Sandia National Laboratories

Analysis of Bias-variance Trade-off in Estimators for Variational Inferencing

Niladri Das & Tommie A. Catanach

ISBA 2022 Montreal Canada

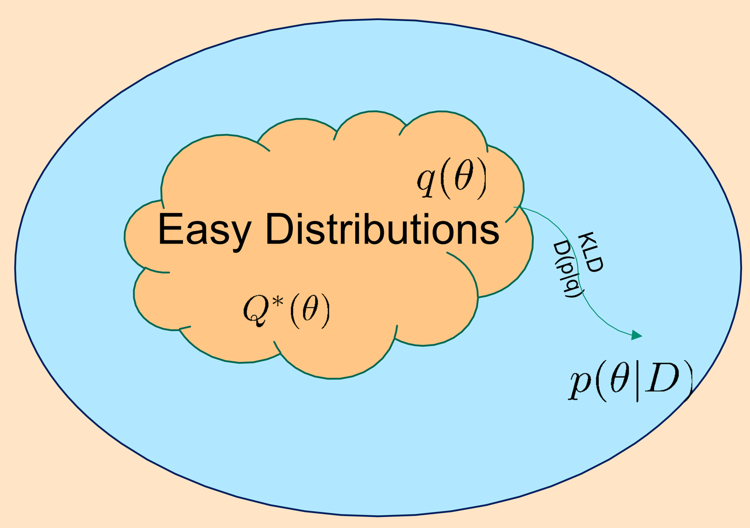
Bayes Rule

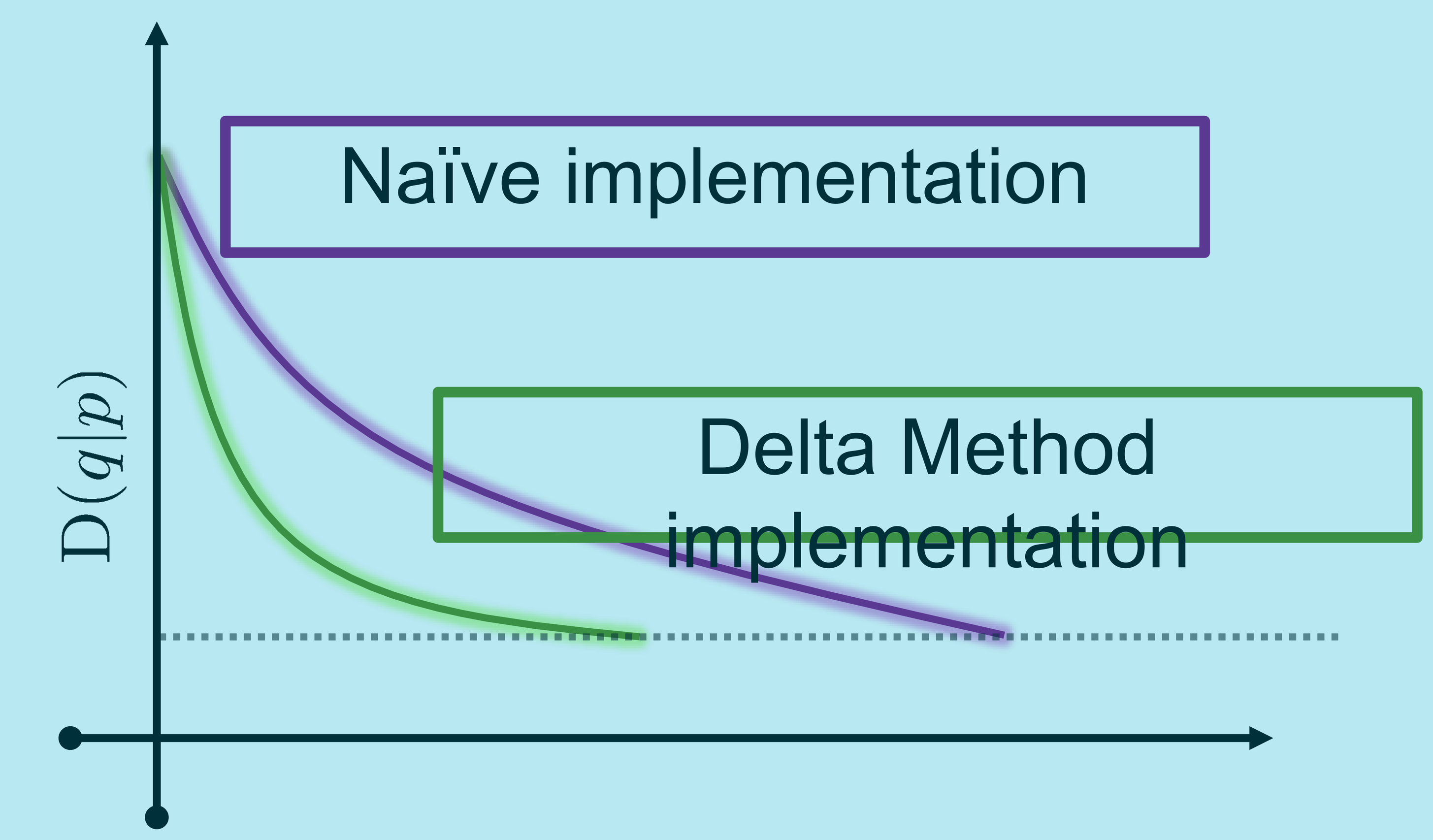
$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{\int_\theta p(\theta)p(D|\theta)d\theta}$$

