# Simulation-Based Inference with Approximately Correct Parameters via Maximum Entropy

Rainier Barrett, Mehrad Ansari, Andrew D White, and Gourab Goshal

thewhitelab.org

## Abstract

Here we show a simple approach to simulation-based inference in the regime of sparse data and approximately correct models, which is common when trying to use an existing model to infer latent variables with observed data. This approach is based on the principle of maximum entropy (MaxEnt) [1] and provably makes the smallest change in the latent joint distribution to accommodate new data. This simple method requires no likelihood or simulator derivatives and its fit is insensitive to prior strength, removing the need to balance observed data fit with prior belief.

## Theory

Given a simulator $f$ and parameters $\vec{\theta}$ we would like to alter a prior distribution $P(\vec{\theta})$ to match some set of observations $g_k$ such that:

$$\int \mathrm{d}\vec{\theta} P'(\vec{\theta}) g_k[f(\vec{\theta})] = E[g_k] = \bar{g}_k \forall k$$

The unique maximum entropy modification to the prior distribution to satisfy $N$ constraints is given by:

$$P'(\vec{\theta}) = \frac{1}{Z'} P(\vec{\theta}) \prod_k^N e^{-\lambda_k g_k[f(\vec{\theta})]}$$

In practice this is approximated by sampling from our prior and calculating weights for each sampled parameter set according to:

$$w_i[P'] = \prod_k^N e^{-\lambda_k g_k[f(\vec{\theta})]}$$

where $\lambda_k$ are found by some optimization method such as stochastic gradient descent, Adam, etc.

If sampling does not provide enough support to match constraints, we can use gradient descent to modify the sampling distribution $P^j(\vec{\theta})$ to minimize the cross-entropy with $P'(\vec{\theta})$:

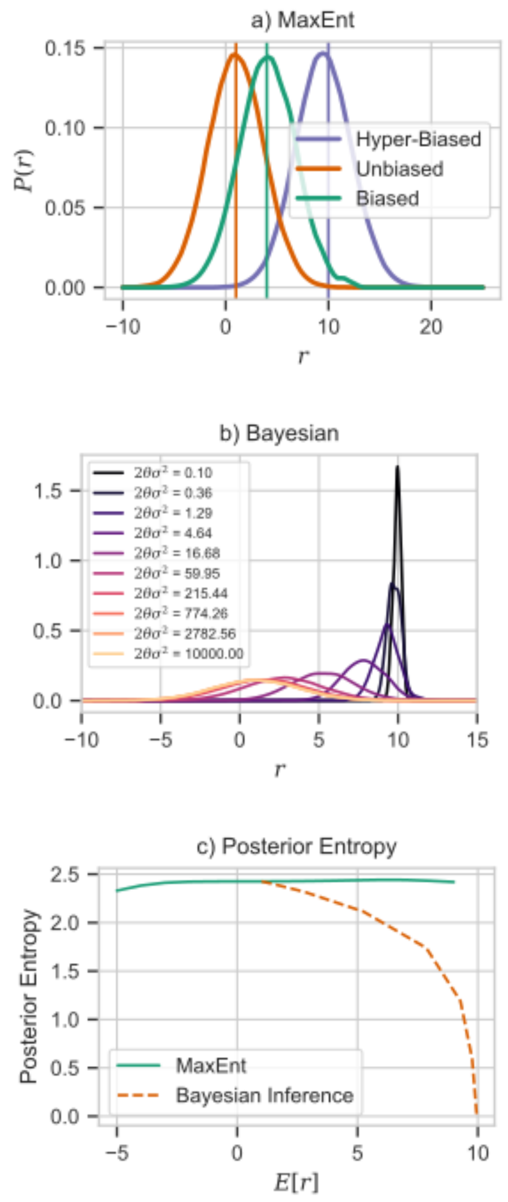
$$P^{j+1}(\vec{\theta}) = P^j(\vec{\theta}) - \eta \frac{\delta}{\delta P^j(\vec{\theta}')} \sum_i w_i[P'] \ln P(\vec{\theta}_i)$$

