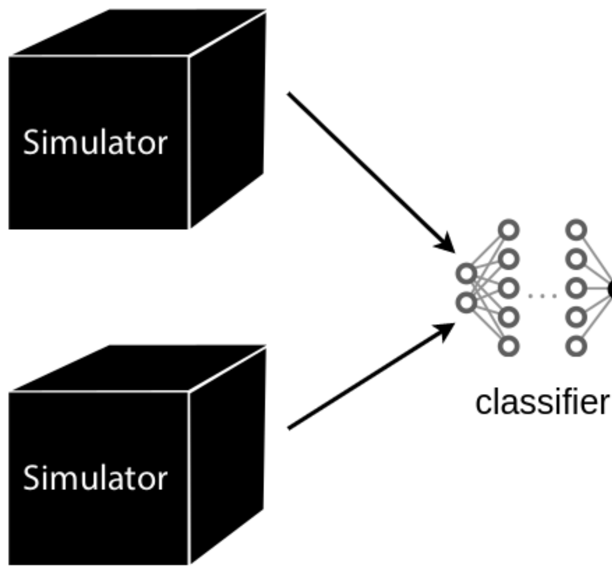

Prior $p(\theta)$
Launch velocity (meters per second)
Launch angle α (radians)
0
2
4
6
8
10
−2
0
2
Posterior $p(\theta|x = 3)$
Launch velocity (meters per second)
Launch angle α (radians)
95%
95%

Neural ratio estimation (NRE)

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:

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

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



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

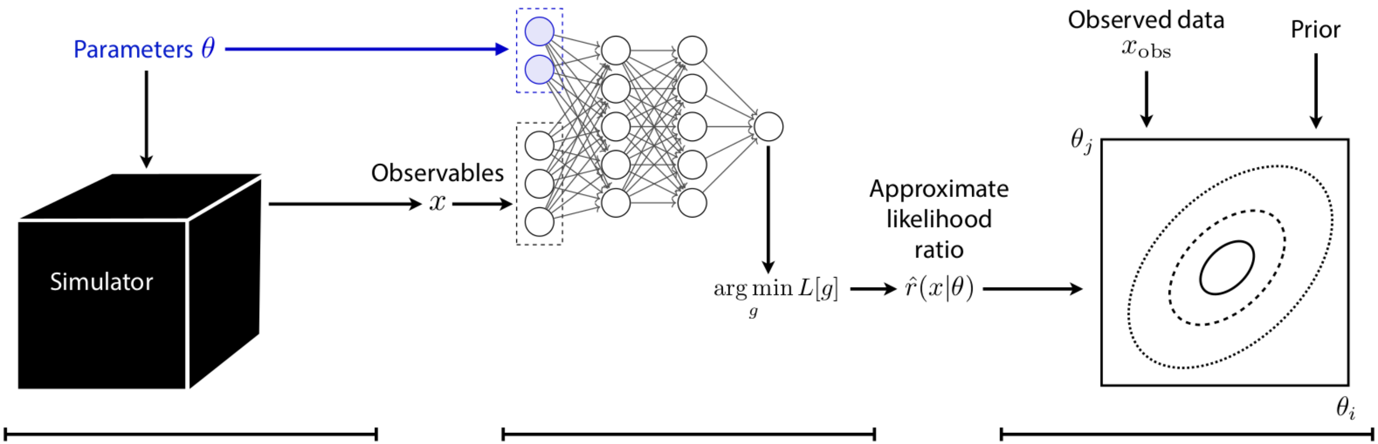
—
Credits: Cranmer et al, 2015; Hermans et al, 2020.

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,

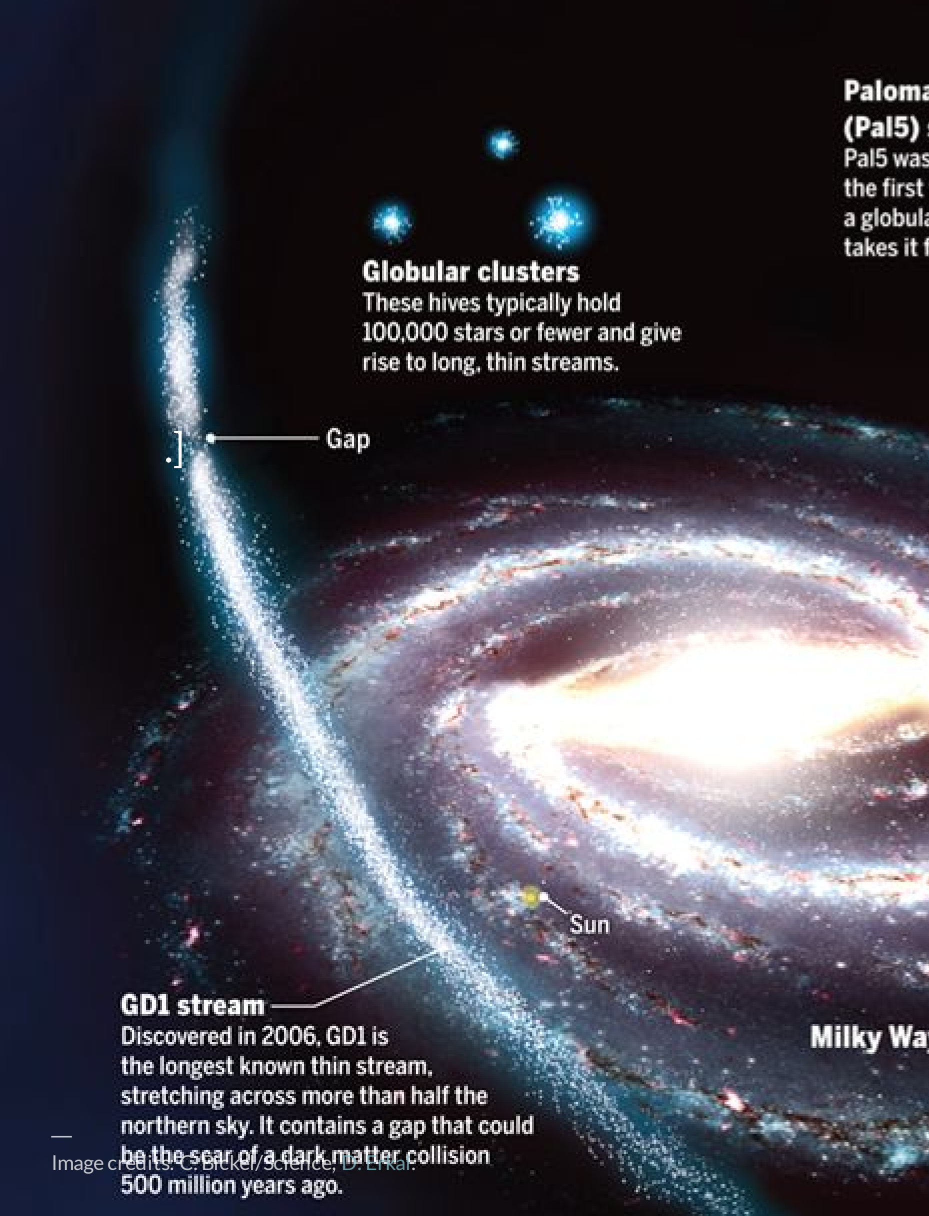
$$r(x|\theta) = \frac{p(x|\theta)}{p(x)} = \frac{p(x,\theta)}{p(x)p(\theta)} \approx \frac{d(x,\theta)}{1 - d(x,\theta)} = \hat{r}(x|\theta).$$



