Structured Semi-Implicit Variational Inference

Iuliia Molchanova[1] Dmitry Molchanov[1,3]
Novi Quadrianto[2] Dmitry Vetrov[1,3]

[1]Samsung AI Center Moscow
[2]University of Sussex
[3]Samsung-HSE Joint Lab

December 08, 2019

p(B|A)yesgroup.ru

SAMSUNG AI Center - Moscow

NATIONAL RESEARCH UNIVERSITY

UNIVERSITY OF SUSSEX

$$q_\phi(z) \approx p(z \mid D) \quad \longleftrightarrow \quad KL(q_\phi(z) \,||\, p(z \mid D)) \to \min_\phi$$

Explicit: $q_\phi(z)$ is given

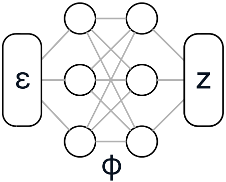
Implicit: $z = g_\phi(\epsilon), \quad \epsilon \sim p(\epsilon)$

$$q_\phi(z) = ?$$

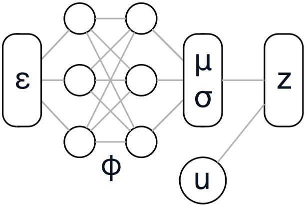
Semi-implicit: $q_\phi(z) = \int q_\phi(z \mid \epsilon) p(\epsilon) d\epsilon$

e.g. $z = \mu_\phi(\epsilon) + u \cdot \sigma_\phi(\epsilon)$

Implicit:



Semi-implicit:



KL-based Variational Inference: overview

Explicit	**Semi-implicit**	**Implicit**
$q_\phi(z)$	$q_\phi(z) = \int q_\phi(z \mid \epsilon) p(\epsilon) d\epsilon$	$z = g_\phi(\epsilon), \quad \epsilon \sim p(\epsilon)$ $q_\phi(z) = ?$
FFG Normalizing flows Neural ODE Autoregressive	(D)SIVI UIVI HVI IWHVI	AVB SSGE KIVI Denoising AE
+ Proper ELBO	+ Proper ELBO (except UIVI)	- Biased ELBO
- Too simple or inefficient architectures	+ Almost no restrictions on $q_\phi(z)$	+ No restrictions on $q_\phi(z)$

Semi-Implicit Variational Inference

$$q_\phi(z) = \int q_\phi(z \mid \epsilon) p(\epsilon) d\epsilon$$

Variational inference:

$$\text{ELBO} = \underbrace{\mathbb{E}_{q_\phi(z)} \log p(D \mid z) + \overbrace{\mathbb{E}_{q_\phi(z)} \log p(z)}_{}}_{\text{No problem: reparameterization}} \overbrace{- \underbrace{\mathbb{E}_{q_\phi(z)} \log q_\phi(z)}_{\text{No access to } q_\phi(z)\ldots}}^{-KL\left(q_\phi(z) \,||\, p(z)\right)} \to \max_\phi$$

$$q_\phi(z) = \int q_\phi(z \mid \epsilon) p(\epsilon) d\epsilon$$

Variational inference:

$$\text{ELBO} = \underbrace{\mathbb{E}_{q_\phi(z)} \log p(D \mid z) + \overbrace{\mathbb{E}_{q_\phi(z)} \log p(z)}^{-KL\left(q_\phi(z) \,||\, p(z)\right)}}_{\text{No problem: reparameterization}} \overbrace{- \underbrace{\mathbb{E}_{q_\phi(z)} \log q_\phi(z)}_{\text{No access to } q_\phi(z)\ldots}} \to \max_\phi$$

Main idea:

$$q_\phi(z) = \mathbb{E}_{p(\epsilon)} q_\phi(z \mid \epsilon) \approx \frac{1}{K+1} \sum_{k=0}^{K} q_\phi(z \mid \epsilon^k)$$

Yin and Zhou. *SIVI* (ICML, 2018)

Semi-Implicit Variational Inference

$$q_\phi(z) = \int q_\phi(z \mid \epsilon) p(\epsilon) d\epsilon$$

Variational inference:

$$\text{ELBO} = \underbrace{\mathbb{E}_{q_\phi(z)} \log p(D \mid z) + \overbrace{\mathbb{E}_{q_\phi(z)} \log p(z)}^{}}_{\text{No problem: reparameterization}} \overbrace{- \underbrace{\mathbb{E}_{q_\phi(z)} \log q_\phi(z)}_{\text{No access to } q_\phi(z)\ldots}}^{-KL\left(q_\phi(z) \,||\, p(z)\right)} \to \max_\phi$$

Main idea:

$$q_\phi(z) = \mathbb{E}_{p(\epsilon)} q_\phi(z \mid \epsilon) \approx \frac{1}{K+1} \sum_{k=0}^{K} q_\phi(z \mid \epsilon^k)$$

Yin and Zhou. *SIVI* (ICML, 2018)

Results in a lower bound!

