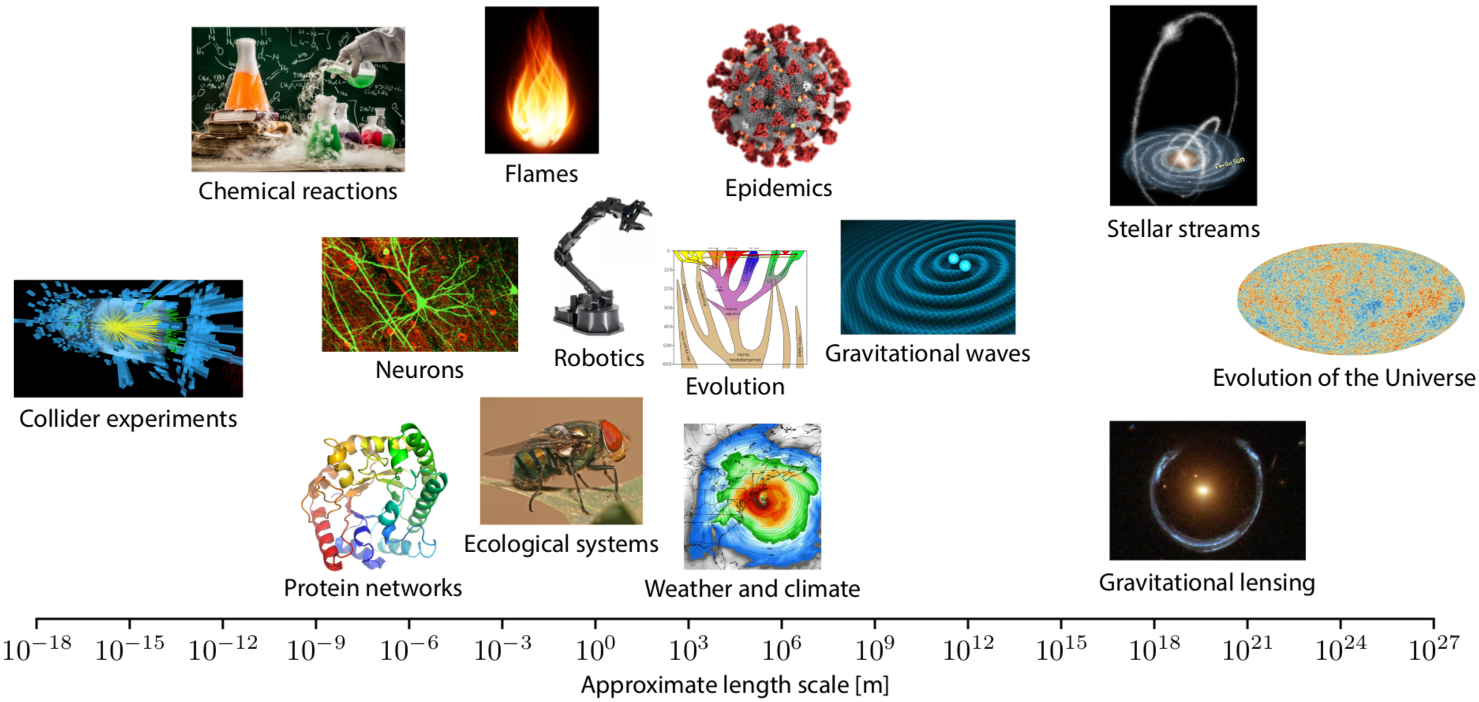
Neural Ratio Estimation for Simulation-Based Inference

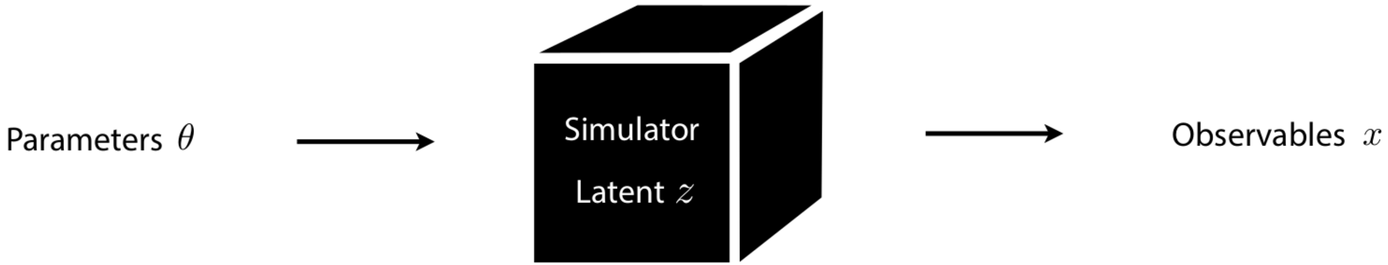
March 2

Computational Methods for Likelihood-Free Bayesian Inference
SIAM Conference on Computational Science and Engineering 2021

