Leveraging Self-Consistency for Data-Efficient Amortized Bayesian Inference

NeurIPS 2023, UniReps workshop





Marvin Schmitt
@MarvinSchmittML

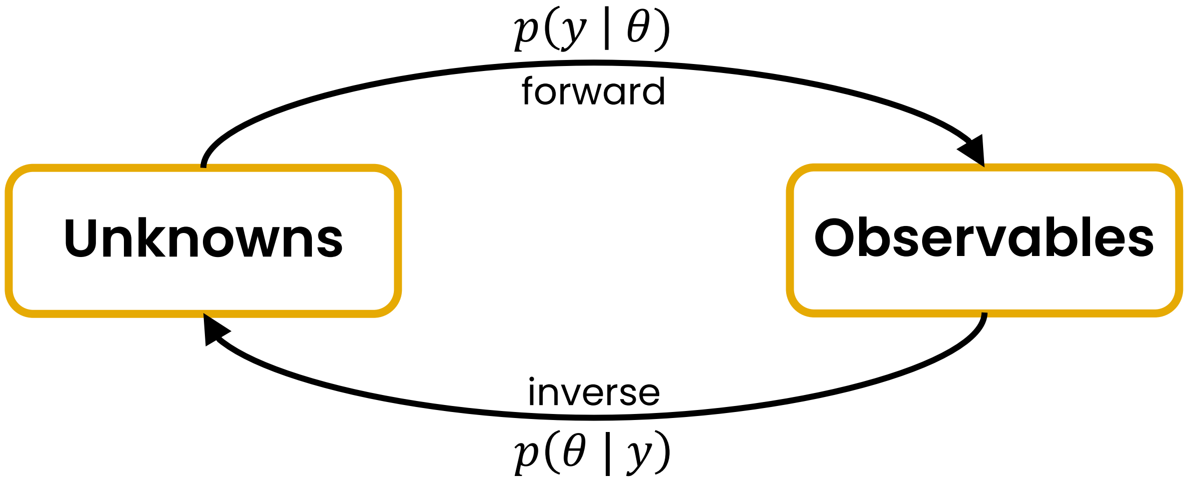
Daniel Habermann
Daniel Habermann

Paul Bürkner
@paulbuerkner

Ullrich Köthe
Ullrich Köthe

Stefan Radev
@stefanradev13

Inverse Problems

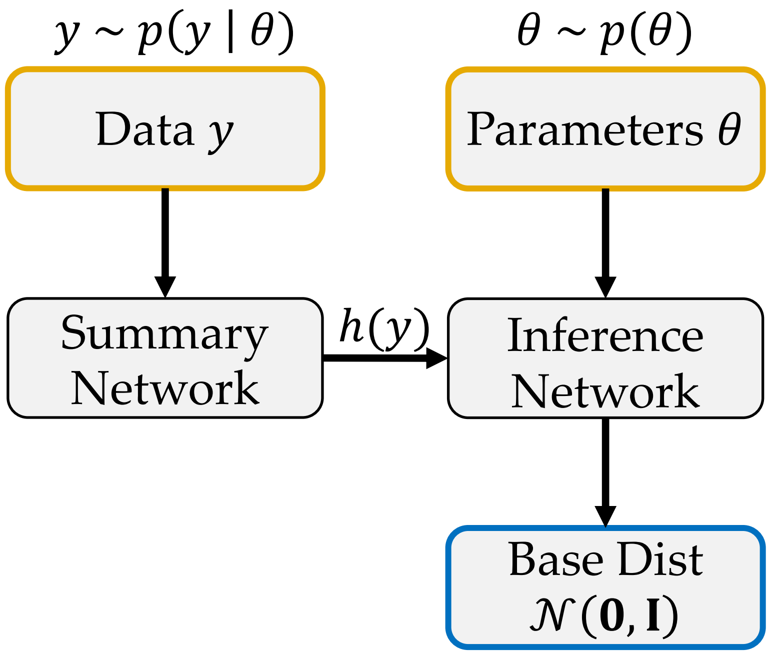


Statistical modeling:	**Parameters θ**	**Data y**
Epidemiology:	Virus attributes	Infection curve (time series)
Image processing:	Crisp image	Blurry image
Psychology:	Cognitive parameters	Reaction times

