Leveraging Self-Consistency for Data-Efficient Amortized Bayesian Inference


Marvin Schmitt
Stuttgart, GER


Desi Ivanova
Oxford, UK

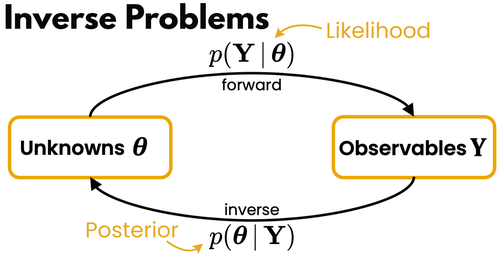

Daniel Habermann
Dortmund, GER


Ullrich Köthe
Heidelberg, GER


Paul Bürkner
Dortmund, GER


Stefan Radev
RPI, US

Inverse Problems

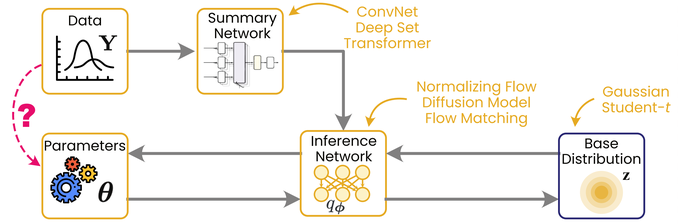


Examples: Gravitational Waves
Decision Making
Climate Modeling

Methods: MCMC, ABC, VI, Flows, ...

Amortized Bayesian Inference

Neural Network Training → Instant Posterior Sampling



Neural Network Training with Maximum Likelihood

$$\mathcal{L}_{\text{NPE}}(\phi) = \mathbb{E}_{p(\mathbf{Y})}\left[\mathrm{KL}(p(\boldsymbol{\theta} \mid \mathbf{Y}) \,\|\, q_\phi(\boldsymbol{\theta} \mid \mathbf{Y}))\right]$$
$$= \mathbb{E}_{p(\boldsymbol{\theta}, \mathbf{Y})}[-\log q_\phi(\boldsymbol{\theta} \mid \mathbf{Y})] + \text{const}$$

Bayes' Theorem

Posterior $p(\boldsymbol{\theta} \mid \mathbf{Y}) = \dfrac{p(\boldsymbol{\theta})\, p(\mathbf{Y} \mid \boldsymbol{\theta})}{p(\mathbf{Y})}$ (Prior, Likelihood, Marginal Likelihood)

re-arrange

$$p(\mathbf{Y}) = \frac{p(\boldsymbol{\theta})\, p(\mathbf{Y} \mid \boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \mathbf{Y})}$$
$$= \text{const}\ \forall\, \theta \in \Theta$$

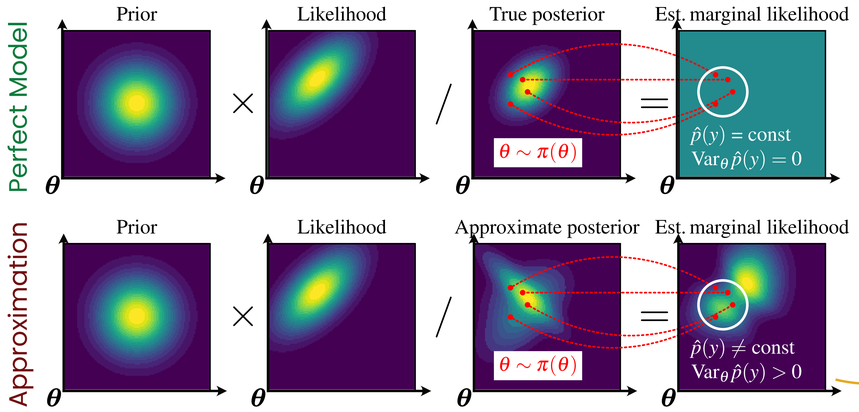
The marginal likelihood is a constant function in θ.