Gilles Louppe
g.louppe@uliege.be

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

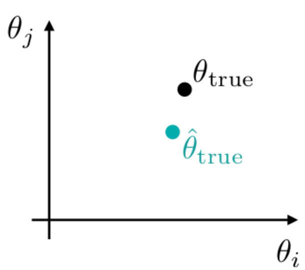
Problem statement(s)

Start with

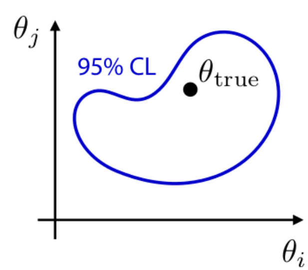
- a simulator that lets you generate N samples $x_i \sim p(x_i|\theta_i)$ (for parameters θ_i of our choice),
- observed data $x_{\text{obs}} \sim p(x_{\text{obs}}|\theta_{\text{true}})$,
- a prior $p(\theta)$.

Then,

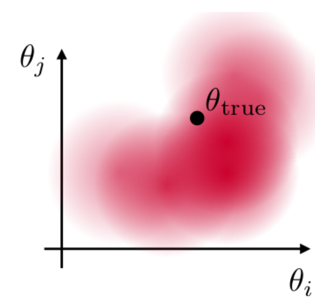
a) estimate θ_{true} (e.g., MLE)



b) construct confidence sets



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



Amortizing Bayes

The Bayes rule can be rewritten as

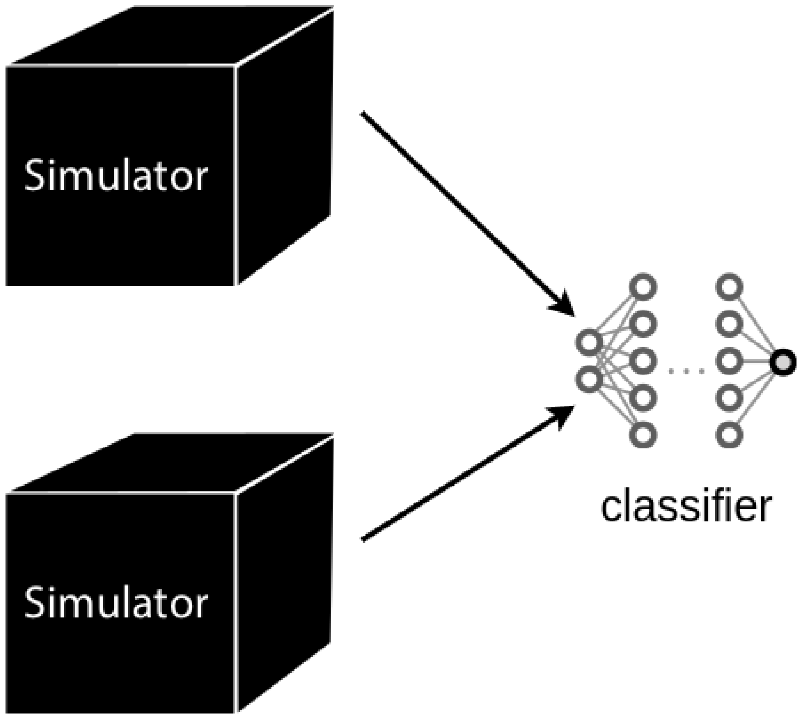
$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = r(x|\theta)p(\theta) \approx \hat{r}(x|\theta)p(\theta),$$

