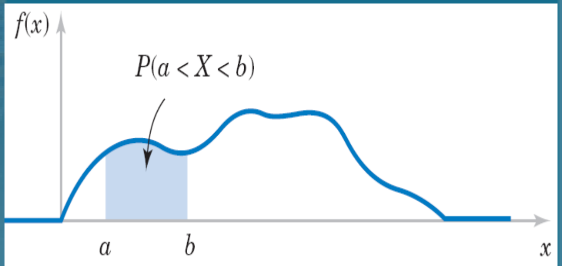
SAND2023-12817C

Sandia National Laboratories

A Study of Bias-Variance in Variational Inferencing Using the Delta Method



Niladri Das & Tommie A. Catanach (P.I.)

March 3, 2023

2023 SIAM CSE, Amsterdam, The Netherlands

U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND****

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

We are going to see a technique that **reduces the variance** in the **estimates** used in **Variational Inferencing** (VI), leading to empirically **faster convergence**.

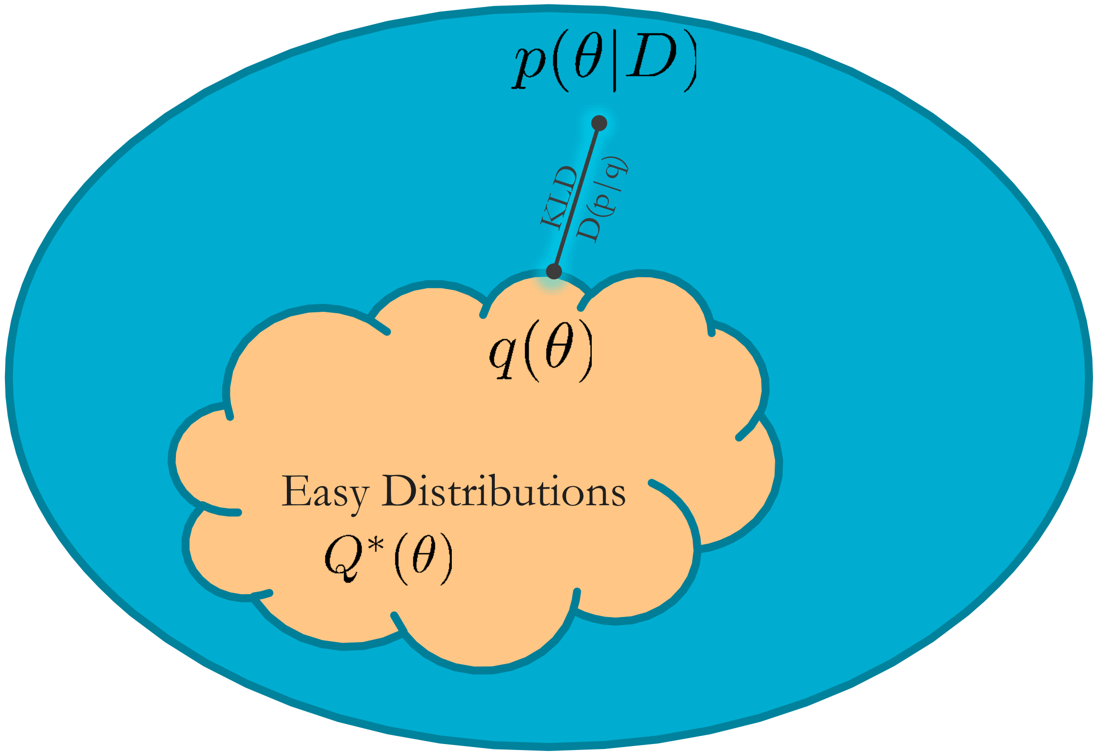
A Study of Bias-Variance in **Variational Inferencing** Using Delta Method

Bayes Rule

Posterior, Prior, Likelihood

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{\int_\theta p(\theta)p(D|\theta)d\theta}$$

