ROBUST AMORTIZED BAYESIAN INFERENCE WITH SELF-CONSISTENCY LOSSES ON UNLABELED DATA

Aayush Mishra[1*], Daniel Habermann[1*], Marvin Schmitt[2], Stefan T. Radev[3], Paul-Christian Bürkner[1]

[1]TU Dortmund University, Germany [2]Independent Scientist [3]Rensselaer Polytechnic Institute, USA [*]equal contribution

ICLR International Conference On Learning Representations

SELF-CONSISTENCY LOSSES

Bayesian self-consistency

Independent of parameters → $p(x)$ = $\frac{p(x \mid \theta^{(1)})\,p(\theta^{(1)})}{p(\theta^{(1)} \mid x)} = \ldots = \frac{p(x \mid \theta^{(L)})\,p(\theta^{(L)})}{p(\theta^{(L)} \mid x)}$ → This ratio remains constant

$q_\phi(\theta \mid x) \approx p(\theta \mid x)$

In ABI, generative neural networks approximate the posterior

$\frac{p(x \mid \theta^{(1)})\,p(\theta^{(1)})}{q(\theta^{(1)} \mid x)} \neq \ldots \neq \frac{p(x \mid \theta^{(L)})\,p(\theta^{(L)})}{q(\theta^{(L)} \mid x)}$

Independence breaks when the parameters are approximate

Unsupervised self-consistency losses

$$(q^*, h^*) = \arg\min_{q,h} \mathbb{E}_{x^* \sim p^*(x)} \left[C \left(\frac{p(x^* \mid \theta)\,p(\theta)}{q(\theta \mid h(x^*))} \right) \right]$$

self-consistency score

posterior approximator

Summary network to extract lower dimensional summary statistics

As the expectation runs over only unlabeled data, no parameters are required and hence the SC loss is unsupervised

The choice of self-consistency score C

$$C\left(\frac{p(x^* \mid \theta)\,p(\theta)}{q(\theta \mid h(x^*))}\right) = \mathrm{Var}_{\theta \sim p_C(\theta)}\left[\log p(x^* \mid \theta) + \log p(\theta) - \log q(\theta \mid h(x^*))\right]$$

Current posterior approximate as the proposal distribution

Semi-supervised losses

$$(q^*, h^*) = \arg\min_{q,h} \mathbb{E}_{(\theta,x)\sim p(\theta,x)}\left[S(q(\theta \mid h(x)), \theta)\right] + \lambda \cdot \mathbb{E}_{x^* \sim p^*(x)}\left[C\left(\frac{p(x^* \mid \theta)\,p(\theta)}{q(\theta \mid h(x^*))}\right)\right]$$

Supervised simulation-based NPE loss

Unsupervised self-consistency loss

PROPOSITIONS

1. Self-consistency losses are strictly proper

Let C be a score that is globally minimised iff its functional argument is constant is constant across the support of the posterior almost everywhere.

$$C\left(\frac{p(x \mid \theta)\,p(\theta)}{q(\theta \mid h(x))}\right)$$

Then C applied to self-consistency ratio with known likelihood is a strictly proper loss. It is globally minimised iff $q(\theta \mid x) \approx p(\theta \mid x)$

2. Variance based SC loss is strictly proper

The variance-based loss

$$\mathrm{Var}_{\theta \sim p_C(\theta)}\left[\log p(x \mid \theta) + \log p(\theta) - \log q(\theta \mid h(x))\right]$$

Is strictly proper provided the support of proposal encompasses the support of posterior.

3. Semi-supervised loss is strictly proper

Under the assumption of Prop. 1, the semi-supervised loss

$$\mathbb{E}_{(\theta,x)\sim p(\theta,x)}\left[S(q,\theta)\right] + \lambda \cdot \mathbb{E}_{x^* \sim p^*(x)}\left[C(\,.\,)\right]$$

Is strictly-proper for any choice of $p^*(x)$.

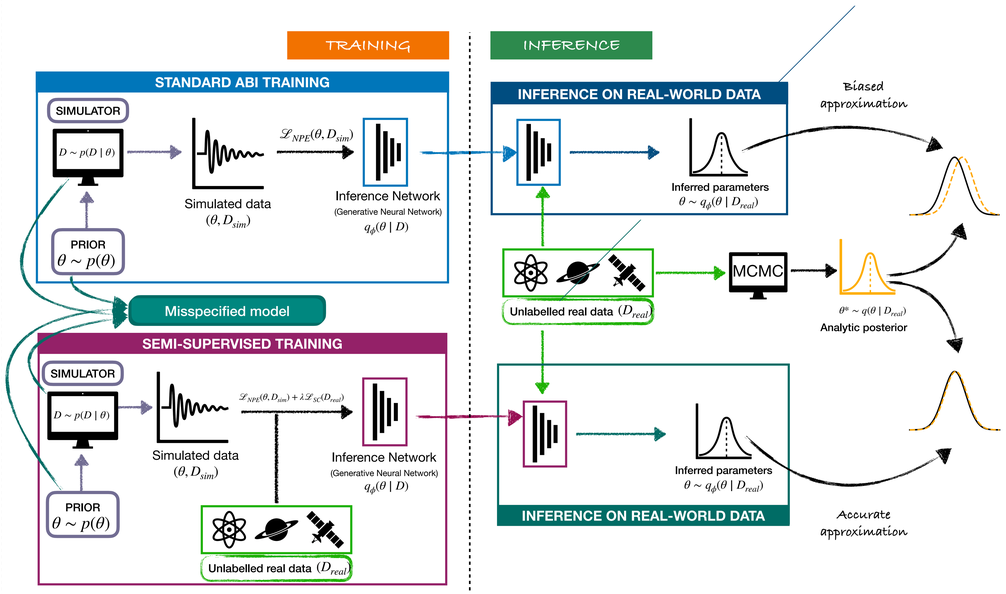
4. Self-consistency for approximate likelihood