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

The likelihood ratio trick

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

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



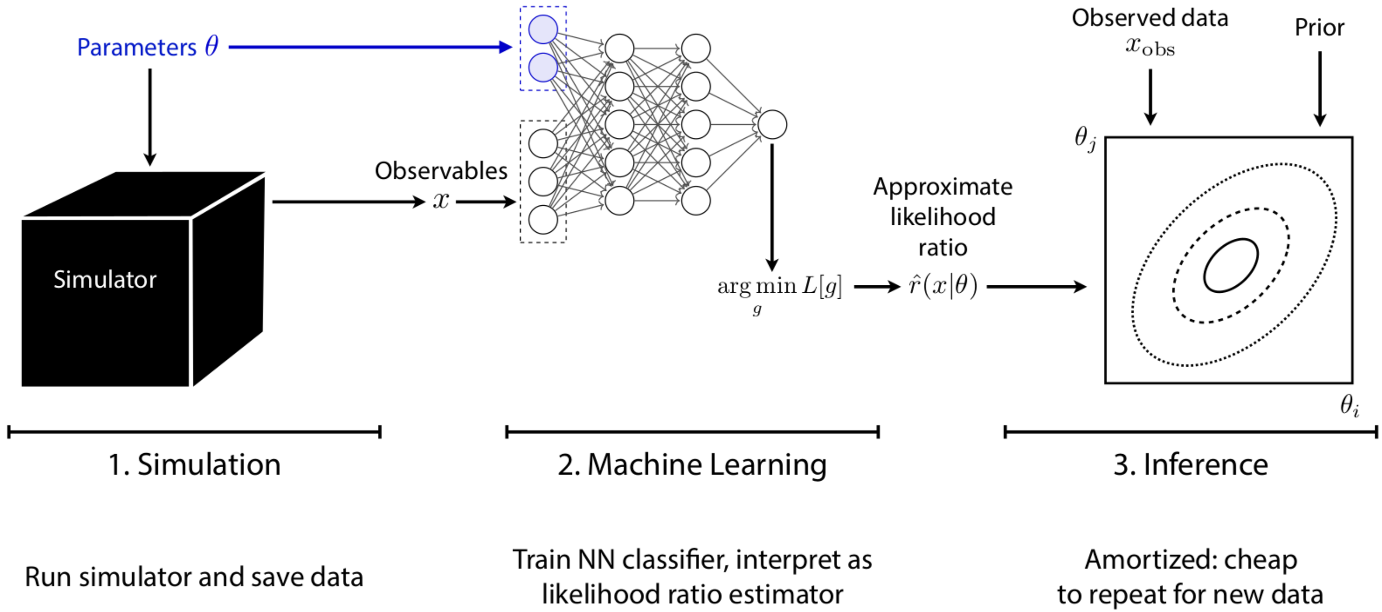
The solution d found after training approximates the optimal classifier

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

Therefore,

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

Inference

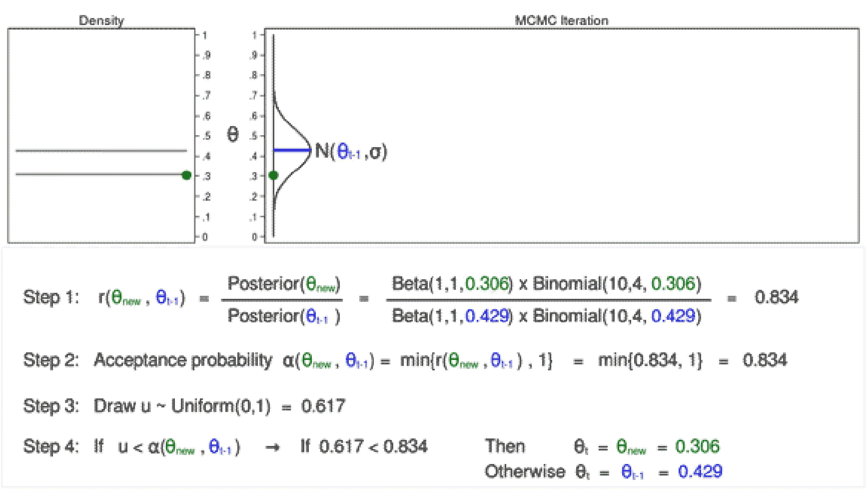

Parameters θ
Simulator
Observables
x
Observed data
x_{obs}
Prior
θ_j
θ_i
Approximate
likelihood
ratio
$\arg\min_g L[g]$
$\hat{r}(x|\theta)$
1. Simulation
2. Machine Learning
3. Inference
Run simulator and save data
Train NN classifier, interpret as
likelihood ratio estimator
Amortized: cheap
to repeat for new data