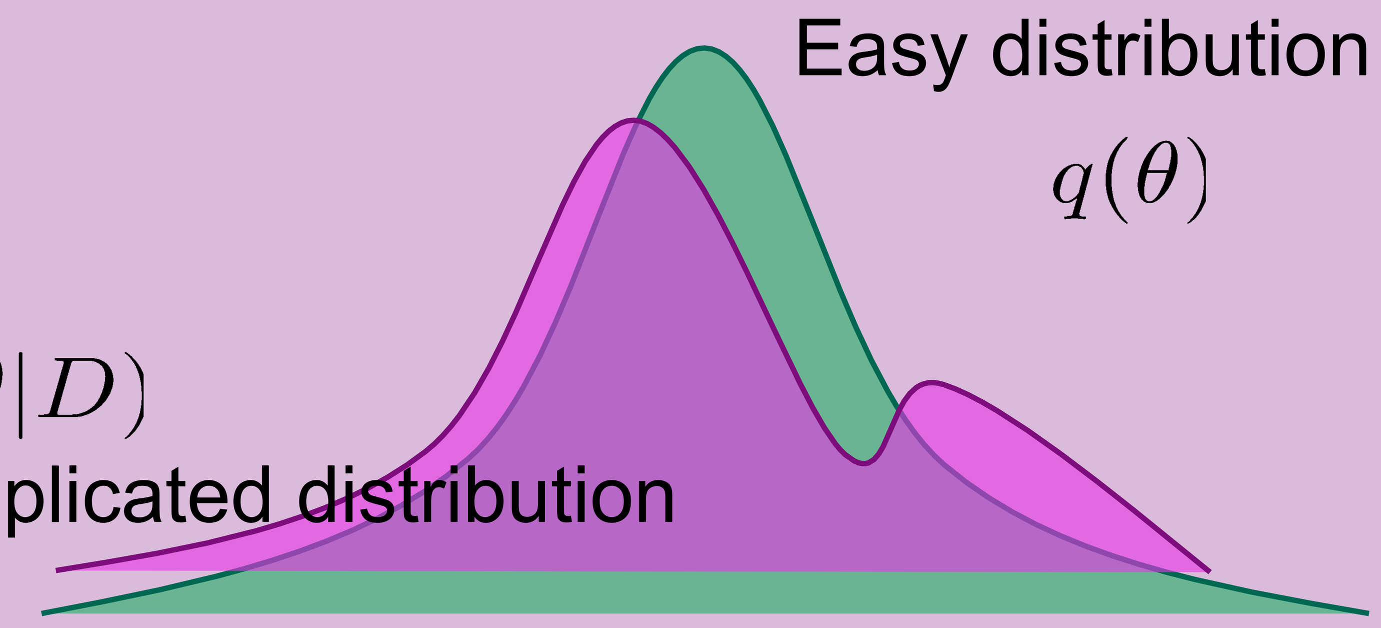
At even moderately high dimensions of the amount numerical operations explode.

