Variational Bayesian Inference for Parametric and Non-Parametric Regression with Missing Predictor Data

Christel Faes, John Ormerod and Matt Wand

August 23, 2010

Bayesian inference

- For parametric regression: long history

  (e.g. Box and Tiao, 1973; Gelman, Carlin, Stern and Rubin, 2004)
- For non-parametric regression: e.g. mixed model representations of penalized splines

  (e.g. Ruppert, Wand and Carroll, 2003)
- For dealing with missingness in data: allows incorporation of standard missing data models

  (e.g. Little and Rubin, 2004; Daniels and Hogan, 2008)
- Easy via MCMC, but can be costly in processing time

Variational Bayes inference

- Part of mainstream Computer Science methodology
  (e.g. Bishop, 2006)
- Recently, used in statistical problems
  (e.g. Teschendorff et al. 2005; McGrory & Titterington, 2007; Ormerod & Wand, 2010)
- Deterministic approach that yields approximate inference
- Involves approximation of posterior densities by other densities for which inference is more tractable

Faes, Ormerod and Wand (2010): develop and investigate variational Bayes for regression analysis with missing data

Elements of Variational Bayes

- Bayesian inference is based on the posterior density function

$$p(\theta|y) = \frac{p(y,\theta)}{p(y)}$$

- For an arbitrary density function q over Θ, the following inequality holds

$$p(y) \geq \underline{p}(y;q) = \exp\left(\int q(\theta) log \left\{\frac{p(y,\theta)}{q(\theta)}\right\} d\theta\right)$$

- Variational Bayes relies on product density restrictions:

$$q(\theta) = \prod_{i=1}^{M} q_i(\theta_i) \text{ for some partition } \{\theta_1, \ldots, \theta_M\} \text{ of } \theta$$

- The optimal densities (with minimum KL divergence) can be shown to satisfy

$$q_i^*(\theta_i) \propto \exp\left\{E_{-\theta_i} \log p(\theta_i|\text{rest})\right\}$$

Simple Linear Regression with Missing Predictor Data

- Assume the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

Cough this in Bayesian framework by taking $\beta_0, \beta_1 \sim N(0, \sigma_\beta^2)$ and $\sigma_\epsilon^2 \sim IG(A_\epsilon, B_\epsilon)$.

- Suppose that predictors are susceptible to missingness and assume

$$x_i \sim N(\mu_x, \sigma_x^2)$$

with hyperpriors $\mu_x \sim N(0, \sigma_{\mu_x}^2)$ and $\sigma_x^2 \sim IG(A_x, B_x)$

- Let R_i be the missingness indicators and consider the missingness mechanisms:
  1. $P(R_i = 1) = p$: MCAR
  2. $P(R_i = 1) = \Phi(\phi_0 + \phi_1 y_i)$ for $\phi_0, \phi_1 \sim N(0, \sigma_\phi^2)$: MAR
  3. $P(R_i = 1) = \Phi(\phi_0 + \phi_1 x_i)$ for $\phi_0, \phi_1 \sim N(0, \sigma_\phi^2)$: MNAR
- Use auxiliary variables $a_i|\phi \sim N((Y\phi)_i, 1)$ or $a_i|\phi \sim N((X\phi)_i, 1)$ for the probit regression components

Approximate Inference via Variational Bayes

- We impose the product density restrictions:

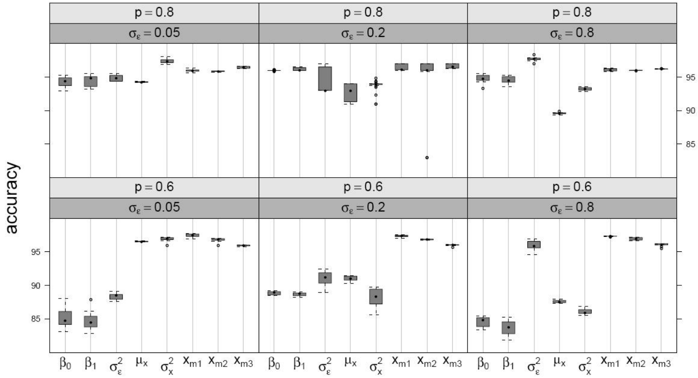
$$\begin{aligned}
\text{MCAR:} \quad & q(\beta, \sigma_\epsilon^2, x_{\text{mis}}, \mu_x, \sigma_x^2) = q(\beta, \mu_x) q(\sigma_\epsilon^2, \sigma_x^2) q(x_{\text{mis}}) \\
\text{MAR:} \quad & q(\beta, \sigma_\epsilon^2, x_{\text{mis}}, \mu_x, \sigma_x^2, \phi, a) = q(\beta, \mu_x, \phi) q(\sigma_\epsilon^2, \sigma_x^2) q(x_{\text{mis}}) q(a) \\
\text{MNAR:} \quad & q(\beta, \sigma_\epsilon^2, x_{\text{mis}}, \mu_x, \sigma_x^2, \phi, a) = q(\beta, \mu_x, \phi) q(\sigma_\epsilon^2, \sigma_x^2) q(x_{\text{mis}}) q(a)
\end{aligned}$$