Likelihood-free MCMC

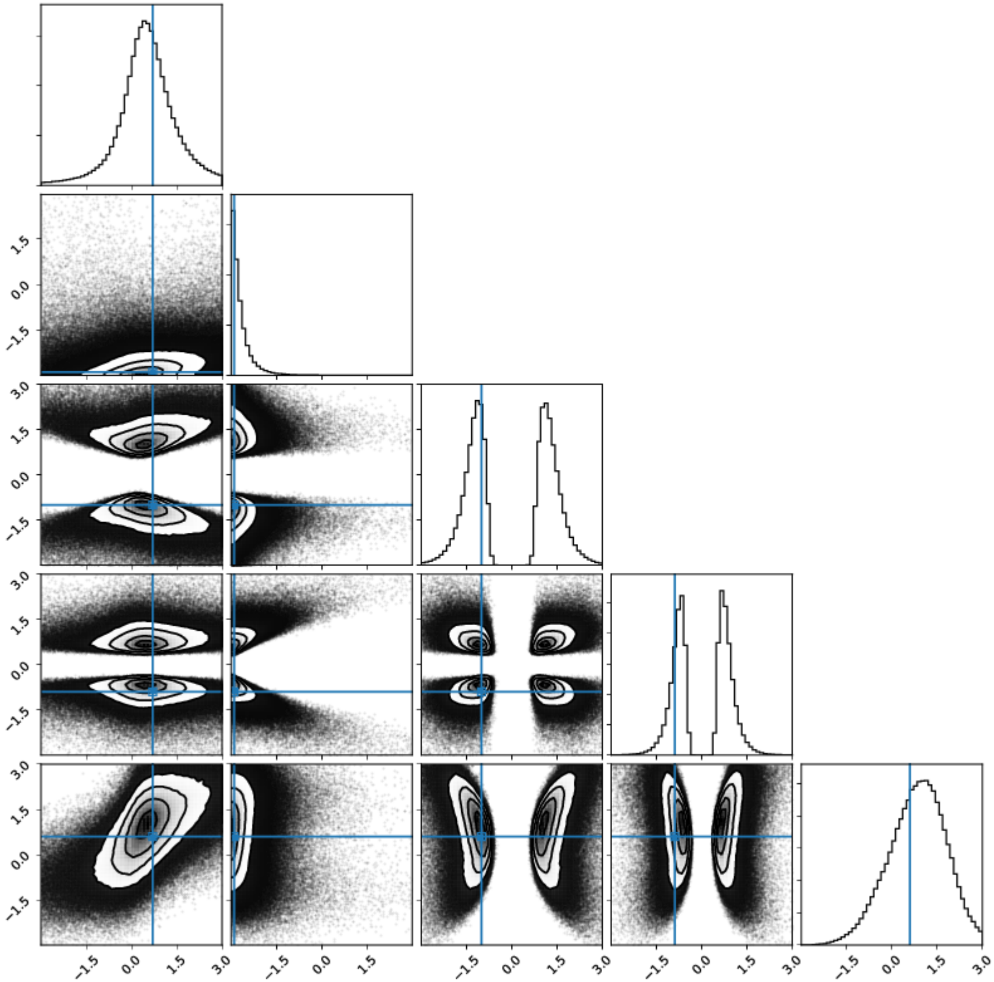
MCMC samplers require the evaluation of the posterior ratios:

$$
\begin{aligned}
\frac{p(\theta_\text{new}|x)}{p(\theta_{t-1}|x)} &= \frac{p(x|\theta_\text{new})p(\theta_\text{new})/p(x)}{p(x|\theta_{t-1})p(\theta_{t-1})/p(x)} \\
&= \frac{r(x|\theta_\text{new})}{r(x|\theta_{t-1})}\frac{p(\theta_\text{new})}{p(\theta_{t-1})}.
\end{aligned}
$$

Extensions with HMC is possible since $\nabla_\theta p(x|\theta) = \frac{\nabla_\theta r(x|\theta)}{r(x|\theta)}$.



Diagnostics



How to assess that the approximate posterior is not wrong?

Coverage

- For every $x, \theta \sim p(x, \theta)$ in a validation set, compute the $1 - \alpha$ credible interval based on $\hat{p}(\theta|x) = \hat{r}(x|\theta)p(\theta)$.
- The fraction of samples for which θ is contained within the interval corresponds to the empirical coverage probability.
- If the empirical coverage is larger that the nominal coverage probaiblity $1 - \alpha$, then the ratio estimator $\hat{r}$ passes the diagnostic.

