Variational Inference with Implicit Models

Dmitry Molchanov

Samsung AI Center, Samsung-HSE Laboratory

Variational inference vs variational learning

- Variational inference:

  Given the joint $p(x, z) = p(x|z)p(z)$, find the posterior $p(z|x)$

$$\mathcal{L} = \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log \frac{p(x|z)p(z)}{q_\phi(z|x)} \to \max_\phi$$

- Variational learning:

  Approximately maximize the marginal log-likelihood $\log \mathrm{p}(\mathrm{x}|\theta_{lh}, \theta_p)$:

$$\log \mathrm{p}\big(\mathrm{x}\big|\theta_{lh}, \theta_p\big) \geq \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log \frac{p_{\theta_{lh}}(x|z) p_{\theta_p}(z)}{q_\phi(z|x)} \to \max_{\phi, \theta_{lh}, \theta_p}$$

This talk

This talk **is** about:

- How to perform variational inference and / or learning in implicit models?

This talk **is not** about:

- How to apply variational inference and / or learning?
- How to apply implicit models?
- Fancy experiments
- Particular models

Implicit variational learning

$$\log \mathrm{p}(\mathrm{x}|\theta_{lh}, \theta_p) \geq \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log \frac{p_{\theta_{lh}}(x|z) p_{\theta_p}(z)}{q_\phi(z|x)} \rightarrow \max_{\phi, \theta_{lh}, \theta_p}$$

We need:

- $\nabla_z \log q_\phi(z|x)$, reparameterization
- $\nabla_{\theta_p} \log p_{\theta_p}(z), \nabla_z \log p_{\theta_p}(z)$
- $\nabla_{\theta_{lh}} \log p_{\theta_{lh}}(x|z), \nabla_z \log p_{\theta_{lh}}(x|z)$

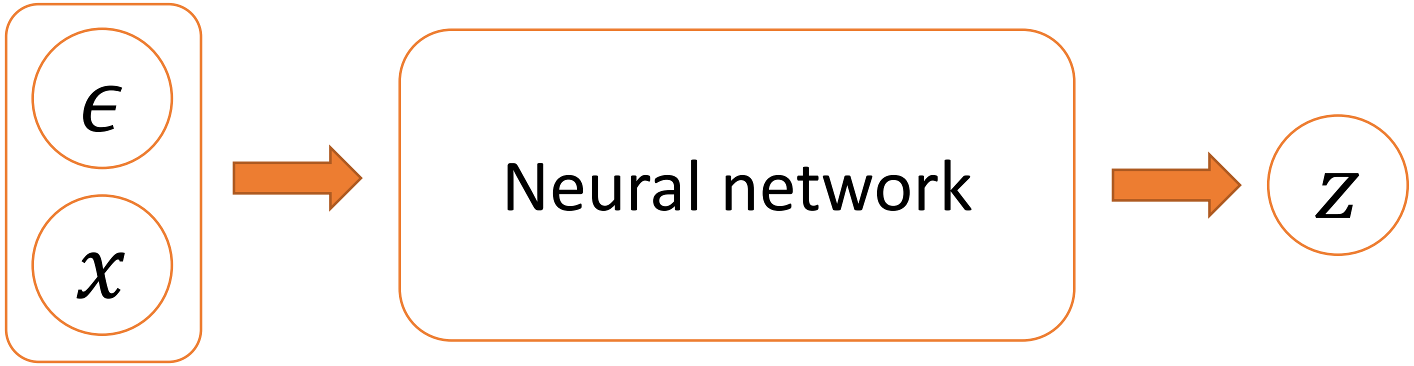
Enough for
variational inference
(not learning)

Implicit models: worst case

We only can **sample** from $q_\phi(z|x), p_{\theta_p}(z), p_{\theta_{lh}}(x|z)$

Implicit posterior

- Explicit posteriors are too simple (e.g. a fully-factorized Gaussian)
  - Exception: normalizing flows – alternative to implicit models
- Arbitrary implicit generator:



