# Variational Bayesian Inference for Multivariate Normal Distribution

Karla Henricksen, Nasim Parsinejad, Sohel Rana, University of Nevada, Reno


## Abstract

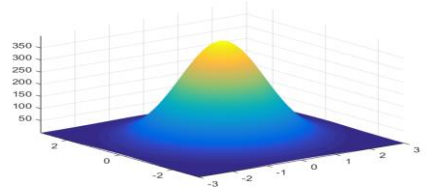
Often times, when performing Bayesian inference, the posterior distribution is difficult or impossible to represent in a closed form. In these cases, one option is to use variational inference to approximate the posterior. In this study we are interested in the performance of the mean compared to the true posterior mean for large dimensions in the multivariate normal distribution. For this reason we utilize the mean-field family for approximation, which does not approximate the covariance well, but is concerned with the mean. We found that as the dimension increases, so does the bias between the estimated and the true mean.


## Background

- Multivariate Normal Distribution
  - Probability density function:

$$f(x) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp(-\tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)),$$

where d is the dimension, μ is the d-dimensional mean vector, and Σ is the d by d covariance matrix.



- Variational Inference [1]
  - Start with the usual Bayesian set up: $p(\theta|y) = \frac{p(\theta,y)}{p(y)}$
  - Posit a family of approximate distributions, Q
  - Minimize Kullback-Leibler Divergence (measure of "distance" between two distributions) to the exact posterior
    - Equivalent to maximizing Evidence Lower Bound (ELBO): $ELBO(q) = E[\log p(y|\theta)] - KL(q(\theta)||p(y))$

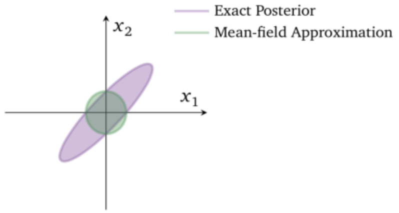
## Background

- Automatic Differential Variation in Stan [2]:
  - Automatic inference algorithm for differentiable probability models

**Transformation of Parameters**
Transform latent variables such that their support is $\mathbf{R}^k$

**Gaussian Approximation**
Posit a mean-field (latent variables are independent) Gaussian family

**Maximize ELBO**
Use stochastic optimization and automatic differentiation



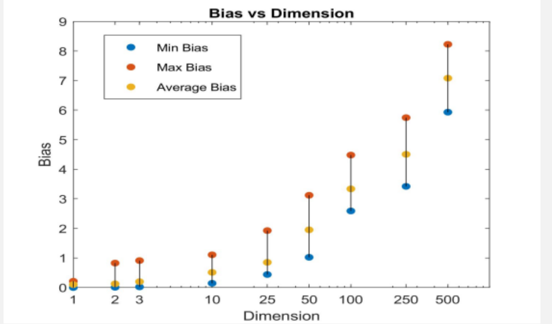

## Data and Design

For each dimension:

- Generate 20 random data points from standard normal distribution with known covariance
- Generate mean and covariance for conjugate prior normal distribution
- Obtain the true posterior mean analytically
- For i in 1:100 :
  - Employ ADVI using statistical software, RStan, to obtain 10,000 samples from the approximate posterior
  - Calculate means and bias for the sample
- Average across all sample means
- Compute statistics for bias

## Results

In the following table, we show the sample mean obtained from ADVI and the true posterior mean along with the average bias for each dimension. We also show the change in the minimum, maximum and average bias in a plot (right).




| Dimension | 1 | 2 | 3 | 10 | 25 | 50 | 100 | 250 | 500 |
|---|---|---|---|---|---|---|---|---|---|
| Sample Mean | 0.0751 | $\begin{pmatrix}0.5538\\-0.9633\end{pmatrix}$ | $\begin{pmatrix}1.8130\\1.7399\\0.6503\end{pmatrix}$ | $\begin{pmatrix}0.2136\\-0.7749\\\vdots\\1.7517\\1.5453\end{pmatrix}$ | $\begin{pmatrix}-1.4439\\-1.4150\\\vdots\\0.5068\\-0.1483\end{pmatrix}$ | $\begin{pmatrix}0.9145\\0.6468\\\vdots\\0.8765\\1.8945\end{pmatrix}$ | $\begin{pmatrix}0.2102\\-0.2628\\\vdots\\0.0596\\-0.0792\end{pmatrix}$ | $\begin{pmatrix}0.3591\\0.1498\\\vdots\\0.0439\\-0.5412\end{pmatrix}$ | $\begin{pmatrix}-0.3107\\-0.3379\\\vdots\\0.3702\\-0.5708\end{pmatrix}$ |
| True Mean | 0.0831 | $\begin{pmatrix}-0.5344\\-0.9802\end{pmatrix}$ | $\begin{pmatrix}1.8709\\1.8023\\0.7338\end{pmatrix}$ | $\begin{pmatrix}0.2608\\-0.6897\\\vdots\\1.8934\\1.6513\end{pmatrix}$ | $\begin{pmatrix}-1.3771\\-1.3185\\\vdots\\0.5303\\-0.1140\end{pmatrix}$ | $\begin{pmatrix}1.0750\\0.8389\\\vdots\\1.1646\\2.1412\end{pmatrix}$ | $\begin{pmatrix}-0.0344\\-0.5321\\\vdots\\0.0872\\-0.0769\end{pmatrix}$ | $\begin{pmatrix}0.4267\\0.2254\\\vdots\\-0.0670\\-0.6387\end{pmatrix}$ | $\begin{pmatrix}-0.3009\\-0.3213\\\vdots\\0.4887\\-0.4407\end{pmatrix}$ |
| Ave. Bias | 0.1005 | 0.1369 | 0.1998 | 0.5141 | 0.8531 | 1.9529 | 3.3381 | 4.5072 | 7.0823 |

## Discussion

We have performed ADVI for the multivariate normal case and analyzed the performance of the mean. We can see an increase in the bias as the dimension increases. However, it is important to note that the biases for larger dimensions are inherently larger than in smaller dimensions due to having more components to include when computing distances, so it is difficult to say whether variational inference plays a strong part in this increase. Nonetheless, it is clear that the sample mean does not perform as well in large dimensions. We conclude by saying that there are other metrics with which to determine the quality of the approximated posterior. One such metric is the parameter from Pareto Smoothed Importance Sampling for which details can be found in reference [3].


References

[1] Blei, D., Variational Inference: A Review for Statisticians. 2018.
[2] Kucukelbir, A., Automatic Variational Inference in Stan. 2015.
[3] Yao, Y. Yes, but Did it Work?: Evaluating Variational Inference.