VARIATIONAL APPROXIMATIONS IN SEMIPARAMETRIC REGRESSION

Matt Wand

University of Wollongong, AUSTRALIA

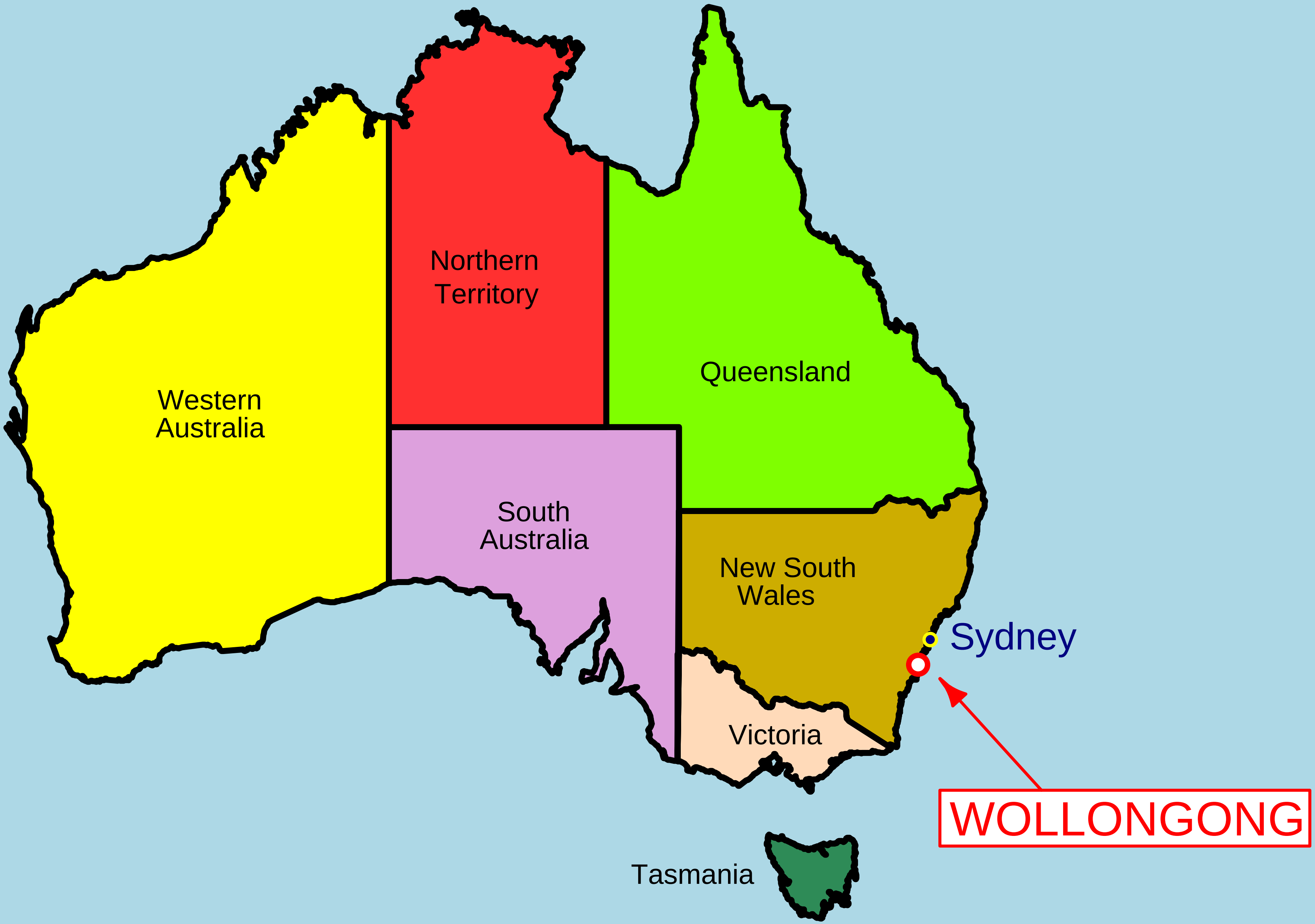
VARIATIONAL APPROXIMATIONS IN SEMIPARAMETRIC REGRESSION

Matt Wand

University of Wollongong, AUSTRALIA

(joint with Peter Hall (Uni. Melbourne)
and John Ormerod (Uni. Wollongong))


Northern
Territory
Queensland
Western
Australia
South
Australia
New South
Wales
Sydney
Victoria
Tasmania


Northern
Territory
Queensland
Western
Australia
South
Australia
New South
Wales
Sydney
Victoria
WOLLONGONG
Tasmania

First...

Primer on Variational Approximation

‘Undergraduate’ Variational Approximation

'Undergraduate' Variational Approximation

Consider the

Bayesian Poisson regression model

$$[y_i|\beta] \overset{\text{ind.}}{\sim} \text{Poisson}(\exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}))$$

Prior on regression coefficients: $(\beta_0, \ldots, \beta_k) \sim N(0, F)$.

‘Undergraduate’ Variational Approximation

Consider the

Bayesian Poisson regression model

$$[y_i|\beta] \overset{\text{ind.}}{\sim} \text{Poisson}(\exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}))$$

Prior on regression coefficients: $(\beta_0, \ldots, \beta_k) \sim N(0, F)$.

Matrix notation:

$$p(y|\beta) = \exp\{y^T X\beta - 1^T \exp(X\beta) - 1^T \log(y!)\}, \qquad \beta \sim N(0, F)$$

The log marginal likelihood is (ignoring constants):

$$\log p(y) = \log \int_{\mathbb{R}^p} p(y|\beta)p(\beta)\, d\beta$$

The log marginal likelihood is (ignoring constants):

$$\begin{aligned}\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\ &= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}}\end{aligned}$$

The log marginal likelihood is (ignoring constants):

$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\quad \times \frac{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}}
\end{aligned}
$$

The log marginal likelihood is (ignoring constants):

$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}) \, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X} \widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2} \widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1} \widetilde{\boldsymbol{\beta}}\} \, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X} \widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2} \widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1} \widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})\}}{(2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})\}} \, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \left[\frac{\exp\{\boldsymbol{y}^T \boldsymbol{X} \widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2} \widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1} \widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\mu})\}} \right]
\end{aligned}
$$

The log marginal likelihood is (ignoring constants):

$$\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right] \\
&\geq \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left(\log\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]\right)
\end{aligned}$$

The log marginal likelihood is (ignoring constants):

$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} \boldsymbol{p}(\boldsymbol{y}|\boldsymbol{\beta})\boldsymbol{p}(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right] \\
&\geq \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left(\log\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]\right) \\
&= \boldsymbol{y}^T\boldsymbol{X}\boldsymbol{\mu} - \mathbf{1}^T\exp\{\boldsymbol{X}\boldsymbol{\mu} + \tfrac{1}{2}\text{diagonal}(\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^T)\} - \tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\qquad -\tfrac{1}{2}\{\text{tr}(\boldsymbol{F}^{-1}\boldsymbol{\Sigma}) + \log|\boldsymbol{\Sigma}|\}
\end{aligned}
$$

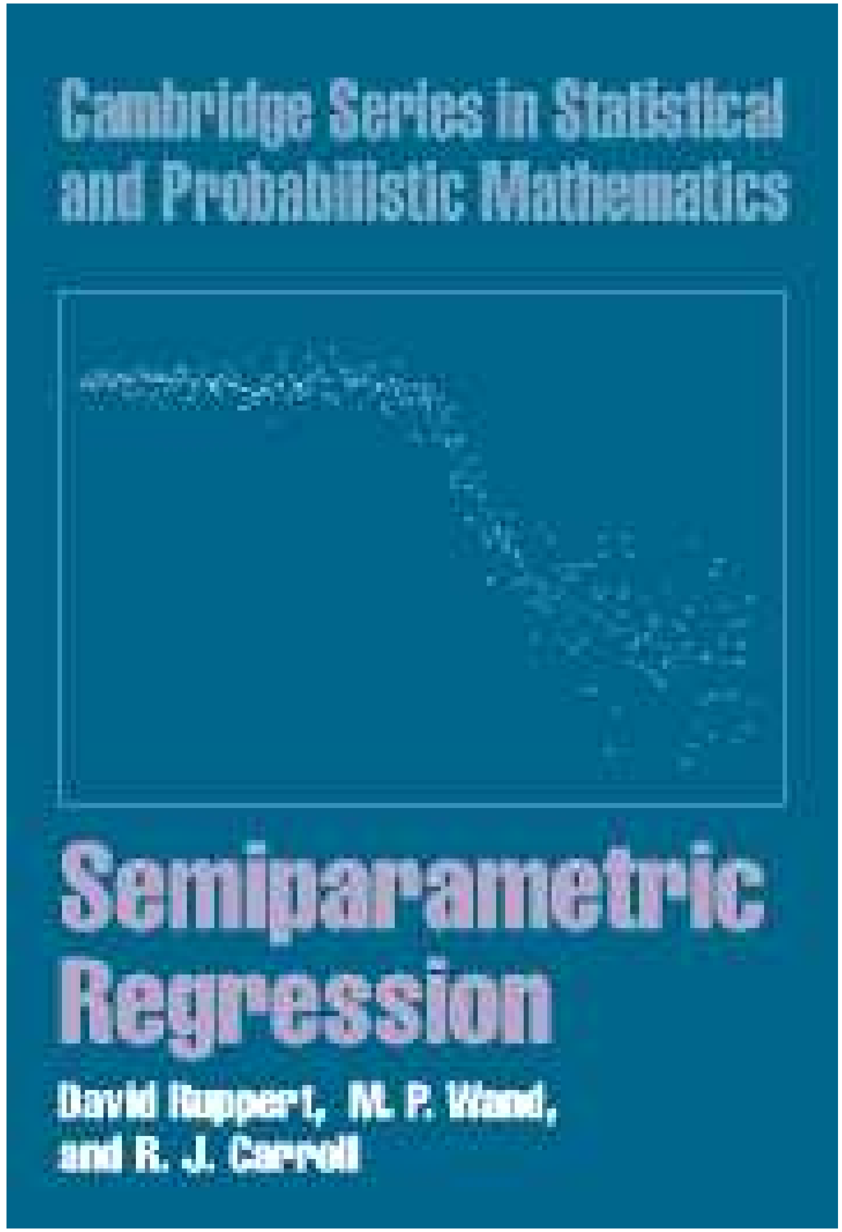
The log marginal likelihood is (ignoring constants):

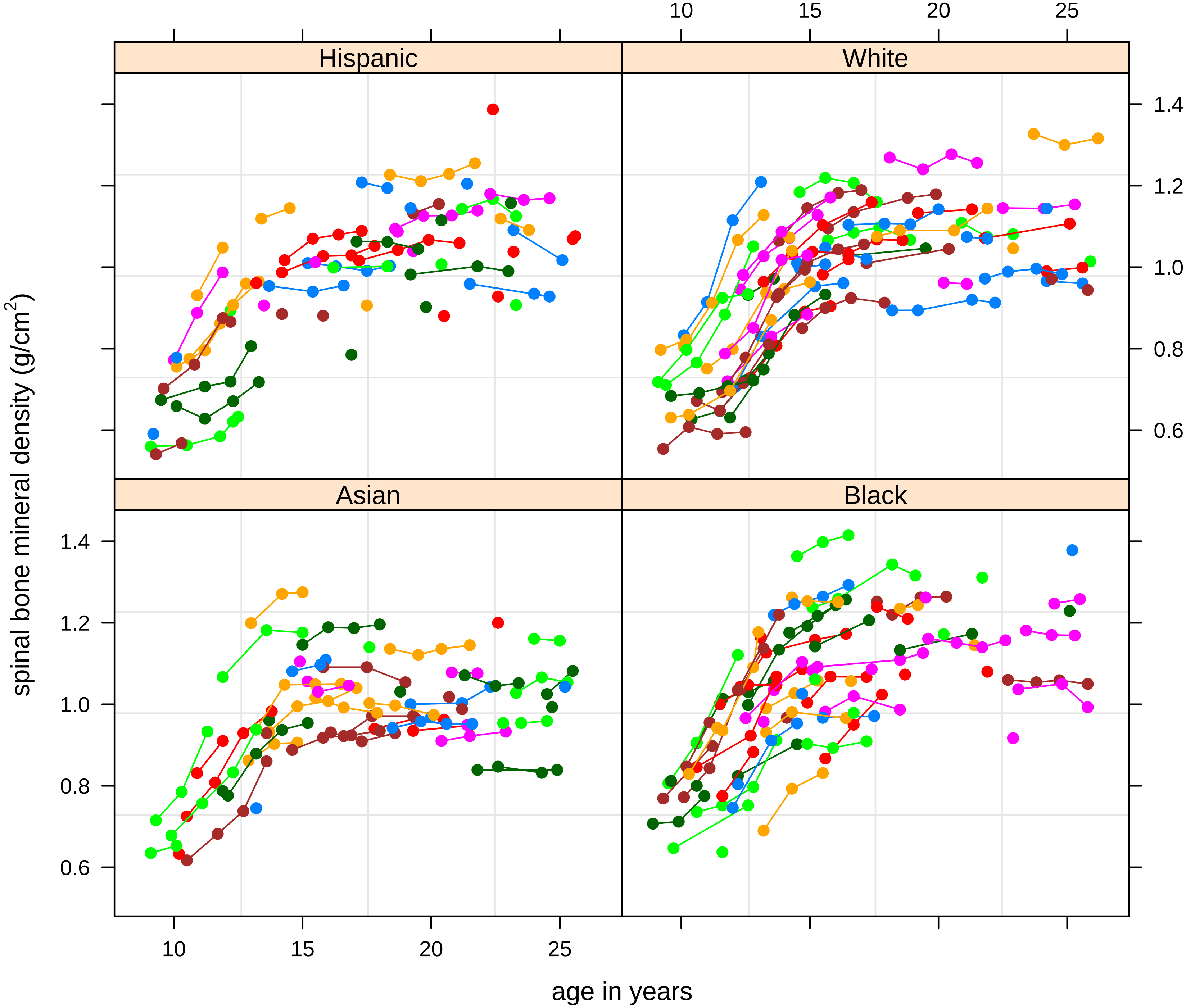
$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right] \\
&\geq \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left(\log\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]\right) \\
&= \boldsymbol{y}^T\boldsymbol{X}\boldsymbol{\mu} - \mathbf{1}^T\exp\{\boldsymbol{X}\boldsymbol{\mu} + \tfrac{1}{2}\text{diagonal}(\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^T)\} - \tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\qquad -\tfrac{1}{2}\{\text{tr}(\boldsymbol{F}^{-1}\boldsymbol{\Sigma}) + \log|\boldsymbol{\Sigma}|\} \\
&= \log\ \underline{p(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\Sigma})} \quad \text{for all } \boldsymbol{\mu}(p\times 1) \text{ and symmetric positive definite } \boldsymbol{\Sigma}(p\times p).
\end{aligned}
$$

