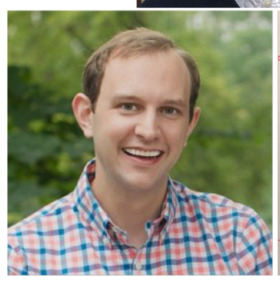
Fast Bayesian Inference in GLMs with Low Rank Data Approximations

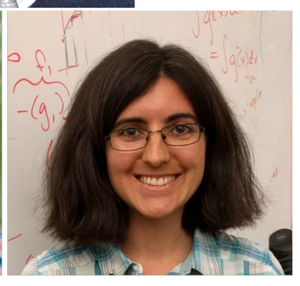
MIT · HARVARD UNIVERSITY

Brian L. Trippe Jonathan H. Huggins Tamara Broderick

Brian Trippe



Jonathan Huggins



Tamara Broderick

Overview

- Scientists, engineers & social scientists are often interested in the relationship between a large set of features and a response.
- For example, a biologist may wish to understand the effect of natural variations of certain genes on the presence of a disease.
- Bayesian generalized linear models (GLMs) provide coherent uncertainty quantification but can be slow to learn.
- We propose a **low rank approximation of data** -- as a form of likelihood approximation.
- We show **improved dimension dependence in time and memory** scaling of inference.
- We provide **theoretical guarantees** and experiments providing a **10x speed-up** with minimal approximation error.

Our Approach

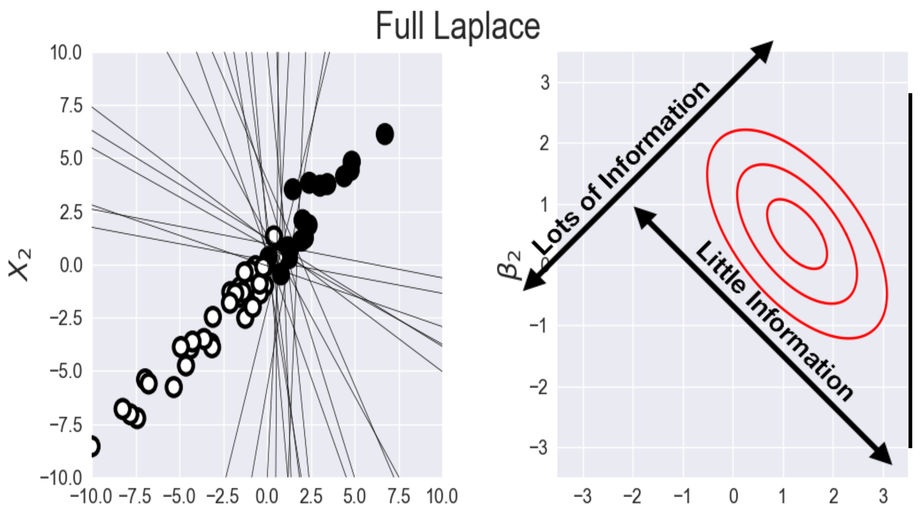
1. Perform an M-truncated SVD of covariates $X \approx U\mathrm{diag}(\lambda)V^T$.
2. Define an approximation to the likelihood:
$$p(y_i|x_i^T\beta) \approx \tilde{p}(y_i|x_i^T UU^T\beta)$$
3. Do approximate Bayesian inference at a fraction of cost:
$$\tilde{p}(\beta|X,Y) \propto p(\beta)\prod_{i=1}\tilde{p}(y_i|x_i^T UU^T\beta)$$

Inference Method	Naive	Our Approach
MCMC (per iteration)	O(DN)	O([D+N]M)
Laplace Approximation	O([N+D]D²)	O([N+M]DM)