- For the MCAR, this leads to optimal densities of the form

$$\begin{aligned}
q^*(\beta) &= \text{Bivariate normal density} \\
q^*(\mu_x) &= \text{Univariate normal density} \\
q^*(\sigma_\epsilon^2) &= \text{Inverse Gamma density} \\
q^*(\sigma_x^2) &= \text{Inverse Gamma density} \\
q^*(x_{\text{mis}}) &= \text{product of Univariate Normal densities}
\end{aligned}$$

- For MAR and MNAR situation, derivations of optimal densities for ϕ and a have easy expressions as well
- Non-parametric regression give rise to non-standard forms and numerical integration is required (we use numerical integration via quadrature)

Simulation Simple Linear Regression with predictor MCAR

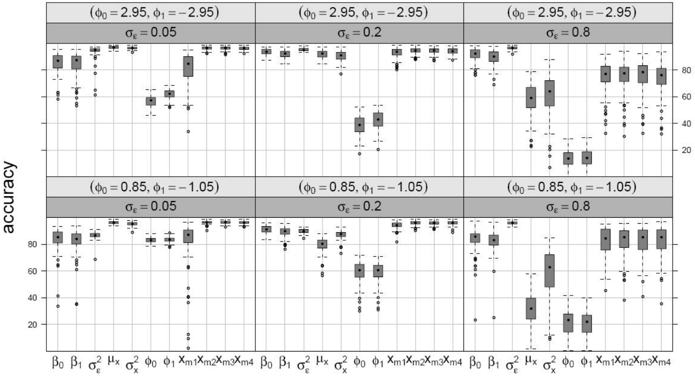


Accuracy measure defined as

$$\text{accuracy}(q^*) = 1 - (IAE(q^*)/sup_q IAE(q)) = 1 - \frac{1}{2} IAE(q^*)$$

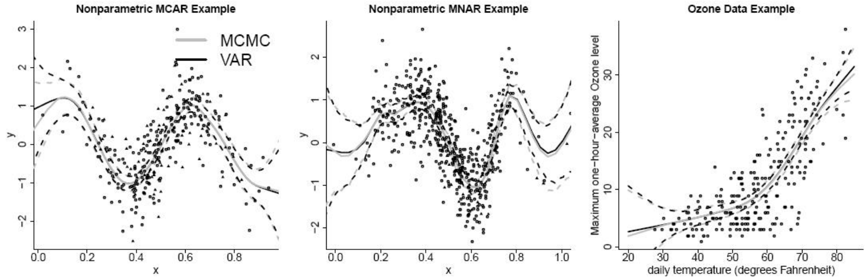
with *IAE* the integrated absolute error of q^*

Simulation Simple Linear Regression with predictor MNAR



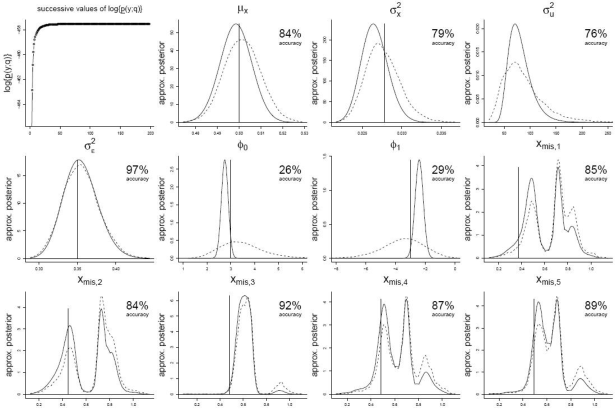
- Accuracy drops when amount of missing data is large and when data are noisy
- Accuracy of missing covariates is high in all situations
- Poor performance for missing mechanism parameters (due to strong correlation between ϕ and a)

Nonparametric Regression with Missing Predictor Data



- Good agreement between variational Bayes and MCMC in fitted functions
- Time needed: 75 seconds for variational Bayes, 15.5 hours for MCMC

Nonparametric Regression with Missing Predictor Data



- Variational Bayes are able to handle the multimodality of posteriors of the x_{mis} (coming from periodic nature of f)
- Good to excellent performance for all parameters (except for missing mechanism parameters)

Conclusions

- Variational Bayes inference achieves good to excellent accuracy for main parameters of interest
- Poor accuracy is realized for the missing data mechanism parameters
- Better accuracy maybe achieved with a more elaborate variational scheme – in situations where they are of interest
- Variational Bayes approximates multimodal posterior densities with high degree of accuracy
- Speed-up in the order of several hundreds

Contact Information

Christel Faes
I-BioStat, Center for Statistics
Hasselt University
Diepenbeek, Belgium

link to paper: http://www.uow.edu.au/ mwand/papers.html