Variational Inference with Continuously-Indexed Normalizing Flows

ICML 2020 INNF+ Workshop

Anthony Caterini, Rob Cornish, Dino Sejdinovic, Arnaud Doucet

University of Oxford

July 18, 2020

Variational Inference

Given a joint target $p_{X,Z}(x, z)$ over data $x \in \mathcal{X}$ and latent variables $z \in \mathcal{Z}$, variational inference (VI) seeks to approximate the intractable posterior $p_{Z|X}$

Variational Inference

Given a joint target $p_{X,Z}(x, z)$ over data $x \in \mathcal{X}$ and latent variables $z \in \mathcal{Z}$, variational inference (VI) seeks to approximate the intractable posterior $p_{Z|X}$

VI methods introduce an approximate posterior q_Z, and are trained to maximize the evidence lower bound (ELBO) objective

Variational Inference

Given a joint target $p_{X,Z}(x, z)$ over data $x \in \mathcal{X}$ and latent variables $z \in \mathcal{Z}$, variational inference (VI) seeks to approximate the intractable posterior $p_{Z|X}$

VI methods introduce an approximate posterior q_Z, and are trained to maximize the evidence lower bound (ELBO) objective

Can also amortize inference across observations

- e.g. Encoder in a VAE [Kingma and Welling, 2014]

Explicit VI with Normalizing Flows

Explicit VI involves specifying q_Z with tractable sampling and density evaluation

Explicit VI with Normalizing Flows

Explicit VI involves specifying q_Z with tractable sampling and density evaluation

Normalizing flows (NFs) [Rezende and Mohamed, 2015] define q_Z explicitly according to the generative process

$$W \sim q_W, \qquad Z := g(W)$$

- g is a diffeomorphism, q_W is a simple base density

Explicit VI with Normalizing Flows

Explicit VI involves specifying q_Z with tractable sampling and density evaluation

Normalizing flows (NFs) [Rezende and Mohamed, 2015] define q_Z explicitly according to the generative process

$$W \sim q_W, \qquad Z := g(W)$$

- g is a diffeomorphism, q_W is a simple base density

Although powerful, NFs can suffer from expressiveness issues, even becoming pathological when the topologies of q_W and $p_{Z|X}$ don't match [Cornish et al., 2020]

Implicit VI with Continuously-Indexed Normalizing Flows

Main contribution: We can gain expressiveness in VI by instead using continuously indexed flows (CIFs) [Cornish et al., 2020], with generative process

$$W \sim q_W, \qquad U \mid W \sim q_{U|W}(\cdot \mid W), \qquad Z := G(W; U),$$

where $q_{U|W}$ is trainable, and each $G(\cdot; u)$ is a normalizing flow

Implicit VI with Continuously-Indexed Normalizing Flows

Main contribution: We can gain expressiveness in VI by instead using continuously indexed flows (CIFs) [Cornish et al., 2020], with generative process

$$W \sim q_W, \qquad U \mid W \sim q_{U|W}(\cdot \mid W), \qquad Z := G(W; U),$$

where $q_{U|W}$ is trainable, and each $G(\cdot; u)$ is a normalizing flow

Any existing normalizing flow g can be used to construct G, e.g.

$$G(w; u) := e^{s(u)} \odot g(w) + t(u)$$

Implicit VI with Continuously-Indexed Normalizing Flows

Main contribution: We can gain expressiveness in VI by instead using continuously indexed flows (CIFs) [Cornish et al., 2020], with generative process

$$W \sim q_W, \qquad U \mid W \sim q_{U|W}(\cdot \mid W), \qquad Z := G(W; U),$$

where $q_{U|W}$ is trainable, and each $G(\cdot; u)$ is a normalizing flow

Any existing normalizing flow g can be used to construct G, e.g.

$$G(w; u) := e^{s(u)} \odot g(w) + t(u)$$

Now, $q_Z(z) := \int q_{Z,U}(z, u) \, \mathrm{d}u$ is defined implicitly over auxiliary variables $u \in \mathcal{U}$, rendering the original ELBO intractable