## Acknowledgement

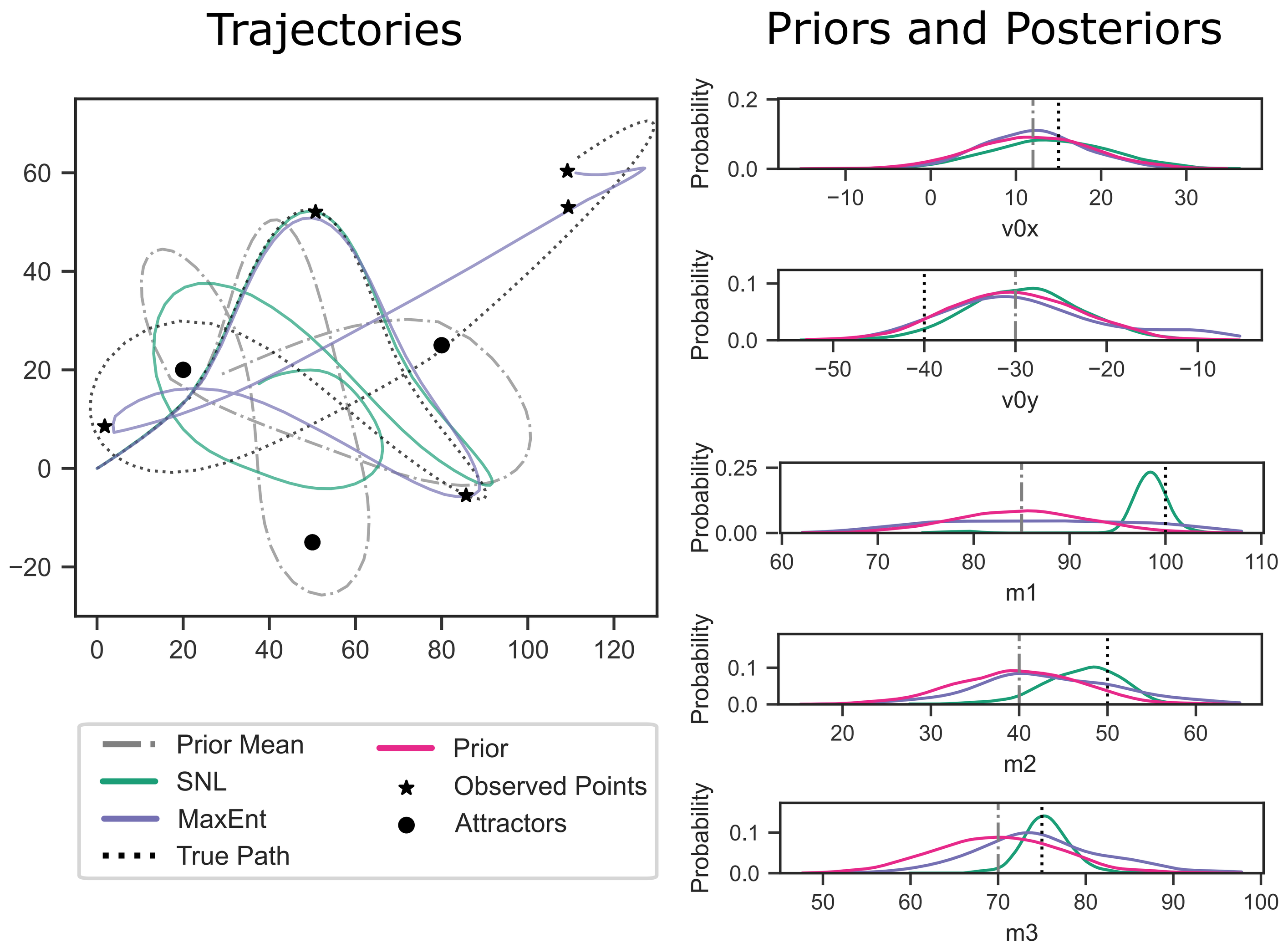
This work was supported by National Science Foundation under grant No. **2029095**

## Linear Model

Here we illustrate MaxEnt on a linear model with Gaussian noise, and compare against Bayesian inference to illustrate a key feature of MaxEnt reweighting, which is that it minimally alters the prior distribution to agree with observed data. "Minimally" here means with respect to entropy of the posterior. a) shows the effect of MaxEnt biasing with and without hyperparameter biasing. b) shows the effect of Baysian inference in the same range. c) shows a comparison of the posterior entropy produced by the two methods at different target values.