Implicit prior

- Hierarchical prior induces an implicit prior

$$p(z) = \int p(z|\psi)p(\psi)d\psi$$

- Incremental learning with implicit posteriors

$$p_{t+1}(z) = q_t(z)$$

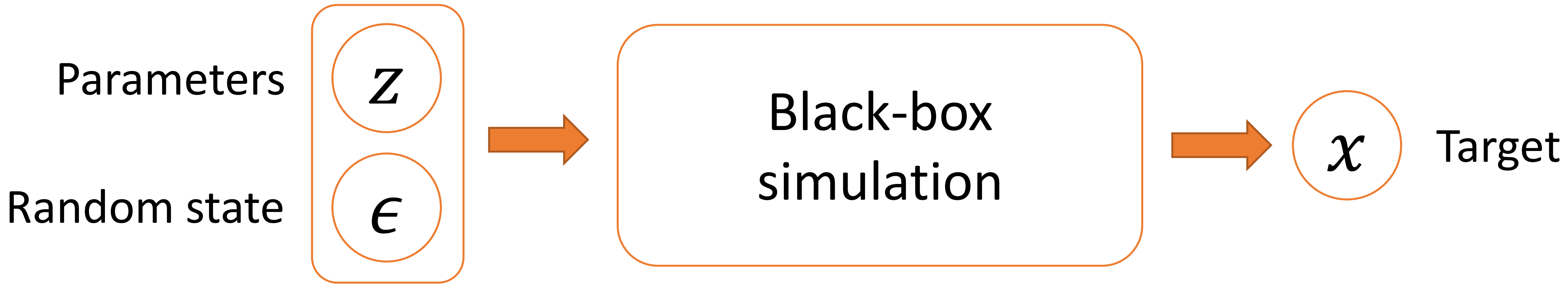
- Optimal prior may be implicit

$$p_{optimal}(z) = \int q_\phi(z|x)p(x)dx$$

Implicit likelihood

(a.k.a. ABC, approximate Bayesian computation)

- We have a black-box simulator that samples $x \sim p_{\theta_{lh}}(x|z)$
- $\log p_{\theta_{lh}}(x|z)$ and its ∇ are unknown



- How to find parameters z?

A common scenario in practical applications (physics, biology, etc.)

Approches to implicit modelling

- Discriminator-based density ratio estimation
- Approaches based on reverse models
  - Hierarchical variational inference
  - Unbiased implicit variational inference
- Denoising-based inference
- Other approaches (not mentioned here):
  - SIVI (semi-implicit VI)
  - KIVI (kernel implicit VI)
  - OPVI (operator VI)
  - …

Why these approaches?

- Discriminator-based density ratio estimation
  - The most wide-spread
  - The most general approach
- Hierarchical variational inference
  - The most simple approach
- Unbiased implicit variational inference
  - Finally feels like “the right way to do it”
- Denoiser-based gradient estimation
  - An unexpected beautiful result

Discriminator-based density ratio estimation

$$\mathcal{D}(z) \approx \frac{q(z)}{p(z)}$$

Minimize discriminator loss:

$$\mathbb{E}_{z\sim p(z)} \log\big(1 + \mathcal{D}(z)\big) \; - \mathbb{E}_{z\sim q(z)} \log \frac{\mathcal{D}(z)}{1 + \mathcal{D}(z)} \to \min_{\mathcal{D}}$$

$$\int \left[p(z) \log\big(1 + \mathcal{D}(z)\big) - q(z) \log \frac{\mathcal{D}(z)}{1 + \mathcal{D}(z)}\right] dz \to \min_{\mathcal{D}}$$

$$p(z) \log\big(1 + \mathcal{D}(z)\big) - q(z) \log \frac{\mathcal{D}(z)}{1 + \mathcal{D}(z)} \to \min_{\mathcal{D}}$$

$$\frac{p(z)}{1 + \mathcal{D}(z)} - \frac{q(z)}{\mathcal{D}(\mathrm{z})} + \frac{q(z)}{1 + \mathcal{D}(z)} = 0 \Leftrightarrow p(z)\mathcal{D}(z) - q(z)\big(1 + \mathcal{D}(z)\big) + q(z)\mathcal{D}(z) = 0$$

$$p(z)\mathcal{D}(\mathrm{z}) - \mathrm{q}(\mathrm{z}) = 0$$

Discriminator-based density ratio estimation

- Variational learning / inference is optimization of ELBO
- ELBO consists of expected log-density-ratios
- Intractable $\Rightarrow$ train a discriminator to approximate density ratio