Probabilistic Symmetries

From here on, fix $\mathbf{Y}$. The marginal likelihood can be computed as a function of $\boldsymbol{\theta}$.

$$p(\mathbf{Y}) = \frac{p(\boldsymbol{\theta_1})\, p(\mathbf{Y} \mid \boldsymbol{\theta_1})}{p(\boldsymbol{\theta_1} \mid \mathbf{Y})} = \ldots = \frac{p(\boldsymbol{\theta_K})\, p(\mathbf{Y} \mid \boldsymbol{\theta_K})}{p(\boldsymbol{\theta_K} \mid \mathbf{Y})} \text{ for } \boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_K} \in \Theta$$

It is **constant** in a perfect joint model, but not under imperfect approximations!



Frame as a loss!

Self-Consistency Loss

Neural Posterior Estimation

$$\mathcal{L}_{\text{SC-NPE}}(\boldsymbol{\phi}) = \mathbb{E}_{p(\mathbf{Y})}\Big[\mathbb{E}_{p(\boldsymbol{\theta} \mid \mathbf{Y})}\big[-\log q_\phi(\boldsymbol{\theta} \mid \mathbf{Y})\big] + \lambda\, \text{Var}\Big(\log \frac{p(\boldsymbol{\theta})\, p(\mathbf{Y} \mid \boldsymbol{\theta})}{q_\phi(\boldsymbol{\theta} \mid \mathbf{Y})}\Big)\Big]$$

Self-Consistency Loss with Explicit Likelihood

Neural Posterior and Likelihood Estimation

$$\mathcal{L}_{\text{SC-NPLE}}(\boldsymbol{\phi}) = \mathbb{E}_{p(\mathbf{Y})}\Big[\mathbb{E}_{p(\boldsymbol{\theta} \mid \mathbf{Y})}\big[-\log q_\phi(\boldsymbol{\theta} \mid \mathbf{Y}) - \log q_\eta(\mathbf{Y} \mid \boldsymbol{\theta})\big] + \lambda\, \text{Var}\Big(\log \frac{p(\boldsymbol{\theta})\, q_\eta(\mathbf{Y} \mid \boldsymbol{\theta})}{q_\phi(\boldsymbol{\theta} \mid \mathbf{Y})}\Big)\Big]$$

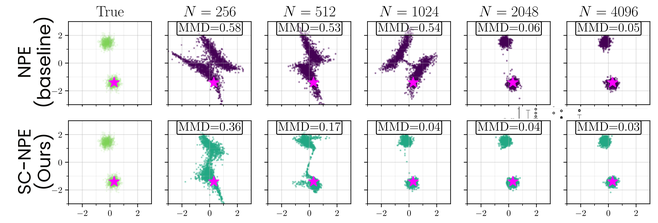
Self-Consistency Loss with Learned Likelihood

Experiments

Gaussian Mixture Model

Posterior estimation (NPE) using explicit likelihood

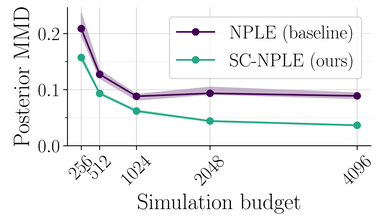
Better posterior samples



Two Moons

Posterior and likelihood estimation (NPLE)

Closer to true posterior (lower MMD)



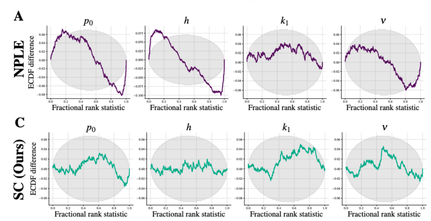
Other results

- Higher likelihood of ground-truths
- Sharper marginal likelihood estimates

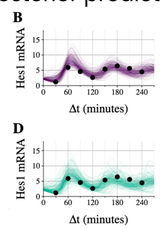
Hes1 Expression

Posterior and likelihood estimation (NPLE)

Better simulation-based calibration



Posterior predictive



More experiments in the paper

ellis European Laboratory for Learning and Intelligent Systems
University of Stuttgart Germany
SimTech
Cyber Valley
University of Oxford
dortmund university
Universität Heidelberg
Informatics for Life
ICML International Conference On Machine Learning