The log marginal likelihood is (ignoring constants):

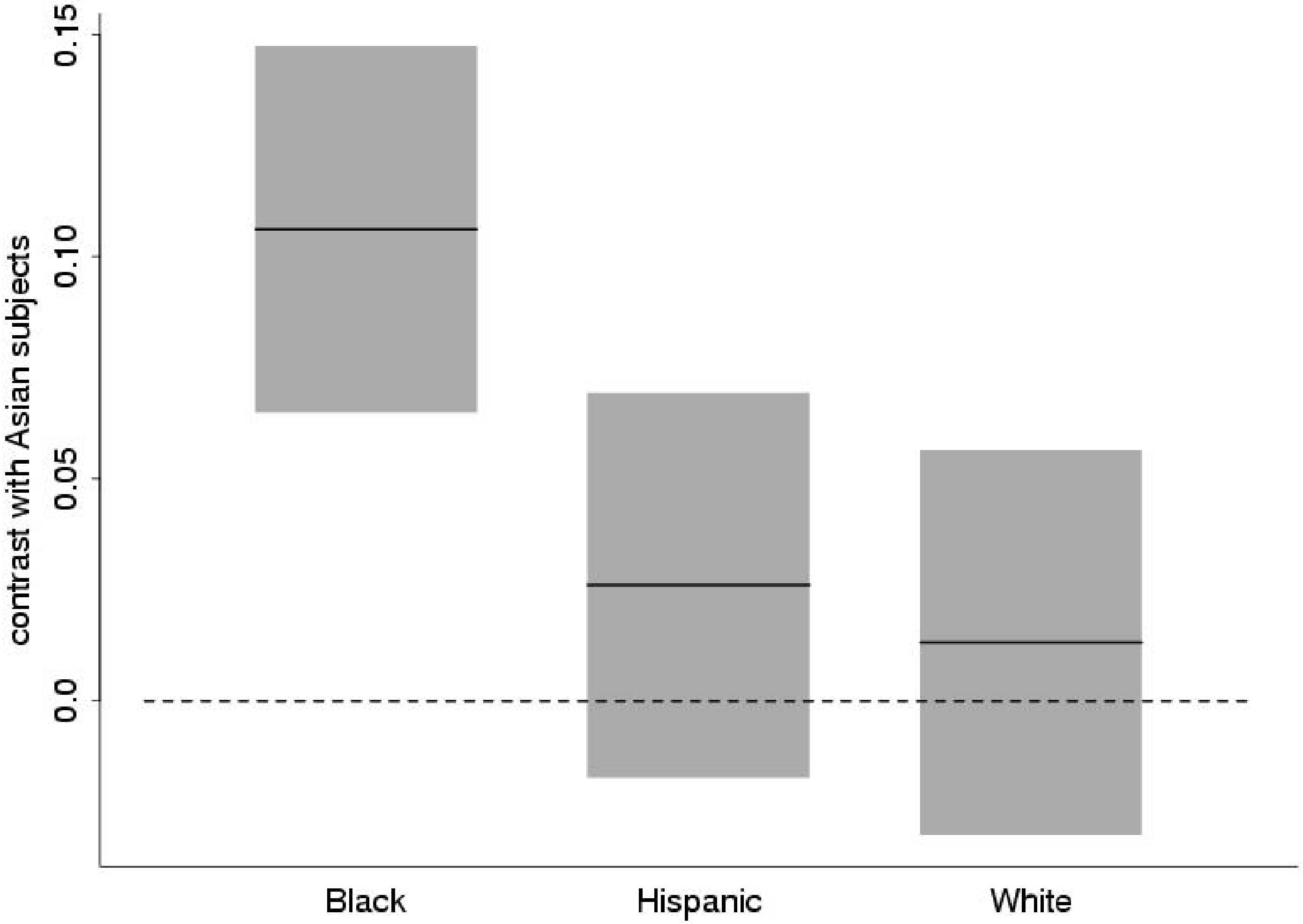
$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right] \\
&\geq \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left(\log\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]\right) \\
&= \boldsymbol{y}^T\boldsymbol{X}\boldsymbol{\mu} - \boldsymbol{1}^T\exp\{\boldsymbol{X}\boldsymbol{\mu} + \tfrac{1}{2}\text{diagonal}(\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^T)\} - \tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\qquad -\tfrac{1}{2}\{\text{tr}(\boldsymbol{F}^{-1}\boldsymbol{\Sigma}) + \log|\boldsymbol{\Sigma}|\} \\
&= \log \underline{p(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\Sigma})} \quad \text{for all } \boldsymbol{\mu}(p\times 1) \text{ and symmetric positive definite } \boldsymbol{\Sigma}(p\times p).
\end{aligned}
$$

A Bit About Semiparametric Regression


Cambridge Series in Statistical and Probabilistic Mathematics
Semiparametric Regression
David Ruppert, M. P. Wand,
and R. J. Carroll


Hispanic
White
Asian
Black
spinal bone mineral density (g/cm^2)
age in years
10
15
20
25
0.6
0.8
1.0
1.2
1.4

Approximate 95% Conf. Int. for Contrasts



Mixed Model Framework

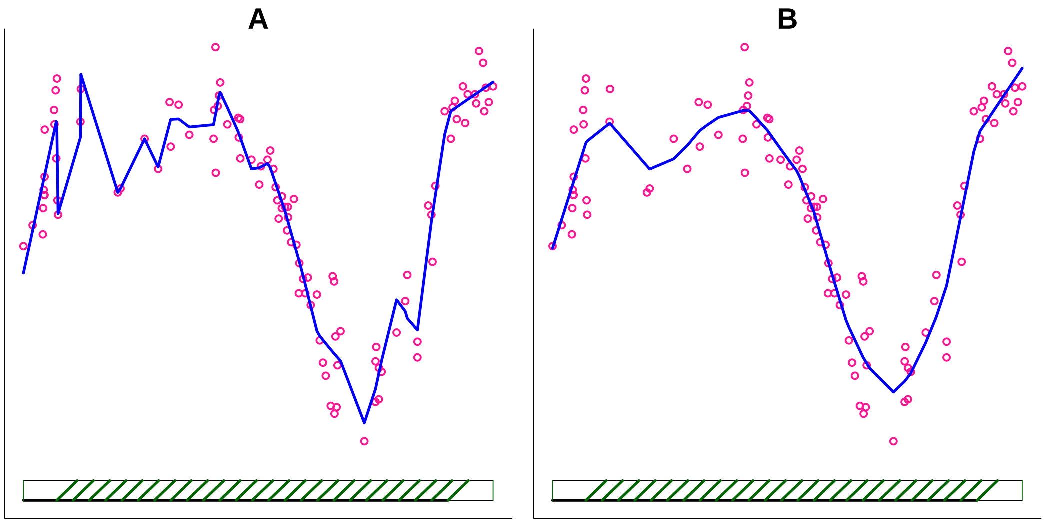
Previous analysis was done completely using
linear mixed models

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim \boldsymbol{N} \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \right)$$

Tricking Mixed Models to do Smoothing

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k (x_i - \kappa_k)_+ + \varepsilon_i$$



A: u_k's fixed

B: u_k i.i.d. $N(0, \sigma_u^2)$

Other Bases

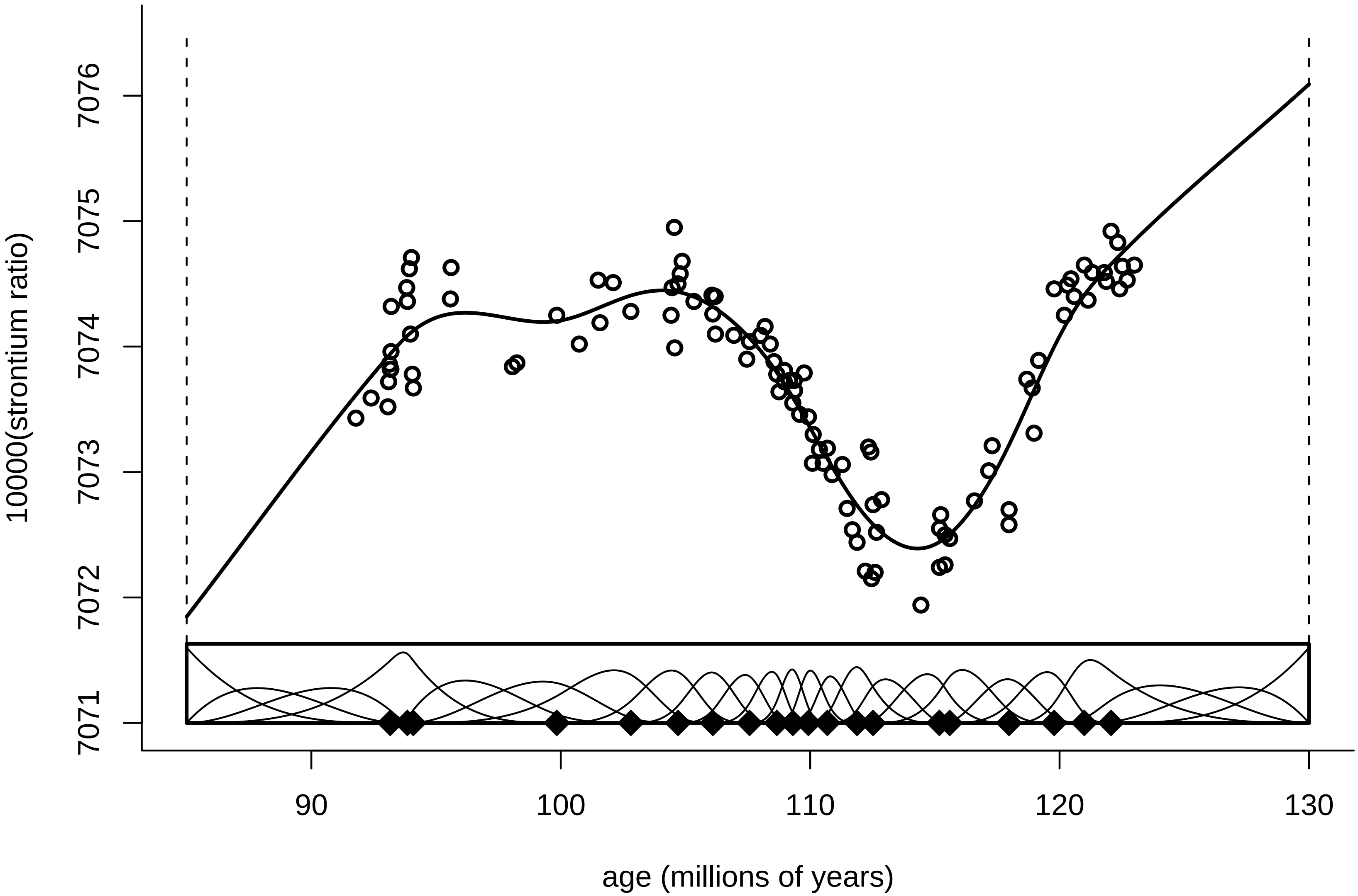
We often replace $(x - \kappa_k)_+$ by nicer $z_k(x)$:

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k z_k(x)$$

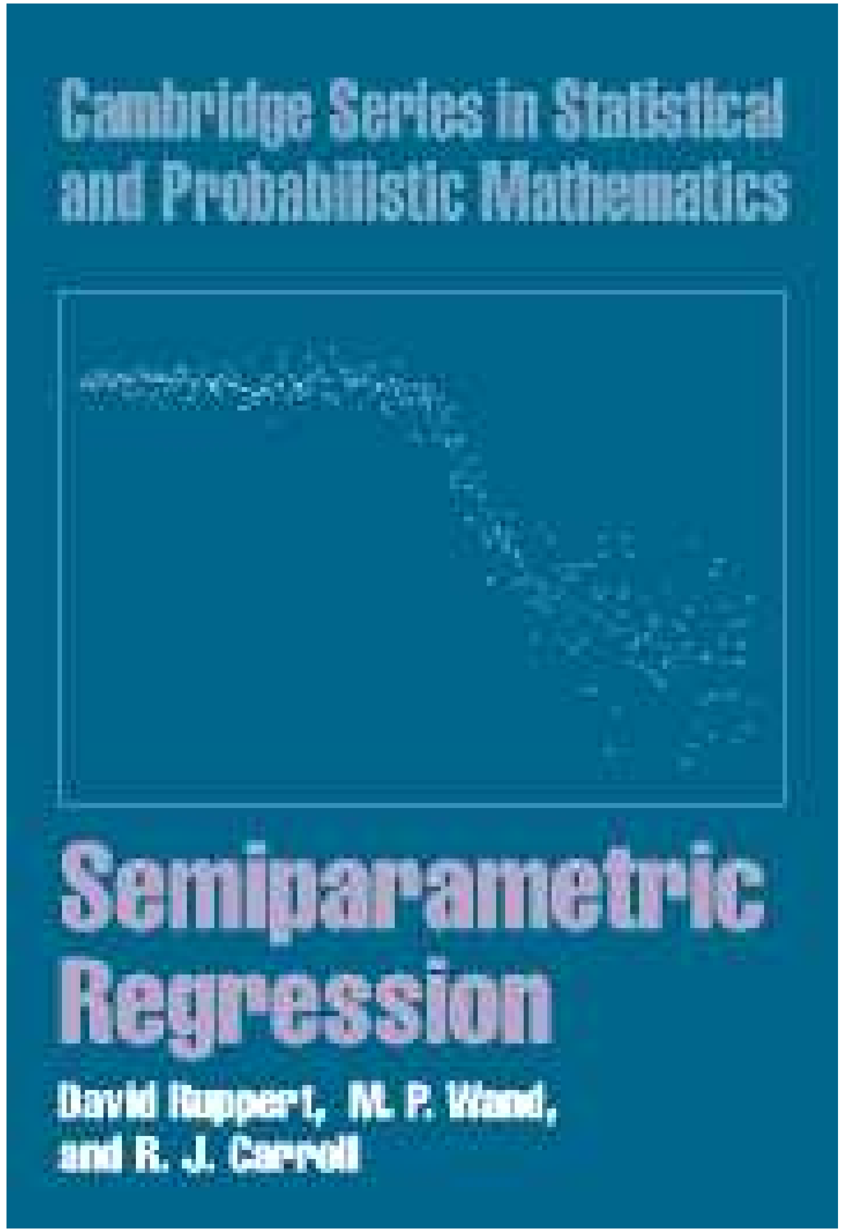
with

$$u_k \text{ i.i.d. } N(0, \sigma_u^2)$$

Particularly nice $z_k(x)$ are those arising from O'Sullivan Statist. Sci. (1986) (see e.g. Wand & Ormerod, Aust. N.Z. J. Statist., 2008).



10000(strontium ratio)
7071
7072
7073
7074
7075
7076
90
100
110
120
130
age (millions of years)


Cambridge Series in Statistical and Probabilistic Mathematics
Semiparametric Regression
David Ruppert, M. P. Wand,
and R. J. Carroll

Keep Updated!

Semiparametric Regression During 2003–2007.

D. RUPPERT, M.P. WAND & R.J. CARROLL

J. American Statist. Assoc. (under review)

Available now on Wand web-site!

Question

Is it possible to do an

entire semiparametric regression analysis

without touching the keyboard?

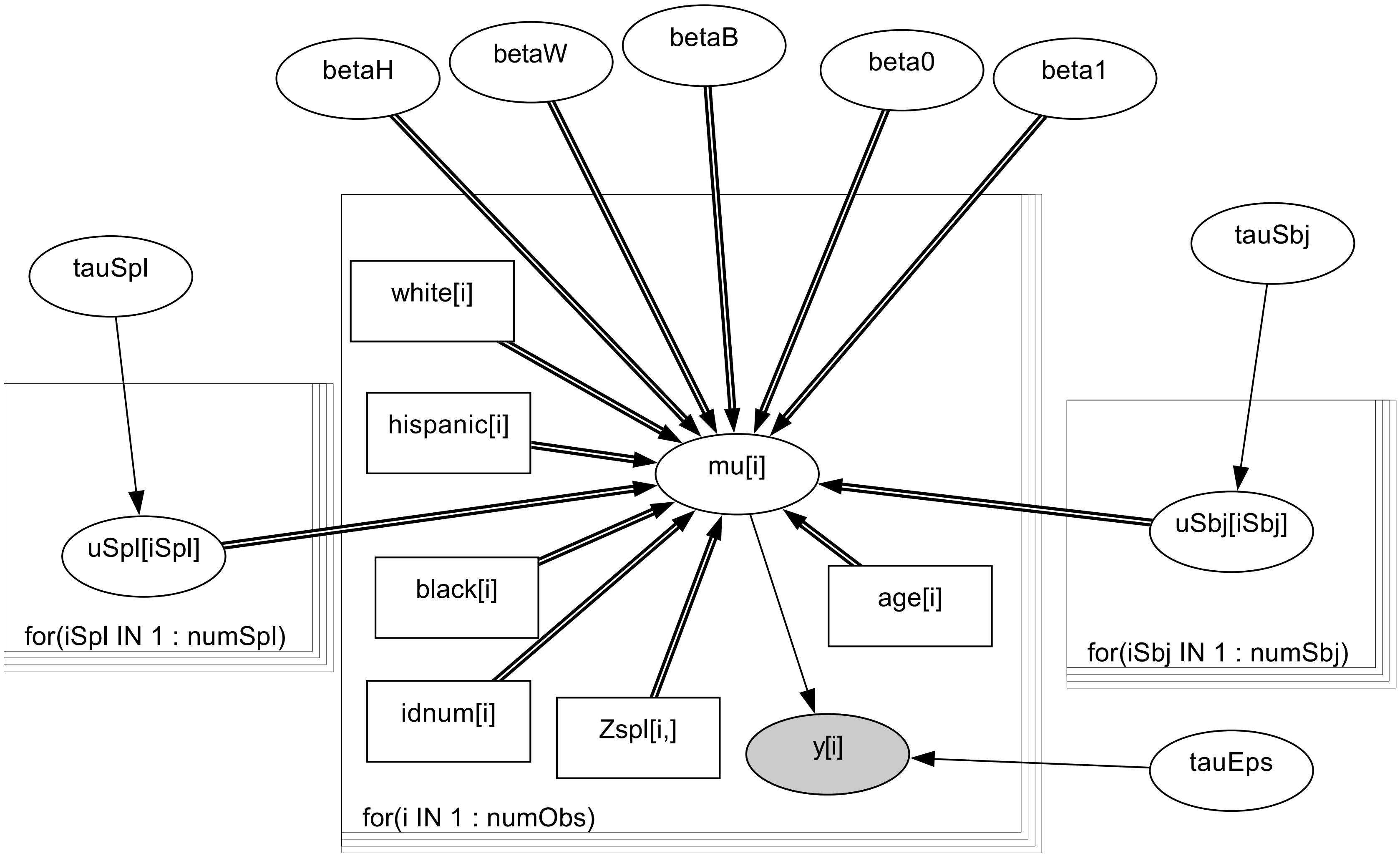
Answer

YES

via a graphical models approach

(and WinBUGS)

name:	y[i]	type:	stochastic	density:	dnorm		
mean	mu[i]	precision	tauEps	lower bound		upper bound	



The Graphical Models

viewpoint of

Semiparametric Regression

A recent trend in semiparametric regression is increased use of

hierarchical Bayesian modelling

Bayesian Hierarchical Model for Spinal Bone Mineral Data

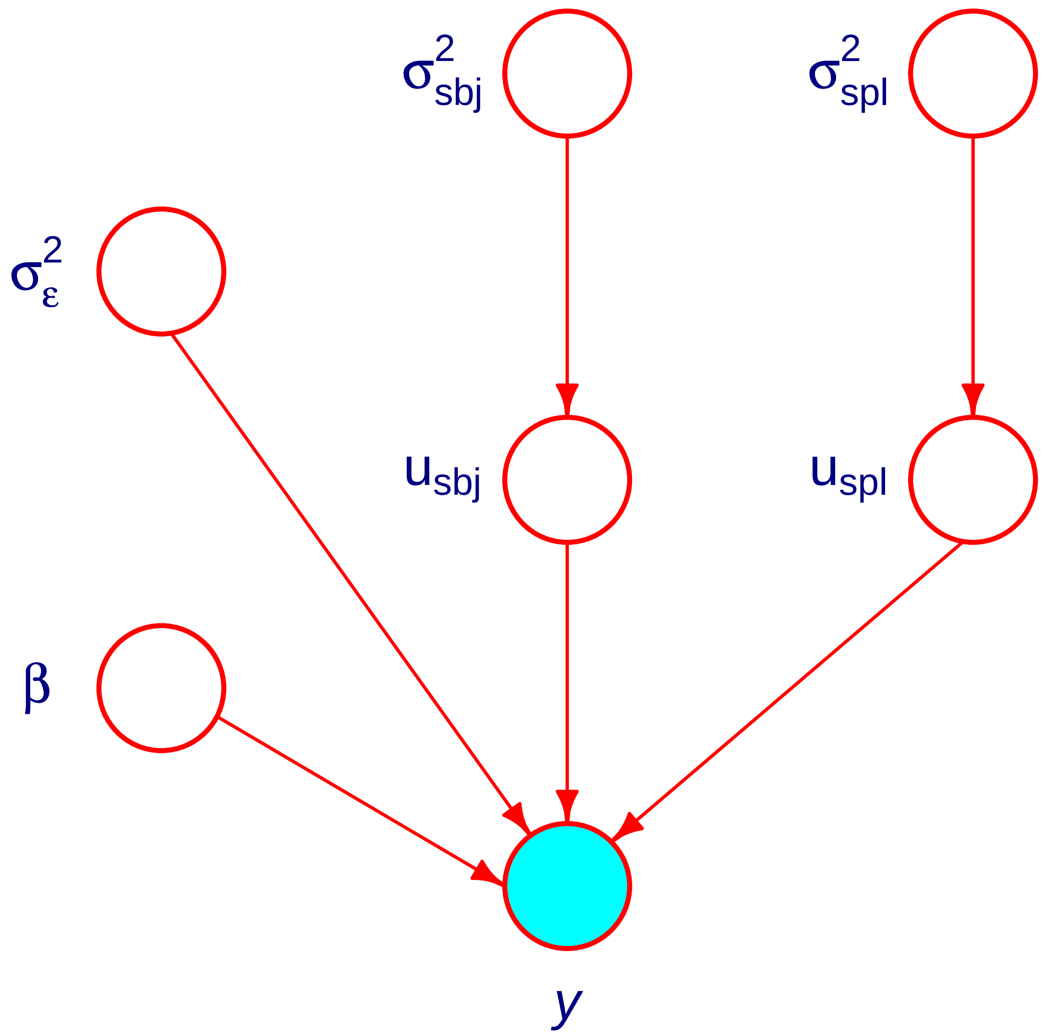
$$[y_{ij}|\beta, u_{\text{sbj}}, u_{\text{spl}}, \sigma^2_{\text{sbj}}, \sigma^2_{\text{spl}}, \sigma^2_{\varepsilon}] \overset{\text{ind.}}{\sim} N\left(\beta^T x_i + u_{i,\text{sbj}} + f(\text{age}_{ij}; \sigma^2_{\text{spl}}), \sigma^2_{\varepsilon}\right),$$

$$\left[u_{\text{sbj}}|\sigma^2_{\text{sbj}}\right] \sim N(0, \sigma^2_{\text{sbj}} I), \quad [u_{\text{spl}}|\sigma^2_{\text{spl}}] \sim N(0, \sigma^2_{\text{spl}} I),$$

$$[\beta] \sim N(0, \sigma^2_{\beta} I), \ \left[1/\sigma^2_{\text{sbj}}\right] \sim \text{Gamma}(A_{\text{sbj}}, B_{\text{sbj}}),$$

$$\left[1/\sigma^2_{\text{spl}}\right] \sim \text{Gamma}(A_{\text{spl}}, B_{\text{spl}}), \ \left[1/\sigma^2_{\varepsilon}\right] \sim \text{Gamma}(A_{\varepsilon}, B_{\varepsilon}).$$

