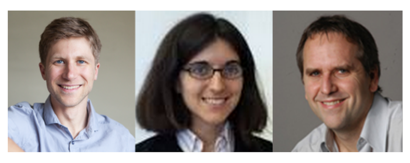
Robust Inference with Variational Bayes



Ryan Giordano **Tamara Broderick** **Michael Jordan**

Motivation:

- Bayesian inference always **requires a likelihood and prior**.
- A **robust posterior** does not depend strongly on the choice of the likelihood or prior.
- When a range of models or priors are reasonable, one needs **quantitative measures of robustness**.
- One measure of **local robustness** is the derivative of a posterior expectation with respect to the prior [2].
- Due in part to the difficulty of calculating local robustness from MCMC draws, **robust Bayes methods are not commonly used in practice.**

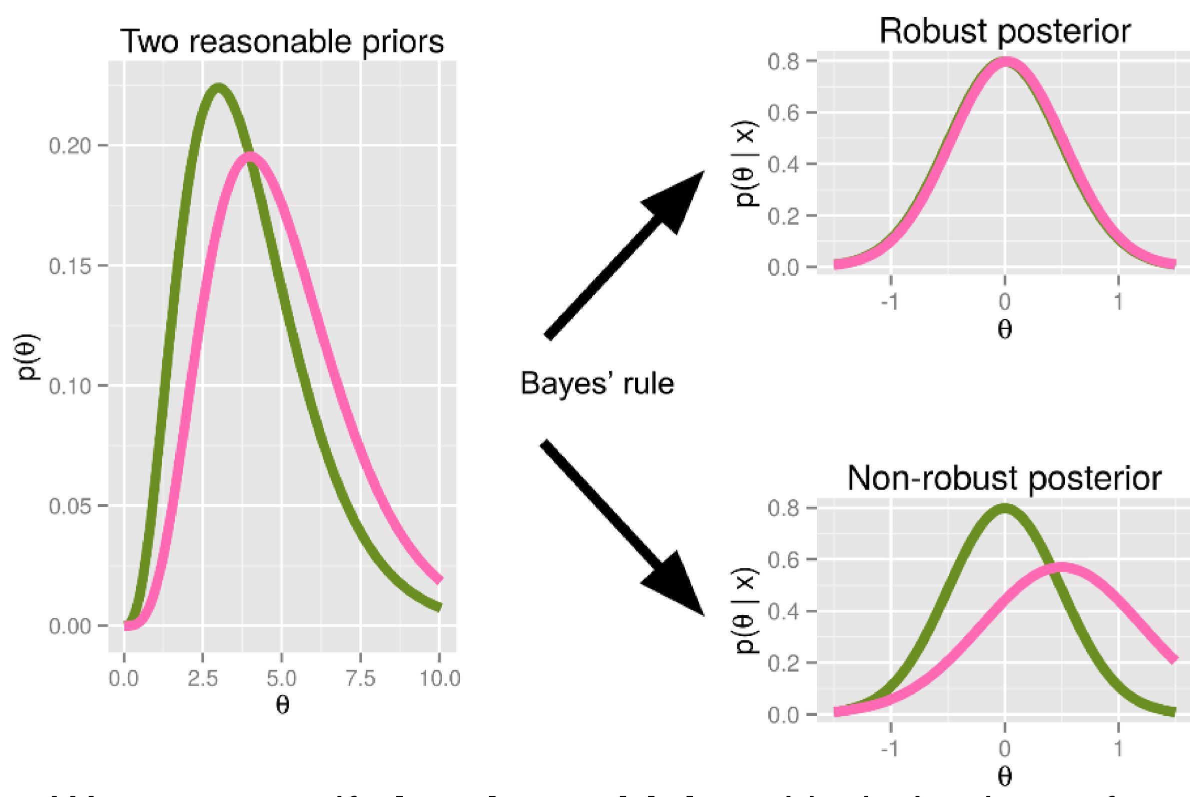
We describe:

- **Easy-to calculate, closed-form local robustness measures** for posteriors estimated with mean field variational Bayes.
- Our robustness measures require little computation beyond what is needed for **linear response variational Bayes** [1] (LRVB).
- We provide estimates for:
  - Sensitivity to **prior parameters** when the prior is in a parametric family
  - **Influence functions** for arbitrary perturbations
- A demonstration on a non-conjugate **hierarchical model from development economics** with comparison to Markov chain Monte Carlo (MCMC).