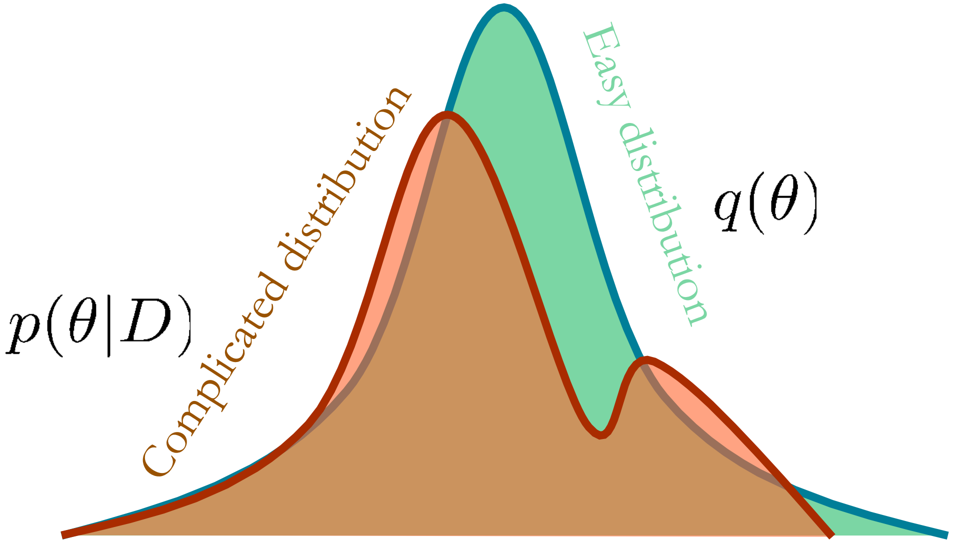
$\mathbb{R}^{100 \ldots}$

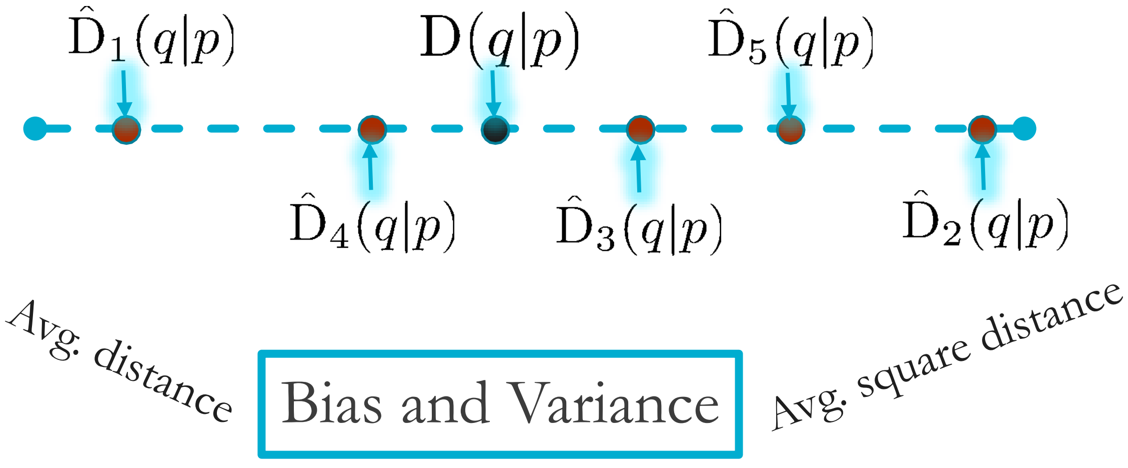


At even moderately high dimensions of the θ number of numerical operations explode.

$q(\theta)$: Variational Distribution

A Study of **Bias-Variance** in Variational Inferencing Using **Delta Method**





D(p|q) = KLD

↓

Sample Estimate

A Study of Bias-Variance in Variational Inferencing Using Delta Method

$$D(q|p) = \int q(\theta \mid \phi) \log \frac{q(\theta \mid \phi)}{p(\theta)} d\theta - \int q(\theta \mid \phi) \log p(\mathcal{D} \mid \theta) d\theta + \log \int p(\mathcal{D} \mid \theta)\, p(\theta)\, d\theta$$

Minimize

Log Evidence , independent of VI params. ϕ

$$\text{ELBO} = -\int q(\theta \mid \phi) \log \frac{q(\theta \mid \phi)}{p(\theta)} d\theta + \int q(\theta \mid \phi) \log p(\mathcal{D} \mid \theta) d\theta$$

Maximize

$$= -\frac{1}{N}\Sigma\Big(\dots\Big) \qquad + \frac{1}{N}\Sigma\Big(\dots\Big)$$

A Study of **Bias-Variance** in Variational Inferencing Using **Delta Method**

$$D(q|p) = \int q(\theta \mid \phi) \log \frac{q(\theta \mid \phi)}{p(\theta)} d\theta - \int q(\theta \mid \phi) \log p(\mathcal{D} \mid \theta) d\theta + \log \int p(\mathcal{D} \mid \theta)\, p(\theta)\, d\theta$$