Amortized Bayesian Inference

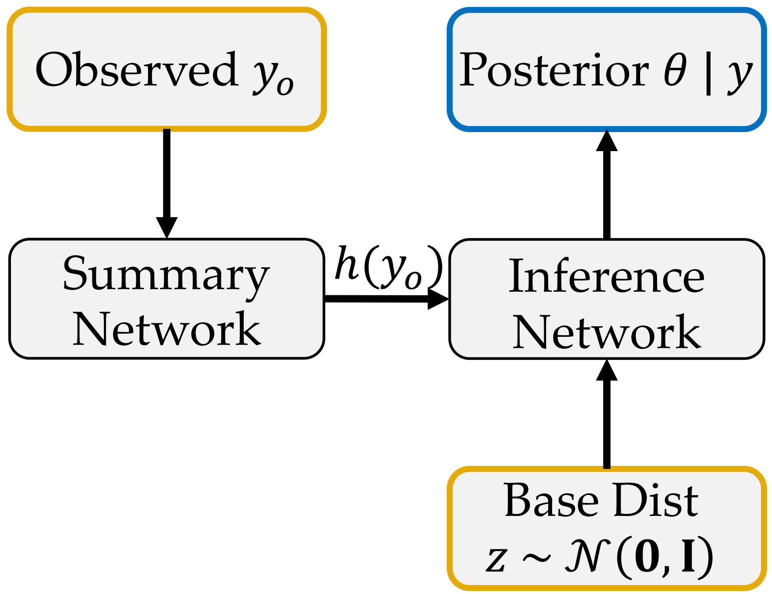
Stage 1: Training

potentially expensive



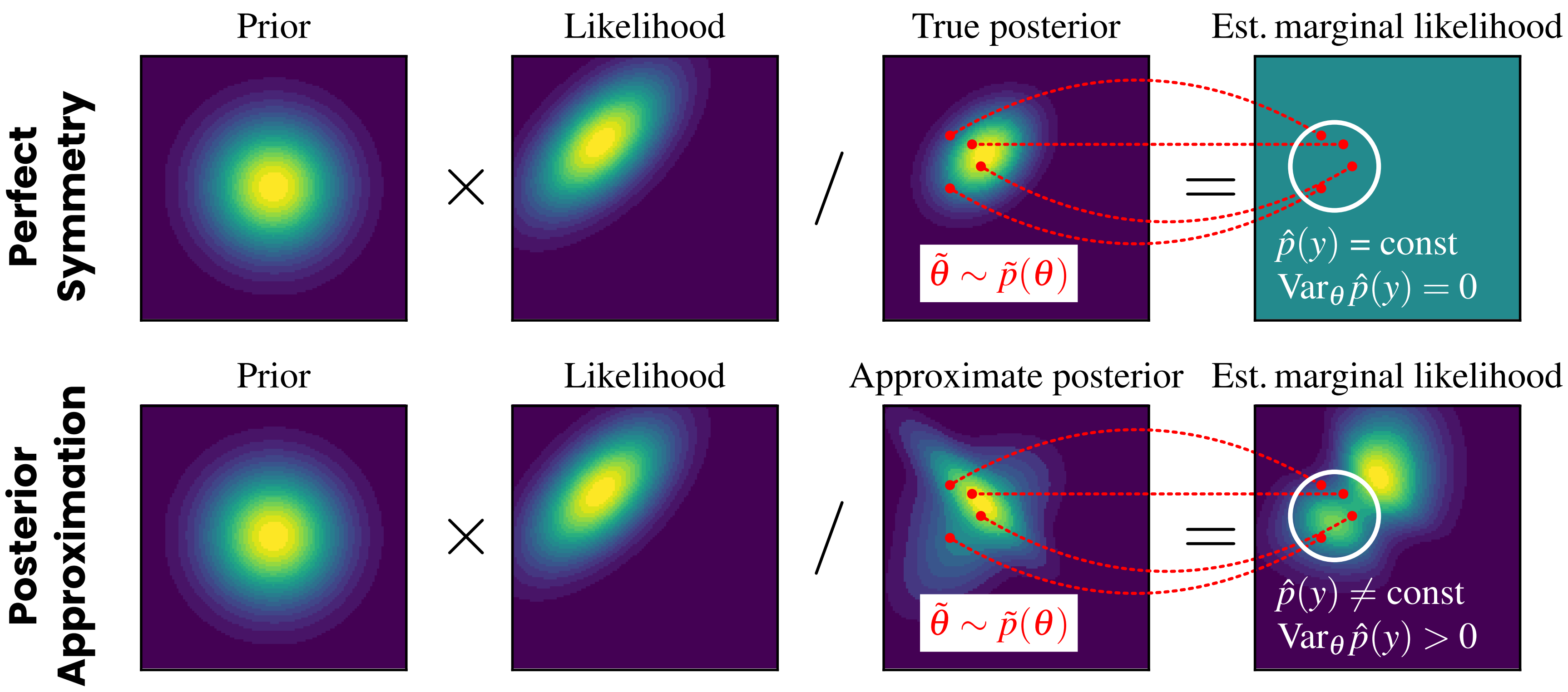
Stage 2: Inference

amortized over many y_o



Self-Consistency Property

$$\boxed{p(y) = p(\theta)\,p(y\,|\,\theta)/p(\theta\,|\,y)} \implies \boxed{\frac{p(\tilde{\theta}_1)\,p(y\,|\,\tilde{\theta}_1)}{p(\tilde{\theta}_1\,|\,y)} = \ldots = \frac{p(\tilde{\theta}_K)\,p(y\,|\,\tilde{\theta}_K)}{p(\tilde{\theta}_K\,|\,y)} \quad \tilde{\theta}_1,\ldots,\tilde{\theta}_K \in \Theta}$$



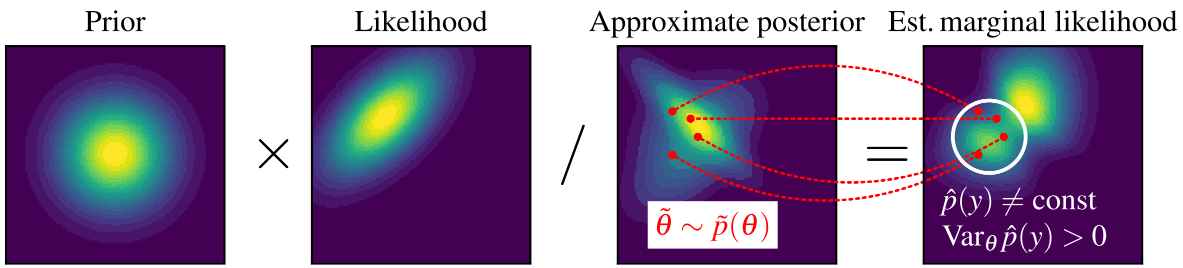
Self-Consistency Loss

- **Idea:** Violations of self-consistency property as loss function

$$\mathcal{L}_{\text{SC}} := \mathbb{E}_{p(y)}\left[\operatorname{Var}_{\tilde{\theta}\sim\tilde{p}(\theta)}\left(\log p(\tilde{\theta}) + \log p(y \mid \tilde{\theta}) - \log q_\phi(\tilde{\theta} \mid y)\right)\right]$$