Theoretical Background

Definitions

$$x, \theta = \text{Observed data and parameters of interest, respectively}$$
$$p(\theta|\alpha) = \text{Prior as a function of } \alpha (\alpha \text{ may be vector- or function-valued})$$
$$p_x^\alpha(\theta) = p(\theta|x,\alpha) = \frac{p(x|\theta)\,p(\theta|\alpha)}{p(x)} = \text{Posterior of } \theta \text{ with prior parameter } \alpha$$



- We can quantify **local sensitivity** with derivatives of expectations:

$$\left.\frac{d\mathbb{E}_{p_x^\alpha}[\theta]}{d\alpha}\right|_\alpha \Delta\alpha = \text{Local sensitivity of } \mathbb{E}_{p_x^\alpha}[\theta] \text{ in the prior perturbation direction } \Delta\alpha$$

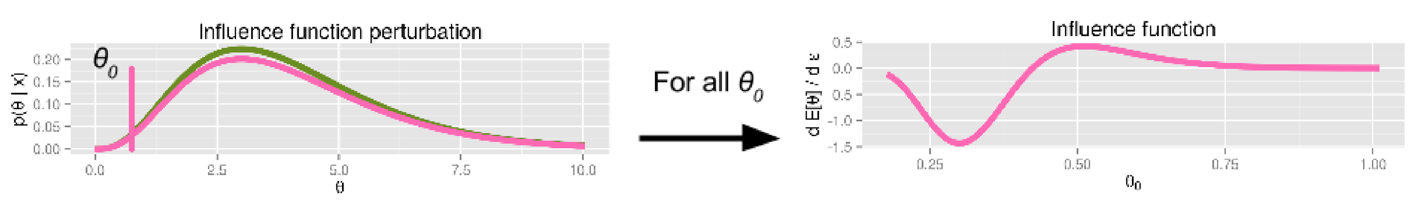
- To estimate this, we can **form a variational approximation** to p_x^α:

$$\mathcal{Q} = \text{A set of exp. family posteriors with natural sufficient statistic } \theta$$
$$q_x^\alpha = \text{argmax}_{q\in\mathcal{Q}}\{\mathbb{E}_q[\log p(x|\theta)] + \mathbb{E}_q[\log p(\theta|\alpha)] - \mathbb{E}_q[\log q(\theta)]\} =$$
$$= \text{The variational poseterior}$$
$$\hat{\Sigma} := (I - \Sigma_q H)^{-1}\Sigma_q = \text{LRVB covariance estimate}$$

- By Taylor expanding the log prior, LRVB gives a **closed form expression for the variational local sensitivity**:

$$\left.\frac{d\mathbb{E}_{q_x^\alpha}[\theta]}{d\alpha}\right|_\alpha \Delta\alpha = \hat{\Sigma}\nabla_m f \qquad \text{where} \qquad f(m) := \frac{d}{d\alpha^T}\mathbb{E}_q[\log(p(\theta|\alpha))]\,\Delta\alpha$$

Influence functions



We can measure the effect of adding arbitrary prior mass $p_c(\theta)$ to marginals of the prior $p(\theta|\alpha)$. Let δ be the Dirac delta function. Then:

- Define our perturbation as: $p(\theta_i|\alpha_i,\epsilon) = (1-\epsilon)p(\theta_i|\alpha_i) + \epsilon\delta(\theta_i - \theta_{i0})$
- Then the "influence function" of θ_{i0} is:

$$\frac{d\mathbb{E}_q[\theta]}{d\epsilon} = \frac{q_x^\alpha(\theta_{i0})}{p(\theta_{i0}|\alpha)}(I - \Sigma_q H)^{-1}\begin{pmatrix}\theta_{i0} - m_i \\ 0\end{pmatrix}$$

- Linear combinations of delta functions describe arbitrary perturbations.

Experiments

Microcredit model

We apply our results to a hierarchical model of microcredit interventions in development economics [4]. The goal is to combine multiple causal studies to gain statistical power.

Defintions

$$y_{ik} = \text{Profit of business } i \text{ in site } k$$
$$T_{ik} = \text{Indicator of whether business } i \text{ in site } k \text{ was in the control or treatment}$$
$$\mu_k = \text{Average profitability in site } k$$
$$\tau_k = \text{Average intervention effect in site } k$$

We observe y_{ik} and T_{ik}, where $k = 1, ..., 7$, and are interested in the posterior distribution of τ_k. The model is:

$$y_{ik}|\mu_k, \tau_k, T_{ik}, \sigma_k \sim N\left(\mu_k + T_{ik}\tau_k, \sigma_k^2\right) \text{ where } \begin{pmatrix}\mu_k \\ \tau_k\end{pmatrix} \sim N\left(\begin{pmatrix}\mu \\ \tau\end{pmatrix}, C\right)$$

We placed a normal prior on (μ, τ) and a non-conjugate LKJ prior [3] on C:

$$\begin{pmatrix}\mu \\ \tau\end{pmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \Lambda^{-1}\right) \qquad \begin{array}{l} C =: SRS \text{ where } S \text{ is diagonal,} \\ R \text{ is a covariance matrix, and} \\ \log p(R) = (\eta - 1)\log|R| + Constant \end{array}$$

Note that the variational means match the MCMC posterior closely, so the LRVB assumptions hold. The worst-estimated parameter is C^{-1}.