Prior-contrastive vs joint-contrastive inference

- VAE ELBO:

$$\mathcal{L} = \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log \frac{p(x|z)p(z)}{q_\phi(z|x)}$$

- Prior-contrastive formulation:

$$\mathcal{L} = \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log p(x|z) - \mathbb{E}_{p(x)} \underbrace{\mathbb{E}_{q_\phi(z|x)} \log \frac{q_\phi(z|x)}{p(z)}}_{\mathrm{KL}(q(z|x) \| p(z))}$$

- Joint-contrastive formulation:

$$\mathcal{L} = -\mathbb{E}_{p(x)} \underbrace{\mathbb{E}_{q_\phi(z|x)} \log \frac{q_\phi(z|x)p(x)}{p(x,z)}}_{\mathrm{KL}(q(z|x)p(x) \| p(x,z))} - \mathcal{H}[p(x)]$$

Prior-constrastive adversarial

- Consider the prior-contrastive VAE formulation:

$$\mathcal{L} = \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log p(x|z) - \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log \frac{q_\phi(z|x)}{p(z)}$$

- Approximate density ratio $\mathcal{D}(z) \approx \frac{q_\phi(z|x)}{p(z)}$
- Approximate ELBO:

$$\mathcal{L} \approx \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log p(x|z) - \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \boxed{\log \mathcal{D}(z)}$$

Implicitly depends on ϕ and θ_p!

Prior-constrastive adversarial

- Approximate ELBO:
$$\mathcal{L} = \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log p(x|z) - \mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}(z)$$
- Need explicit likelihood
- Implicit posterior and prior
- How to optimize?
  - $\mathcal{D}(z) = \mathcal{D}_{\phi,\theta_p}(z)$ depends on ϕ and θ_p
  - $\nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi,\theta_p}(z) |_{\phi=\phi_0} = \nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi_0,\theta_p}(z) |_{\phi=\phi_0}$

Prior-contrastive adversarial

$$\nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi,\theta_p}(z)\Big|_{\phi=\phi_0} = \nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi_0,\theta_p}(z)\Big|_{\phi=\phi_0}$$

Assume the optimal discriminator:

$$\nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi,\theta_p}(z)\Big|_{\phi=\phi_0} = \nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \frac{q_\phi(z|x)}{p(z)}\Big|_{\phi=\phi_0} =$$

$$= \nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \frac{q_\phi(z|x) q_{\phi_0}(z|x)}{p(z) q_{\phi_0}(z|x)}\Big|_{\phi=\phi_0} =$$

$$\nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \frac{q_{\phi_0}(z|x)}{p(z)}\Big|_{\phi=\phi_0} + \nabla_\phi \mathrm{KL}(q_\psi \| q_{\psi_0})\Big|_{\phi=\phi_0} =$$

$$\nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \frac{q_{\phi_0}(z|x)}{p(z)}\Big|_{\phi=\phi_0} = \nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi_0,\theta_p}(z)\Big|_{\phi=\phi_0}$$

Prior-constrastive adversarial

- Approximate ELBO:
$$\mathcal{L} = \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log p(x|z) - \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}(z)$$
- Need explicit likelihood
- Implicit posterior and prior
- How to optimize?
  - $\mathcal{D}(z) = \mathcal{D}_{\phi,\theta_p}(z)$ depends on ϕ and θ_p
  - $\nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi,\theta_p}(z)\,|_{\phi=\phi_0} = \nabla_\phi \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}_{\phi_0,\theta_p}(z)\,|_{\phi=\phi_0}$
  - Differentiate through SGD to obtain ∇_{θ_p}!

