Posterior Computation with Intractable Likelihoods

Elliot Maceda, NCSU

Joint work with Emily Hector, Amanda Lenzi, and Brian Reich

August 3, 2025

NC STATE UNIVERSITY

A motivating example[1]

One Brazil family's haunting diagnosis

DISEASE

Mother's Zika virus followed by microcephaly

tories.
The symptoms tend to be mild and it's estimated just one in four people show signs of infection. There is no vaccine or treatment.
Microcephaly is a condition where babies are born with

aly grow up to have normal intelligence and development. Others suffer serious developmental delays, dwarfism or seizures. Dr. Gustavo Malinger, an expert, said babies he saw faced severe mental handicaps with "no chances of intellect."

HEALTH

Zika virus sexually transmitted in Texas

ZIKA VIRUS

Outbreak of defects a threat, says WHO

HEALTH

Heat helps spread of Zika virus

A Health Secretary employee fumigates against mosquito Aedes Aegypti mosquito outside houses of Cali, Colombia, on Thursday. The Zika virus is "spreading explosively" in the Americas and the region may see up to four million cases of the disease strongly suspected of causing birth defects, the World Health Organization announced Thursday. LUIS ROBAYO/AFP/GETTY IMAGES

Zika virus spreading at 'alarming' rate: WHO

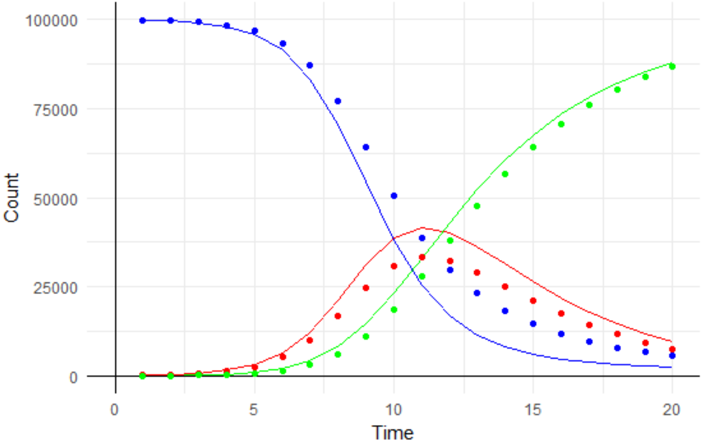
HEALTH

U of A lab joins Zika virus fight

[1]Metro World News

Modeling the problem

- The Susceptible-Infected-Recovered (SIR) model is a common approach to epidemiological problems.



Modeling the problem

- However, the Zika Virus spread across Brazilian states.
- It would be more realistic if we make this a spatial model.
- This can be done with a spatial stochastic differential equation.

The spatial SIR model[2]

- The probability of a new infection in location s within the time interval $(t, t+\delta)$ is approximately

$$\beta \frac{\delta}{N_s} X_s(t) Y_s(t) + \phi \frac{\delta}{N_s} \sum_{k \neq s} X_k(t) Y_s(t)$$

  Y_s represents the infected count,
  X_s represents the susceptible count,
  N_s is the population in location s.

- The probability of a recovery in the same time interval is

$$\eta \delta Y_s(t)$$

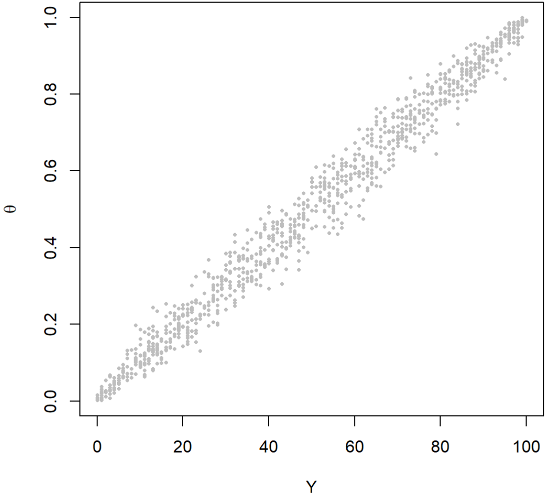
- The likelihood of $\theta = (\beta, \phi, \eta)$ is intractable.