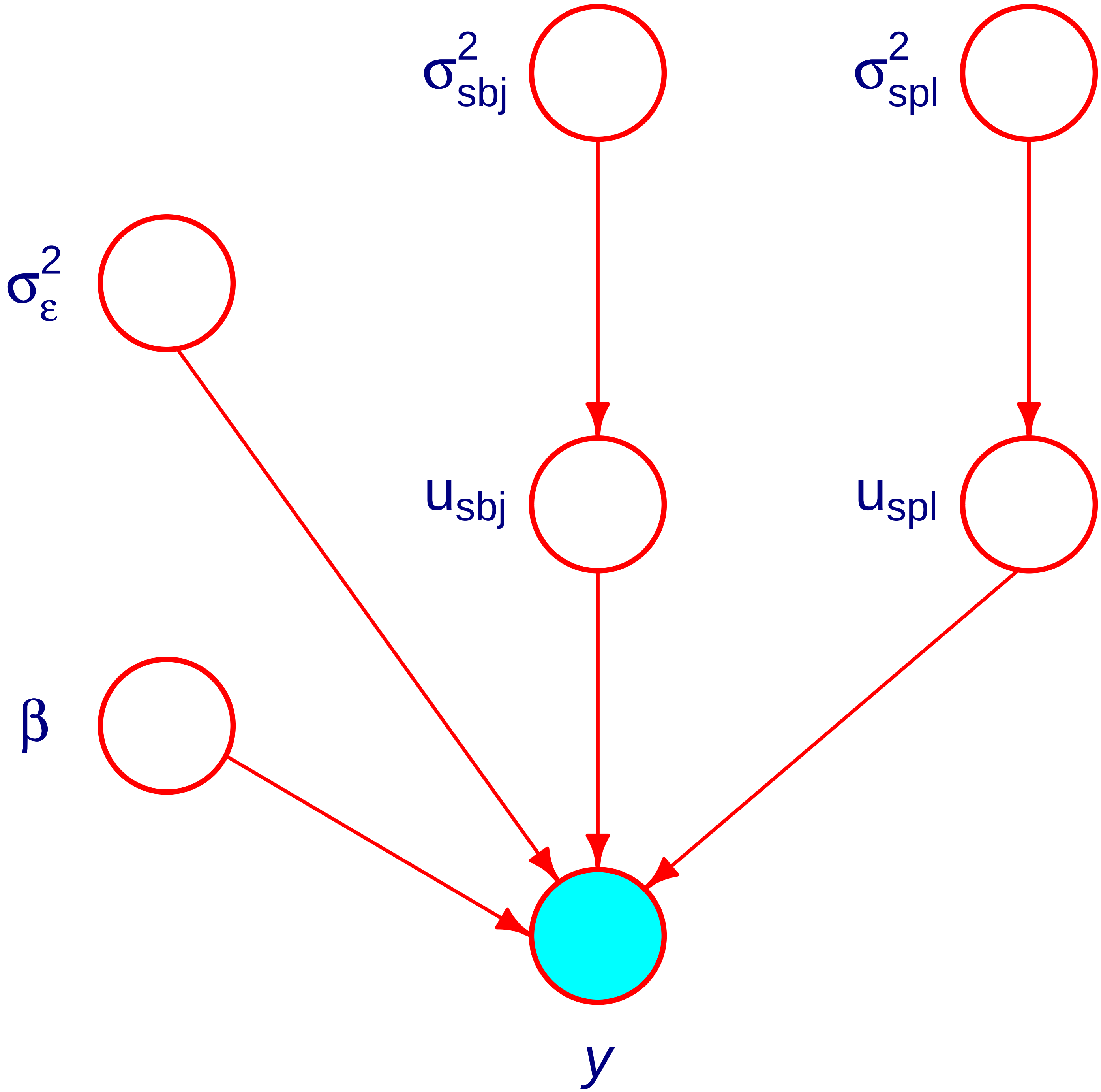
Directed Acyclic Graph (DAG) Representation



Inference Problem in Graphical Models Jargon

$$\mathcal{E} = \text{evidence nodes} = \{\boldsymbol{y}\}$$

$$\mathcal{H} = \text{hidden nodes} = \{\boldsymbol{\beta}, \boldsymbol{u}_{\text{sbj}}, \boldsymbol{u}_{\text{spl}}, \boldsymbol{\sigma}^2_{\text{sbj}}, \boldsymbol{\sigma}^2_{\text{spl}}, \boldsymbol{\sigma}^2_{\boldsymbol{\varepsilon}}\}$$



σ^2_{sbj}
σ^2_{spl}
σ^2_{ε}
u_{sbj}
u_{spl}
β
y

Probability Calculus Problem

$$p(\mathcal{H}|\mathcal{E}) = \frac{p(\mathcal{H}, \mathcal{E})}{p(\mathcal{E})}$$

Probability Calculus Problem

$$p(\mathcal{H}|\mathcal{E}) = \frac{p(\mathcal{H}, \mathcal{E})}{p(\mathcal{E})}$$

For current problem:

$$p(\beta, u_{\text{sbj}}, u_{\text{spl}}, \sigma^2_{\text{sbj}}, \sigma^2_{\text{spl}}, \sigma^2_{\varepsilon}|y) = \frac{p(\beta, u_{\text{sbj}}, u_{\text{spl}}, \sigma^2_{\text{sbj}}, \sigma^2_{\text{spl}}, \sigma^2_{\varepsilon}, y)}{p(y)}$$

The MCMC Solution

Most common method for solving probability calculus problems is

Monte Carlo Markov Chain (MCMC).

The MCMC Solution

Most common method for solving probability calculus problems is

Monte Carlo Markov Chain (MCMC).

Software packages **WinBUGS**

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

and **BRugs**

Ligges, U., Thomas, A., Spiegelhalter, D., Best, N., Lunn, D., Rice, K. & Sturtz, S. (2007). `BRugs 0.4`.

provide an effective means of fitting.