- **Integration** into standard neural posterior estimation loss

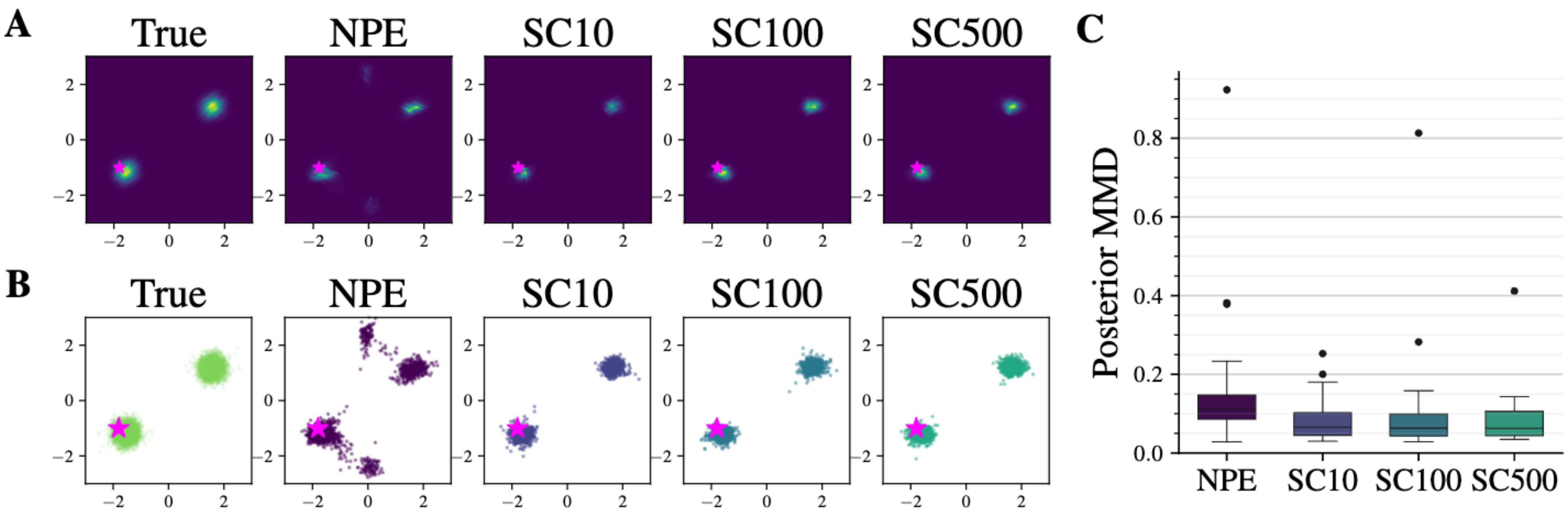
$$\mathcal{L}_{\text{NPE-SC}} := \mathbb{E}_{p(y)}\Bigg[\underbrace{\mathbb{E}_{p(\theta \mid y)}\left[-\log q_\phi(\theta \mid y)\right]}_{\text{NPE loss}} + \underbrace{\lambda \operatorname{Var}_{\tilde{\theta}\sim\tilde{p}(\theta)}\left(\log p(\tilde{\theta}) + \log p(y \mid \tilde{\theta}) - \log q_\phi(\tilde{\theta} \mid y)\right)}_{\text{self-consistency loss } \mathcal{L}_{\text{SC}} \text{ with weight } \lambda \in \mathbb{R}_+}\Bigg]$$



Experiment 1: Gaussian Mixture

Posterior estimation, N = 1024 training budget

- Model: $\theta \sim \mathcal{N}(\theta \,|\, \mathbf{0}, \mathbf{I}), \qquad y \sim 0.5\,\mathcal{N}(y \,|\, \theta, \mathbf{I}/2) + 0.5\,\mathcal{N}(y \,|\, -\theta, \mathbf{I}/2)$
- **Results:** Better density and sampling compared to default (NPE)