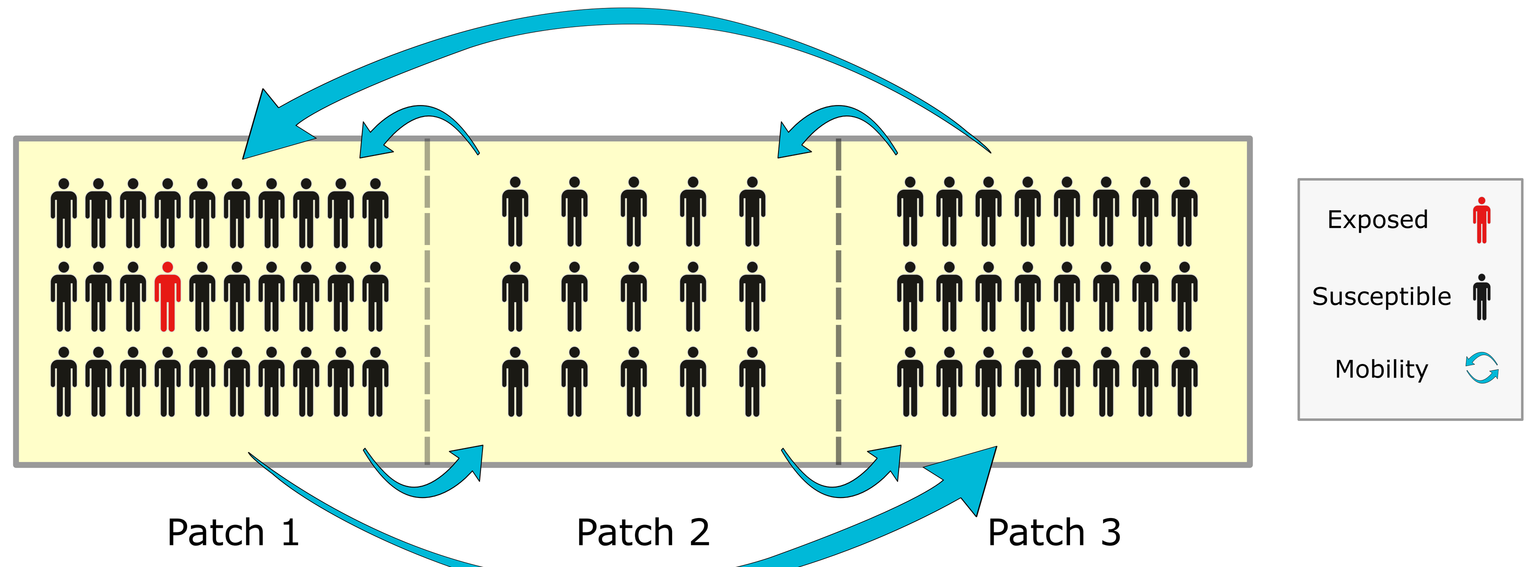

## Gravitational Field Simulation



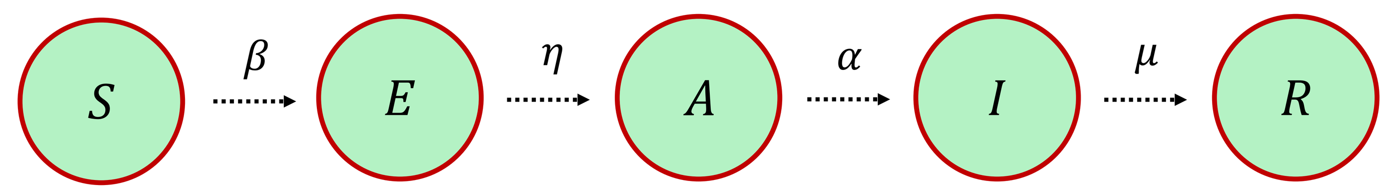

Comparison of MaxEnt and Sequential Neural Likelihood Estimation (SNL) [2] methods for SBI. This model was a one-body, unit-mass gravitational field simulation with three fixed-position attractors. The parameters were the initial velocity of the particle, and the masses of the three attractors. Left: weighted mean paths from both methods alongside the path generated by the mean of the prior distribution and the ground truth path. Right: kernel density estimate of the parameter posterior distributions alongside their respective priors.