1. Simulation	2. Machine Learning	3. Inference
Run simulator and save data	Train NN classifier, interpret as likelihood ratio estimator	Amortized: cheap to repeat for new data

$$p(\theta|x) \approx r(x|\theta)p(\theta)$$

Constraining dark matter with stellar streams (Bayesian)



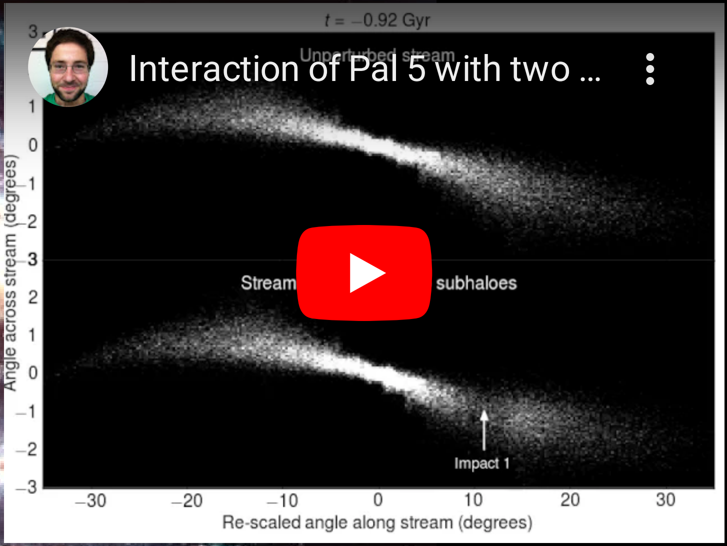
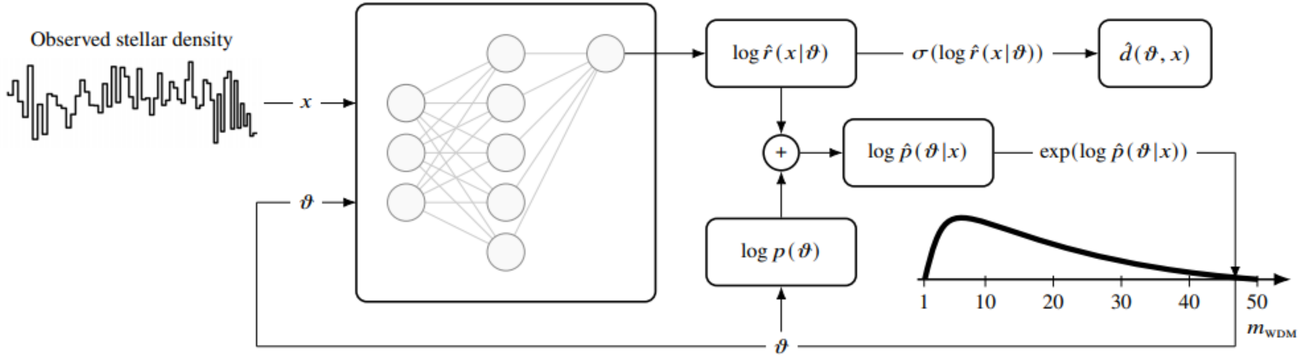
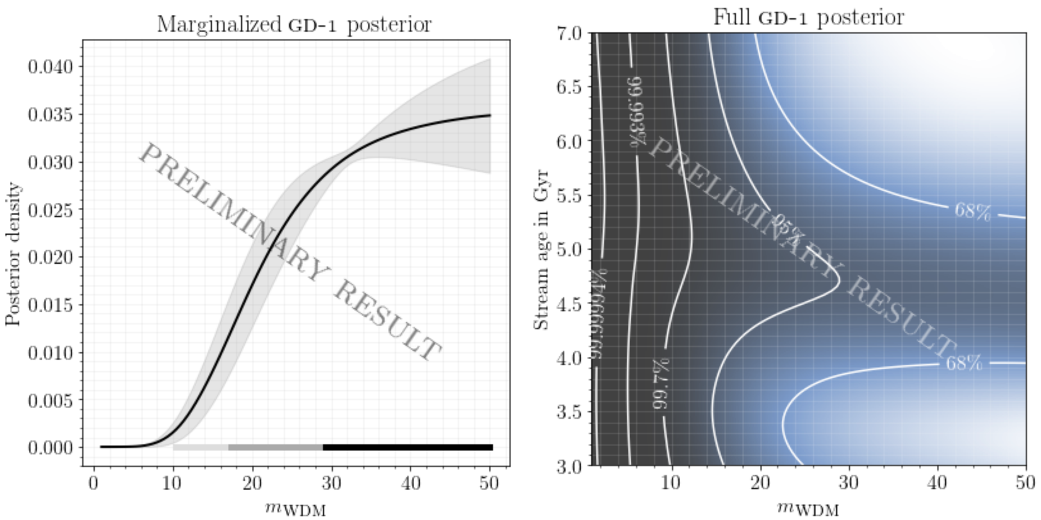


Image credits: C. Bickel/Science; D. Erkal.