If posterior is known and likelihood is approximated, then

$$C\left(\frac{q(x \mid \theta)\,p(\theta)}{p(\theta \mid h(x))}\right)$$

is strictly proper - globally minimised iff $q(x \mid \theta) \approx p(x \mid \theta)$

Closing the Simulation Gap — with unlabeled real data you already have



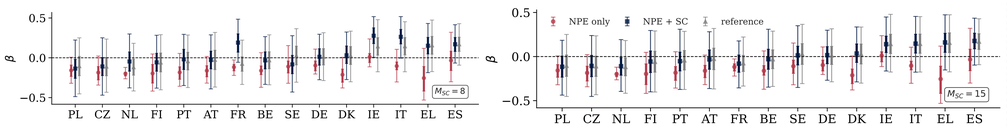
MOTIVATION

- Amortized Bayesian Inference (ABI) relies heavily on simulated training data but in practice, simulators are often misspecified.
- Standard ABI methods lead to overconfident and biased posteriors under simulation gap.
- We propose a semi-supervised training method for ABI using self-consistency losses on unlabeled real-world data.
- Our approach provides reliable posterior estimates under model misspecification as shown by our empirical results.
- We also generalise the self-consistency losses beyond their variance based form and prove their strict-properness.

AIR PASSENGER TRAFFIC: AN AUTOREGRESSIVE MODEL WITH PREDICTORS

$$y_{j,t+1} \sim \mathcal{N}(\alpha_j + y_{j,t}\beta_j + u_{j,t}\gamma_j + w_{j,t}\delta_j, \sigma_j)$$

Data: annual air passenger counts, 15 European countries → USA, 2004–2019

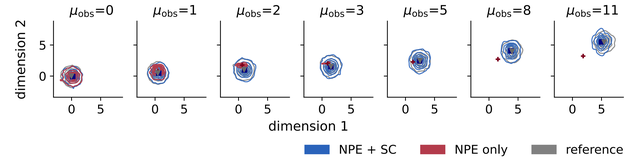


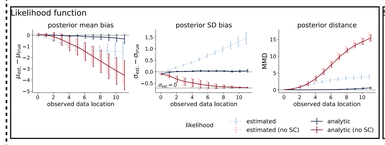
MULTIVARIATE NORMAL MODEL

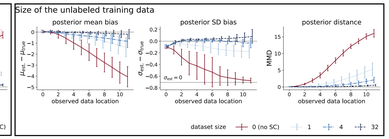
Prior: $\theta \sim \mathcal{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}} I_D)$ Likelihood: $x^{(k)} \sim \mathcal{N}(\theta, \sigma^2_{\text{lik}} I_D)$

For self-consistency training: $x_m^{*(k)} \sim \mathcal{N}(\mu^*, I_D)$, where $\mu^* = 3.0$

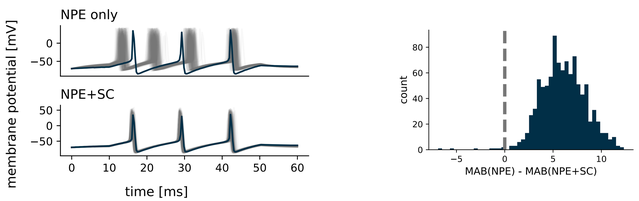
Under prior misspecification $\mu_{\text{prior}} \neq \mu_{\text{obs}}$



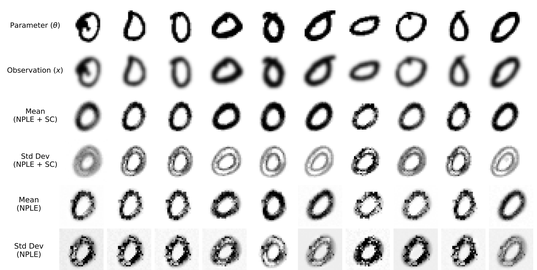




HODGKIN-HUXLEY MODEL OF NEURON ACTIVATION



BAYESIAN DENOISING OF MNIST IMAGES



ACKNOWLEDGEMENTS

Daniel Habermann and Paul Bürkner acknowledge support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projects 508399956 and 528702768. Paul Bürkner further acknowledges support of the DFG Collaborative Research Center 391 (Spatio-Temporal Statistics for the Transition of Energy and Transport) – 520388526. Stefan T. Radev is supported by NSF under Grant No. 2448380. Aayush Mishra acknowledges the computing time provided on the Linux HPC cluster at Technical University Dortmund (LiDO3).