Surrogate Likelihoods for Variational Annealed Importance Sampling

Martin Jankowiak   Du Phan

ICML 2022

Consider a global latent variable model:

$$p_\theta(\mathcal{D}, \mathbf{z}) = p_\theta(\mathbf{z}) \prod_{n=1}^{N} p_\theta(y_n | \mathbf{z}, \mathbf{x}_n)$$

dataset   latent variable

How do we approximate the posterior?

$$p_\theta(\mathbf{z} | \mathcal{D})$$

How do we compute the evidence?

$$p_\theta(\mathcal{D}) = \int d\mathbf{z}\, p_\theta(\mathcal{D}, \mathbf{z})$$

Variational Inference		MCMC	
Asymptotically **biased**	😢	Asymptotically **unbiased**	😍
Does estimate **evidence**	😍	Does **not** estimate **evidence**	😢
Does support **amortization**	😍	Does **not** support **amortization**	😢
Does support **subsampling**	😍	Does **not** support **subsampling**	😢
Does support **model learning**	😍	Does **not** support **model learning**	😢

How can we get the best of both worlds?

Can we support data subsampling (mini-batching)?

A principle road block to supporting data subsampling in variational MCMC methods is reliance on the **gradient of the full model log density**

$$\nabla_{\mathbf{z}} \log p_\theta(\mathcal{D}, \mathbf{z})$$

expensive to compute!

Basic strategy: learn a simple parametric *deterministic* approximation to the likelihood that is cheap to evaluate

surrogate log likelihood

$$\hat{\Psi}_{\mathrm{L}}(\mathbf{z}) = \sum_n \omega_n \log p(\tilde{y}_n | \mathbf{z}, \tilde{\mathbf{x}}_n) \approx \log p_\theta(\mathcal{D}|\mathbf{z})$$

learnable weights

surrogate data points

full log likelihood

- Conceptually similar to Bayesian coresets
- Weights can be learned jointly with other variational parameters
- Never need to evaluate the full model log density

Combine surrogate likelihood with VI+MCMC algorithm

DAIS ≡ UHA

Asymptotically **unbiased (?)**

Does estimate **evidence**

Does support **amortization**

Does support **model learning**

Does **not** support **subsampling**

Differentiable Annealed Importance Sampling and the Perils of Gradient Noise

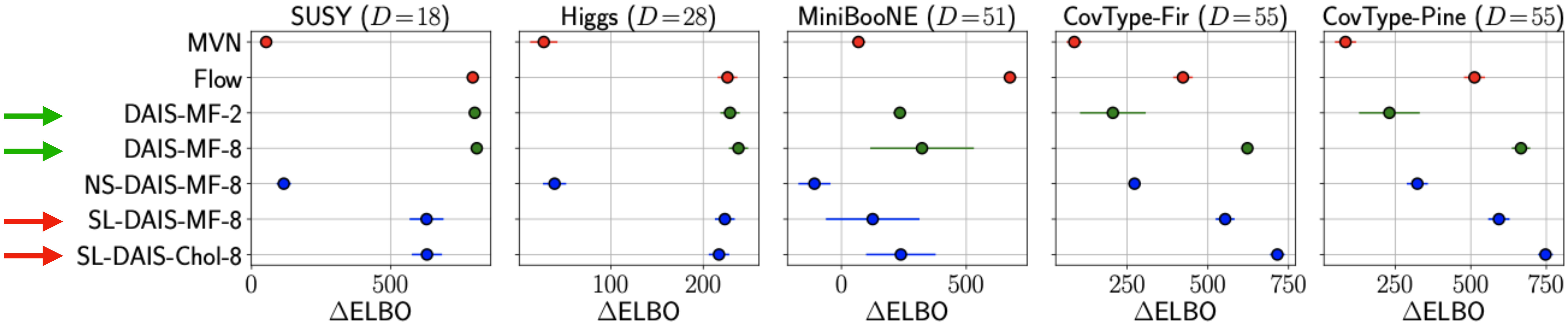
Guodong Zhang[1,2], Kyle Hsu[3], Jianing Li[1], Chelsea Finn[3], Roger Grosse[1,2]
[1]University of Toronto, [2]Vector Institute, [3]Stanford University
{gdzhang, rgrosse}@cs.toronto.edu
{kylehsu, cbfinn}@cs.stanford.edu, jrobert.li@mail.utoronto.ca

MCMC Variational Inference via Uncorrected Hamiltonian Annealing

Tomas Geffner
College of Information and Computer Science
University of Massachusetts, Amherst
Amherst, MA
tgeffner@cs.umass.edu

Justin Domke
College of Information and Computer Science
University of Massachusetts, Amherst
Amherst, MA
domke@cs.umass.edu

Experiment: logistic regression with 50k data points



ELBO improvement w.r.t. mean-field baseline

Summary & Outlook

- Surrogate likelihoods are a useful ingredient in constructing variational inference algorithms that leverage gradient-based MCMC
- Offer a trade-off between posterior fidelity and computational cost
- Can also be applied to local latent variable models: see paper
- Resulting algorithms are easily automated in a PPL like NumPyro
- The design space for further elaborations of variational inference + MCMC remains open!

https://num.pyro.ai/en/stable/autoguide.html