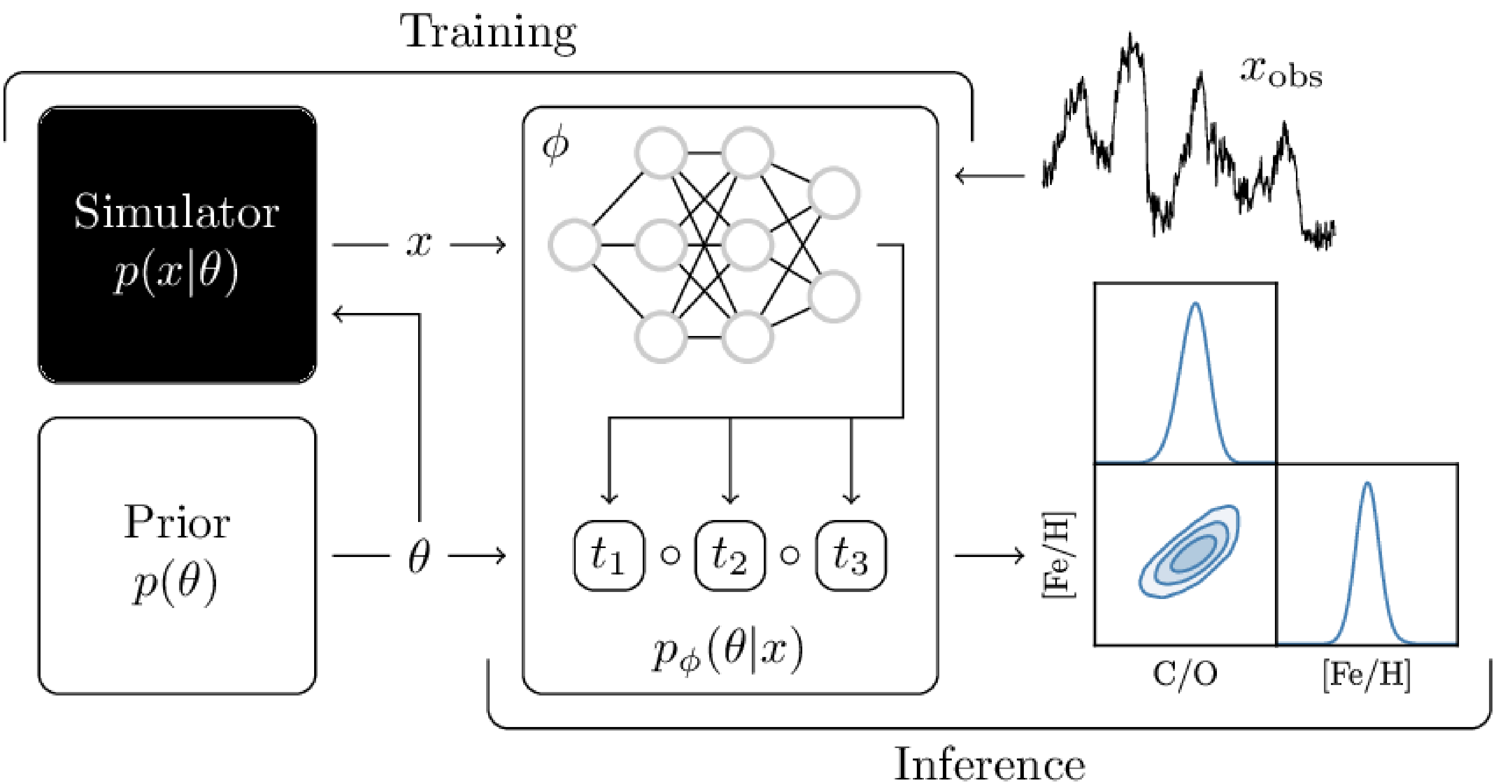
Observed stellar density
x
ϑ
log r̂(x|ϑ)
σ(log r̂(x|ϑ))
d̂(ϑ, x)
log p̂(ϑ|x)
exp(log p̂(ϑ|x))
log p(ϑ)
1
10
20
30
40
50
m_WDM



Marginalized GD-1 posterior
Posterior density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
10
20
30
40
50
mWDM
PRELIMINARY RESULT
Full GD-1 posterior
Stream age in Gyr
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
99.993%
99.99994%
99.7%
95%
68%
68%
PRELIMINARY RESULT

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

Neural Posterior Estimation (NPE)