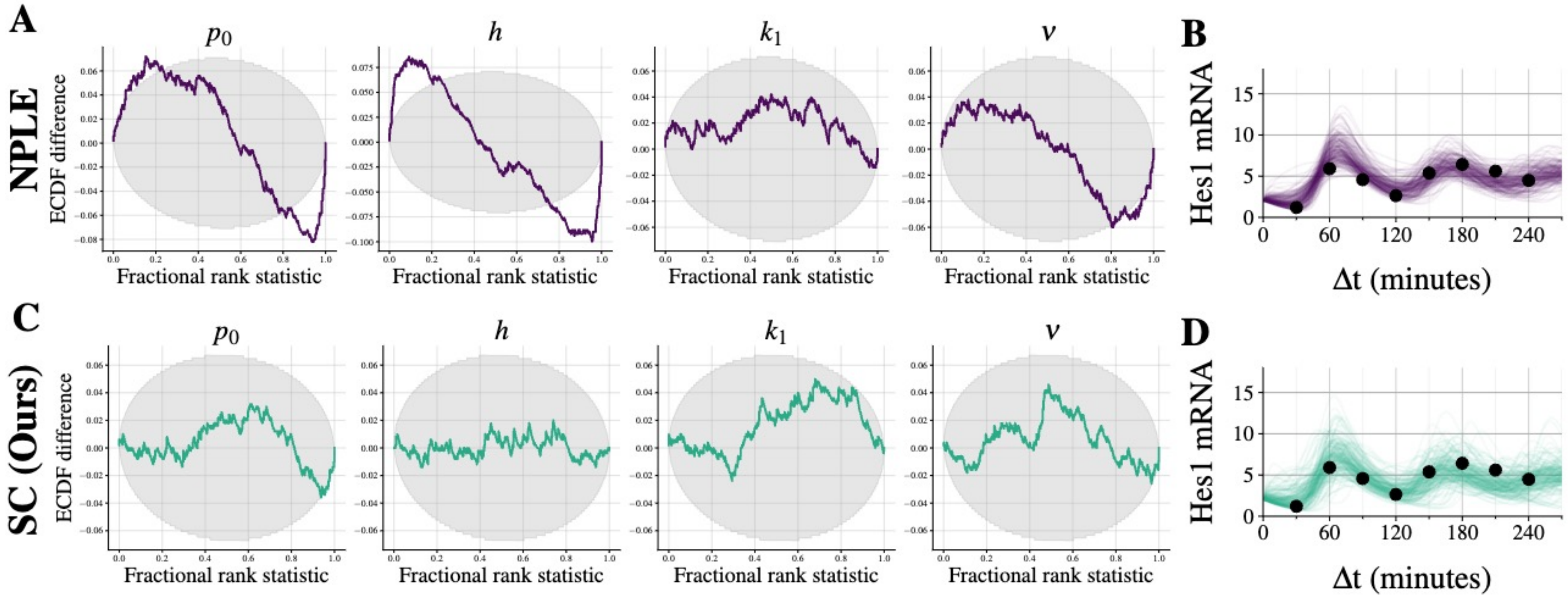


Experiment 2: Hes1 Expression Model

Posterior and likelihood estimation, N = 512 training budget

Results:

- Better simulation-based calibration (SBC; Talts et al., 2018)
- Similar posterior predictive results



Conclusion & Next Steps

Achievements

- Better posterior estimation in low-data settings
- Straightforward integration of the additive self-consistency loss

Next steps

- Explore better and more stable self-consistency training schemes
- Push the envelope on training/inference without ground-truths
- Potential for out-of-distribution settings?

Contact

Marvin Schmitt

@MarvinSchmittML

Daniel Habermann

Daniel Habermann

Paul Bürkner

@paulbuerkner

Ullrich Köthe

Ullrich Köthe

Stefan Radev

@stefanradev13

Partially funded by

Cyber Valley

SimTech Cluster of Excellence

Informatics for Life

STRUCTURES CLUSTER OF EXCELLENCE