Minimize

Accurate Calculation

$$= \frac{1}{N}\Sigma(\dots) \quad \mathcal{X}$$

$$= \frac{1}{N}\Sigma(\dots) \quad \log(\mathcal{Y})$$

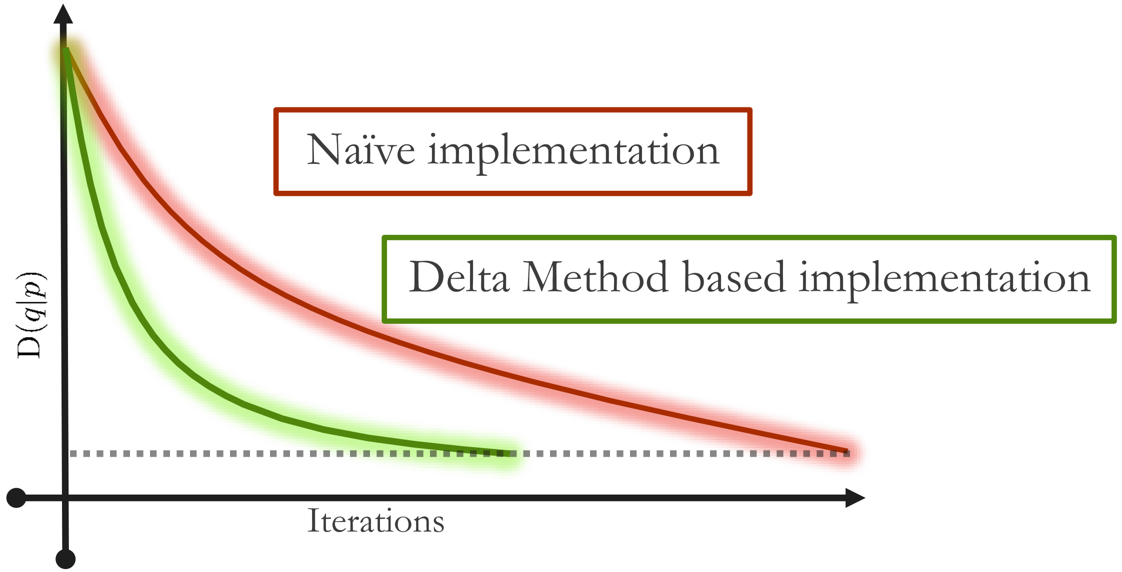
$$\mathcal{V}ar(\hat{\mathcal{X}} + \log(\hat{\mathcal{Y}})) = \frac{\Sigma_{\mathcal{X}}}{N} + \frac{\Sigma_{\mathcal{Y}}}{N\mathcal{Y}^2} + 2\frac{\Sigma_{\mathcal{X}\mathcal{Y}}}{N\mathcal{Y}}$$

A Study of **Bias-Variance** in Variational Inferencing Using **Delta Method**

Delta Method : Approx. probability distribution of $\mathrm{D}(q|p)$

Technique : Reduce variance in approx. of $\mathrm{D}(q|p)$

Performance : Faster convergence to $q(\theta)$



A Study of **Bias-Variance** in Variational Inferencing Using **Delta Method**

$$D(q|p) = \int q(\theta \mid \phi) \log \frac{q(\theta \mid \phi)}{p(\theta)} d\theta - \int q(\theta \mid \phi) \log p(\mathcal{D} \mid \theta) d\theta + \log \int \frac{p(\mathcal{D} \mid \theta)\, p(\theta)}{q(\theta \mid \phi)} q(\theta \mid \phi)\, d\theta$$

$$= f(\phi) - \int \boxed{\frac{q(\theta(\zeta) \mid \phi)}{r(\theta(\zeta) \mid \varphi)}} \log p(\mathcal{D} \mid \theta(\zeta)) p(\zeta) d\zeta + \log \int \boxed{\frac{p(\mathcal{D} \mid \theta(\zeta))\, p(\theta(\zeta))}{r(\theta(\zeta) \mid \varphi)}} p(\zeta) d\zeta$$

Importance Distribution

$r(\theta(\zeta) \mid \varphi)$

Re-parameterization params.