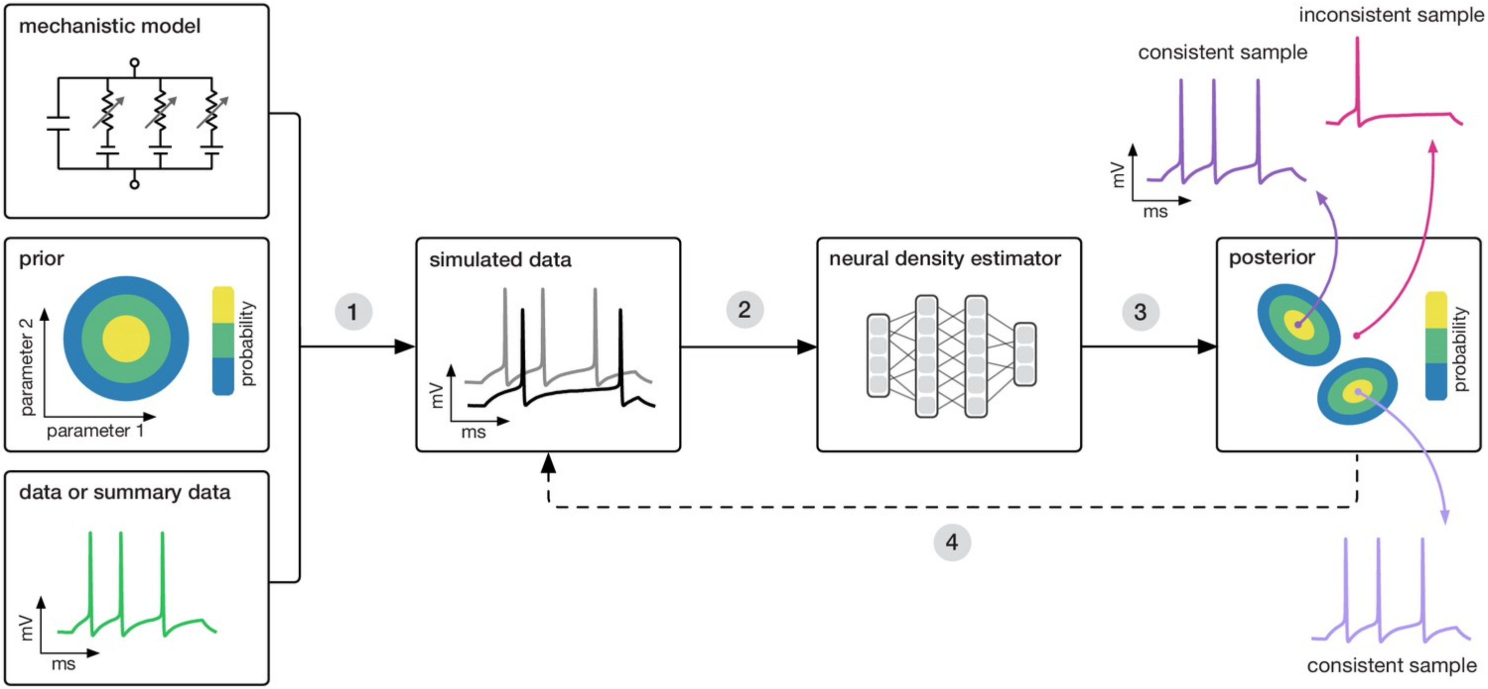


Use variational inference to directly estimate the posterior, by solving

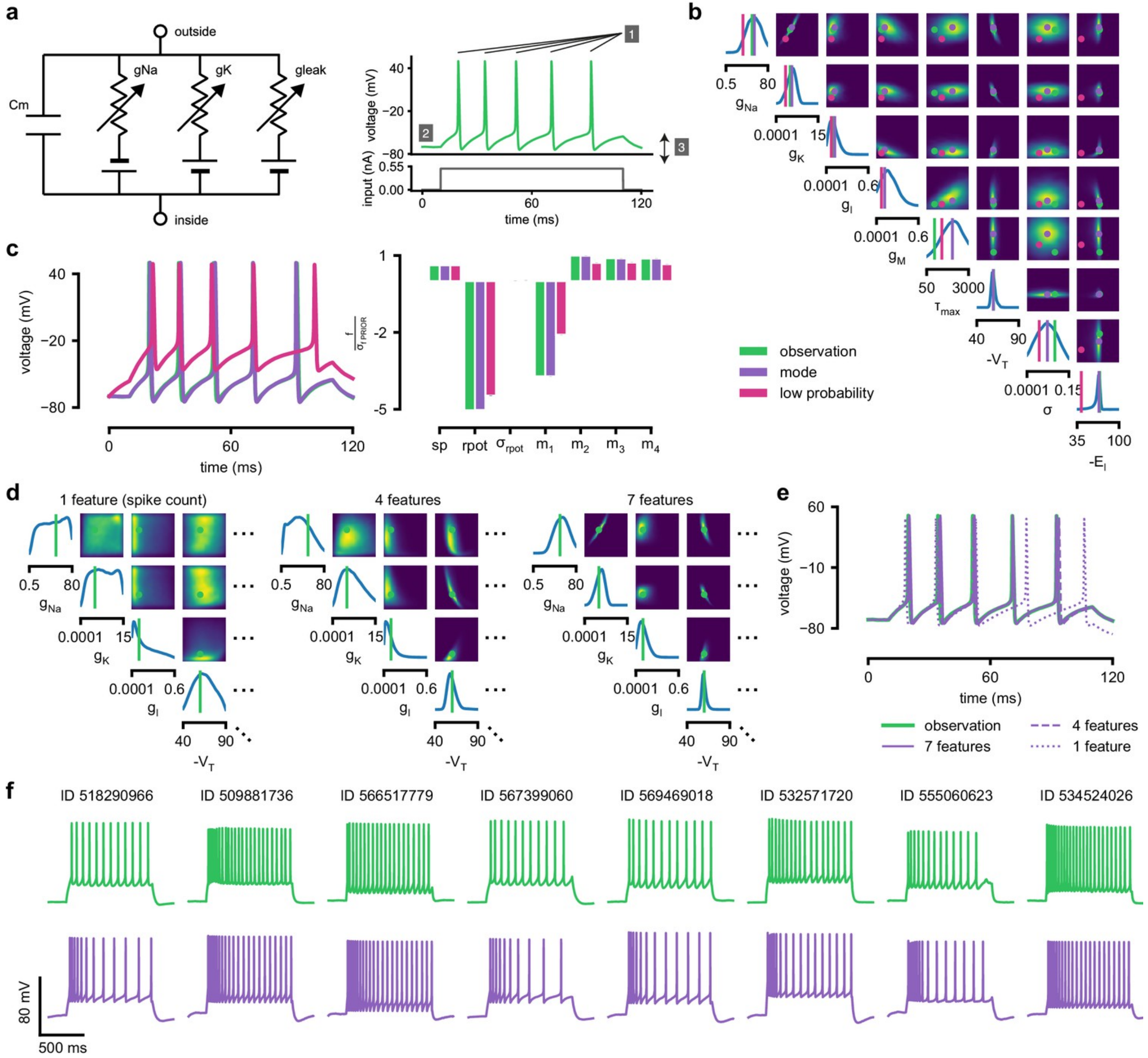
$$
\begin{aligned}
&\min_{q_\phi} \mathbb{E}_{p(x)} \left[\mathrm{KL}(p(\theta|x) || q_\phi(\theta|x)) \right] \\
&= \min_{q_\phi} \mathbb{E}_{p(x)} \mathbb{E}_{p(\theta|x)} \left[\log \frac{p(\theta|x)}{q_\phi(\theta|x)} \right] \\
&= \max_{q_\phi} \mathbb{E}_{p(x,\theta)} \left[\log q_\phi(\theta|x) \right]
\end{aligned}
$$

where q_ϕ is a neural density estimator, such as a normalizing flow.

Mechanistic models of neural dynamics (Gonçalves et al, 2022)