Joint-contrastive adversarial

- Joint-contrastive formulation:

$$\mathcal{L} = -\mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log \frac{q_\phi(z|x)p(x)}{p(x,z)} - \mathcal{H}[p(x)]$$

- Approximate density ratio:

$$\mathcal{D}(x,z) \approx \frac{q_\phi(z|x)p(x)}{p(x,z)}$$

- Approximate ELBO:

$$\mathcal{L} \mathrel{\tilde{\propto}} - \mathbb{E}_{p(x)}\mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}(x,z)$$

Joint-contrastive adversarial

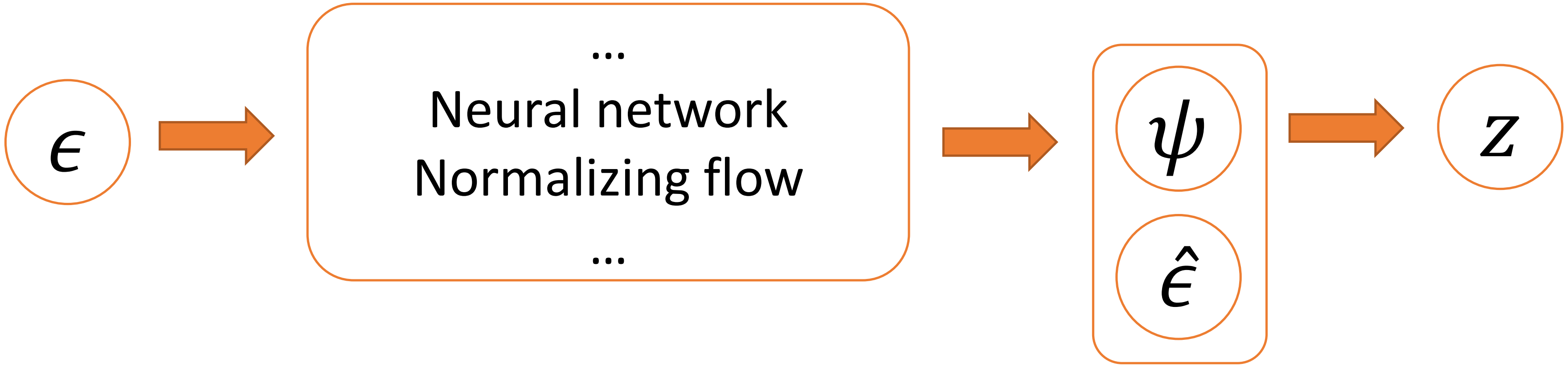
- Approximate ELBO:
$$\mathcal{L} \approx -\mathbb{E}_{p(x)} \mathbb{E}_{q_\phi(z|x)} \log \mathcal{D}(x, z)$$
- All distributions may be implicit
- High-dimensional DRE is difficult
- How to optimize?
  - Same trick holds for ϕ
  - Need to differentiate through SGD to obtain ∇_{lh}, ∇_p

Semi-implicit formulation

- Semi-implicit distribution:

$$q(z) = \int q(z|\psi)q(\psi)d\psi$$

- Even if $q(z|\psi)$ and $q(\psi)$ are explicit, $q(z)$ may be implicit
- If $q(\psi)$ is implicit, $q(z)$ can model any implicit distribution
    - Consider $q(z|\psi) = \delta(z - \psi)$



Hierarchical variational inference

- Semi-implicit distribution:

$$q_\phi(z) = \int q_\phi(z|\psi) q_\phi(\psi) d\psi$$

- Both $q(z|\psi)$ and $q(\psi)$ are explicit
  - Example: $q(z|\psi) = \mathcal{N}(\psi, \sigma^2)$
  - Example: $q(\psi) = \mathrm{NF}(\epsilon)$
- ELBO:

$$\mathcal{L} = \mathbb{E}_{q_\phi(z)} \log p(x|z)p(z) - \mathbb{E}_{q_\phi(z)} \log q_\phi(z)$$