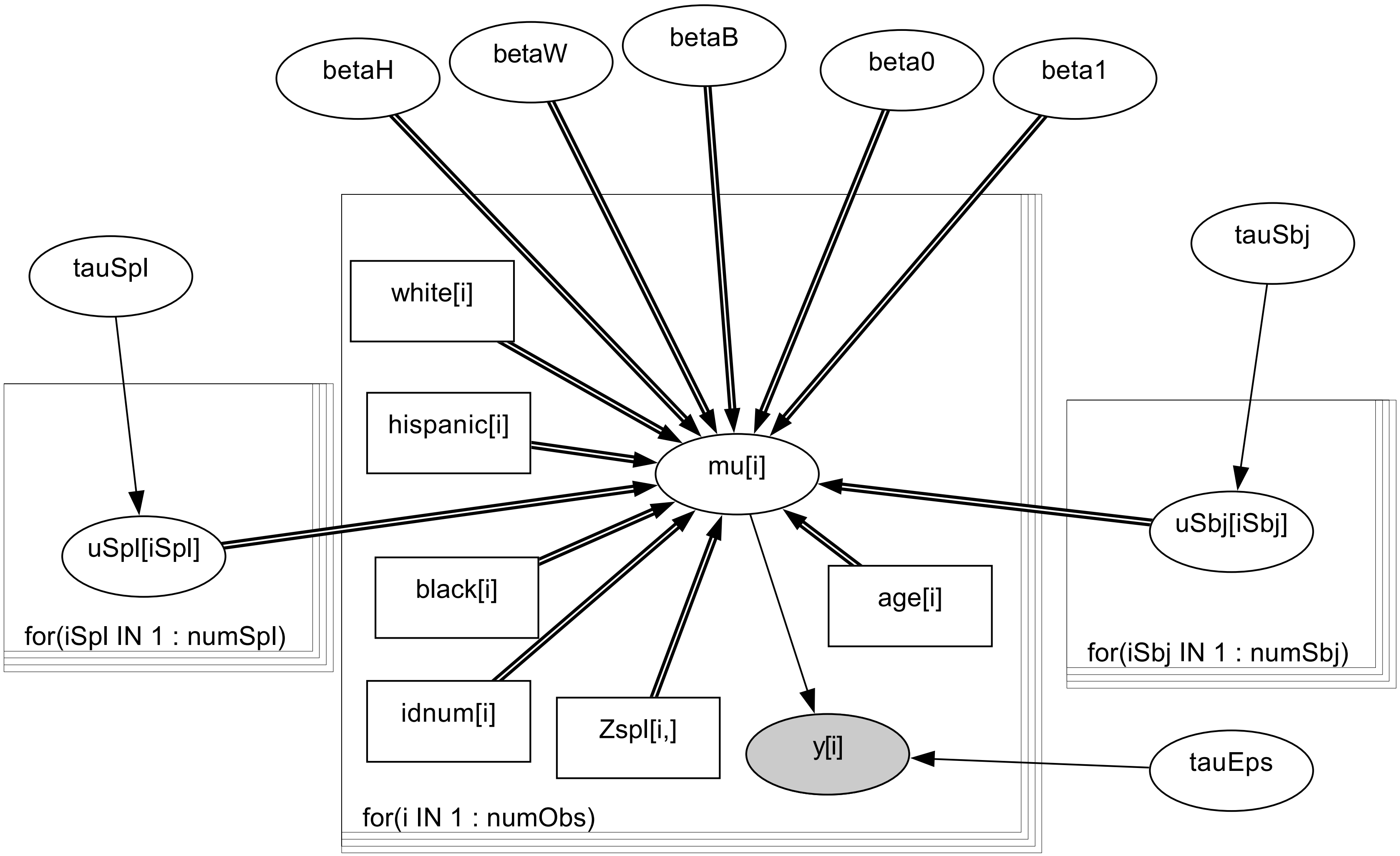
WinBUGS Code

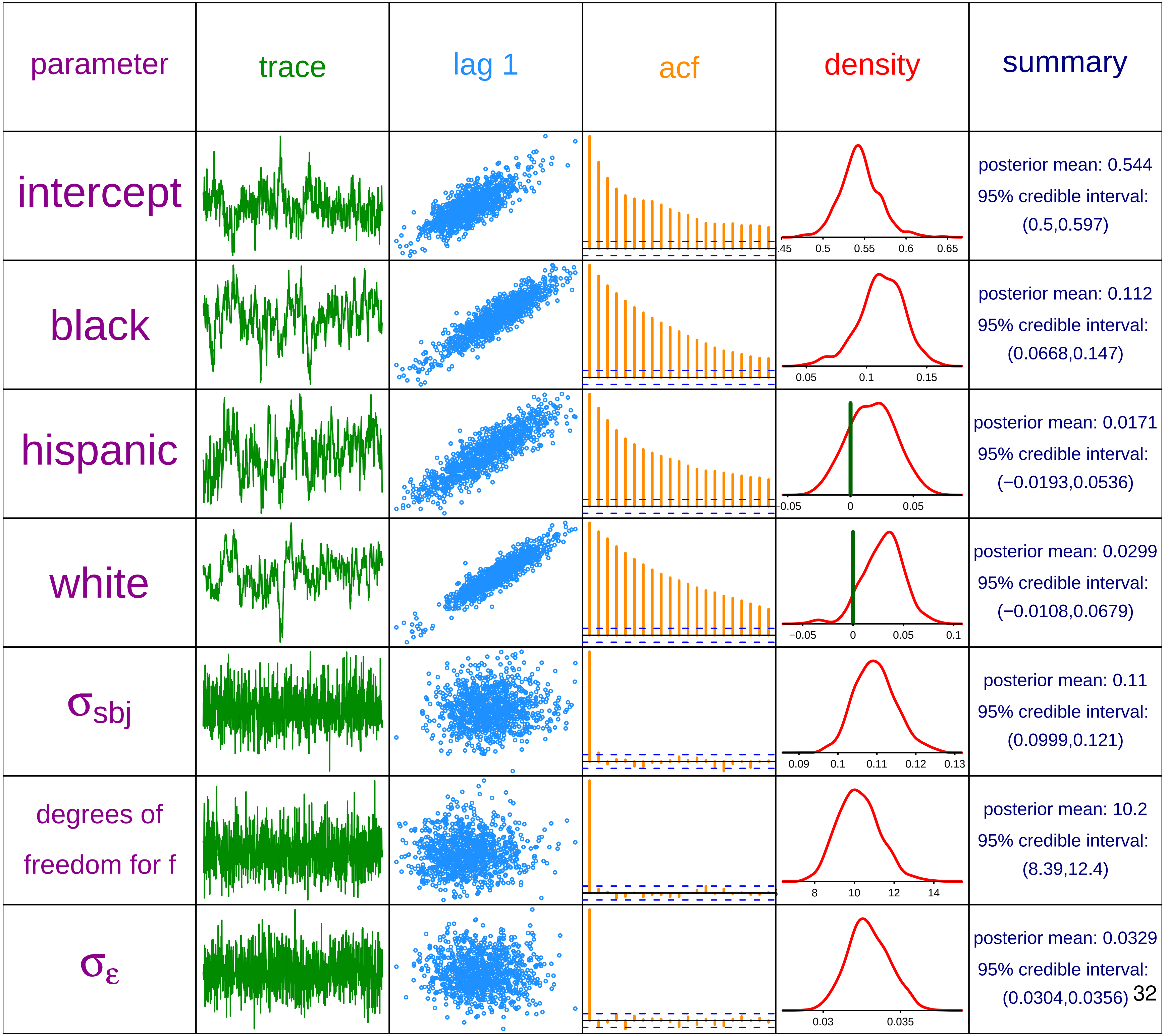
```
model
{
   for(i in 1:numObs)
   {
      mu[i] <- beta0 + uSbj[idnum[i]] + betaB*black[i] + betaH*hispanic[i]
               + betaW*white[i] + betaAge*sage[i] + inprod(uSpl[],Zspl[i,])
      sSBMD[i] ~ dnorm(mu[i],tauErr)
   }
   for (iSbj in 1:numSbj)
   {
       uSbj[iSbj] ~ dnorm(0,tauSbj)
   }
   for (iSpl in 1:numSpl)
   {
       uSpl[iSpl] ~ dnorm(0,tauSpl)
   }
   beta0  ~ dnorm(0,1.0E-8)   ; betaB  ~ dnorm(0,1.0E-8)
   betaH  ~ dnorm(0,1.0E-8)   ; betaW  ~ dnorm(0,1.0E-8)
   betaAge ~ dnorm(0,1.0E-8)  ; tauSbj ~ dgamma(0.01,0.01)
   tauSpl ~ dgamma(0.01,0.01) ; tauErr ~ dgamma(0.01,0.01)
   sigSbj <- 1/sqrt(tauSbj) ; sigSpl <- 1/sqrt(tauSpl)
   sigErr <- 1/sqrt(tauErr)
}
```

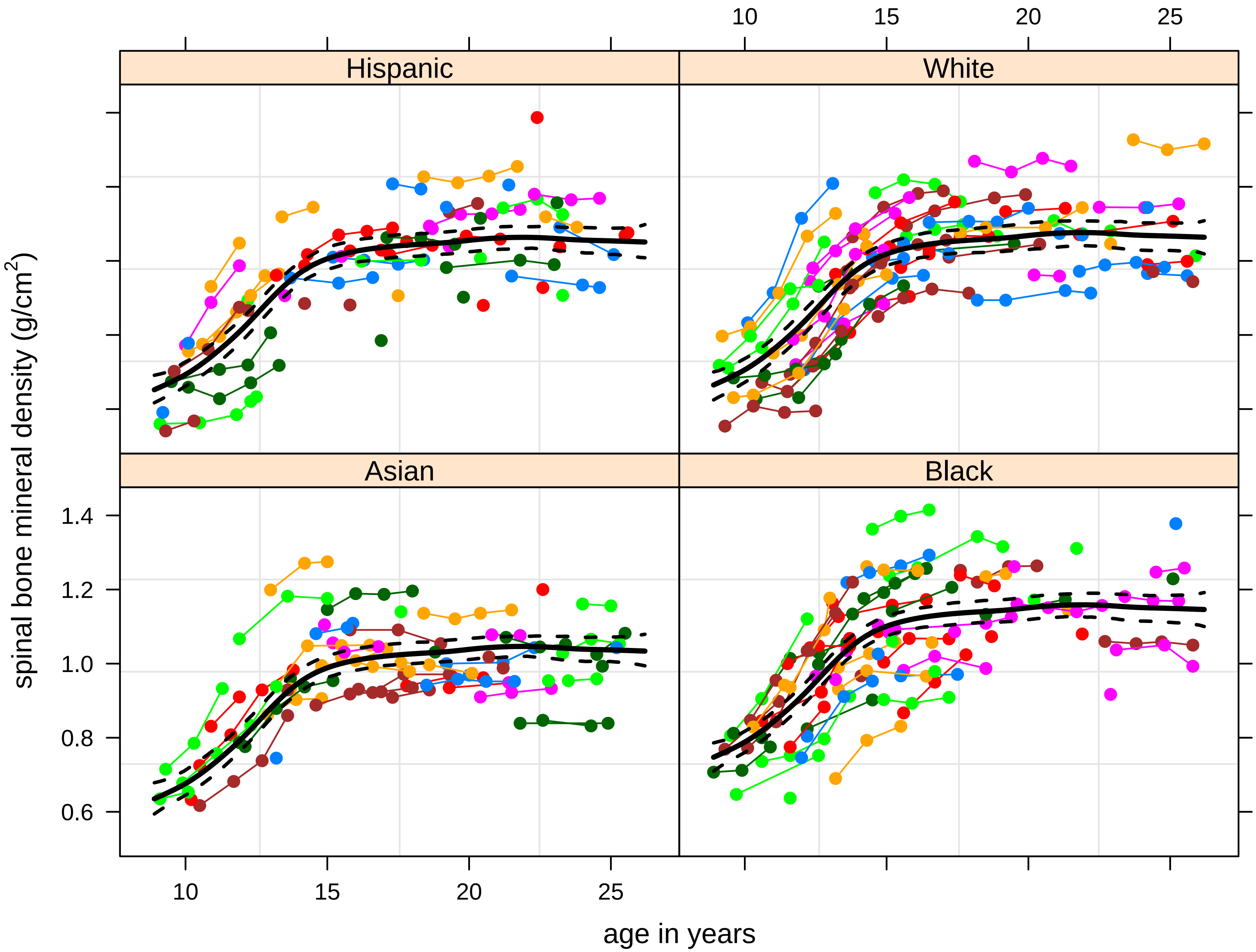
Alternatively, we can specify model in WinBUGS using its

graphical model drawing facility

name:	y[i]	type:	stochastic	density:	dnorm		
mean	mu[i]	precision	tauEps	lower bound		upper bound	




parameter
trace
lag 1
acf
density
summary
intercept
posterior mean: 0.544
95% credible interval:
(0.5,0.597)
black
posterior mean: 0.112
95% credible interval:
(0.0668,0.147)
hispanic
posterior mean: 0.0171
95% credible interval:
(−0.0193,0.0536)
white
posterior mean: 0.0299
95% credible interval:
(−0.0108,0.0679)
σ_{sbj}
posterior mean: 0.11
95% credible interval:
(0.0999,0.121)
degrees of freedom for f
posterior mean: 10.2
95% credible interval:
(8.39,12.4)
σ_{ε}
posterior mean: 0.0329
95% credible interval:
(0.0304,0.0356)



Hispanic
White
Asian
Black
spinal bone mineral density (g/cm^2)
age in years
10
15
20
25
0.6
0.8
1.0
1.2
1.4

Non-standard Semiparametric Regression

Graphical models approach to semiparametric regression is

more advantageous

when situation is non-standard.

Examples:

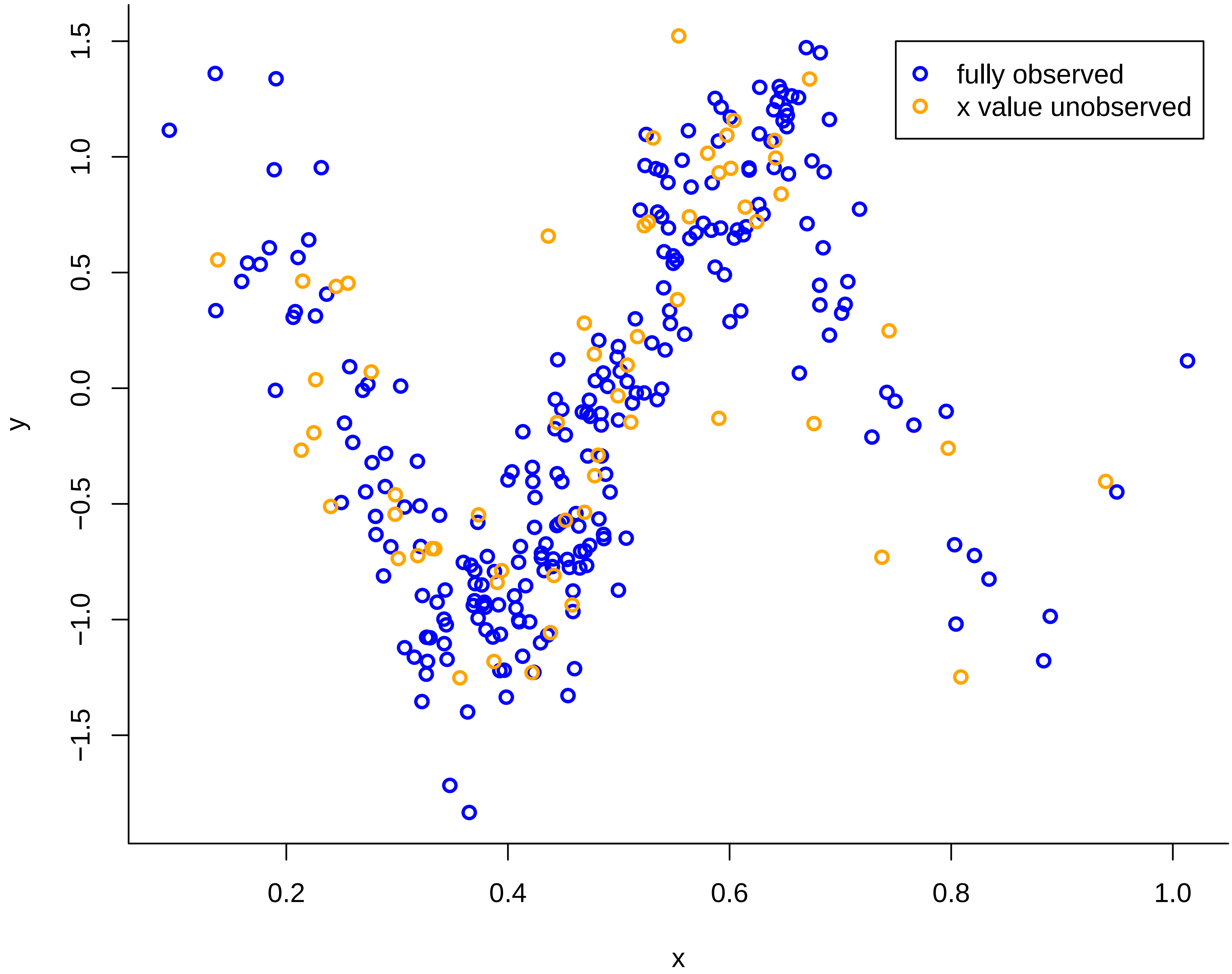
- Missing data.
- Measurement error.