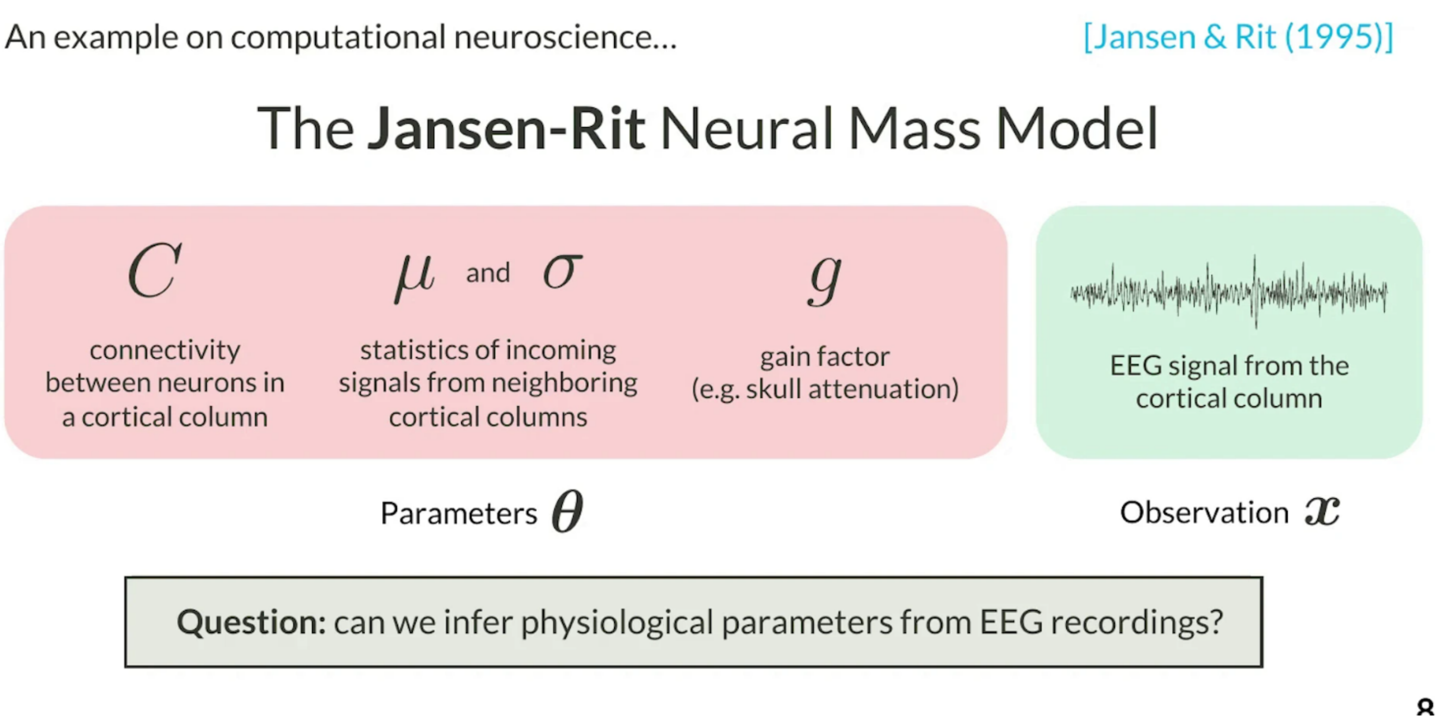
—
Image credits: Gonçalves et al, 2020.



—

Image credits: Gonçalves et al, 2020.

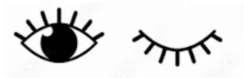
HNPE: Leveraging global parameters for NPE (Rodrigues et al, 2021)



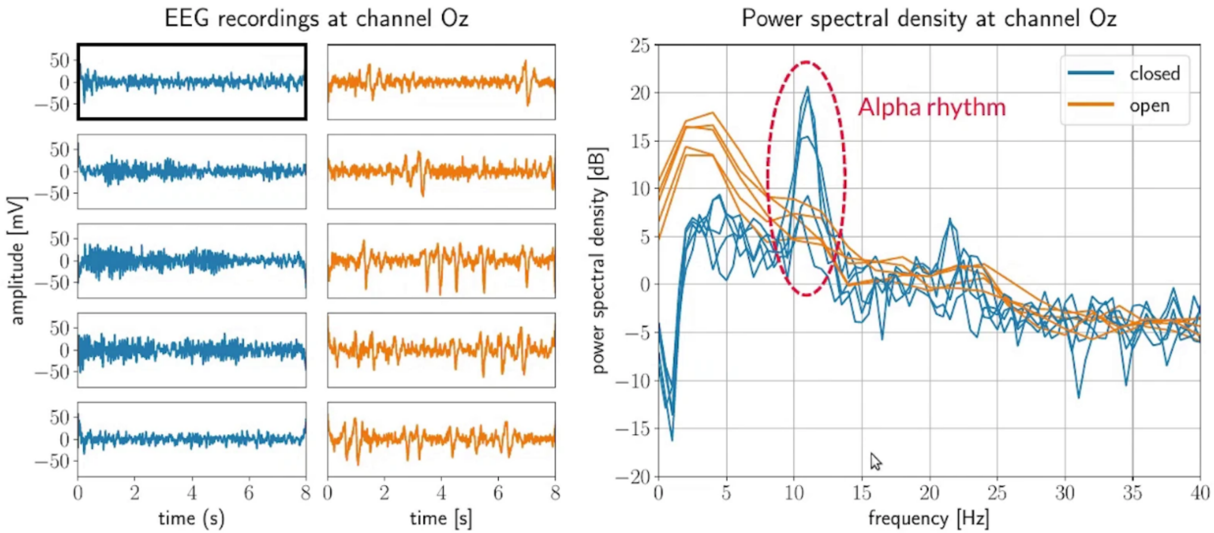
—

Image credits: Rodrigues et al, 2020.

An example on computational neuroscience...