Molchanov et al. *DSIVI* (AISTATS, 2019)

$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) \geq -\mathbb{E}_{\epsilon^{0..K}} \mathbb{E}_{z \mid \epsilon^0} \log \frac{1}{K+1} \sum_{k=0}^{K} q_\phi(z \mid \epsilon^k)$$

Semi-Implicit Variational Inference

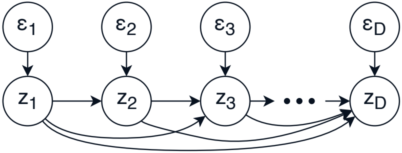
+ Simple procedure; only one hyperparameter K

+ A broader family of $q_\phi(z)$:
  Unlike HVI, UIVI and IWHVI, allows for implicit (and even discrete) $p(\epsilon)$!

- MoG approximation does not work well in high dimensions...

$$q_\phi(z) = \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}) =$$

$$= \prod_{i=1}^{D} \mathbb{E}_{p(\epsilon_i)} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i)$$



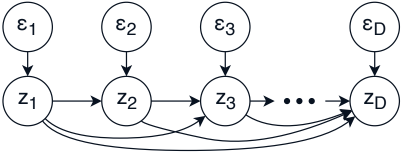
Straight-forward (joint) SIVI entropy bound:

$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) \geq -\mathbb{E}_{\epsilon^{0..K}} \mathbb{E}_{z \mid \epsilon^0} \log \frac{1}{K+1} \sum_{k=0}^{K} q_\phi(z \mid \epsilon^k)$$

Structured Semi-Implicit Variational Inference

$$q_\phi(z) = \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}) =$$

$$= \prod_{i=1}^{D} \mathbb{E}_{p(\epsilon_i)} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i)$$



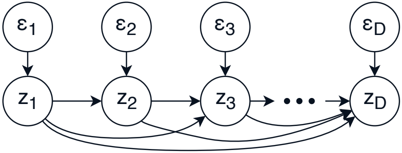
Straight-forward (joint) SIVI entropy bound:

$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) \geq -\mathbb{E}_{\epsilon^{0..K}} \mathbb{E}_{z \mid \epsilon^0} \log \frac{1}{K+1} \sum_{k=0}^{K} \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

Structured Semi-Implicit Variational Inference

$$q_\phi(z) = \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}) =$$

$$= \prod_{i=1}^{D} \mathbb{E}_{p(\epsilon_i)} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i)$$



Straight-forward (joint) SIVI entropy bound:

$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) \geq -\mathbb{E}_{\epsilon^{0..K}} \mathbb{E}_{z \mid \epsilon^0} \log \frac{1}{K+1} \sum_{k=0}^{K} \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

Structured SIVI entropy bound:

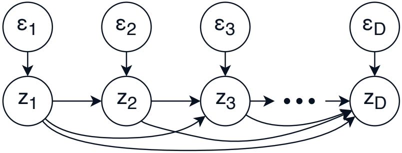
$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) = -\sum_{i=1}^{D} \mathbb{E}_{z_{1..i}} \log q_\phi(z_i \mid z_{1..i-1}) \geq$$

$$\geq -\sum_{i=1}^{D} \mathbb{E}_{z_{1..i-1}} \mathbb{E}_{\epsilon_i^{0..K}} \mathbb{E}_{z_i \mid z_{1..i-1}, \epsilon_i^0} \log \frac{1}{K+1} \sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

Structured Semi-Implicit Variational Inference

$$q_\phi(z) = \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}) =$$

$$= \prod_{i=1}^{D} \mathbb{E}_{p(\epsilon_i)} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i)$$



Straight-forward (joint) SIVI entropy bound:

$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) \geq -\mathbb{E}_{\epsilon^{0..K}} \mathbb{E}_{z \mid \epsilon^0} \log \frac{1}{K+1} \sum_{k=0}^{K} \prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