[2] Parker Trostle, Joseph Guinness, Brian J Reich (2024)

Let’s start with a simpler example

```
> S     <- 100000 # Number of MC samples
> n     <- 100
> theta <- runif(S,0,1) # Draw from prior
> Y     <- rbinom(S,n,theta) # Draw from likelihood
> plot(theta,Y)
```

MC Estimate of the joint distribution of (θ, Y)

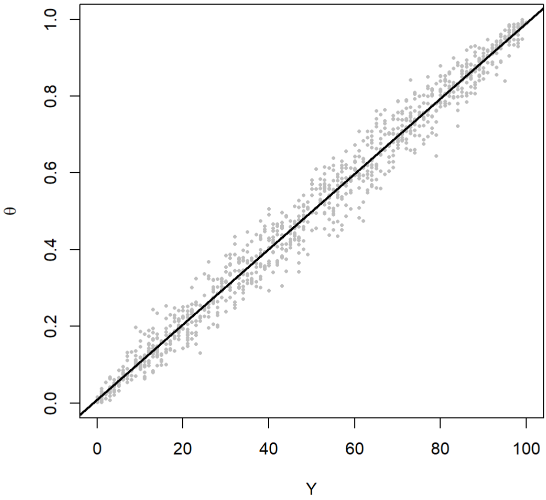


An idea– modeling $\theta|Y$

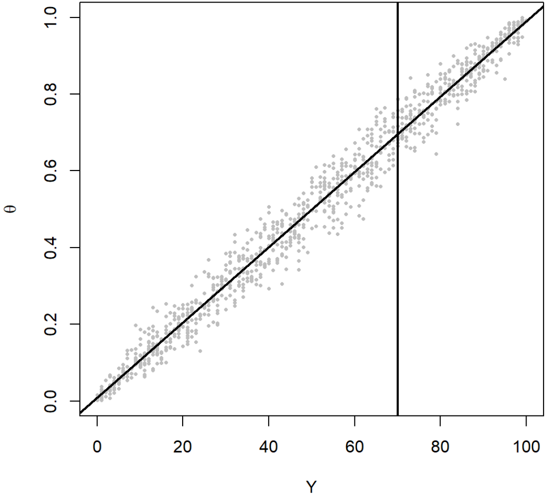
- Bayesian computation targets the distribution of $\theta|Y$.
- Model the samples of θ as the outcome and Y as the predictor!
- For this straightforward example, try

$$\theta = a + bY + e$$

Linear regression fit



The posterior at $Y = 70$ is the prediction at $Y = 70$



Using machine learning to approximate the posterior

- Usually the relationship between θ and Y is not so straightforward.
- In this case, $\theta|Y$ could be modeled using machine learning.
- Neural-network based methods are called "neural posterior estimators".