We consider a database of EEG recordings from subjects with eyes open or closed



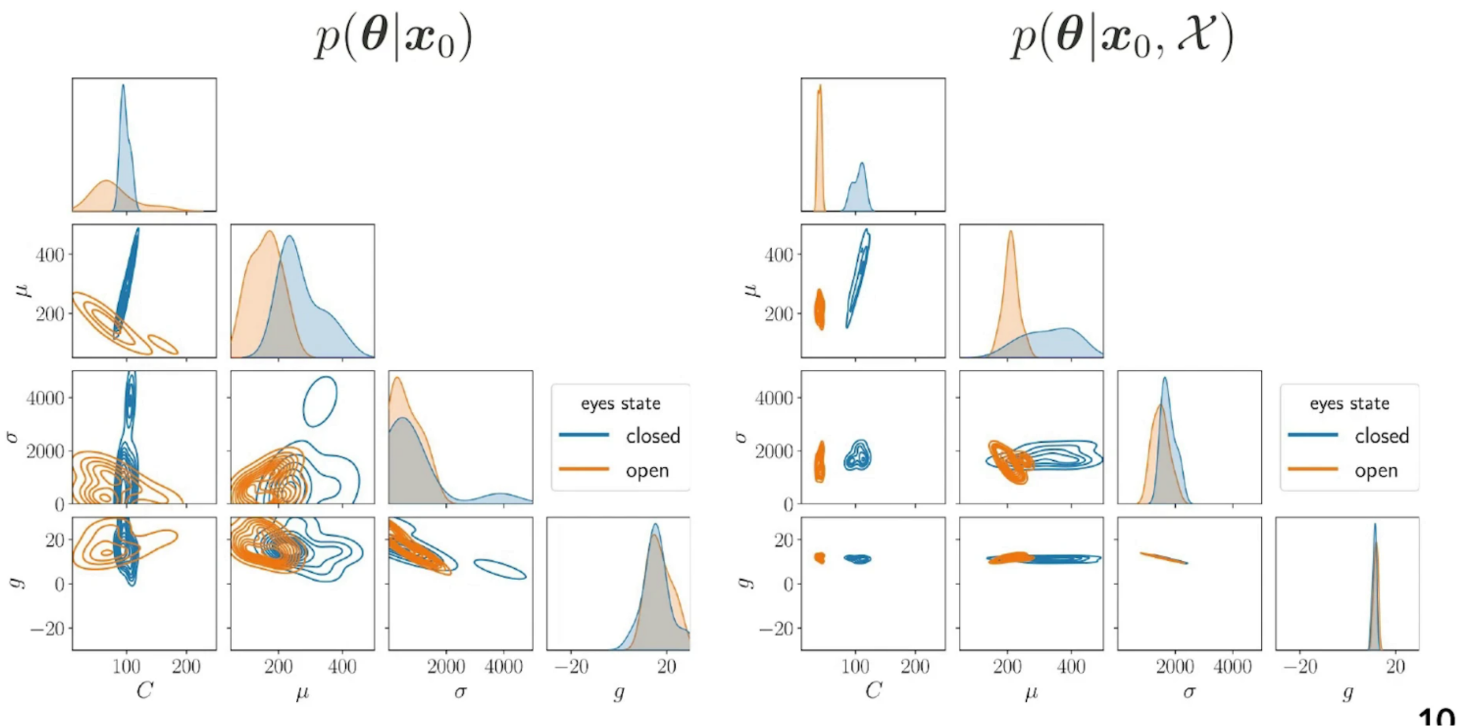
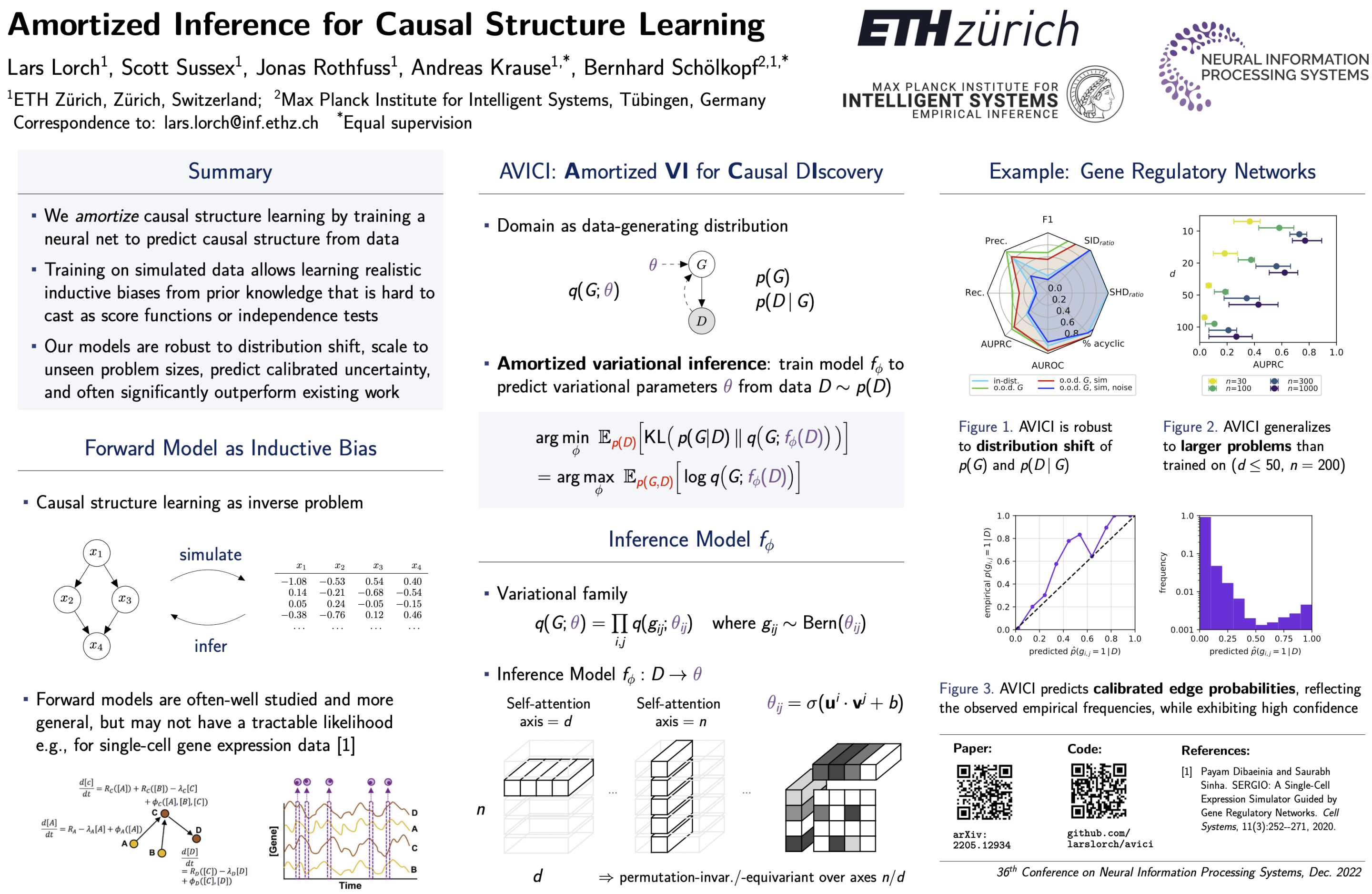


Image credits: Rodrigues et al, 2020.



Amortized Inference for Causal Structure Learning

Lars Lorch[1], Scott Sussex[1], Jonas Rothfuss[1], Andreas Krause[1,*], Bernhard Schölkopf[2,1,*]

[1]ETH Zürich, Zürich, Switzerland; [2]Max Planck Institute for Intelligent Systems, Tübingen, Germany

Correspondence to: lars.lorch@inf.ethz.ch [*]Equal supervision

ETH zürich — MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS EMPIRICAL INFERENCE — NEURAL INFORMATION PROCESSING SYSTEMS

Summary

- We *amortize* causal structure learning by training a neural net to predict causal structure from data
- Training on simulated data allows learning realistic inductive biases from prior knowledge that is hard to cast as score functions or independence tests
- Our models are robust to distribution shift, scale to unseen problem sizes, predict calibrated uncertainty, and often significantly outperform existing work

Forward Model as Inductive Bias

- Causal structure learning as inverse problem

simulate / infer

x_1	x_2	x_3	x_4
−1.08	−0.53	0.54	0.40
0.14	−0.21	−0.68	−0.54
0.05	0.24	−0.05	−0.15
−0.38	−0.76	0.12	0.46
…	…	…	…

- Forward models are often-well studied and more general, but may not have a tractable likelihood e.g., for single-cell gene expression data [1]

