Simulation-efficient marginal posterior estimation with *swyft*:

Stop wasting your precious time!

Benjamin Kurt Miller, Alex Cole, Gilles Louppe, Christoph Weniger

github.com/undark-lab/swyft

What's the big idea?

- Likelihood-free inference facilitates posterior estimation on data where we have access to a simulator. (Old news)
- Nested Ratio Estimation approximates the likelihood-to-evidence ratio by zeroing in on regions of high likelihood.
- By using a point process, we never reject simulations and reuse previous ones. This is highly simulator efficient.
- The above allows for extremely efficient estimation of any marginal posterior of interest: Effective for constraining parameters within uncertainty bounds. We call it *swyft*.

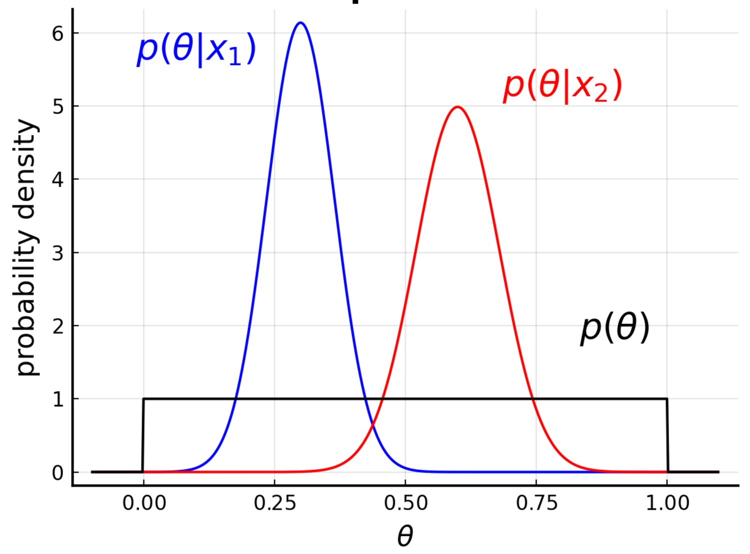
Likelihood-free Inference

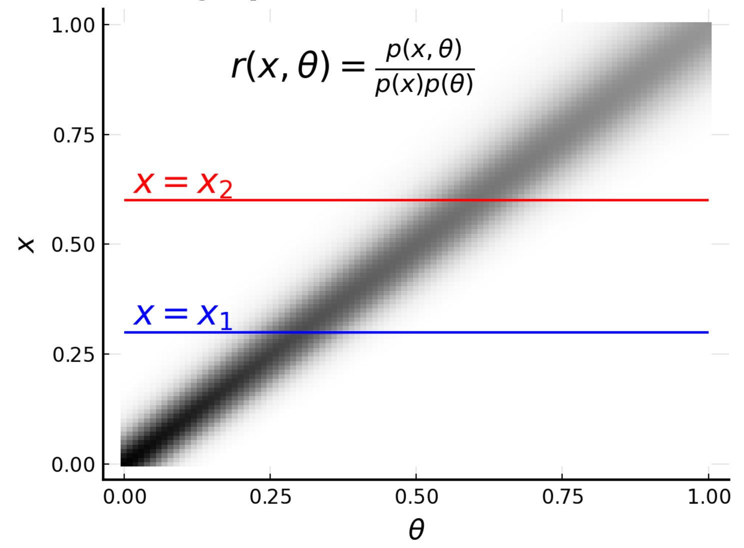
Likelihood-free (simulation-based) inference employs, $\mathbf{g}$, a non-linear *simulator* mapping a stochastic latent state $\mathbf{z}$ and a parameter vector $\boldsymbol{\theta}$ to an observation $\mathbf{x} = \mathbf{g}(\boldsymbol{\theta}, \mathbf{z})$.