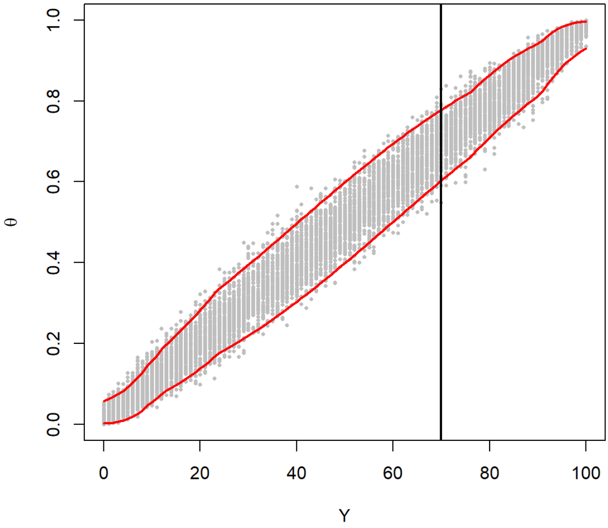
VaNBayes: Our Contribution

- We impose a parametric assumption on the distribution of $\theta|Y$.
- The neural network estimates the parameters of the parametric family, similar to variational Bayes.
- We target low-dimension summaries of the posteriors, avoiding high-dimensional density estimation.
- Unlike modern neural posterior estimators, VaNBayes can estimate discrete posteriors.

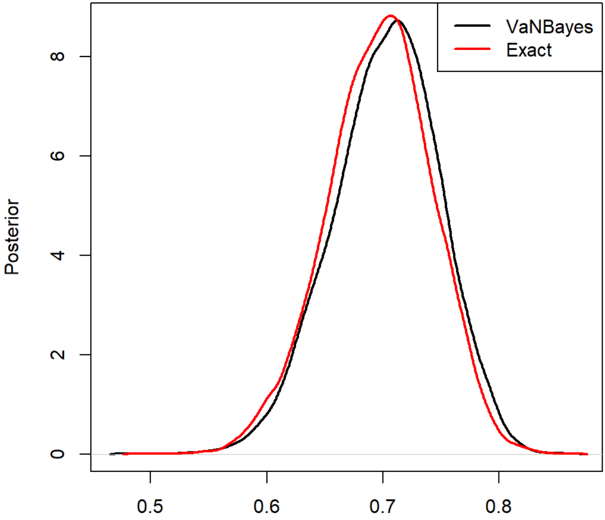
Implementing VaNBayes on the simple example

- Let $(f_1(Y), f_2(Y))$ be a neural network with two outputs.
- Model the posterior as $\theta|Y \sim N(f_1(Y), e^{f_2(Y)})$.
- Fit the neural network using the simulated draws of (θ, Y).
- Once trained, the posterior is found by plugging in our data.

VaNBayes is amortized



Approximate and exact posterior for $Y = 70$ ($n = 100$)



Dealing with higher dimensions

- Statistical modeling is difficult in higher dimensions.
- We are usually interested in a low dimension summary of the parameters:

$$\boldsymbol{\gamma} = \tau(\boldsymbol{\theta})$$

- If Y is large, we can reduce to summary statistics:

$$\mathsf{Z} = T(\mathsf{Y})$$

- The statistical modeling problem is now $\boldsymbol{\gamma}|\mathsf{Z}$.

The VaNBayes Algorithm

for $i = 1, ..., N$ **do**

  sample parameters from prior $\boldsymbol{\theta}^{(i)} \sim \pi(\boldsymbol{\theta})$

  sample data from likelihood $\mathsf{Y}^{(i)}|\boldsymbol{\theta}^{(i)} \sim f(\mathsf{y}|\boldsymbol{\theta}^{(i)})$

  calculate parameters of interest $\boldsymbol{\gamma}^{(i)} = \tau(\boldsymbol{\theta}^{(i)})$

  calculate summary statistics $\mathsf{Z}^{(i)} = T(\mathsf{Y}^{(i)})$

end for

Train an ML model to approximate the posterior $p(\boldsymbol{\gamma}|\mathsf{Z})$

Return $p(\boldsymbol{\gamma}|\mathsf{Z}_0)$ for Z_0, the observed data's summary statistics

Other VaNBayes Properties

- We can generate data from a proposal distribution instead of the prior distribution, similar to importance sampling.