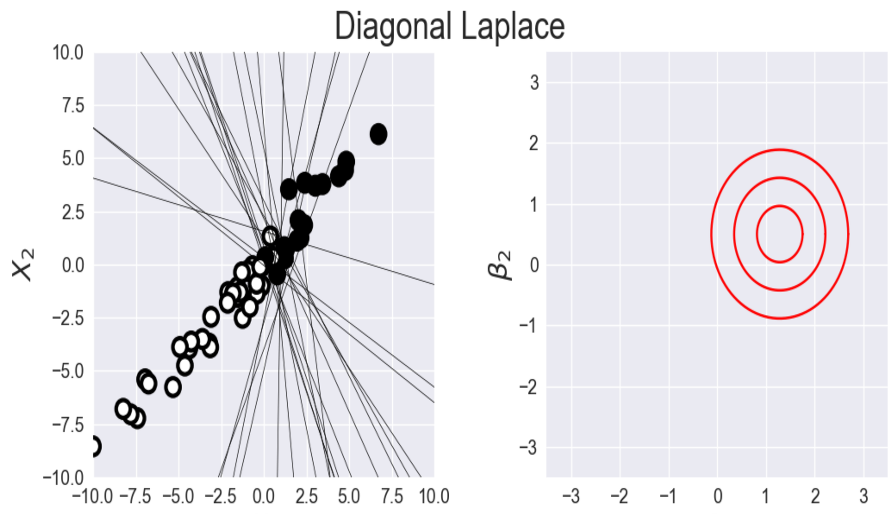
When data are exactly low rank, our approach is exact; otherwise it is an approximation.

Logistic Regression with Laplace Approximations

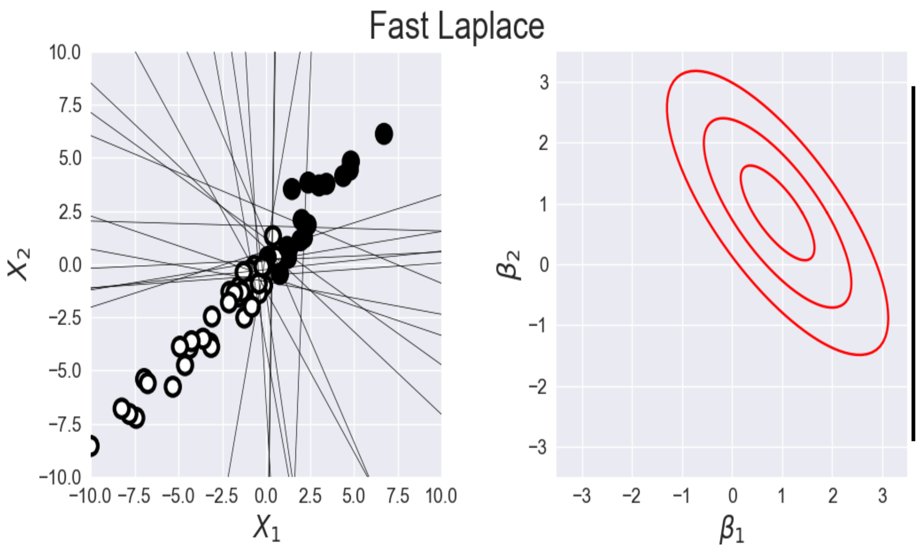
We compare of several Gaussian approximations to the exact Bayesian posterior in a toy 2D logistic regression model. Points represent the dataset; lines represent posterior samples of the decision boundary.



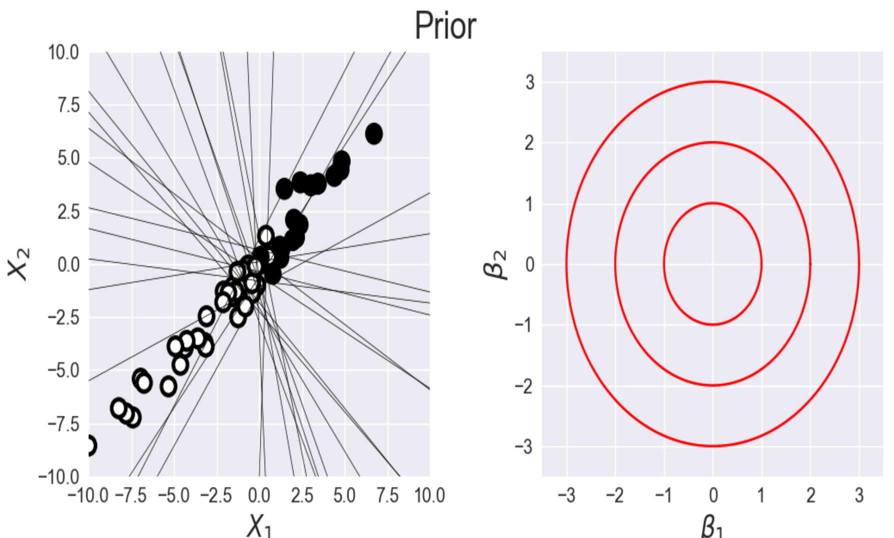
The usual Laplace approximation closely captures the exact posterior.



