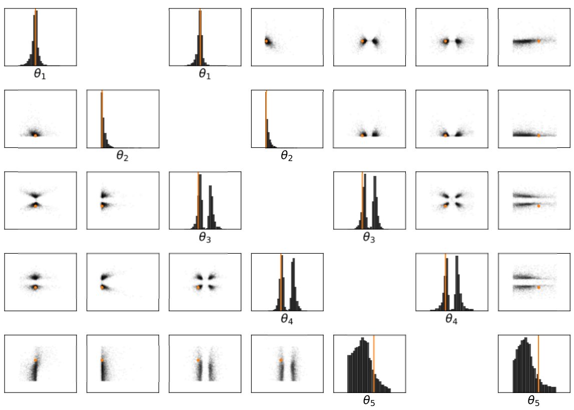
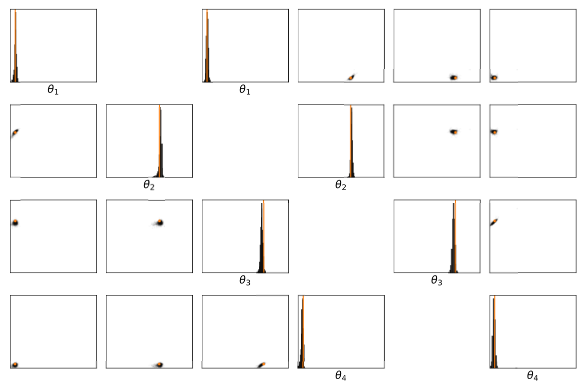
On Contrastive Learning for Likelihood-free Inference
Supplementary Material

A. Additional experimental results

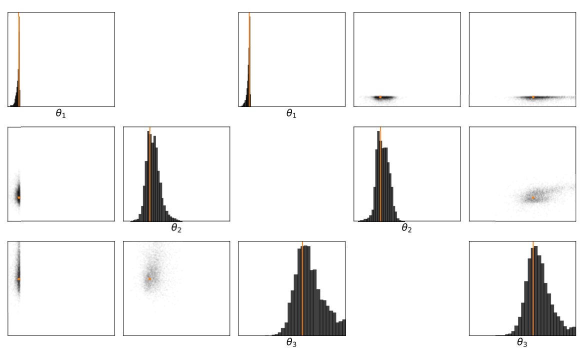
A.1. SRE vs SNPE-C



(a) Nonlinear Gaussian

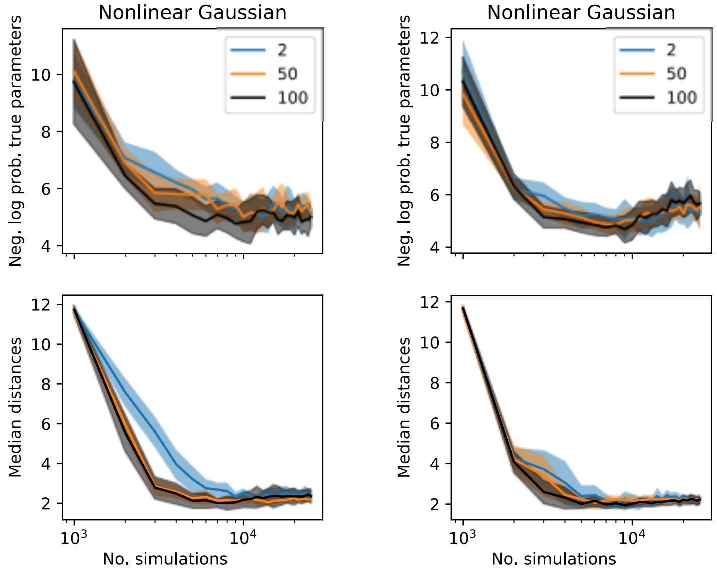


(b) Lotka-Volterra



(c) M/G/1

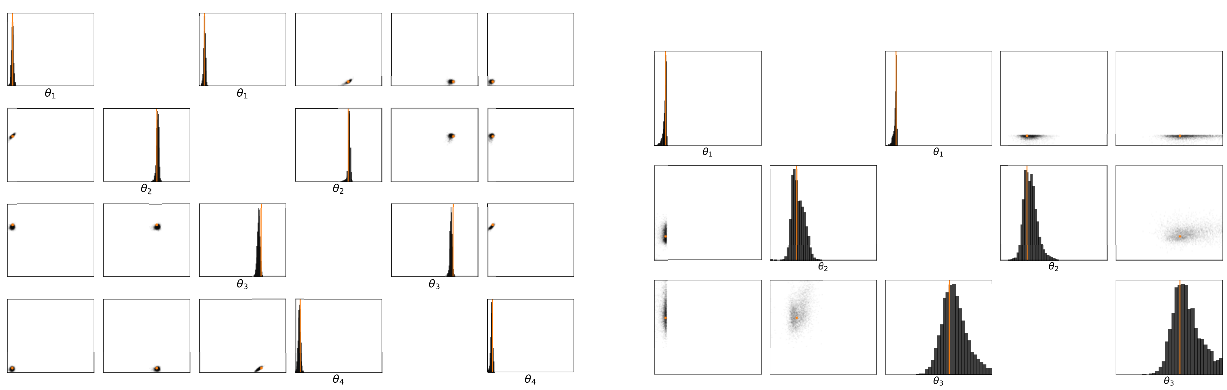
Figure 4: Comparison of posterior samples for SRE (sub-figure left) and SNPE-C (sub-figure right) on each task. For both methods, we use $K = 100$ to generate the contrasting set.