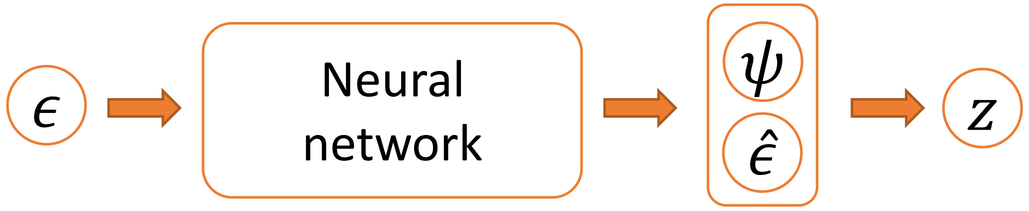
HVI: bounding the entropy

$$-\mathbb{E}_{q(z)} \log q(z) = -\mathbb{E}_{q(z,\psi)} \log \frac{q(z|\psi)q(\psi)}{q(\psi|z)} =$$
$$= -\mathbb{E}_{q(z,\psi)} \log q(z|\psi)q(\psi) + \mathbb{E}_{q(z)q(\psi|z)} \log q(\psi|z) \geq$$
$$\geq -\mathbb{E}_{q(z,\psi)} \log q(z|\psi)q(\psi) + \mathbb{E}_{q(z)q(\psi|z)} \log r_\theta(\psi|z)$$

- $r_\theta(\psi|z)$ is a parametric **reverse model**
- New lower bound:

$$\mathcal{L} \geq \mathcal{L}_{HVI} = \mathbb{E}_{q_\phi(z)} \log p(x|z)p(z) -$$
$$-\mathbb{E}_{q_\phi(z,\psi)} \log q_\phi(z|\psi)q_\phi(\psi) + \mathbb{E}_{q_\phi(z)q_\phi(\psi|z)} \log r_\theta(\psi|z) \to \max_{\phi,\theta}$$

Unbiased implicit variational inference



- Semi-implicit distribution:

$$q_\phi(z) = \int q_\phi(z|\psi) q_\phi(\psi) d\psi$$

- $q(z|\psi)$ is explicit
- What if $q(\psi)$ is implicit?
- Equivalent reformulation:

$$z \sim q_\phi(z) = \int q_\phi\big(z \big| \psi = f_\phi(\epsilon)\big) p(\epsilon) d\psi = \int q_\phi(z|\epsilon) p(\epsilon) d\epsilon$$

- $p(\epsilon)$ is explicit!

Unbiased implicit variational inference

- Now we can perform HVI with implicit $q_\phi(\psi)$!
- … Or go even further

- ELBO:

$$\mathcal{L} = \mathbb{E}_{q_\phi(z)} \log p(x|z)p(z) - \mathbb{E}_{q_\phi(z)} \log q_\phi(z)$$

- $\mathcal{L}$ is intractable
- But $\nabla_\phi \mathcal{L}$ can be estimated efficiently!

Unbiased implicit variational inference

Provides an unbiased gradient estimate for the ELBO

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z) = \nabla_\phi \mathbb{E}_{p(\epsilon)} \log q_\phi\left(f_\phi(\epsilon)\right) =$$

$$= \mathbb{E}_{p(\epsilon)} \nabla_\phi \log q_\phi(z)\Big|_{z=f_\phi(\epsilon)} + \mathbb{E}_{p(\epsilon)} \nabla_z \log q_\phi(z)\Big|_{z=f_\phi(\epsilon)} \cdot \nabla_\phi f_\phi(\epsilon) =$$

$$= \underbrace{\mathbb{E}_{q_\phi(z)} \nabla_\phi \log \mathrm{q}_\phi(\mathrm{z})}_{=0} + \mathbb{E}_{p(\epsilon)} \nabla_z \log q_\phi(z)\Big|_{z=f_\phi(\epsilon)} \cdot \nabla_\phi f_\phi(\epsilon)$$

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z) = \mathbb{E}_{p(\epsilon)} \color{red}{\nabla_z \log q_\phi(z)}\Big|_{z=f_\phi(\epsilon)} \cdot \nabla_\phi f_\phi(\epsilon)$$