Structured SIVI entropy bound:

$$-\mathbb{E}_{q_\phi(z)} \log q_\phi(z) = -\sum_{i=1}^{D} \mathbb{E}_{z_{1..i}} \log q_\phi(z_i \mid z_{1..i-1}) \geq$$

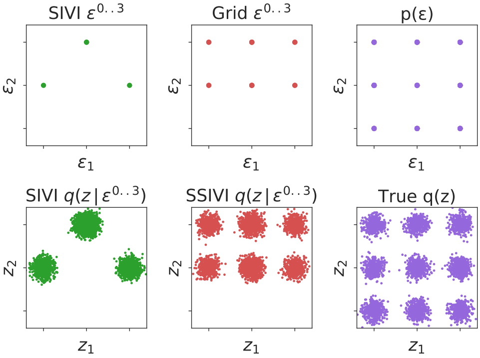
$$\geq -\mathbb{E}_{\epsilon^{0..K}} \mathbb{E}_{z \mid \epsilon^0} \log \frac{1}{(K+1)^D} \prod_{i=1}^{D} \sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

Structured Semi-Implicit Variational Inference

SIVI approximation:

$$q_\phi(z) \approx \frac{1}{K+1}\sum_{k=0}^{K} q_\phi(z \mid \epsilon^k) = \frac{1}{K+1}\sum_{k=0}^{K}\prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

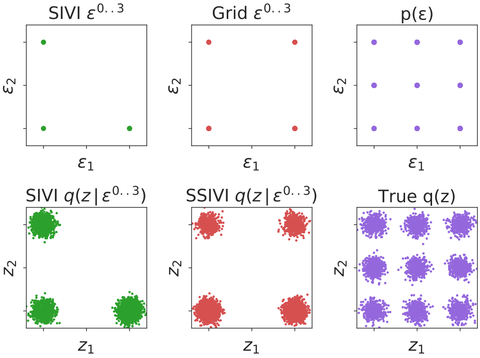
SSIVI approximation: $$q_\phi(z) \approx \frac{1}{(K+1)^D}\prod_{i=1}^{D}\sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$



SIVI approximation:

$$q_\phi(z) \approx \frac{1}{K+1}\sum_{k=0}^{K} q_\phi(z \mid \epsilon^k) = \frac{1}{K+1}\sum_{k=0}^{K}\prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

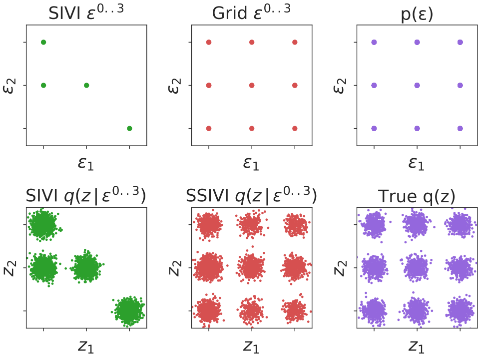
SSIVI approximation: $$q_\phi(z) \approx \frac{1}{(K+1)^D}\prod_{i=1}^{D}\sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$



SIVI approximation:

$$q_\phi(z) \approx \frac{1}{K+1}\sum_{k=0}^{K} q_\phi(z \mid \epsilon^k) = \frac{1}{K+1}\sum_{k=0}^{K}\prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

SSIVI approximation: $$q_\phi(z) \approx \frac{1}{(K+1)^D}\prod_{i=1}^{D}\sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$



Structured Semi-Implicit Variational Inference

SIVI approximation:

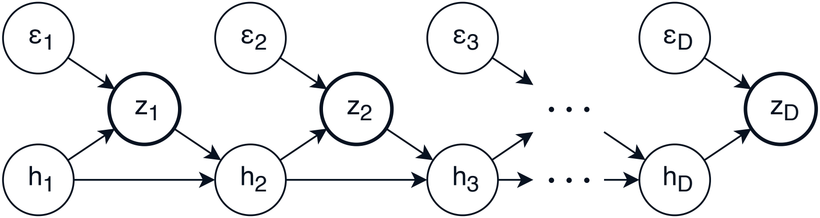
$$q_\phi(z) \approx \frac{1}{K+1}\sum_{k=0}^{K} q_\phi(z \mid \epsilon^k) = \frac{1}{K+1}\sum_{k=0}^{K}\prod_{i=1}^{D} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