The factorized Laplace approximation underestimates uncertainty.



Fast Laplace approximation uses a rank 1 approximation to X.



For reference, the prior is highly uncertain.

As expected, our approximation yields greater uncertainty than the usual Laplace approximation.

Acknowledgements

We thank Raj Agrawal for helpful comments and discussion.

Background

Generalized Linear Models (GLMs)

- Consider a regression of N, D-dimensional covariates, X, on N responses, Y.
- GLMs are a widely used class of interpretable models w. parameter $\beta \in \mathbb{R}^D$.
- Accommodate different response types (counts, binary, heavy-tailed)
- Characterized by likelihoods of the form: $y_i|x_i,\beta \sim p(y_i|x_i^T\beta)$

Conjugate Gaussian Bayesian Regression

Generative Model

$$\beta \sim \mathcal{N}(0, \sigma_\beta I)$$
for $i = 1, 2, \ldots, N$:
$$y_i \sim \mathcal{N}(x_i^T\beta, \tau^{-1})$$

Analytic Posterior

$$p(\beta|Y,X) = \mathcal{N}(\beta|\mu_N, \Sigma_N)$$
$$\Sigma_N := \left(\Sigma_\beta^{-1} + \tau X^T X\right)^{-1}$$
$$\mu_N := \tau\Sigma_N X^T Y$$

****Posterior has an analytic form, but inference takes O(ND²+D³) time****

Conjugate Regression when X is Rank *M<D*

- We can write the SVD of X as: $X = U\mathrm{diag}(\lambda)V^T$,

for some $U \in \mathbb{R}^{D,M}$, $V \in \mathbb{R}^{N,M}$ with $M < D, N$

- And then:
$$\Sigma_N = \sigma_\beta^2\left\{I - U\mathrm{diag}\left(\frac{\tau\lambda^2}{\sigma_\beta^{-2} + \tau\lambda^2}\right)U^T\right\} \quad \text{and} \quad \mu_N = U\frac{\tau\lambda}{\sigma_\beta^{-2} + \tau\lambda^2}V^T Y.$$

****Exact inference takes O(NDM) time****

Key Theoretical Results

Theorem: In conjugate linear regression, if each $|y_i| < b$, our approximation $\tilde{p}(\beta|X,Y) = \mathcal{N}(\tilde{\mu}_N, \tilde{\Sigma}_N)$, satisfies :
$$\|\tilde{\mu}_N - \mu_N\|_2 \le \sigma_\beta^2\tau\left(\lambda_{M+1}^2\|\mu_N\|_2 + \lambda_{M+1}\sqrt{N}b\right)$$
Also, $\Sigma_N^{-1} - \tilde{\Sigma}_N^{-1} = \tau(X^TX - UU^TX^TXUU^T)$,

hence $\|\Sigma_N^{-1} - \tilde{\Sigma}_N^{-1}\|_2 = \tau\lambda_{N,M+1}^2$.

