Variational Inference (VI)

VI approximates the posterior $p(z|\mathcal{D}) \approx q(z|\lambda_z)$ by maximizing the evidence lower bound:

$$\text{ELBO: } \max_{\lambda_z} \mathcal{L}(\lambda_z) := \mathbb{E}_q \left[ \log p(D, z) - \log q(z|\lambda_z) \right]$$

where $q(z)$ is a tractable distribution parametrized by $\lambda_z$. 

Probabilistic Model

$$\log p(D, z) - \log q(z|\lambda_z)$$
ELBO Optimization

Block-box VI (BBVI):

\[ \lambda_z \leftarrow \lambda_z + \beta \nabla_{\lambda_z} \mathcal{L}(\lambda_z) \]

Natural-gradient VI (NGVI):

\[ \lambda_z \leftarrow \lambda_z + \beta \ F_z(\lambda_z)^{-1} \nabla_{\lambda_z} \mathcal{L}(\lambda_z) \]

where \( F_z(\lambda_z) \) is the Fisher information matrix of \( q(z|\lambda_z) \).

Advantages of NGVI:

- NGVI can be simple and fast when \( q \) is in the exponential family (e.g., Gaussian) (Khan and Lin, AI&Stats 2017).

NGVI for Exp-Family: \( \lambda_z \leftarrow \lambda_z + \beta \nabla_{m_z} \mathcal{L}(\lambda_z) \)

because \( \nabla_{m_z} \mathcal{L}(\lambda_z) = F_z(\lambda_z)^{-1} \nabla_{\lambda_z} \mathcal{L}(\lambda_z) \).
Problem Formulation

Challenges of NGVI when \( q(z) \) is not in the exponential-family:

- Computing \( F_z(\lambda_z)^{-1} \nabla_{\lambda_z} \mathcal{L}(\lambda_z) \) could be complicated.
- \( F_z(\lambda_z) \) can be singular.
- Often no simple update beyond exponential family.

Our goal: perform a simple NGVI update for more flexible variational approximations (e.g., skewness, multi-modality)

(a) Skew Gaussian

(b) Finite Mixture of Gaussians
This Work

Main Contribution: propose a new NGVI update for a class of mixture of exponential family distributions.

We consider the following mixture:

\[ q(z|\lambda) = \int q(z|w, \lambda_z) q(w|\lambda_w) \, dw \]

We propose to use the (joint) Fisher matrix \( F_{wz} \) of \( q(w, z|\lambda) \) since:

\[ \nabla_m \mathcal{L}(\lambda) = F_{wz}(\lambda)^{-1} \nabla_\lambda \mathcal{L}(\lambda) \]

where \( m \) is the proposed expectation parameter.

- Proposed NGVI update: \( \lambda \leftarrow \lambda + \beta \nabla_m \mathcal{L}(\lambda) \)
Proposed NGVI

Advantage of the proposed NGVI:
▶ Has the same cost as BBVI if computing $\nabla_m \mathcal{L}(\lambda)$ is easy.
▶ Is faster than BBVI.

Variational approximations:
▶ Finite mixture of exp-family distributions:
  Mixture of Gaussians (multi-modality)
  Birnbaum-Saunders distribution (non-Gaussian mixture)
▶ Gaussian compound distribution:
  Skew Gaussian (skewness)
  Normal inverse-Gaussian (heavy tails)
Summary & Poster Presentation

Conclusion:
a simple NGVI update for approximations outside the exp-family.

Poster Presentation:

▶ This work:
Poster #217, Pacific Ballroom, Today, 6:30 PM

▶ New gradient estimators via Stein’s lemma:
“Stein’s Lemma for the Reparameterization Trick with Exponential-family Mixtures”, the workshop on Stein’s method, Saturday