Nonparametric Regression with Missingness in Predictor

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \text{ i.i.d. } N(0, \sigma_\varepsilon^2), \quad 1 \leq i \leq n$$

$$x_i \overset{\text{ind.}}{\sim} N(\mu_x, \sigma_x^2),$$ but some are missing

(completely at random).


fully observed
x value unobserved
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
y
0.2
0.4
0.6
0.8
1.0
x

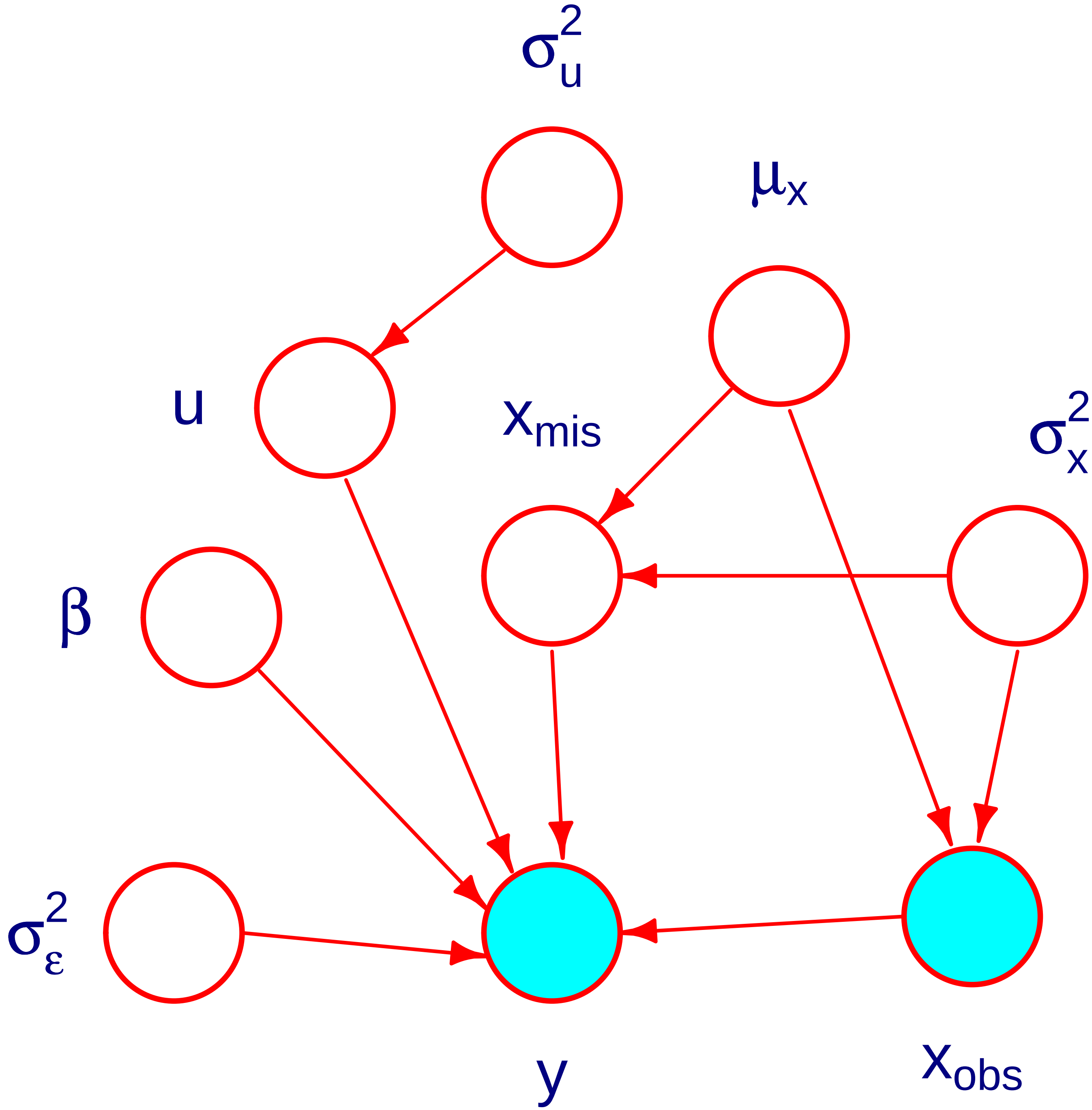
Hierarchical Bayes Model for Missingness Example

$$[y_i|x_i,\beta,u,\sigma_\varepsilon^2] \stackrel{\text{ind.}}{\sim} N\left(\beta_0+\beta_1 x_i+\sum_{k=1}^{K} u_k z_k(x_i),\sigma_\varepsilon^2\right),$$

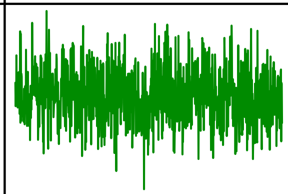
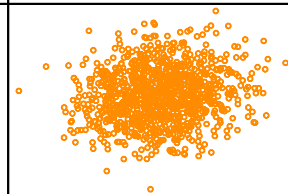
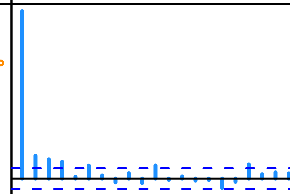
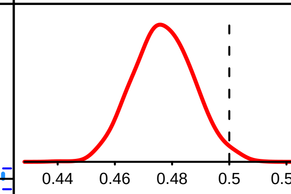
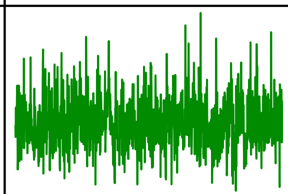
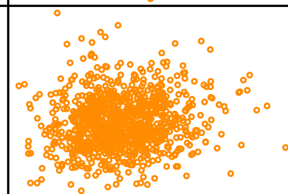
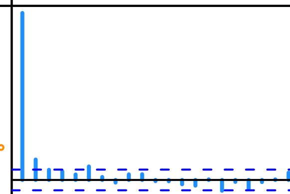
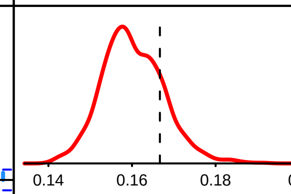
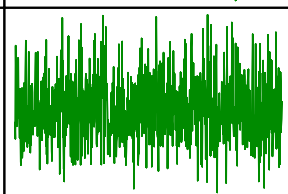
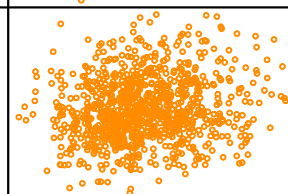
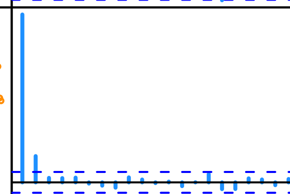
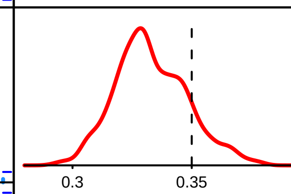
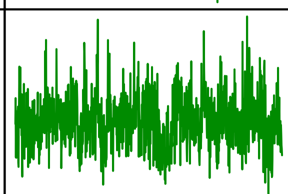
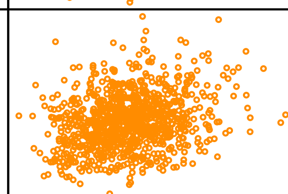
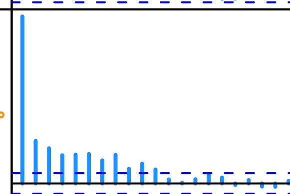
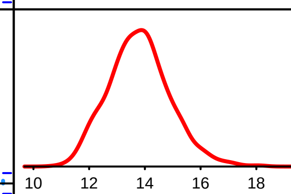
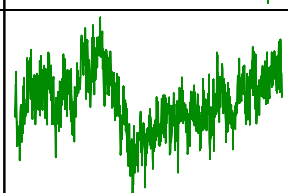
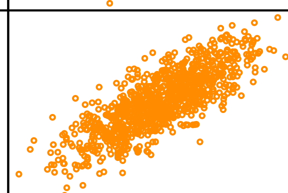
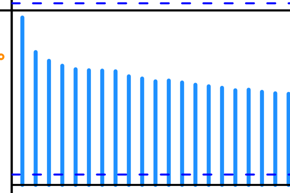
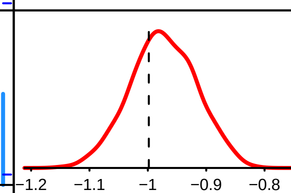
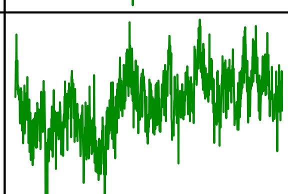
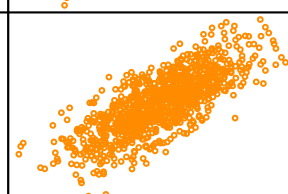
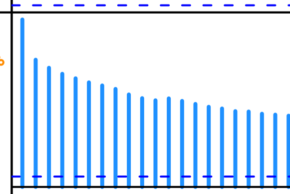
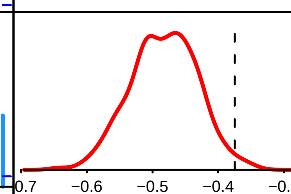
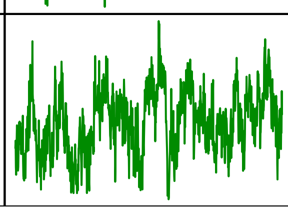
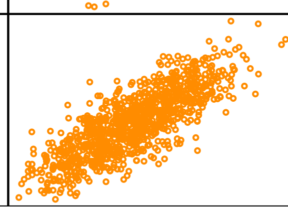
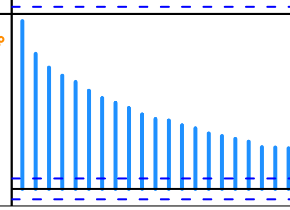
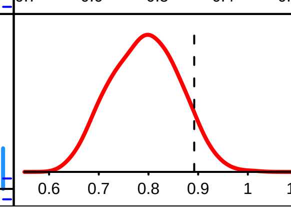
$$\left[u|\sigma_u^2\right] \sim N(0,\sigma_u^2 I), \quad [x_i|\mu_x,\sigma_x^2] \stackrel{\text{ind.}}{\sim} N(\mu_x,\sigma_x^2),$$

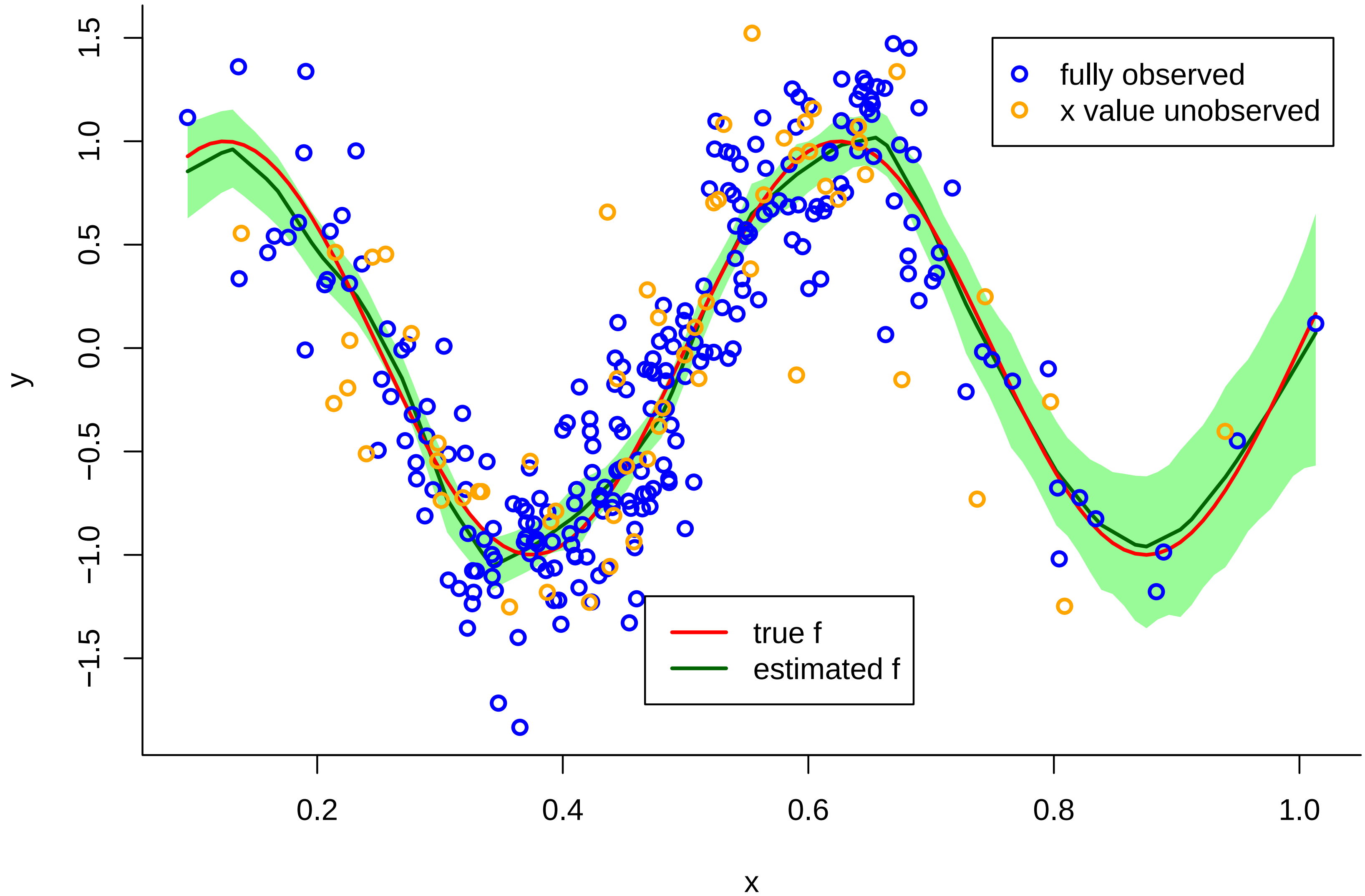
$$[\beta] \sim N(0,\sigma_\beta^2 I), \quad [\mu_x] \sim N(0,\sigma_{\mu_x}^2),$$

$$\left[\sigma_u^2\right] \sim \text{IG}(A_u,B_u), \quad \left[\sigma_\varepsilon^2\right] \sim \text{IG}(A_\varepsilon,B_\varepsilon), \quad [\sigma_x^2] \sim \text{IG}(A_x,B_x).$$