Alternative

$q(\theta)$: Variational Distribution





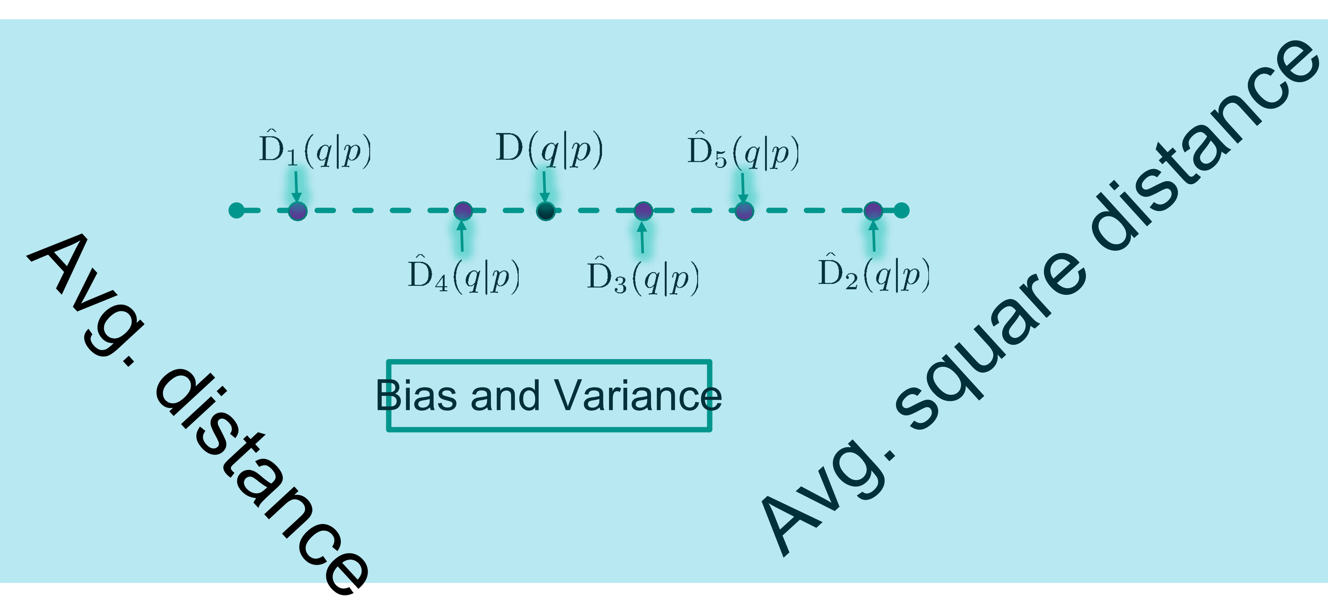


Further improvement: **Importance Distribution**

$$J = \int q(\theta \mid \phi) \log \frac{q(\theta \mid \phi)}{p(\theta)} d\theta - \int q(\theta \mid \phi) \log p(\mathcal{D} \mid \theta) d\theta + \log \int \frac{p(\mathcal{D} \mid \theta)\, p(\theta)}{q(\theta \mid \phi)} q(\theta \mid \phi)\, d\theta \tag{1}$$

$$= f(\phi) - \int \boxed{\frac{q(\theta(\zeta) \mid \phi)}{r(\theta(\zeta) \mid \varphi)}} \log p(\mathcal{D} \mid \theta(\zeta)) p(\zeta) d\zeta + \log \int \boxed{\frac{p(\mathcal{D} \mid \theta(\zeta))\, p(\theta(\zeta))}{r(\theta(\zeta) \mid \varphi)}} p(\zeta) d\zeta \tag{2}$$

where $r(\theta \mid \varphi)$ is an importance distribution.



Two step optimization:

1. **Variational distribution parameter.**
2. **Importance distribution parameter**

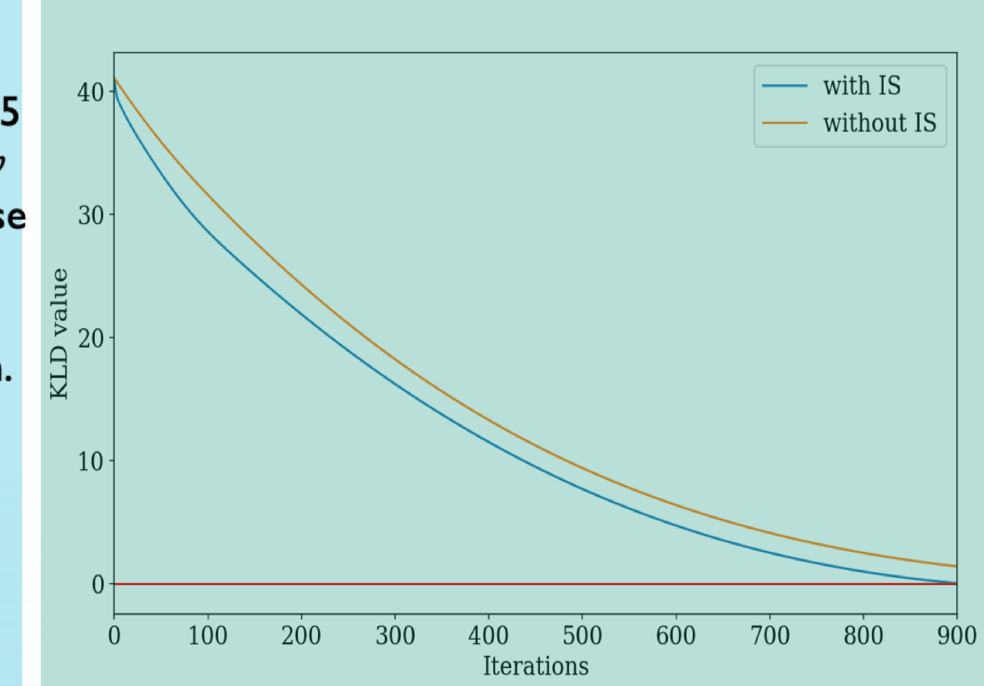
Delta Method : Approx. probability distribution of $D(q|p)$

Technique : Reduce variance in approx. of $D(q|p)$

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

Numerical Ex. :

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



Niladri Das | Postdoctoral Appointee | ndas@sandia.gov
Tommie A. Catanach | P.I. | tacatan@sandia.gov
08739 Extreme-scale DS & Analytics