AVICI: Amortized VI for Causal DIscovery

- Domain as data-generating distribution

$q(G;\theta)$ $p(G)$ $p(D\,|\,G)$

- **Amortized variational inference**: train model f_ϕ to predict variational parameters θ from data $D \sim p(D)$

$$\arg\min_\phi \; \mathbb{E}_{p(D)}\Big[\mathrm{KL}\big(p(G|D)\,\|\,q(G; f_\phi(D))\big)\Big]$$
$$= \arg\max_\phi \; \mathbb{E}_{p(G,D)}\Big[\log q\big(G; f_\phi(D)\big)\Big]$$

Inference Model f_ϕ

- Variational family

$$q(G;\theta) = \prod_{i,j} q(g_{ij};\theta_{ij}) \quad \text{where } g_{ij} \sim \mathrm{Bern}(\theta_{ij})$$

- Inference Model $f_\phi : D \rightarrow \theta$

Self-attention axis $= d$ — Self-attention axis $= n$ — $\theta_{ij} = \sigma(\mathbf{u}^i \cdot \mathbf{v}^j + b)$

$\Rightarrow$ permutation-invar./-equivariant over axes n/d

Example: Gene Regulatory Networks

Figure 1. AVICI is robust to **distribution shift** of $p(G)$ and $p(D\,|\,G)$

Figure 2. AVICI generalizes to **larger problems** than trained on ($d \leq 50$, $n = 200$)

Figure 3. AVICI predicts **calibrated edge probabilities**, reflecting the observed empirical frequencies, while exhibiting high confidence

Paper: arXiv: 2205.12934

Code: github.com/larslorch/avici

References:

[1] Payam Dibaeinia and Saurabh Sinha. SERGIO: A Single-Cell Expression Simulator Guided by Gene Regulatory Networks. *Cell Systems*, 11(3):252–271, 2020.

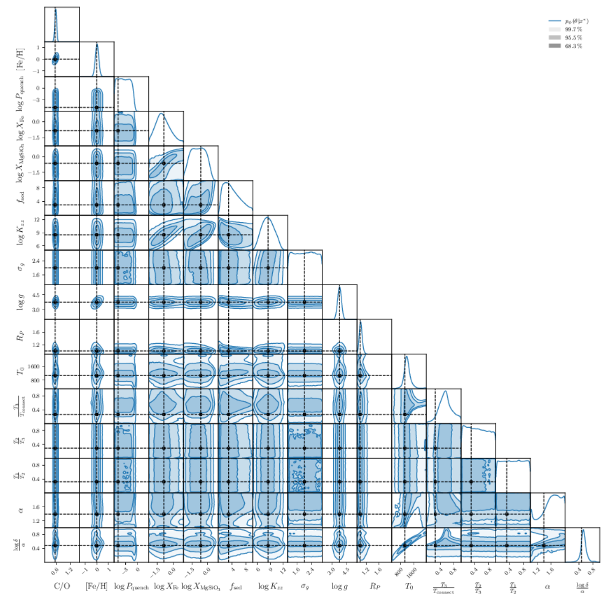
36th Conference on Neural Information Processing Systems, Dec. 2022

Image credits: Lorch et al, 2022.

Computational faithfulness

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

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

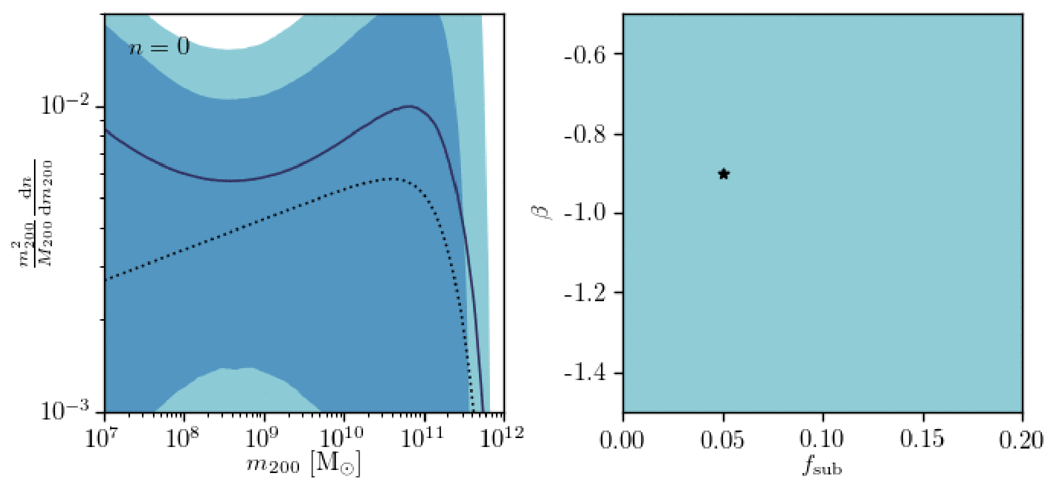


How do we know this is good enough?

Mode convergence:

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

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



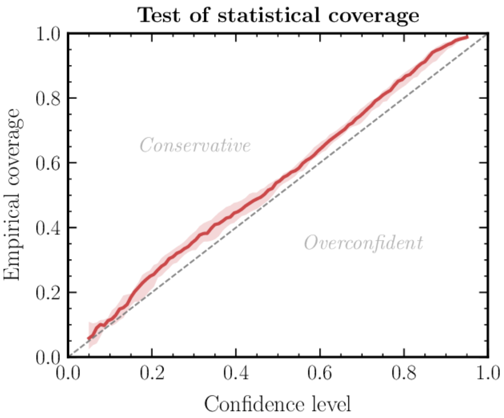
—
Credits: Brehmer et al, 2019.