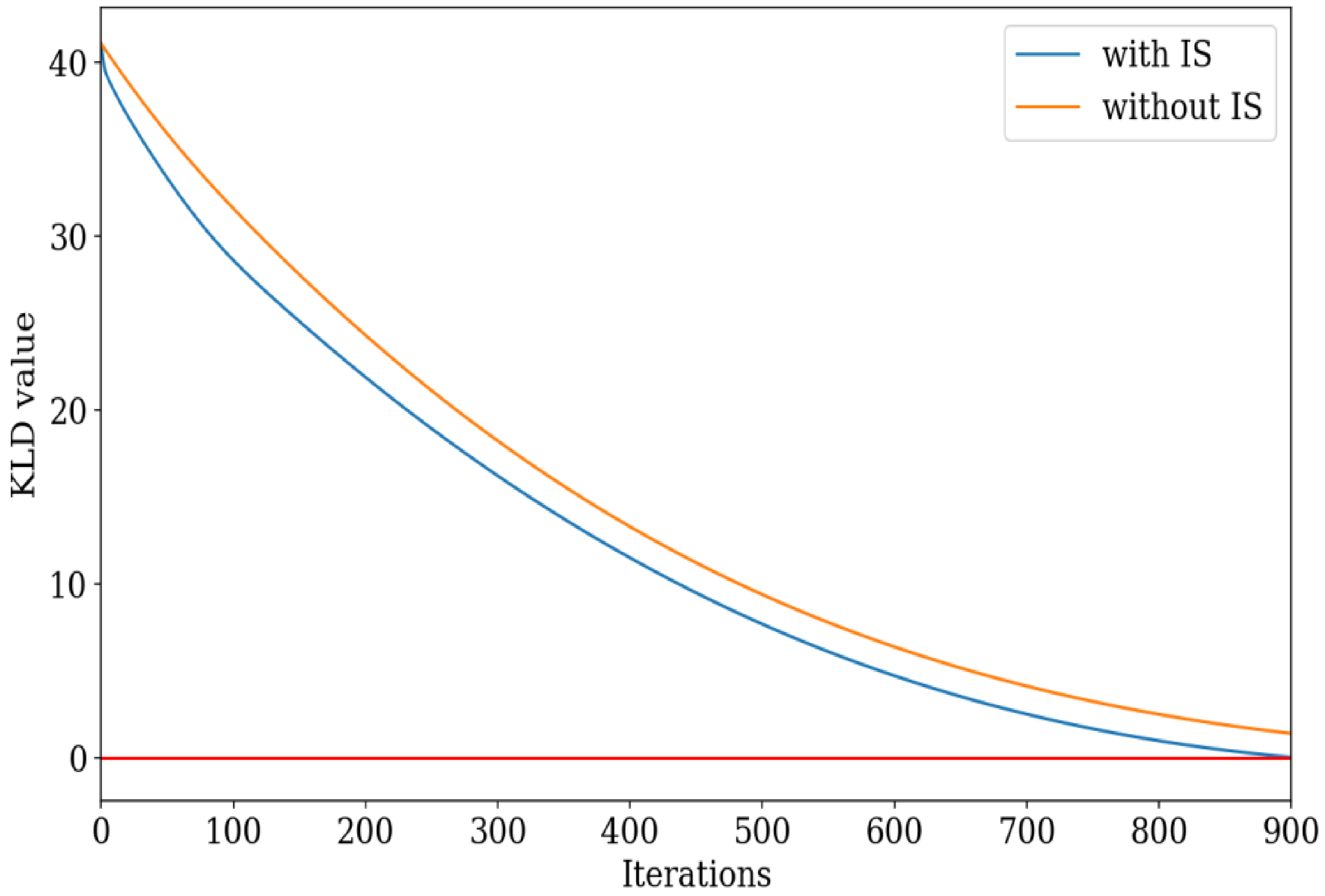
ζ

Two Step Optimization:

- **VI params. Optimization** ϕ
- **IS params. Optimization** φ

A Study of Bias-Variance in Variational Inferencing Using Delta Method

- Dimension of the problem: **5**
- Linear model : $\mathcal{Y} = X + v$
- v : Zero mean Gaussian noise
- Gaussian Importance distribution initialized with initial variational distribution.
- The VI parameters (μ, L) where $LL^T = \Sigma$
- Sample size **1000** for both VI and IS.



A Study of Bias-Variance in Variational Inferencing Using Delta Method

Key Takeaways:

- Reduce the variance of the estimate of the KLD by introducing important sampling.
- Solve two optimization problems.
- Take larger optimization steps for fast convergence.

Contact Information

Niladri Das | Postdoctoral Appointee | ndas@sandia.gov

Tommie A. Catanach | P.I. | tacatan@sandia.gov

08739 | Extreme-scale DS & Analytics

Thank You