Convergence towards the nominal value

If the approximation $\hat{r}$ is correct, then the posterior $\hat{p}(\theta|\mathcal{X})$ should concentrate around θ^* as the number of observations

$$\mathcal{X} = \{x_1, ..., x_n\},$$

for $x_i \sim p(x|\theta^*)$, increases.

ROC AUC score

The ratio estimator $\hat{r}(x|\theta)$ is only exact when samples x from the reweighted marginal model $p(x)\hat{r}(x|\theta)$ cannot be distinguished from samples x from a specific likelihood $p(x|\theta)$.

Therefore, the predictive ROC AUC performance of a classifier should be close to 0.5 if the ratio is correct.

Constraining dark matter with stellar streams

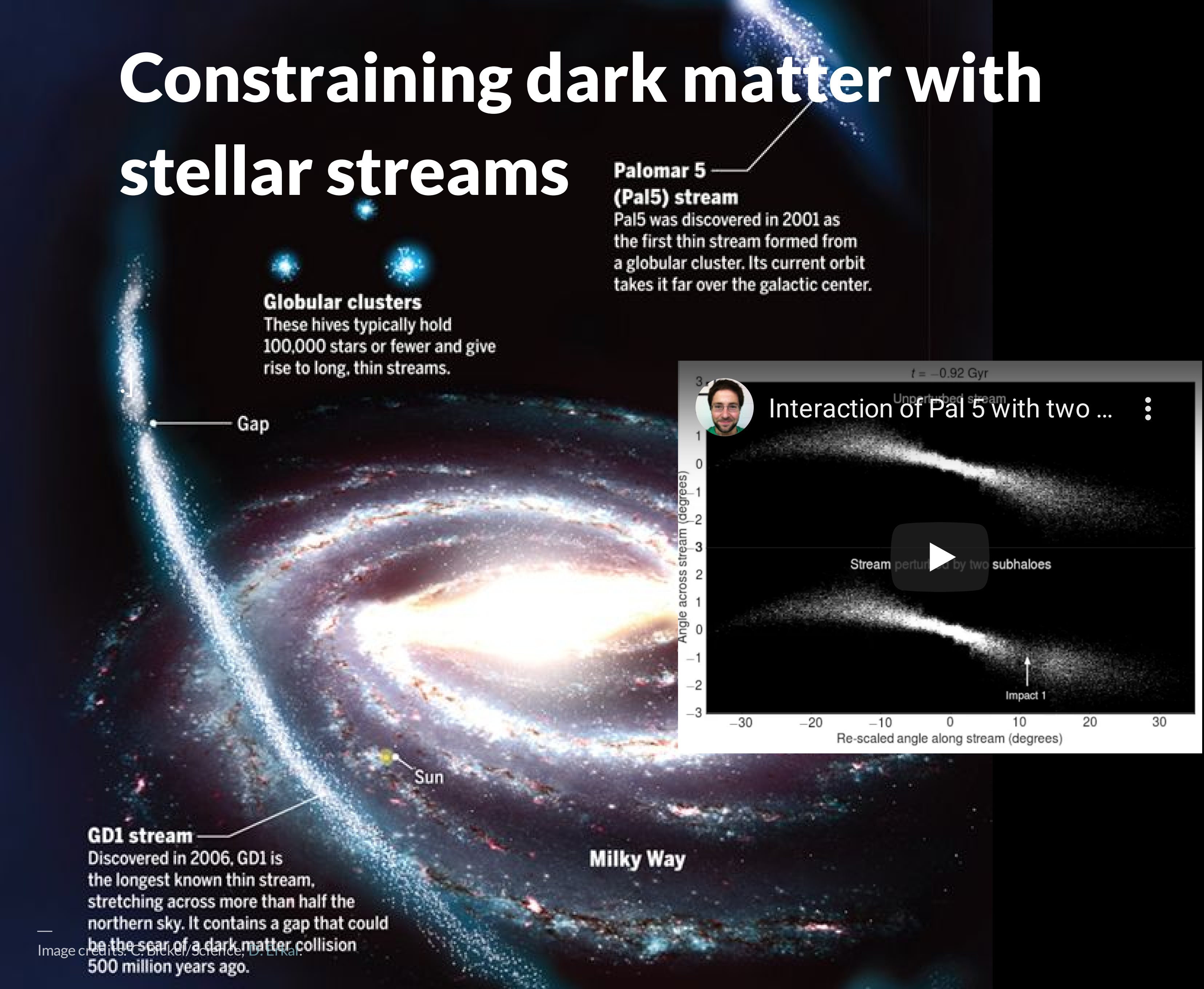
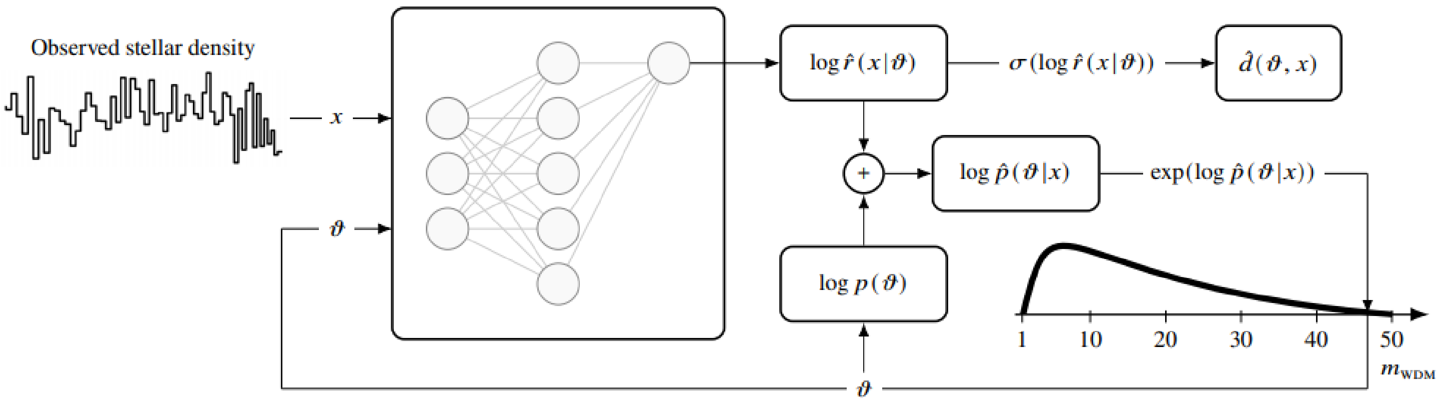
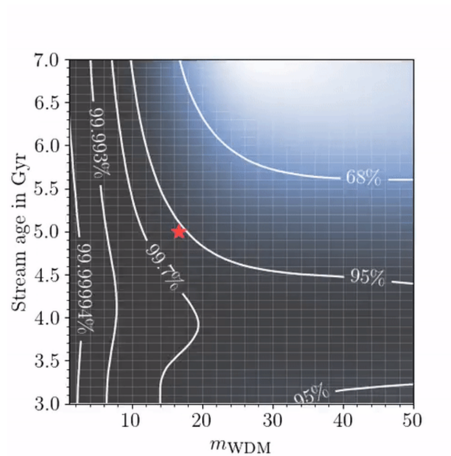
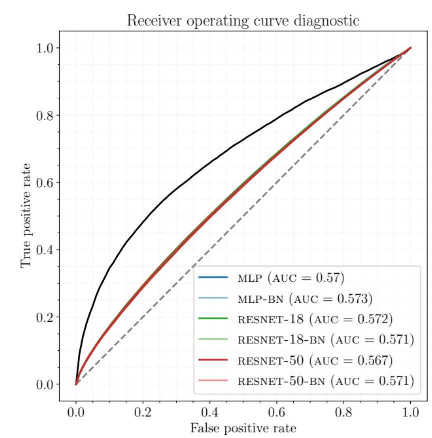


Image credits: C. Bickel/Science, D. Erkal.

