EPI scaling

Sean Bittner

September 17, 2020

1 Introduction

A relative strength of EPI is its scalability in parameter dimension $|z|$. This write-up compares the performance of EPI to SMC-ABC and SNPE as $|z|$ is increased. The application is to condition arbitrary rank-2 RNNs on a regime of stable amplification.

SMC-ABC: Traditional approaches to likelihood-free inference – approximate Bayesian computation (ABC) methods – randomly sample parameters z until a suitable set is obtained. State-of-the-art ABC methods leverage sequential monte-carlo (SMC) sampling techniques to obtain parameter sets more efficiently. To obtain more parameter samples, SMC-ABC must be run from scratch again. ABC methods do not confer log probabilities of samples.

SNPE: Like EPI, sequential neural posterior estimation (SNPE) uses deep learning to produce flexible posterior approximations. Like traditional Bayesian inference methods, SNPE conditions directly on the statistics of data. This differs from EPI, where posteriors are conditioned on emergent properties (moment constraints on the posterior predictive distribution). Peculiarities of SNPE (density estimation approach, two deep networks) make scaling in z prohibitive.

Rank-2 RNN: The model we use is a rank-2 RNN with N neurons:

$$\begin{aligned} U = \begin{bmatrix} u_1 & u_2 \end{bmatrix}, V = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \\ J = UV^\top + g\chi \\ u_1, u_2, v_1, v_2 \in [-1, 1]^N, \quad g = 0.01 \\ \tau \dot{r} = -r + Jr \end{aligned}$$

From [1], we know necessary and sufficient conditions for RNNs to exhibit stable amplification. These two conditions are on the maximal real eigenvalue of J ($\text{real}(\lambda_1)$) and the maximal eigenvalue of $J^s = \frac{J+J^\top}{2}$(λ_1^s):

$$\text{real}(\lambda_1) < 1, \lambda_1^s > 1$$

In our analysis, we seek to condition rank-2 networks of increasing size on a regime of stable amplification. Networks with $\text{real}(\lambda_1) = 0.5 \pm 0.5$ and $\lambda_1^s = 1.5 \pm 0.5$ will yield moderate (quantifiable) amplification.

EPI emergent property: EPI can naturally condition on this emergent property.

$$\mathbb{E}\begin{bmatrix} \text{real}(\lambda_1) \\ \lambda_1^s \\ (\text{real}(\lambda_1) - 0.5)^2 \\ (\lambda_1^s - 1.5)^2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.5 \\ 0.25^2 \\ 0.25^2 \end{bmatrix}$$

SNPE observation: SNPE cannot condition on the variance of observations across posterior. Thus, we condition on an observation x_0 located at the mean of our desired emergent property.

$$x_0 = \begin{bmatrix} \text{real}(\lambda_1) \\ \lambda_1^s \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix}$$

SMC-ABC criteria: ABC methods define tolerance ϵ and distance for observed data x_0.

$$\epsilon = 0.5,\ l2 \text{ distance, and } x_0 = \begin{bmatrix} \text{real}(\lambda_1) \\ \lambda_1^s \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix}$$

2 Results

SMC-ABC was run with pyabc package. SNPE was run with delfi package. EPI was run with revision branch of epi package. All methods use a uniform prior form -1 to 1 for each parameter.

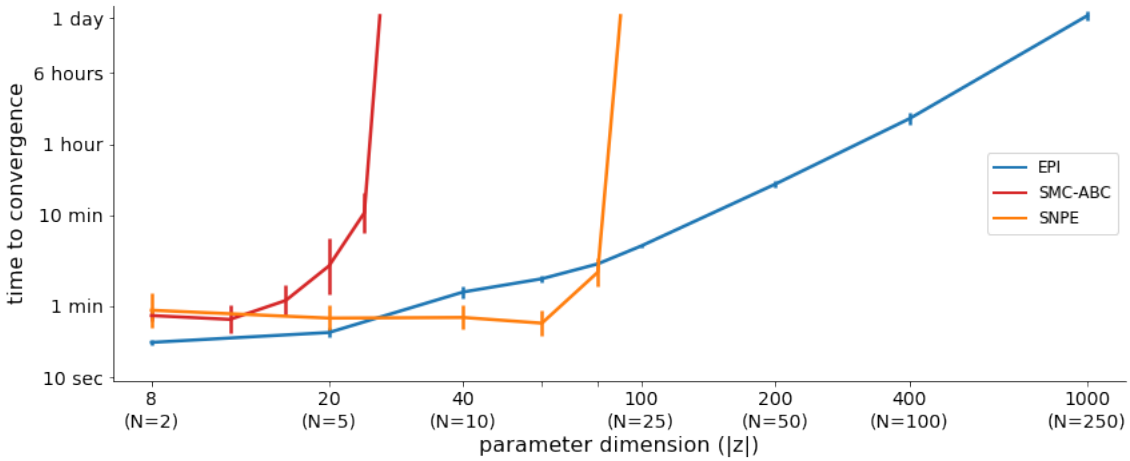

Figure 1: EPI scales with z to high dimensions. Convergence definitions: EPI (blue) - satisfies all moment constraints, SNPE (orange)- produces at least $2/n_{\text{train}}$ parameter samples are in the bounds of emergent property (mean +- 0.5), and SMC-ABC (red) - 100 particles with $\epsilon < 0.5$ are produced.