We're interested in the marginal posterior, $p(\boldsymbol{\vartheta}|\boldsymbol{x})$, where $\boldsymbol{\vartheta}$ are the parameters of interest and $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\eta})$. Namely,

$$p(\boldsymbol{\vartheta}|\boldsymbol{x}) = \int \frac{p(\boldsymbol{x}|\boldsymbol{\vartheta},\boldsymbol{\eta})}{p(\boldsymbol{x})} p(\boldsymbol{\theta}) d\boldsymbol{\eta} = \frac{p(\boldsymbol{x}|\boldsymbol{\vartheta})}{p(\boldsymbol{x})} p(\boldsymbol{\vartheta})$$

The likelihood ratio $r(\boldsymbol{x}, \boldsymbol{\vartheta}) = p(\boldsymbol{x}|\boldsymbol{\vartheta})/p(\boldsymbol{x})$ can be learned by training a classifier to distinguish positive samples drawn jointly $(\boldsymbol{x}, \boldsymbol{\vartheta}) \sim p(\boldsymbol{x}, \boldsymbol{\vartheta})$ from negative samples drawn marginally $(\boldsymbol{x}, \boldsymbol{\vartheta}) \sim p(\boldsymbol{x})p(\boldsymbol{\vartheta})$, i.e. simulation-parameters pairs vs. simulations paired with an arbitrary parameter vector.

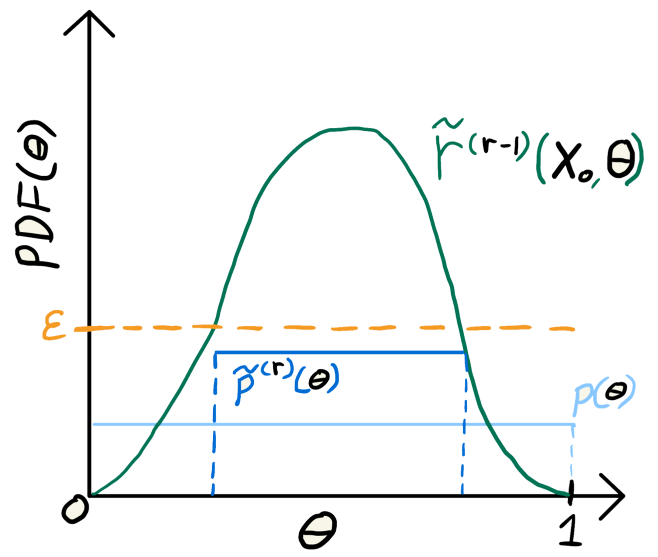




The likelihood ratio allows for computation of any posterior.

Simulate what matters: Targeted Simulation

Our analysis focuses on an observation of interest, $\boldsymbol{x_0}$. Only parameter values which could have plausibly generated $\boldsymbol{x_0}$ will contribute to the marginalization. Nested Ratio Estimation (NRE) finds this region by iteratively constraining the initial prior, $p(\boldsymbol{\theta})$, based on 1-dim marginal posterior estimates from previous iterations.



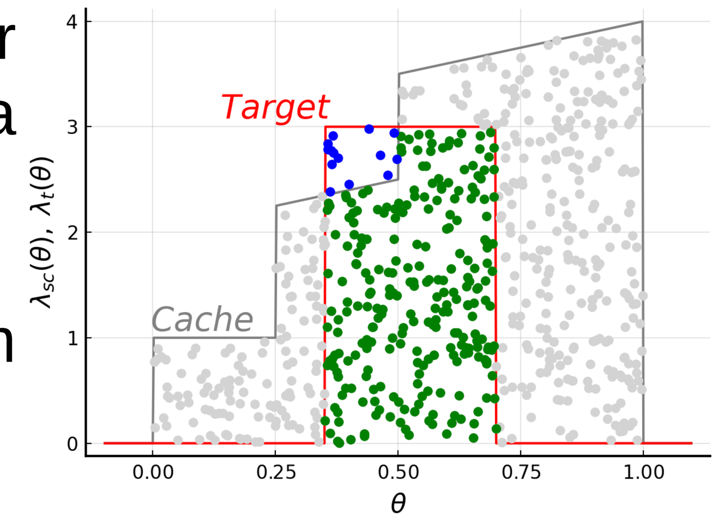
At each round, NRE predicts *all* 1-dim marginals of $\boldsymbol{\theta}$ for the same underlying posterior. This allows for reuse of data and efficient training in a multi-target regime.

The final constrained prior, $\tilde{p}^{(R)}(\boldsymbol{\theta})$, along with previously generated and cached samples, facilitates for hyper-efficient estimation of *any* higher dimensional marginal posterior through likelihood-to-evidence ratio estimation.