## Epidemiology Model

Epidemic spreading in networks can be modeled as a reaction-diffusion process . The reaction corresponds to an infection caused by interactions of subjects within a fully-mixed region or patch of varying granularities (a meta-population), while diffusion corresponds to movement of people (of various infection states) between patches [3].



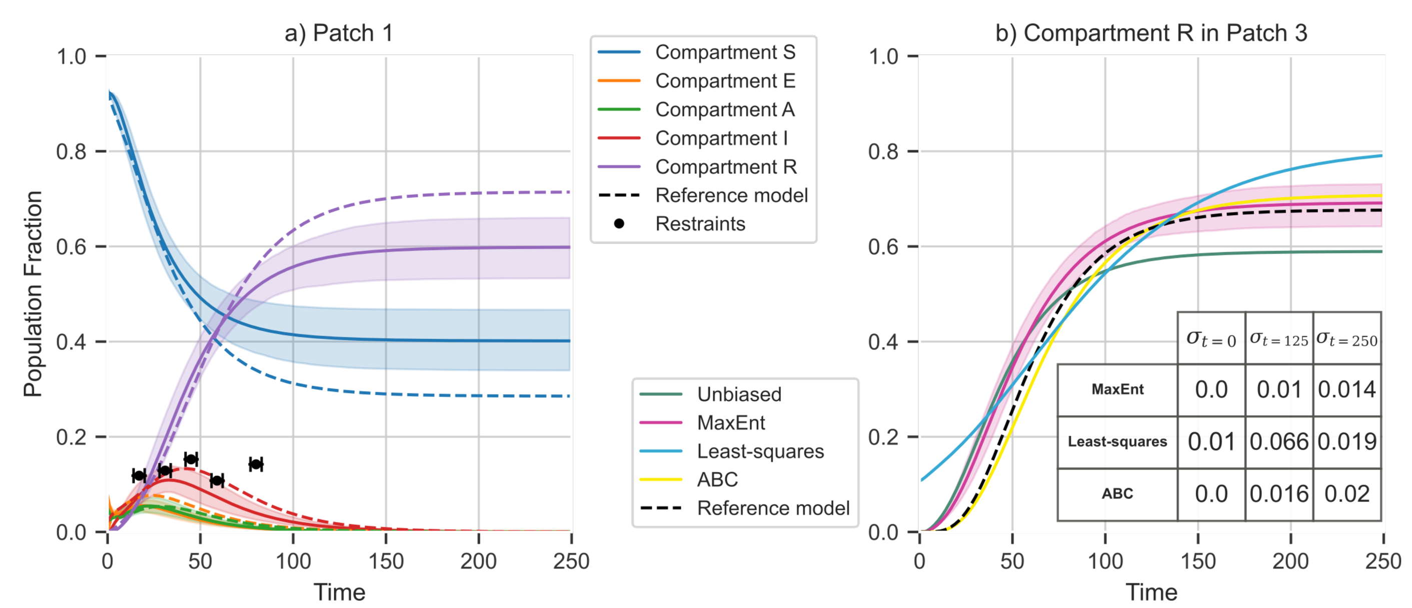

Depending on infection status, individuals are separated into compartments in each pach and a mobility network accounts for all mobility patterns between the patches.




**S:** Susceptible **E:** Exposed **A:** Asymptomatic

**I:** Infected **R:** Recovered

We compare the performance of MaxEnt, least-squares and ABC [4]. Model variances were compared using the predictions for the trajectory of compartment **R** in patch 3.




| | $\sigma_{t=0}$ | $\sigma_{t=125}$ | $\sigma_{t=250}$ |
|---|---|---|---|
| MaxEnt | 0.0 | 0.01 | 0.014 |
| Least-squares | 0.01 | 0.066 | 0.019 |
| ABC | 0.0 | 0.016 | 0.02 |

## Bibliography

1. E.T. Jaynes. Information theory and statistical mechanics. Physical Review, 106(4):620-630, 1957.
2. George Papamakarios, David C. Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. ArXiv, 2019.
3. Jesus Gomez-Gardenes, David Soriano- Panos, and Alex Arenas. Critical regimes driven by recurrent mobility patterns of reaction-diffusion processes in networks. Nature Physics, 14(4):391395, 2018.
4. Mark A Beaumont, Wenyang Zhang, and David J Blading. Approximate Bayesian computation in population genetics. Genetics, 162(4): 2025-2035, 2002.