Figure 1 points:

1. EPI is capable of scaling to high dimensional parameter spaces.
2. SMC has many hyperparameters, of which pyABC selects automatically by running some initial diagnostics upon initialization. In concurrence with the literacture, SMC-ABC fails to converge around 25-30 dimensions, since it's proposal samples never get close enough to the target statistics.
3. We searched over many SNPE hyperparameter choices: $n_{\text{train}} \in [2,000, 10,000, 100,000]$ is the number of simulations run per training epoch, and $n_{\text{mades}} \in [2, 3]$ is the number of masked autoregressive density estimators in the deep parameter distribution architecture. The greater n_{train}, the longer each epoch will take, but the more likely SNPE may converge during that epoch. Greater n_{mades} increases the flexibility of the deep parameter distribution of SNPE, but slows optimization. For the timing plot, we show the fastest among all of these choices, and for the convergence plot, we show the best convergence among all of these choices. During optimization, we used n_{atom}=100 atomic proposals as is recommended.

Figure 2 points:

1. EPI consistently produces the same distribution posterior predictive independent of the dimensionality.

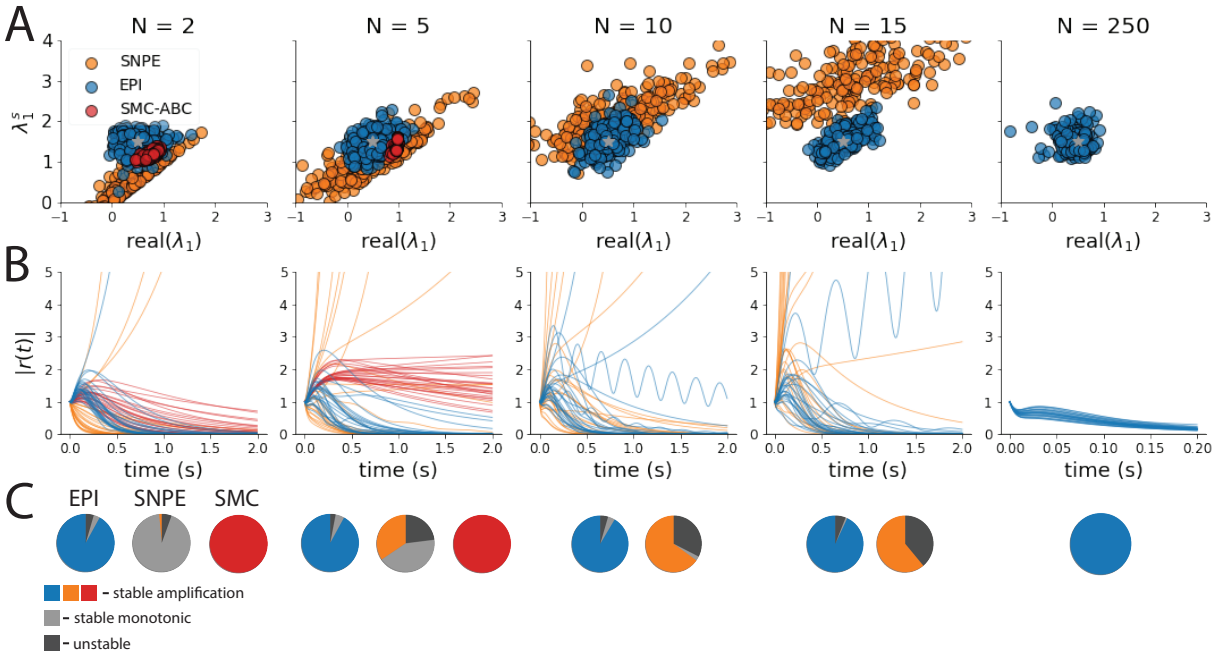

Figure 2: Emergent property fidelity. Top row: Posterior predictive distributions of EPI (blue), SNPE (orange), and SMC-ABC (red). Gray star indicates emergent property mean, and gray dashed lines indicate two standard deviations corresponding to the variance constraint. For $N <= 6$ where SMC-ABC converges, samples are not diverse (path degeneracies). For $N >= 25$, SNPE does not produce a posterior approximation yielding parameters with simulations near x_0. Bottom row: simulations of network parameters resulting from each method ($\tau = 100ms$). Each trace corresponds to simulation of one z. Almost all EPI networks, few SNPE networks, and all SMC-ABC networks exhibit stable amplification. There is little variation in SMC-ABC network simulations.

2. SMC produces a limited variety of parameters due to the nature of its proposal generation algorithm, yet all parameters obtained produce stable amplification.

3. SNPE's posterior predictive distribution is not necessarily close to the conditioning point, and is very dependent on dimensionality.

References

[1] Giulio Bondanelli and Srdjan Ostojic. Coding with transient trajectories in recurrent neural networks. *PLoS computational biology*, 16(2):e1007655, 2020.