(a) SRE (b) SNPE-C

Figure 5: Comparison of SRE and SNPE-C metrics on Nonlinear Gaussian task. Before recovering the multimodal posterior, the posteriors can sometimes become too confident in certain parameter settings, leading to the observed negative log likelihood behaviour.

A.2. SRE vs SNL



(a) Lotka-Volterra (b) M/G/1

Figure 6: Comparison of posterior samples for SNL (sub-figure left) and SRE (sub-figure right) on each task.

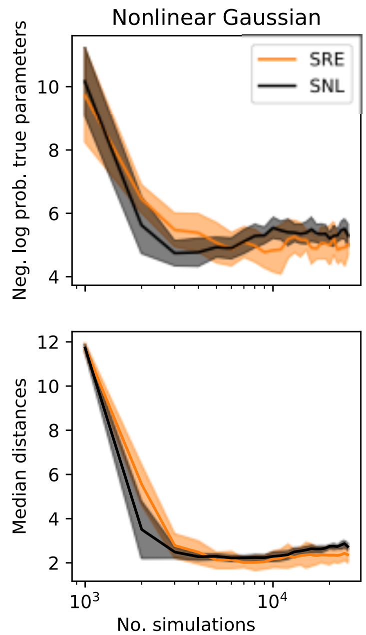


Figure 7: Comparison of Nonlinear Gaussian metrics for SRE and SNL.

A.3. SNPE-C MCMC

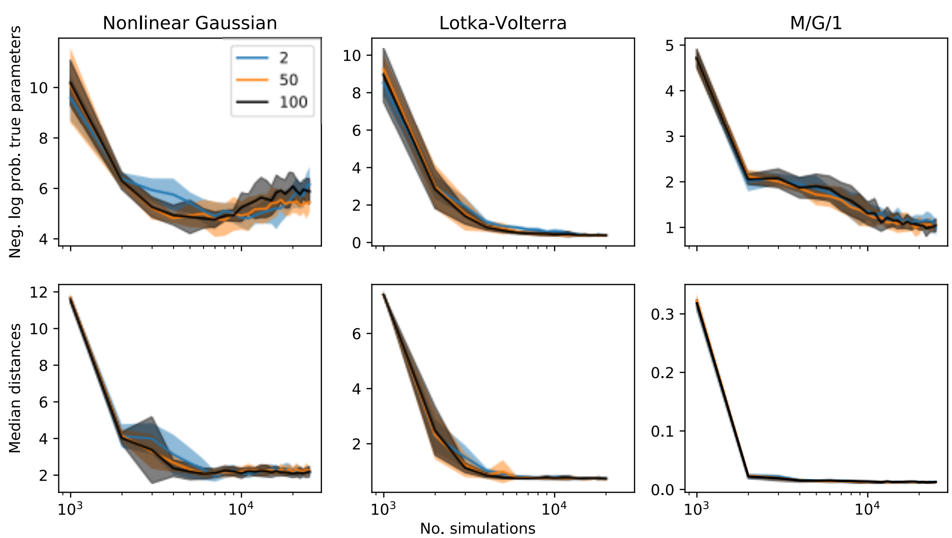


Figure 8: Metrics for Nonlinear Gaussian, Lotka-Volterra, and M/G/1 using SNPE-C with MCMC instead of i.i.d. sampling in each round.

B. Correctness of SNPE-C for arbitrary proposals $\tilde{p}(\boldsymbol{\theta})$

Assuming optimality of the classifier, we have

$$
\begin{aligned}
f_{\boldsymbol{\phi}}(\mathbf{x}, \boldsymbol{\theta}) &= \log \frac{\tilde{p}(\boldsymbol{\theta} \,|\, \mathbf{x})}{\tilde{p}(\boldsymbol{\theta})} + \tilde{c}(\mathbf{x}) && (17)\\
&= \log \frac{p(\boldsymbol{\theta} \,|\, \mathbf{x})}{p(\boldsymbol{\theta})} + c(\mathbf{x}), \text{ where } c(\mathbf{x}) = \tilde{c}(\mathbf{x}) + \log \frac{p(\mathbf{x})}{\tilde{p}(\mathbf{x})}. && (18)
\end{aligned}
$$

Now, since $f_{\boldsymbol{\phi}}(\mathbf{x}, \boldsymbol{\theta}) = \log \frac{q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \,|\, \mathbf{x})}{p(\boldsymbol{\theta})}$, we have

$$
\begin{aligned}
\log \frac{q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \,|\, \mathbf{x})}{p(\boldsymbol{\theta})} &= \log \frac{p(\boldsymbol{\theta} \,|\, \mathbf{x})}{p(\boldsymbol{\theta})} + c(\mathbf{x}) && (19)\\
\iff \log q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \,|\, \mathbf{x}) &= \log p(\boldsymbol{\theta} \,|\, \mathbf{x}) + c(\mathbf{x}). && (20)
\end{aligned}
$$

Exponentiating and then integrating both sides w.r.t. $\boldsymbol{\theta}$ gives

$$
\int q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \,|\, \mathbf{x}) \, \mathrm{d}\boldsymbol{\theta} = \exp(c(\mathbf{x})) \int p(\boldsymbol{\theta} \,|\, \mathbf{x}) \, \mathrm{d}\boldsymbol{\theta} \implies c(\mathbf{x}) = 0. \tag{21}
$$

Thus for the optimal classifier, we have

$$
q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \,|\, \mathbf{x}) = p(\boldsymbol{\theta} \,|\, \mathbf{x}), \tag{22}
$$

and the parameterized conditional density estimator recovers the true posterior for *any* proposal $\tilde{p}(\boldsymbol{\theta})$.