Auxiliary Variational Inference

Since $q_{Z,U}$ is tractable for CIFs, we can train using the framework of auxiliary variational inference [Agakov and Barber, 2004]

Auxiliary Variational Inference

Since $q_{Z,U}$ is tractable for CIFs, we can train using the framework of auxiliary variational inference [Agakov and Barber, 2004]

New objective is the auxiliary ELBO, having auxiliary inference distribution $r_{U|Z}$

Auxiliary Variational Inference

Since $q_{Z,U}$ is tractable for CIFs, we can train using the framework of auxiliary variational inference [Agakov and Barber, 2004]

New objective is the auxiliary ELBO, having auxiliary inference distribution $r_{U|Z}$

CIFs already provide an auxiliary inference distribution

- For multi-layer CIFs, this has the correct factorization structure

Results - Mixture of Gaussians

We compare explicit VI with a neural spline flow (NSF) [Durkan et al., 2019] to a CIF-NSF on a mixture of Gaussians inference problem with varied initial noise σ_0

Results - Mixture of Gaussians

We compare explicit VI with a neural spline flow (NSF) [Durkan et al., 2019] to a CIF-NSF on a mixture of Gaussians inference problem with varied initial noise σ_0

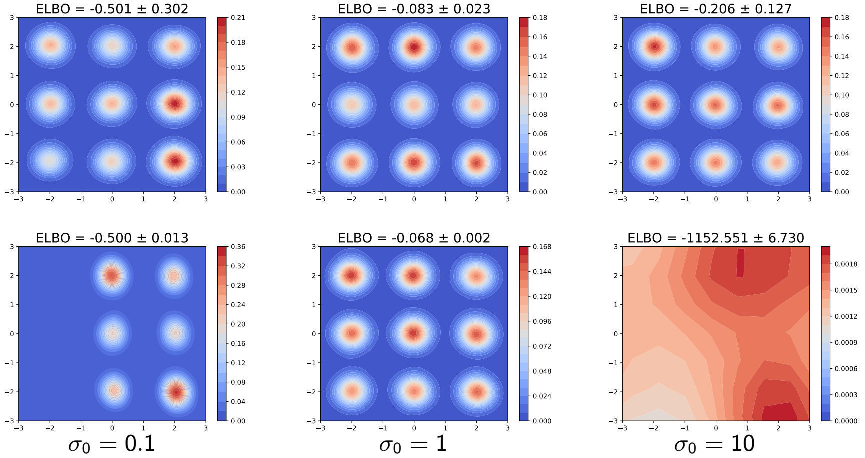

Figure 1: CIFs on top, NSFs on bottom.

Results - Image Datasets

We further compare the performance of generative models which use either NSFs, CIF-NSFs, or baseline VAEs as an amortized inference method

Results - Image Datasets

We further compare the performance of generative models which use either NSFs, CIF-NSFs, or baseline VAEs as an amortized inference method

Table 1: Test-set marginal log-likelihood averaged over three runs.

	VAE	NSF	CIF-NSF
MNIST	$-87.37_{\pm 0.15}$	$-82.95_{\pm 0.11}$	$\mathbf{-82.22_{\pm 0.13}}$
FASHION-MNIST	$-217.82_{\pm 0.07}$	$-215.45_{\pm 0.08}$	$\mathbf{-214.50_{\pm 0.11}}$

Thank you!



Figure 2: Joint work with Rob Cornish, Dino Sejdinovic, and Arnaud Doucet

References

Anthony Caterini, Rob Cornish, Dino Sejdinovic, and Arnaud Doucet. Variational inference with continuously-indexed normalizing flows. In *ICML Workshop on Invertible Neural Networks and Normalizing Flows*, 2020.

Rob Cornish, Anthony L Caterini, George Deligiannidis, and Arnaud Doucet. Relaxing bijectivity constraints with continuously-indexed normalising flows. In *International Conference on Machine Learning*, 2020.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR*, 2014.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pages 1530–1538, 2015.

Felix V Agakov and David Barber. An auxiliary variational method. In *International Conference on Neural Information Processing*, pages 561–566. Springer, 2004.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, pages 7509–7520, 2019.