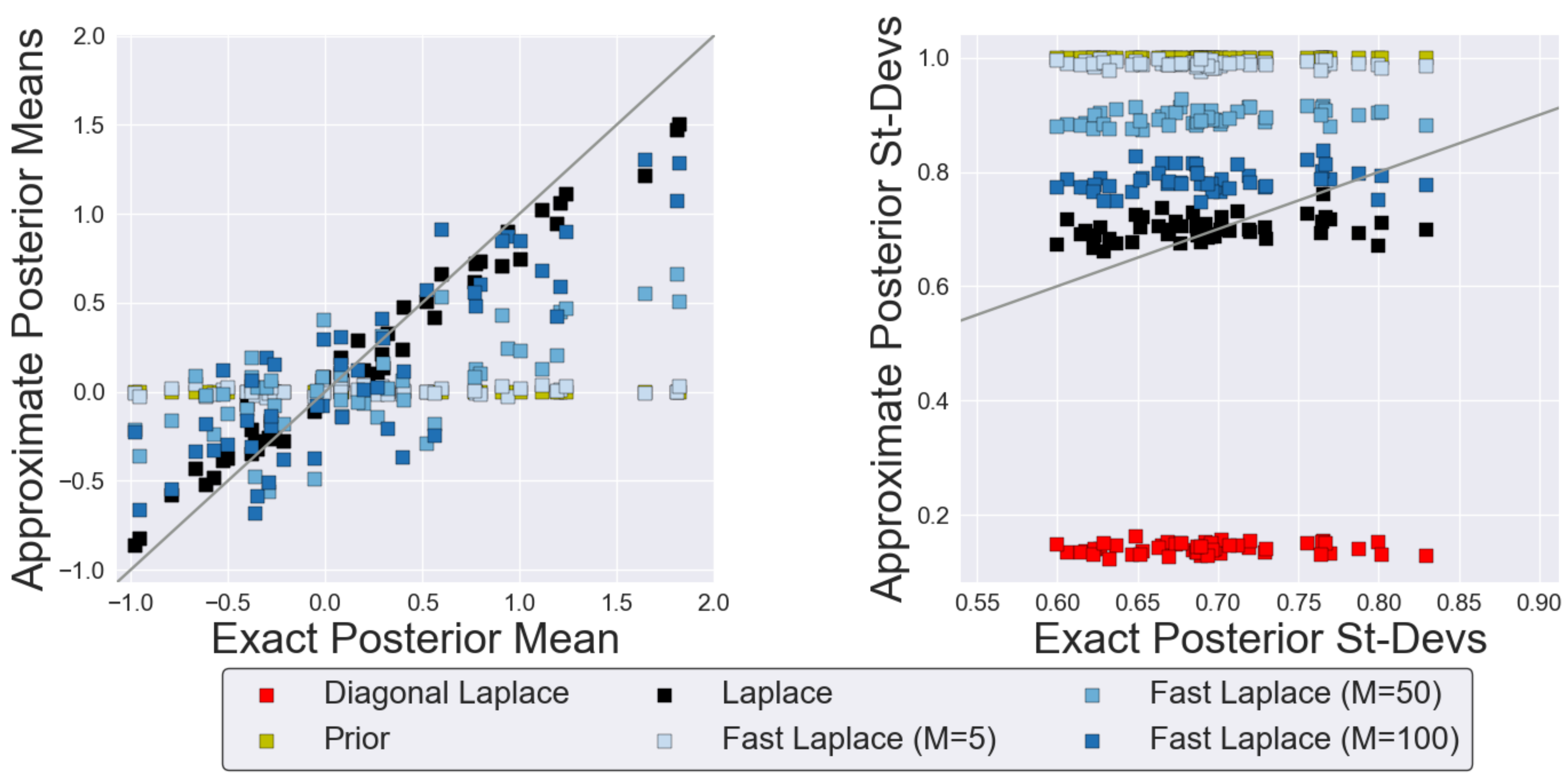
Corollary (consistency):

$\tilde{\mu}_N \xrightarrow{p} \tilde{\mu}$, the maximum a priori vector satisfying $U^T\tilde{\mu} = U^T\beta$

Corollary (conservativeness):

$\tilde{p}(\beta|X,Y)$ is no less uncertain than $p(\beta|X,Y)$; Formally,
$$\tilde{\Sigma}_N \succeq \Sigma_N \text{ and } \mathrm{H}\big(\tilde{p}(\beta|X,Y)\big) \ge \mathrm{H}\big(p(\beta|X,Y)\big).$$

Results for Logistic Regression



Approximate posterior mean and standard deviation across a subset of parameters as M varies. X-axis represents ground truth from running Hamiltonian Monte Carlo.