- The VaNBayes posterior minimizes an average of the reverse KL divergences of all posteriors, weighted by their evidence values[3]

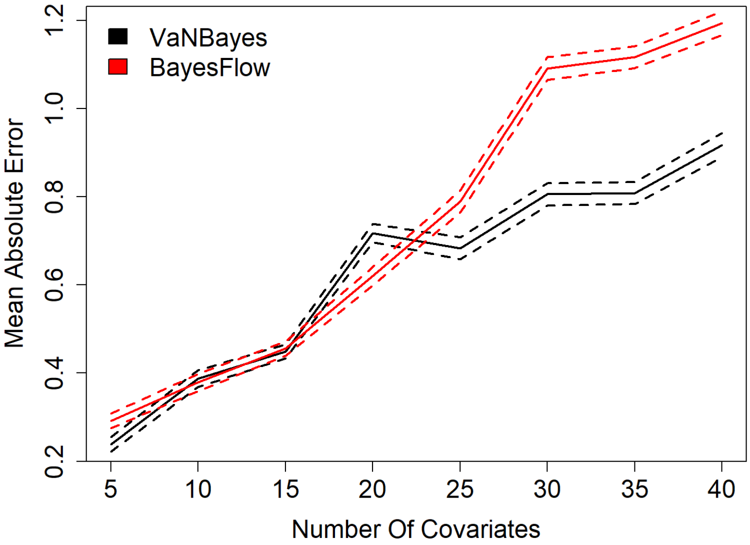
[3] This is equivalent to minimizing the reverse KL Divergence between the true joint and the VaNBayes parametrized joint distributions.

Simulation study with a modern neural posterior estimator

- Bayesflow[4] is a modern normalizing flow-based neural posterior estimator.
- This performs density estimation directly, which is tricky in higher dimensions.
- Let's compare this against VaNBayes in multiple linear regression as the number of covariates increases.
- We'll use the marginal heterogeneous normal model for VaNBayes.
- Compare across 100 datasets for each set of covariates.

[4] Radev et al. (2023a) BayesFlow: Amortized Bayesian Workflows With Neural Networks

VaNBayes seems to perform better as the dimension increases



Analyzing the Zika Virus dataset[5]

- A small adjustment to our spatial SIR model:

- The probability of a new infection in location s within the time interval $(t, t+\delta)$ is approximately

$$exp(\beta_0 + \beta_1 \log X_s(t))\frac{\delta}{N_s}X_s(t)Y_s(t) + \phi\frac{\delta}{N_s}\sum_{k \neq s} X_k(t)Y_s(t)$$

  Y_s represents the infected count,
  X_s represents the susceptible count,
  N_s is the population in location s.

- Let's estimate the posteriors of $\theta = (\beta_0, \beta_1, \phi)$ marginally.

[5]PAHO (2021). Data - BRA Zika Report

VaNBayes setup

- We assume the uninformative priors $\beta_0 \sim \text{Uniform}(-3, 1)$, $\beta_1 \sim \text{Uniform}(-1, 1)$, and $\phi \sim \text{LogNormal}(-2, 1)$.
- We assume a normal distribution posterior for a transformation of each of the parameters.
- The data are weekly infected counts across the 27 Brazilian states over 40 weeks.
- We use PCA to summarize the data across states and time.

Analysis results

- We find the 95% credible intervals $\beta_0 \in (-0.80, -0.75), \beta_1 \in (0.61, 0.80)$, and $\phi \in (0.18, 0.29)$.

- This agrees with intuition: a larger population tends to imply a higher infection rate for the state.

Other applications we've explored

- Max-stable processes
- Spatial autologistic regression
- Sparse linear regression

Moving forward

- Importance sampling allows us to generate (θ, Y) using a "proposal" distribution other than the prior.

- Considering a second-stage/sequential option for VaNBayes:
    - Currently VaNBayes avoids using data to train (amortization).
    - Potential to use data to inform a good proposal distribution.
    - Loses amortization, but posterior may be more accurate.

- At some point in the future make an R package.

Thank you!

This work was partially supported by National Science Foundation grants DMS2152887 and DMS2342344, and a grant from North Carolina State University's Office of Global Engagement.