A common observation at the root of several other diagnostics is to check for the **self-consistency** of the Bayesian joint distribution,

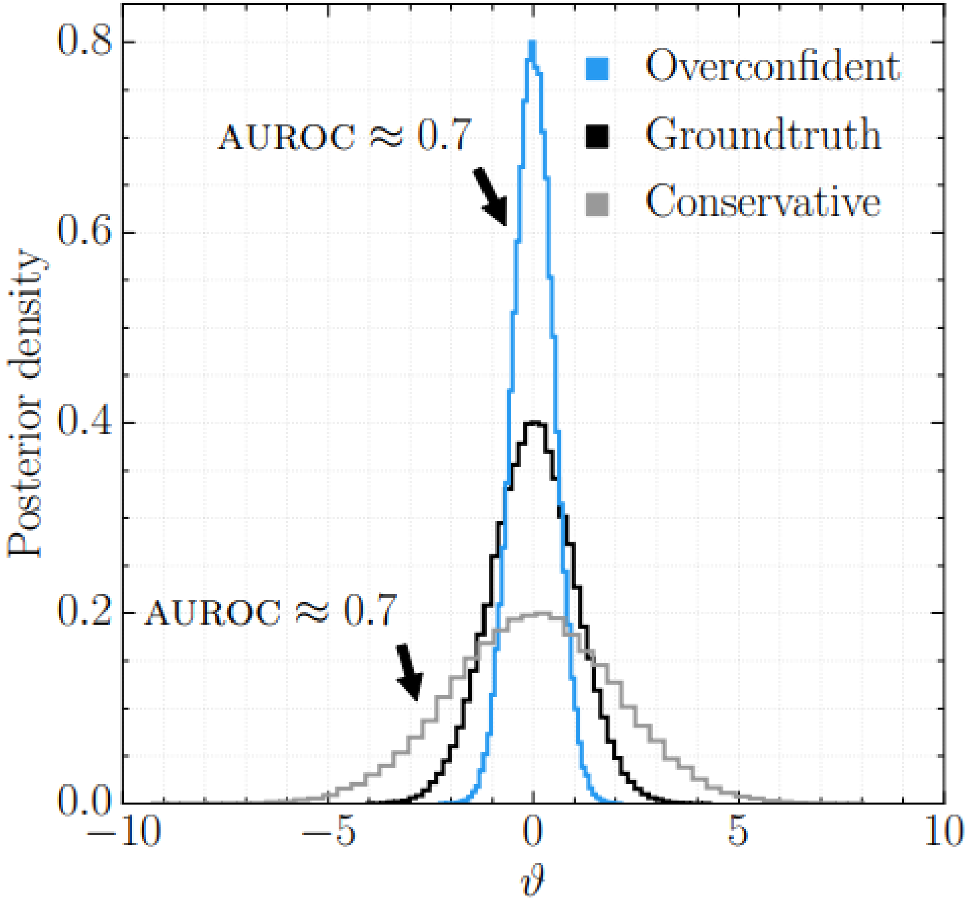
$$p(\theta) = \int p(\theta')p(x|\theta')p(\theta|x)d\theta' \, dx.$$

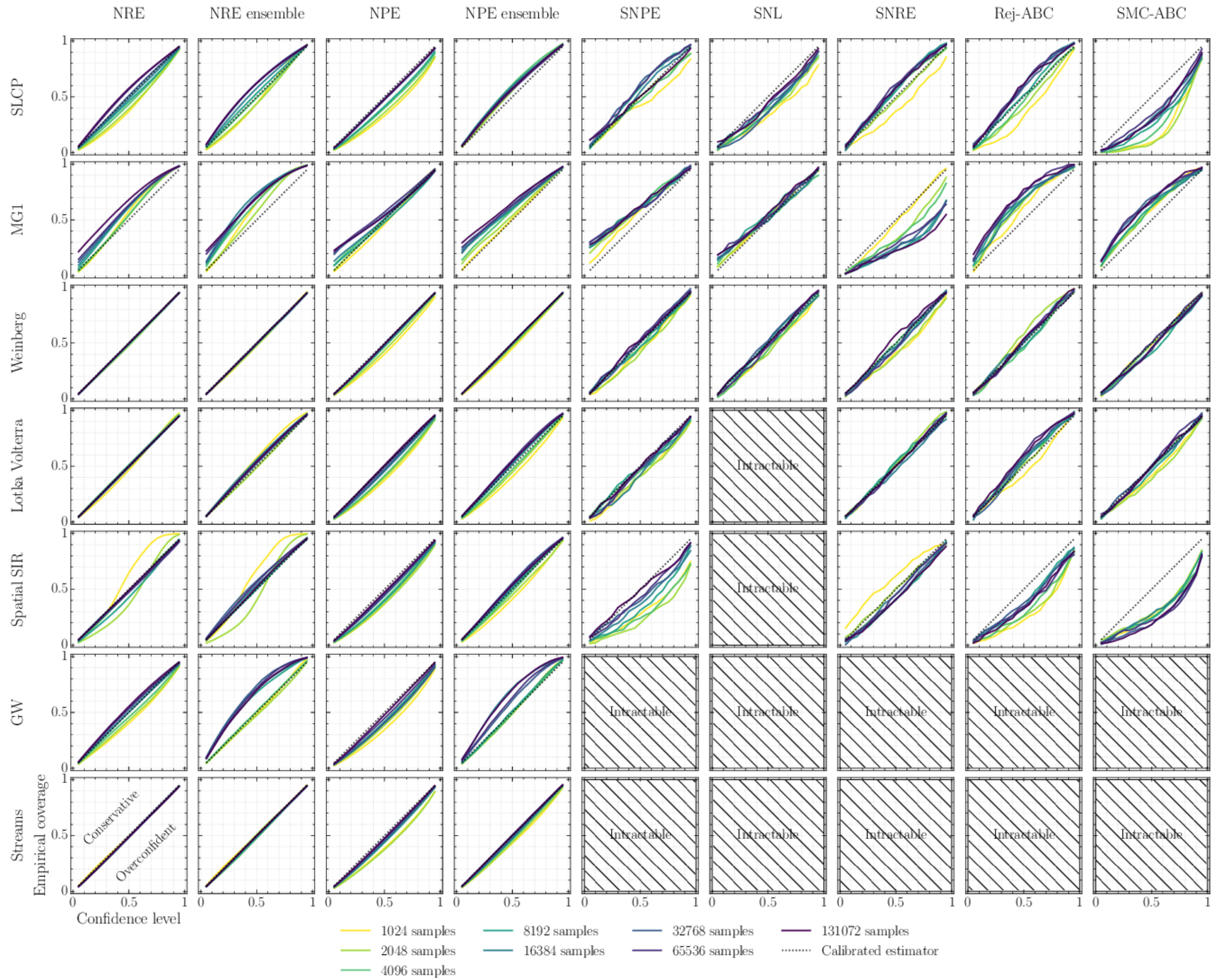
Coverage diagnostic:

- 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.


0.8
0.6
0.4
0.2
0.0
Posterior density
AUROC ≈ 0.7
AUROC ≈ 0.7
Overconfident
Groundtruth
Conservative
−10
−5
0
5
10
ϑ



—

Credits: Hermans et al, 2021.

What if diagnostics fail?

Summary

Simulation-based inference is a major evolution in the statistical capabilities for science, enabled by advances in machine learning.

Need to reliably and efficiently evaluate the quality of the posterior approximations.

The end.