Don't throw work away: Recycle simulations

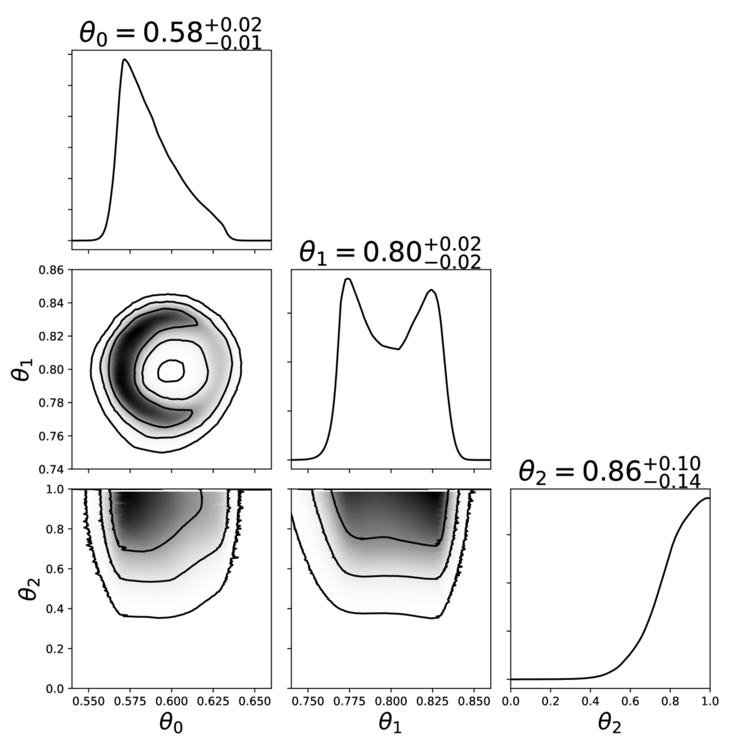
Simulating $\boldsymbol{x}$ can be expensive. We reuse appropriate simulations by framing our samples as being drawn from an inhomogeneous Poisson point process (iP3)—a construction yielding a Poissonian distributed number of samples.

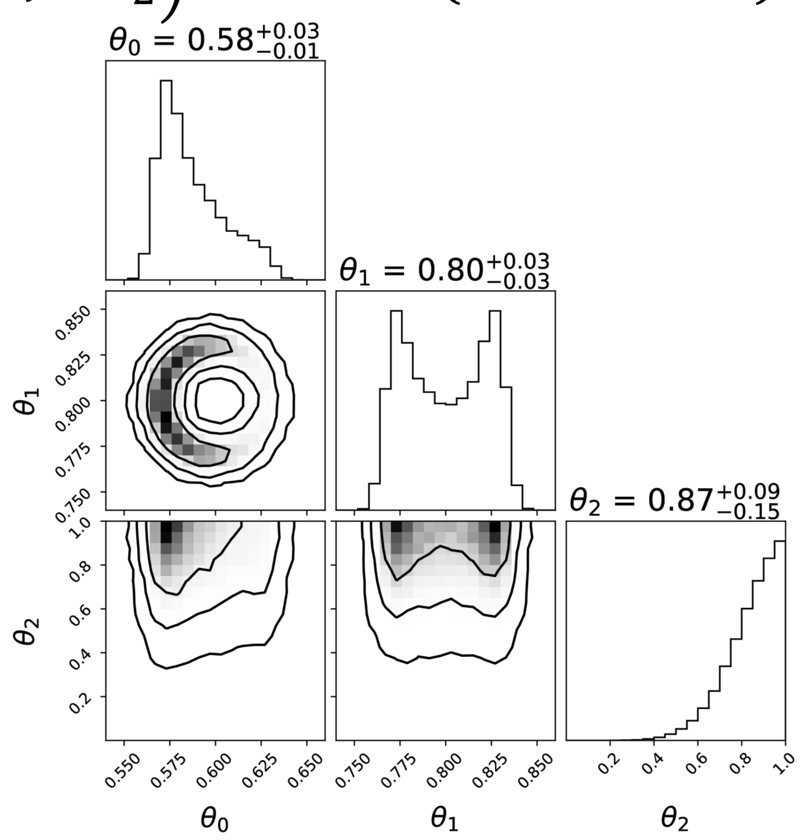
An iP3 is defined by its intensity function. We store samples and intensity functions from previous rounds. When we need more samples than cached. only then do we simulate.