SSIVI approximation: $$q_\phi(z) \approx \frac{1}{(K+1)^D}\prod_{i=1}^{D}\sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)$$

- Same computational complexity! (given the structure of $q_\phi(z)$)
- ELBO gap is **decreased** by

$$\mathbb{E}_{\epsilon^{0..K}} KL\left(\frac{1}{K+1}\sum_{k=0}^{K} q_\phi(z \mid \epsilon^k) \,\middle\|\, \frac{1}{(K+1)^D}\prod_{i=1}^{D}\sum_{k=0}^{K} q_\phi(z_i \mid z_{1..i-1}, \epsilon_i^k)\right)$$

Autoregressive Semi-Implicit Generator



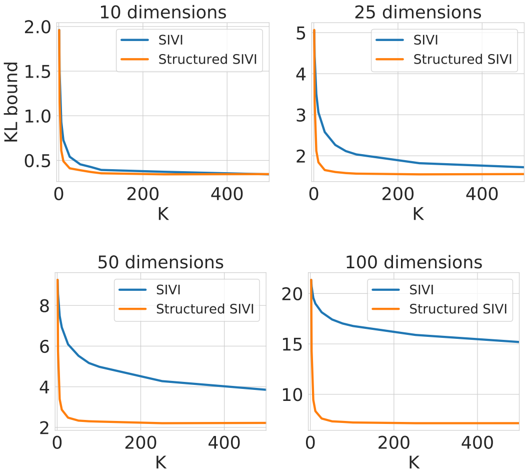
$$h_1 = h(0, 0), \qquad \epsilon_1 \sim \mathcal{N}(\epsilon_1 \mid 0, 1), \qquad z_1 \sim \mathcal{N}(z_1 \mid \mu(h_1, \epsilon_1), \sigma^2(h_1, \epsilon_1)),$$
$$h_i = h(z_{i-1}, h_{i-1}), \qquad \epsilon_i \sim \mathcal{N}(\epsilon_i \mid 0, 1), \qquad z_i \sim \mathcal{N}(z_i \mid \mu(h_i, \epsilon_i), \sigma^2(h_i, \epsilon_i)).$$

Defines a general structure:

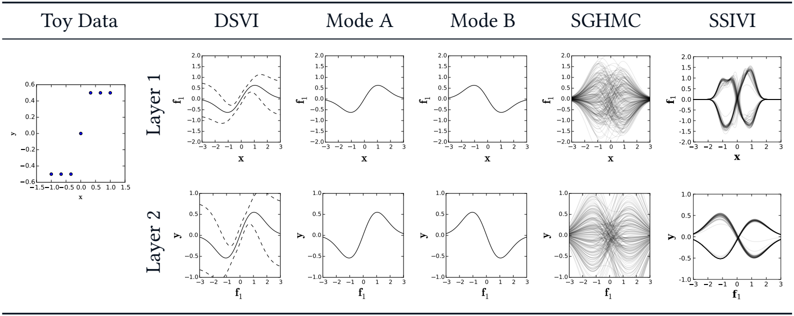
$$q_\phi(z) = q_\phi(z_1) q_\phi(z_2 \mid z_1) \cdot \ldots \cdot q_\phi(z_D \mid z_{1..D-1})$$

Implemented as a recurrent neural network

$$p(z) = \mathrm{Laplace}(z_1 \mid 0, 1) \prod_{i=2}^{d} \mathrm{Laplace}(z_i \mid z_{i-1}, 1); \qquad KL\left(q_\phi(z) \,||\, p(z)\right) \to \min_\phi$$



Experiments: Deep Gaussian Processes



All columns but last are taken from Havasi et al. *SGHMC DGP* (NeurIPS, 2018)

Structured SIVI posterior approximation:

$$q_\phi(u^{1..L}) = q_\phi(u^1) \prod_{l=2}^{L} q_\phi(u^l \mid u^{l-1})$$

$$q_\phi(u^l \mid u^{l-1}) = \int q_\phi(u^l \mid \epsilon^l, u^{l-1}) q_\phi(\epsilon^l) d\epsilon^l$$

Discussion

- Utilizing structure results in a tighter evidence bound!
- Would utilizing structure help in UIVI, IWHVI, AVB and others?
- How would auto regressive semi-implicit models compare to straightforward semi-implicit models?