Unbiased implicit variational inference

$$\nabla_z \log q_\phi(z) = \frac{1}{q_\phi(z)} \nabla_z q_\phi(z) = \frac{1}{q_\phi(z)} \nabla_z \int q_\phi(z|\epsilon') p(\epsilon') d\epsilon' =$$

$$= \frac{1}{q_\phi(z)} \int d\epsilon' p(\epsilon') \nabla_z q_\phi(z|\epsilon') = \frac{1}{q_\phi(z)} \int d\epsilon' p(\epsilon') q_\phi(z|\epsilon') \nabla_z \log q_\phi(z|\epsilon') =$$

$$= \int d\epsilon' q_\phi(\epsilon'|z) \nabla_z q_\phi(z|\epsilon') = \mathbb{E}_{q(\epsilon'|z)} \nabla_\phi \log q_\phi(z|\epsilon')$$

$$\nabla_z \log q_\phi(z) = \mathbb{E}_{q_\phi(\epsilon'|z)} \nabla_\phi \log q_\phi(z|\epsilon')$$

Unbiased implicit variational inference

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z) = \mathbb{E}_{p(\epsilon)} \nabla_z \log q_\phi(z) \Big|_{z=f_\phi(\epsilon)} \cdot \nabla_\phi f_\phi(\epsilon)$$

$$\nabla_z \log q_\phi(z) = \mathbb{E}_{q_\phi(\epsilon'|z)} \nabla_\phi \log q_\phi(z|\epsilon')$$

- Joint density $q_\phi(z, \epsilon') = q_\phi(z|\epsilon')q(\epsilon')$ is tractable!
  - We can now sample $\epsilon' \sim q_\phi(\epsilon'|z)$ using MCMC
  - We can sample $(z, \epsilon) \sim q_\phi(z, \epsilon)$ and use it to start HMC
  - No warm-up needed for MCMC!

Unbiased implicit variational inference

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z) = \mathbb{E}_{p(\epsilon)} \nabla_z \log q_\phi(z) \Big|_{z=f_\phi(\epsilon)} \cdot \nabla_\phi f_\phi(\epsilon)$$

$$\nabla_z \log q_\phi(z) = \mathbb{E}_{q_\phi(\epsilon'|z)} \nabla_\phi \log q_\phi(z|\epsilon')$$

How to estimate $\nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z)$:

1. Sample $\epsilon \sim p(\epsilon), z \sim q(z|\epsilon)$
2. Estimate $\nabla_z \log q_\phi(z)$:
    1. Use (z, ϵ) to start a MC for $q_\phi(\epsilon'|z)$
    2. Perform several MC steps to obtain $\epsilon' \sim q_\phi(\epsilon'|z)$, $\epsilon' \perp\!\!\!\perp \epsilon$
    3. Use $\nabla_\phi \log q_\phi(z|\epsilon') \simeq \nabla_z \log q_\phi(z)$
3. Use $\nabla_\phi \log q_\phi(z|\epsilon') \cdot \nabla_\phi f_\phi(\epsilon) \simeq \nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z)$

Overview

- Density ratio estimation
  - Prior-contrastive
  - Joint-contrastive
  - Any distribution can be made implicit
  - Variational learning is difficult (need to differentiate through SGD)
    - Has not been done in this setting?
  - Stability of DRE is a concern
- IVI using reverse models
  - Relatively simple optimization problem
  - Less broad applicability
    - Both HVI and UIVI should extend to implicit priors
    - Variational learning is not possible

Denoiser-guided learning

Consider an arbitrary distribution $\mathrm{p}(z)$

A curious way to approximate $\nabla_z \log p(z)$:

$$\mathrm{DAE}(z) = \arg\min_{D(z)} \mathbb{E}_{z\sim q(z)} \mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)} \|D(z+\epsilon) - z\|_2^2 =$$
$$= z + \sigma^2 \nabla_z \log p(z) + o(\sigma^2)$$

1. Train a denoising autoencoder $D(z)$ with noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$
2. Approximate $\nabla_z \log p(z) \approx \frac{D(z)-z}{\sigma^2}$

