Comparing MCMC and VaNBayes for Heteroskedastic Models

Adira Cohen* Connor McNeill* Miles Woollacott*

With help from Elliot Maceda

*Department of Statistics, North Carolina State University

December 5, 2024

Outline

Model of Interest

We observe independent data $(Y_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$. $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is a known matrix of covariates such that $Corr(\boldsymbol{X}_{.i}, \boldsymbol{X}_{.j}) = \rho^{|i-j|}$.

We model

$$Y_i = \beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + \text{expit}\left\{\gamma_0 + \boldsymbol{X}_i^T \boldsymbol{\gamma}\right\} \epsilon_i,$$

where:

- $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$
- $\beta_0, \gamma_0 \in \mathbb{R}$
- $\boldsymbol{\beta}$, $\boldsymbol{\gamma} \in \mathbb{R}^p$ is sparse

The Dataset

Use a simulated dataset. Below describes our $\boldsymbol{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$:

- $n = 50$, and $p = 10$
- We will have 10 validation data points $Y_1^*, \ldots, Y_{10}^*$
- $\rho = 0.5$
- $\beta_1 = \beta_2 = \beta_6 = \frac{1}{2}$, otherwise 0
- $\gamma_2 = \gamma_3 = \gamma_{10} = 1$, otherwise 0
- $\beta_0 = \gamma_0 = 0$.

MCMC Approach

What happens if we try to fit the heteroskedastic model using MCMC in JAGS?

- β_0 and $\gamma_0 \sim \mathcal{N}(0, 0.1)$
- Hierarchical prior for $\boldsymbol{\beta}$:

$$\beta_j|\pi_j \sim \begin{cases} \mathcal{N}(0, \tau_\beta) & \text{with prob. } \pi_j \\ 0 & \text{with prob. } 1-\pi_j \end{cases}$$

$$\pi_j \sim \text{Beta}\,(0.5, 0.5)$$

- Similar distributions for $\boldsymbol{\gamma}$
- τ_β and $\tau_\gamma \sim \text{Gamma}(0.1, 0.1)$
- We obtain the Posterior Inclusion Probabilities (PIPs) for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ along with the PPD for Y^*

The Problem with MCMC/ABC

- When the likelihood's derivative is poorly behaved, MCMC may suffer in performance
- Approximate Bayesian Computing (ABC) is a likelihood-free approach, but it converges very slowly in higher dimensions

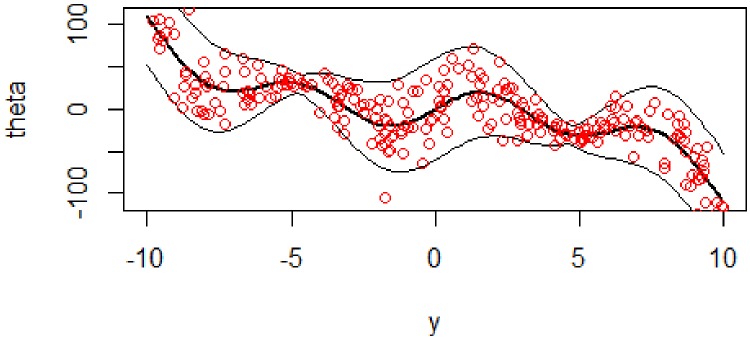
We need a more flexible method with weaker assumptions where the likelihood is not well-behaved, and in higher dimensions.

Introduction to VaNBayes

- A variational Bayes approach
- Idea: leverage advantages of machine learning and Bayesian inference
  1. Learn relationship between many simulated datasets and parameters using a neural net
  2. Plug observed data into the trained neural network

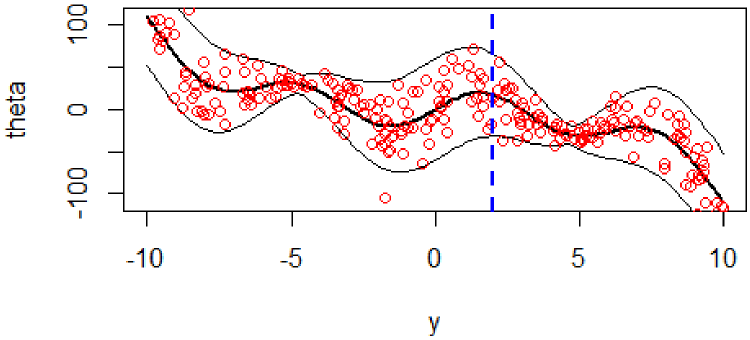
Toy VaNBayes Example

First: neural net learns relationship between simulated data y and parameter θ.



Toy VaNBayes Example

Second: plug observed y (blue) into trained neural network.