Perturbing prior parameters

The LRVB predicted sensitivity matches the effect of manually perturbing and re-running MCMC. As expected, C^{-1} is the worst-estimated.

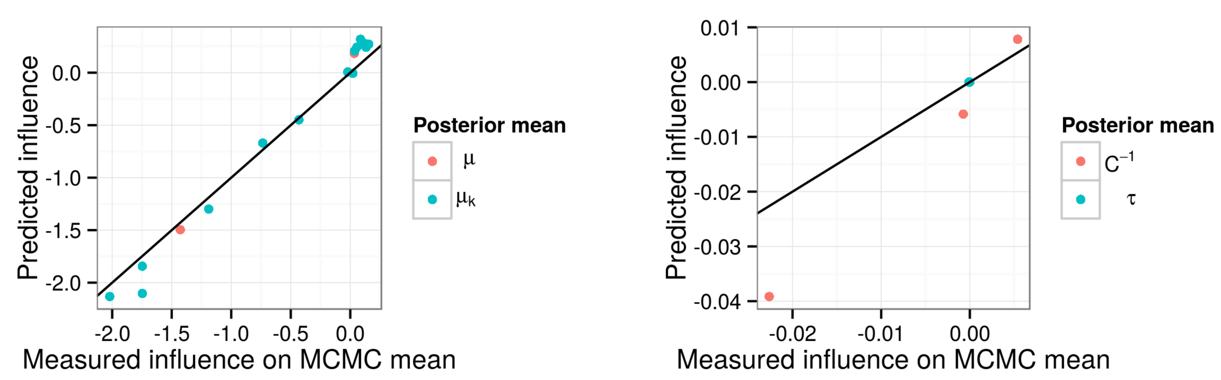


Figure 1: *Predicted vs actual effects of perturbations*

Note that $\mathbb{E}_p\mu$ and $\mathbb{E}_p\tau$ are robust to η but not to Λ. The influence function shows that (μ, τ) are only highly sensitive to prior mass centered tightly around the posterior, which is an unrealistic prior perturbation.

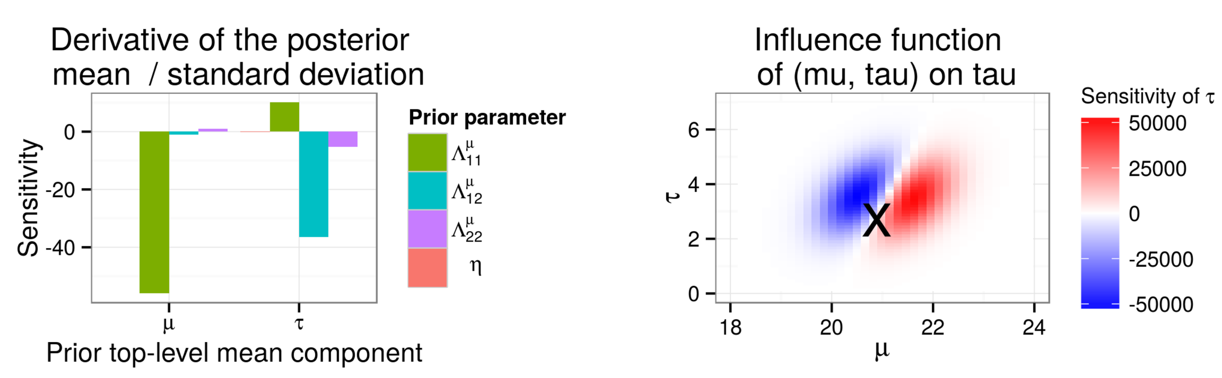


Figure 2: *The sensitivity of μ and τ*

References

[1] R. Giordano, T. Broderick, and M. Jordan. Linear response methods for accurate covariance estimates from mean field variational bayes. *arXiv preprint arXiv:1506.04088*, 2015.

[2] P. Gustafson. Local robustness in bayesian analysis. In D. R. Insua and F. Ruggeri, editors, *Robust Bayesian Analysis*, volume 152. Springer Science & Business Media, 2012.

[3] D. Lewandowski, D. Kurowicka, and H. Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9):1989–2001, 2009.

[4] R. Meager. Understanding the impact of microcredit expansions: A bayesian hierarchical analysis of 7 randomised experiments. *arXiv preprint arXiv:1506.06669*, 2015.

NIPS 2015