Denoiser-guided learning: optimal denoiser

$$L_{DAE} = \int \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}\left[q(z)\|D(z+\epsilon) - z\|_2^2\right] dz = [\tilde{z} = z + \epsilon] =$$

$$= \int \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}\left[p(\tilde{z} - \epsilon)\|D(\tilde{z}) - \tilde{z} + \epsilon\|_2^2\right] dz$$

$$0 = \nabla_D \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}\left[p(\tilde{z} - \epsilon)\|D(\tilde{z}) - \tilde{z} + \epsilon\|_2^2\right] =$$

$$= \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}\left[p(\tilde{z} - \epsilon)\nabla_D\|D(\tilde{z}) - \tilde{z} + \epsilon\|_2^2\right] =$$

$$= \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}[p(\tilde{z} - \epsilon)2(D(\tilde{z}) - \tilde{z} + \epsilon)]$$

$$\mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}[p(\tilde{z} - \epsilon)D(\tilde{z})] = \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}[p(\tilde{z} - \epsilon)(\tilde{z} - \epsilon)]$$

$$D(\tilde{z}) = \frac{\mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}[p(\tilde{z} - \epsilon)(\tilde{z} - \epsilon)]}{\mathbb{E}_{\epsilon \sim \mathcal{N}(0,\sigma^2 I)}[p(\tilde{z} - \epsilon)]}$$

Denoiser-guided learning: intuition

$$D(\tilde{z}) = \frac{\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}-\epsilon)(\tilde{z}-\epsilon)]}{\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}-\epsilon)]} = \tilde{z} - \frac{\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}-\epsilon)\epsilon]}{\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}-\epsilon)]}$$

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}-\epsilon)] \approx \mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}) - \epsilon p'(\tilde{z})] = p(\tilde{z})$$

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}[p(\tilde{z}-\epsilon)\epsilon] \approx \mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2 I)}\left[p(\tilde{z})\epsilon - \epsilon^2 p'(\tilde{z})\right] = -\sigma^2 p'(\tilde{z})$$

$$D(\tilde{z}) \approx \tilde{z} + \sigma^2 \frac{p'(\tilde{z})}{p(\tilde{z})} = \tilde{z} + \sigma^2 \nabla_{\tilde{z}} \log p(\tilde{z})$$

$$\nabla_z \log p(z) \approx \frac{D(z) - z}{\sigma^2}$$

Denoiser-based VI:

- Recall UIVI:

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} \log q_\phi(z) = \mathbb{E}_{p(\epsilon)} \nabla_z \log q_\phi(z) \Big|_{z=f_\phi(\epsilon)} \cdot \nabla_\phi f_\phi(\epsilon)$$

- We can now approximate $\nabla_z \log q_\phi(z)$ using a denoiser!
- Train a separate denoiser to estimate $\nabla_z \log p(z)$ for an implicit prior!
  - Similar to prior-contrastive DRE
- Train a joint denoiser over (x, z) to estimate $\nabla_z \log p(x, z)$!
  - Similar to joint-contrastive DRE

Literature

- Huszár, Ferenc. "Variational inference using implicit distributions." *arXiv preprint arXiv:1702.08235* (2017). https://arxiv.org/abs/1702.08235
- Ranganath, Rajesh, Dustin Tran, and David Blei. "Hierarchical variational models." *International Conference on Machine Learning*. 2016. https://arxiv.org/abs/1511.02386
- Michalis K. Titsias, Francisco J. R. Ruiz. “Unbiased Implicit Variational Inference.” *arXiv preprint arXiv:1808.02078 (2018). https://arxiv.org/abs/1808.02078*
- Salimans, Tim, Diederik Kingma, and Max Welling. "Markov chain monte carlo and variational inference: Bridging the gap." *International Conference on Machine Learning*. 2015. https://arxiv.org/abs/1410.6460
- Louizos, Christos, and Max Welling. "Multiplicative normalizing flows for variational bayesian neural networks." *arXiv preprint arXiv:1703.01961* (2017). https://arxiv.org/abs/1703.01961