Advantages/Disadvantages of VaNBayes

Pros

- Very fast to fit on observed data once trained
- Likelihood-free and very flexible
- As long as we provide enough data, VaNBayes should work
- Convergence should be guaranteed theoretically

Cons

- Slow to train.
- Black-box model limits interpretation.
- Requires careful handling to prevent blow-up

Simulation Setup

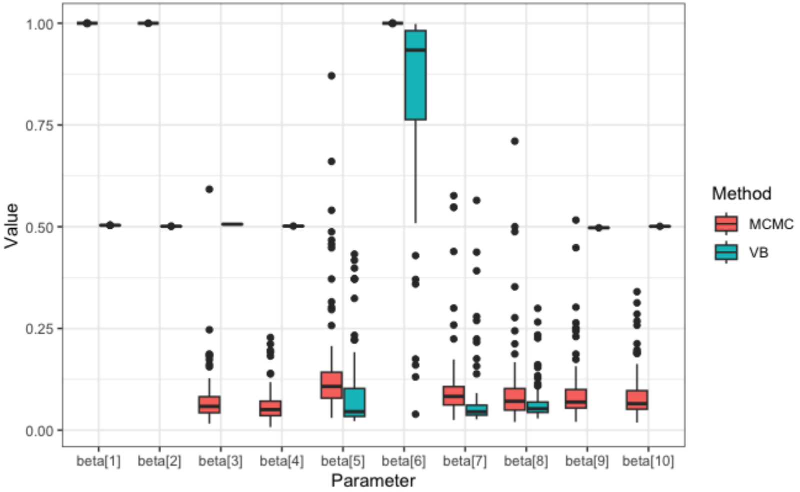
First we train our neural networks for VaNBayes using 100,000 datasets.

Then for 100 simulations:

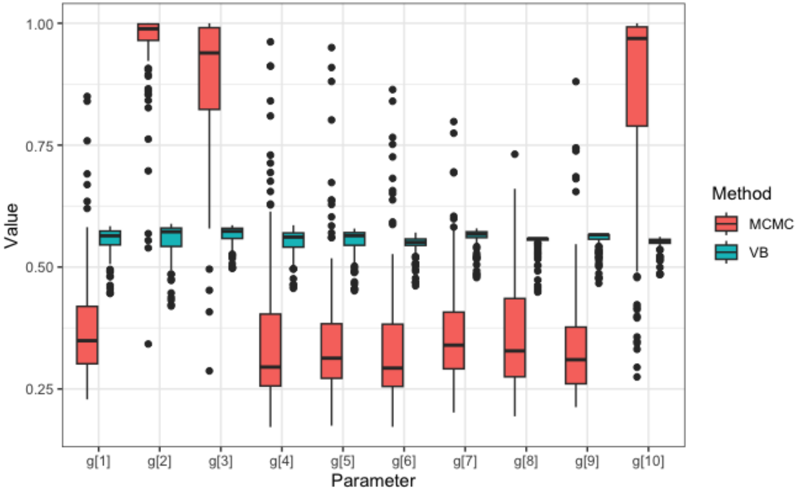
1. Generate data from our model
2. Fit MCMC using JAGS with the generated data
3. Obtain results using the trained neural networks for VaNBayes

Then compare results by looking at the distribution of the PIPs and the median of Y^*.

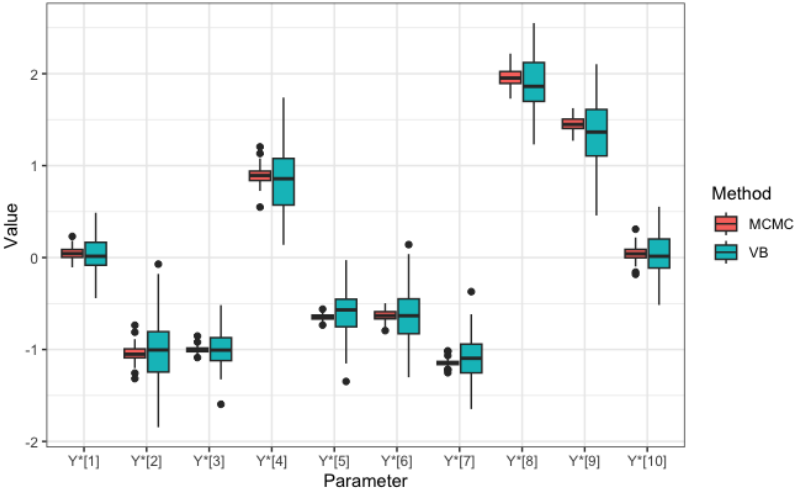
Results for β



Results for γ



Results for $\boldsymbol{Y}^*$



Conclusion

- MCMC worked well in this case, VaNBayes had some issues
- VaNBayes predicts Y^* well, but has higher variance

Future Work:

- Different summary statistics
- Less complex models
- Changing the size of the neural network

References

E. Maceda, E. C. Hector, A. Lenzi, and B. J. Reich. A variational neural bayes framework for inference on intractable posterior distributions, 2024. URL https://arxiv.org/abs/2404.10899.