Efficient inference
Hyper-efficient follow up

We compare *swyft* (L) to MulitNest (R). Let $g(\boldsymbol{\theta}, \boldsymbol{z}) = \left(\theta_0, \sqrt{(\theta_0 - 0.6)^2 + (\theta_1 - 0.8)^2}, \theta_2\right) + \boldsymbol{n}, \boldsymbol{\theta} = (0.57, 0.8, 1.0)$



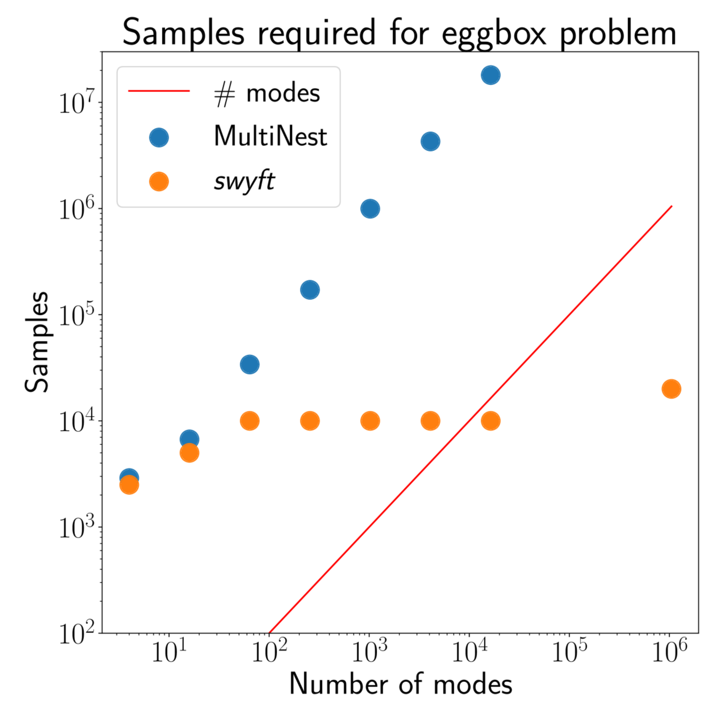


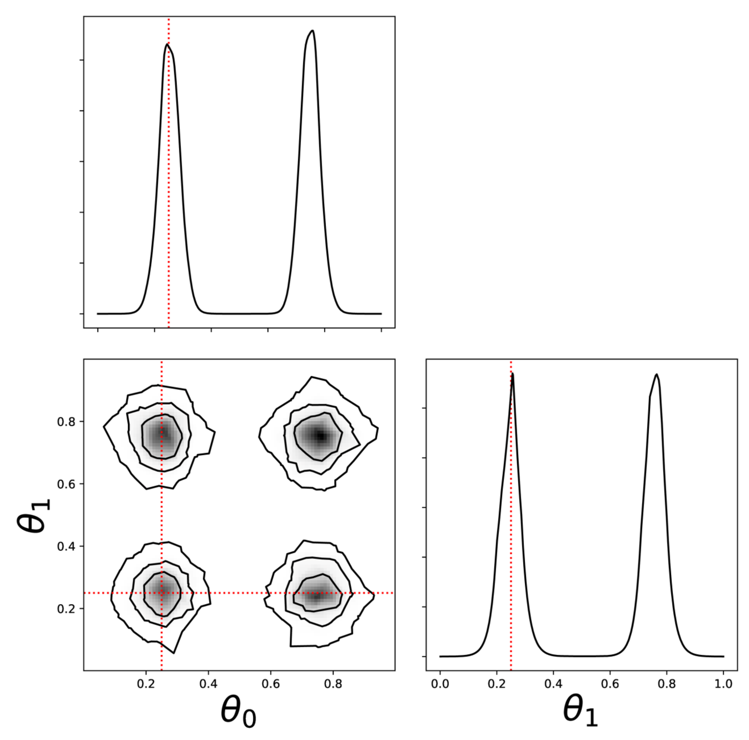
Follow-up analysis (L), on a new $\boldsymbol{\theta}$, only new simulations counted.

	NRE/iP3	MultiNest
Run 1	20011	160722
Run 2	3668	171209

Marginalize over 10^6 modes!

By approximating the marginal posteriors *swyft* is solving a different problem than likelihood-based methods. Let $g_i(\boldsymbol{\theta}_i, \boldsymbol{z}) = \sin(\pi\theta_i) + n_i$. The number of modes grows exponentially with dimension, flummoxing MultiNest at high dimension. *swyft* converges with O(1) simulations.





Cost vs. complexity (L) and a subset of marginals (R).