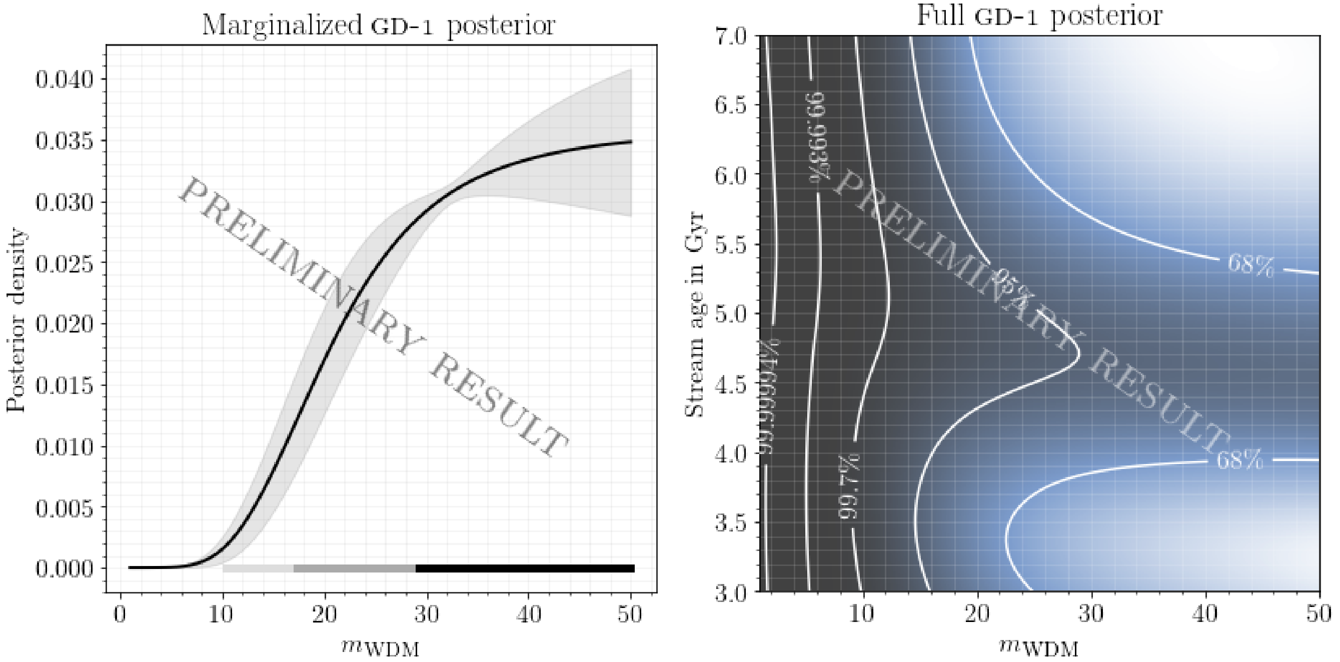


Architecture	68% CR	95% CR
$\hat{r}(x\|\vartheta)$ with $\vartheta \triangleq (m_{\mathrm{WDM}})$		
MLP	$0.685_{\pm 0.004}$	$0.954_{\pm 0.002}$
MLP-BN	$0.687_{\pm 0.006}$	$0.951_{\pm 0.002}$
RESNET-18	$0.667_{\pm 0.004}$	$0.943_{\pm 0.002}$
RESNET-18-BN	$0.672_{\pm 0.004}$	$0.945_{\pm 0.001}$
RESNET-50	$0.671_{\pm 0.005}$	$0.947_{\pm 0.003}$
RESNET-50-BN	$0.678_{\pm 0.004}$	$0.949_{\pm 0.004}$
$\hat{r}(x\|\vartheta)$ with $\vartheta \triangleq (m_{\mathrm{WDM}}, t_{\mathrm{age}})$		
MLP	$0.685_{\pm 0.005}$	$0.953_{\pm 0.002}$
MLP-BN	$0.685_{\pm 0.004}$	$0.952_{\pm 0.003}$
RESNET-18	$0.666_{\pm 0.005}$	$0.945_{\pm 0.002}$
RESNET-18-BN	$0.671_{\pm 0.003}$	$0.945_{\pm 0.003}$
RESNET-50	$0.674_{\pm 0.006}$	$0.944_{\pm 0.002}$
RESNET-50-BN	$0.677_{\pm 0.004}$	$0.947_{\pm 0.003}$





Coverage **Convergence to θ^*** **ROC AUC score**



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

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, such as ratio estimation based on neural networks.
- Amortized estimators are well suited for diagnosing the quality of the resulting posteriors.

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.
- Hermans, J., Begy, V., & Louppe, G. (2019). Likelihood-free MCMC with Approximate Likelihood Ratios. arXiv preprint arXiv:1903.04057.
- Cranmer, K., Pavez, J., & Louppe, G. (2015). Approximating likelihood ratios with calibrated discriminative classifiers. arXiv preprint arXiv:1506.02169.

The end.