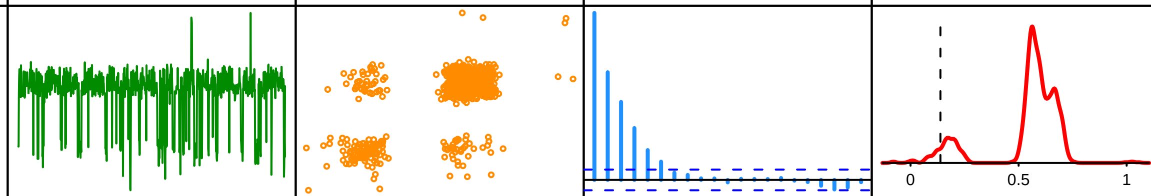
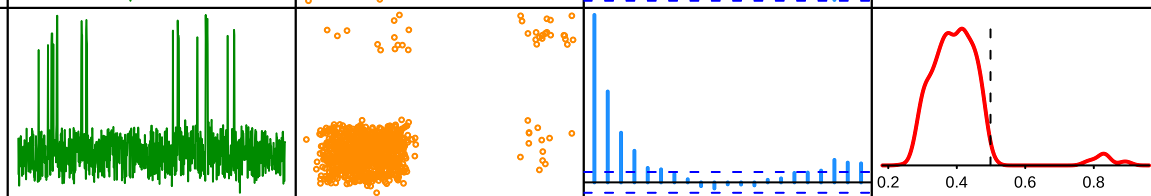
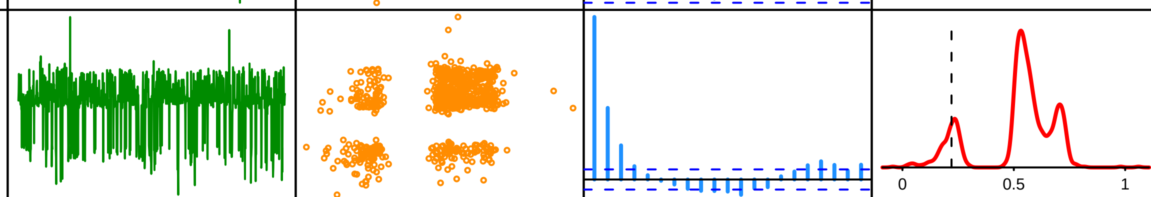
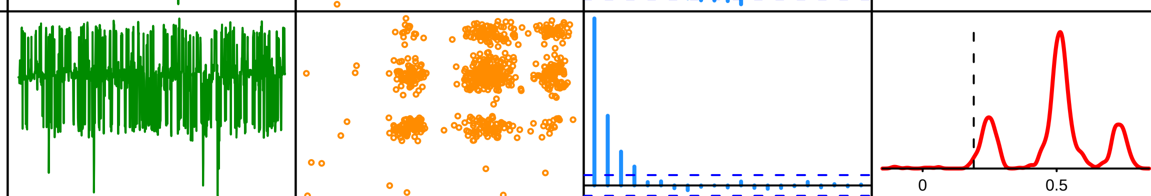
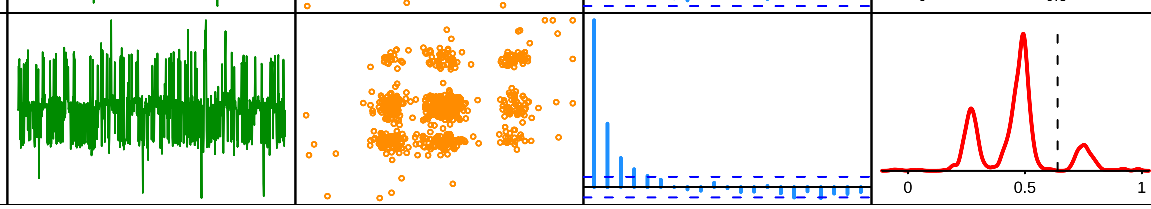


σ^2_u
μ_x
u
x_{mis}
σ^2_x
β
σ^2_ε
y
x_{obs}

parameter	trace	lag 1	acf	density	summary
μ_x					posterior mean: 0.477 95% credible interval: (0.457,0.497)
σ_x					posterior mean: 0.16 95% credible interval: (0.147,0.176)
σ_ε					posterior mean: 0.333 95% credible interval: (0.306,0.366)
degrees of freedom for f					posterior mean: 13.8 95% credible interval: (11.8,16.1)
first quartile of x					posterior mean: −0.973 95% credible interval: (−1.08,−0.863)
second quart. of x					posterior mean: −0.483 95% credible interval: (−0.583,−0.386)
third quartile of x					posterior mean: 0.793 95% credible interval: (0.668,0.922)



fully observed
x value unobserved
true f
estimated f
y
x
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.2
0.4
0.6
0.8
1.0

parameter	trace	lag 1	acf	density	summary
x^{mis}_{10}					posterior mean: 0.537 95% credible interval: (0.118,0.709)
x^{mis}_{18}					posterior mean: 0.406 95% credible interval: (0.288,0.816)
x^{mis}_{27}					posterior mean: 0.515 95% credible interval: (0.148,0.734)
x^{mis}_{44}					posterior mean: 0.492 95% credible interval: (0.204,0.762)
x^{mis}_{59}					posterior mean: 0.462 95% credible interval: (0.231,0.788)

References for last segment...

SEMIPARAMETRIC REGRESSION
AND GRAPHICAL MODELS

Wand, M.P. (2009) Aust. N.Z. J. Statist. (invited)

References for last segment...

SEMIPARAMETRIC REGRESSION
AND GRAPHICAL MODELS

Wand, M.P. (2009) Aust. N.Z. J. Statist. (invited)

NON-STANDARD SEMIPARAMETRIC REGRESSION VIA BRUGS

Marley, J.K. and Wand, M.P. (2009) unpublished manuscript

(both on Wand web-site)

Summary of Talk so Far

Summary of Talk so Far

- Semiparametric regression flexible and powerful body of methodology.

Summary of Talk so Far

- Semiparametric regression flexible and powerful body of methodology.
- Hierarchical Bayesian models and directed acyclic graphs (DAGs) effective general approach to fitting and inference (esp. if situation is non-standard).

Summary of Talk so Far

- Semiparametric regression flexible and powerful body of methodology.
- Hierarchical Bayesian models and directed acyclic graphs (DAGs) effective general approach to fitting and inference (esp. if situation is non-standard).
- Markov Chain Monte Carlo (MCMC) and software packages WinBUGS and BRugs facilitate fitting and inference.

Summary of Talk so Far

- Semiparametric regression flexible and powerful body of methodology.
- Hierarchical Bayesian models and directed acyclic graphs (DAGs) effective general approach to fitting and inference (esp. if situation is non-standard).
- Markov Chain Monte Carlo (MCMC) and software packages WinBUGS and BRugs facilitate fitting and inference.
- Main drawback of MCMC: SLOWNESS!!

Possibly faster alternate approach, (mainly) from Computer Science, is:

Variational Approximation.

Possibly faster alternate approach, (mainly) from Computer Science, is:

Variational Approximation.

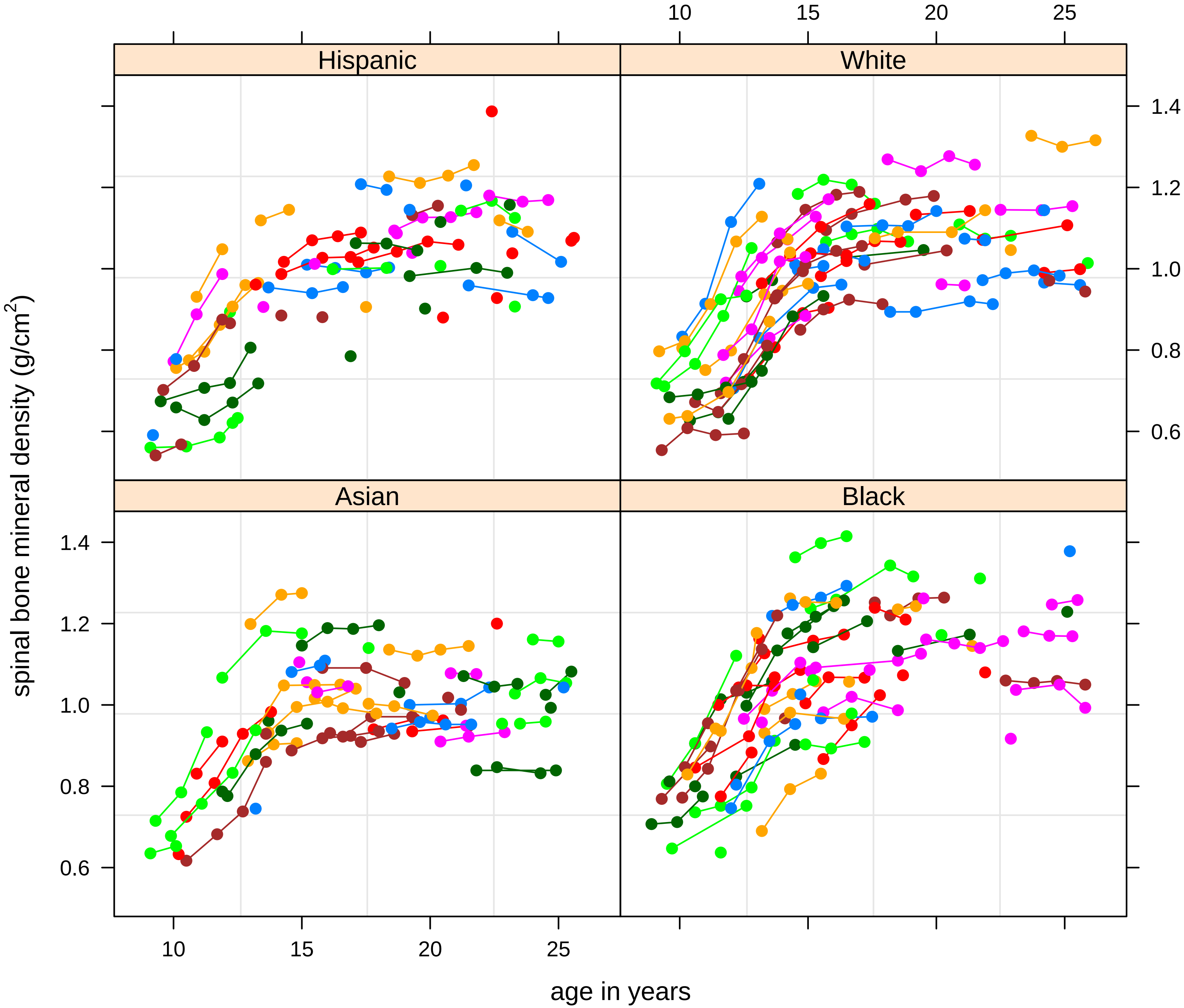
These have led (quite recently!) to:

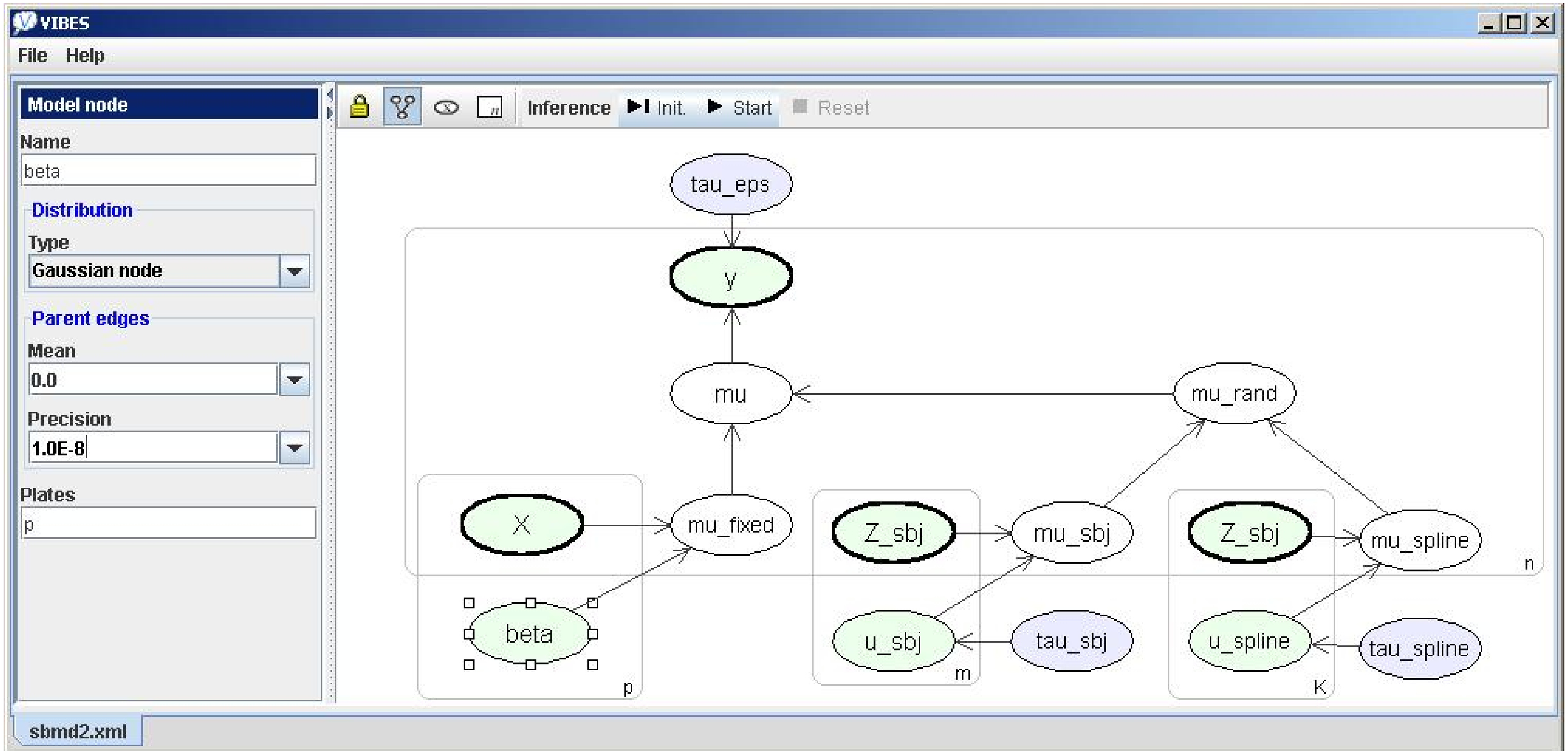
Variational Inference Engines.

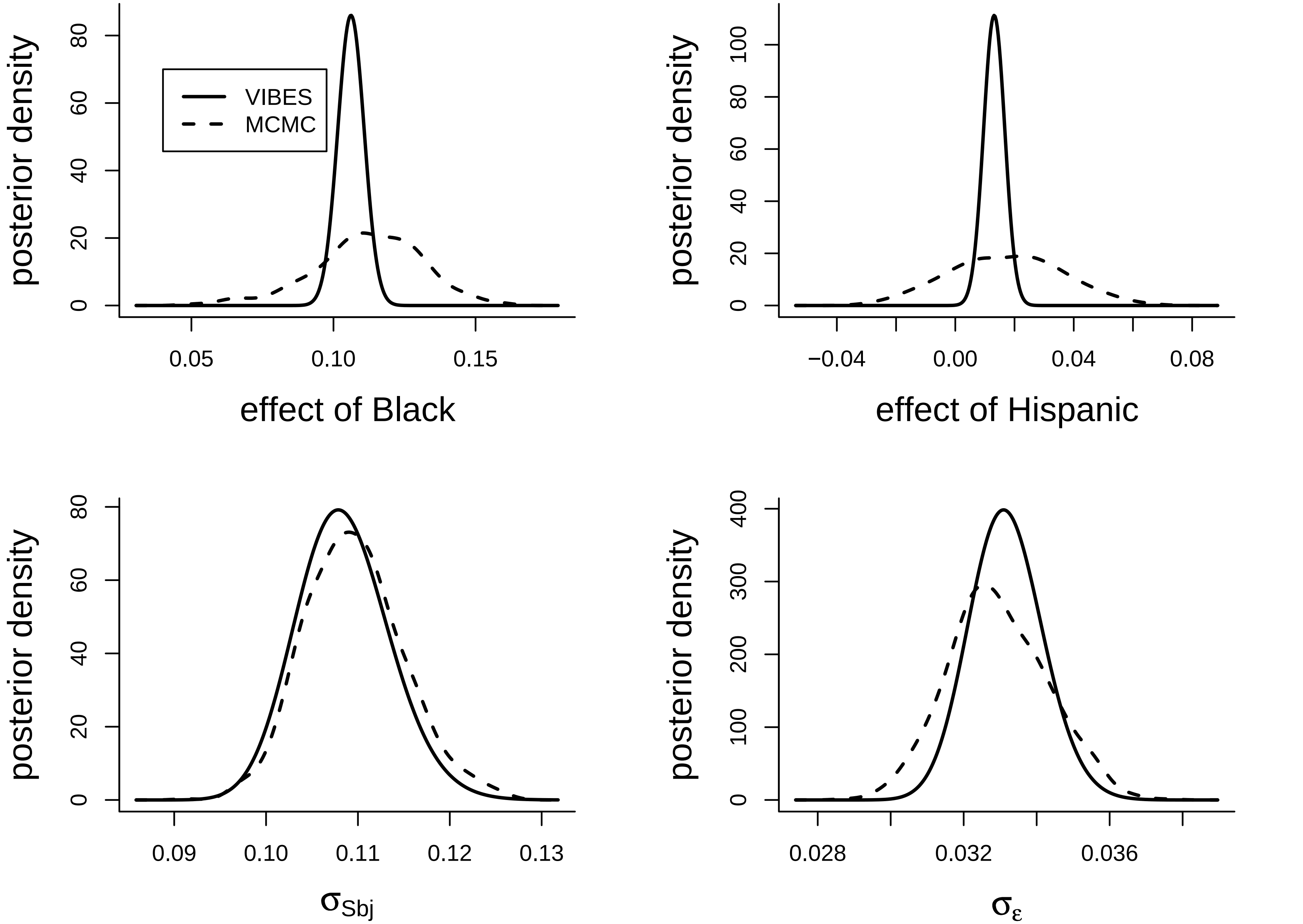
Prototype Package for Variational Inference

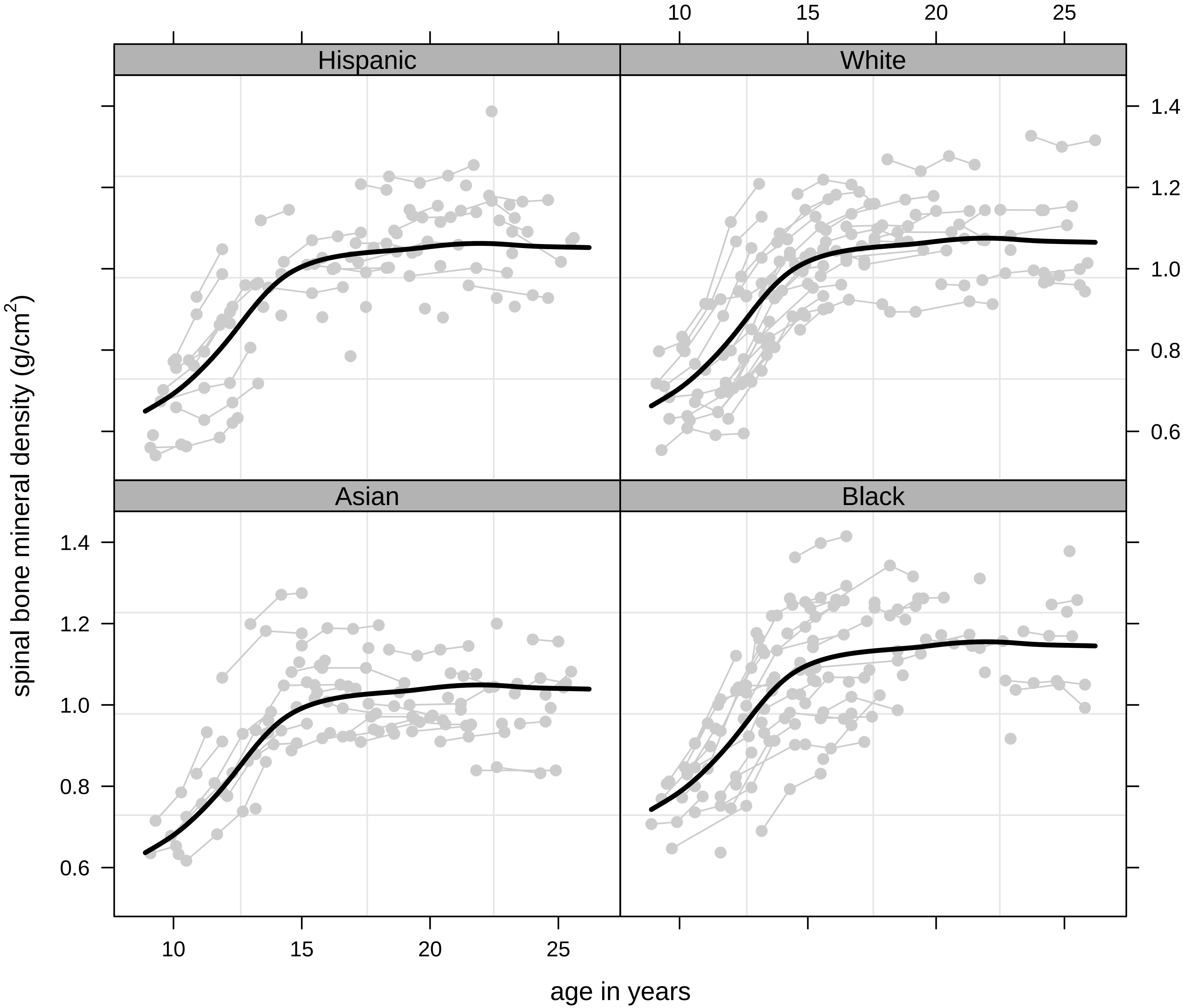
VIBES: A VARIATIONAL INFERENCE ENGINE FOR BAYESIAN NETWORKS

by Bishop, Spiegelhalter & Winn (2002)
NIPs Proceedings


Hispanic
White
Asian
Black
spinal bone mineral density (g/cm²)
age in years
10
15
20
25
0.6
0.8
1.0
1.2
1.4


VIBES
File Help
Model node
Name
beta
Distribution
Type
Gaussian node
Parent edges
Mean
0.0
Precision
1.0E-8
Plates
p
Inference
Init.
Start
Reset
tau_eps
y
mu
mu_rand
X
mu_fixed
Z_sbj
mu_sbj
Z_sbj
mu_spline
n
beta
u_sbj
tau_sbj
u_spline
tau_spline
p
m
K
sbmd2.xml


posterior density
VIBES
MCMC
effect of Black
effect of Hispanic
σ_{Sbj}
σ_{ε}
0.05
0.10
0.15
−0.04
0.00
0.04
0.08
0.09
0.10
0.11
0.12
0.13
0.028
0.032
0.036


Hispanic
White
Asian
Black
spinal bone mineral density (g/cm²)
age in years
10
15
20
25
0.6
0.8
1.0
1.2
1.4

Beyond VIBES

The developers of VIBES (Cambridge, UK) have just released (only 32 days ago!) a

new and improved variational inference engine

named

Infer.NET

(research.microsoft.com/infernet)

These Computer Science guys now even put their

conference talks on the web...

videolectures.net/abi07_winn_ipi

Variational Approximation Research ‘Schools’

location	key researchers
Berkeley, USA	Jordan, Jaakkola (now MIT),...
Cambridge, UK	MacKay, Bishop, Ghahramani, Winn, Minka,...
Glasgow, UK	Titterington, Wang,...

Variational Approximation Research 'Schools'

location	key researchers
Berkeley, USA	Jordan, Jaakkola (now MIT),...
Cambridge, UK	MacKay, Bishop, Ghahramani, Winn, Minka,...
Glasgow, UK	Titterington, Wang,...
Wollongong, Australia	Ormerod, Wand

Illustration of Berkeley for Simple Problem: Bayesian Logistic Regression

$$\text{logit}\{P(y_i = 1)\} = \beta_0 + \beta_1 x_i, \quad 1 \le i \le n; \;\; \beta_0, \beta_1 \sim N(0, 10^8 I).$$

Illustration of Berkeley for Simple Problem: Bayesian Logistic Regression

$$\text{logit}\{P(y_i = 1)\} = \beta_0 + \beta_1 x_i, \quad 1 \le i \le n; \;\; \beta_0, \beta_1 \sim N(0, 10^8 I).$$

Posterior for slope $p(\beta_1 | y)$ depends on intractable integral:

$$\int_{-\infty}^{\infty} \exp\{\beta_0 \mathbf{1}^T y - \mathbf{1}^T b(\beta_0 \mathbf{1} + \beta_1 x) - \beta_0^2/(2 \times 10^8)\} \, d\beta_0$$

where $b(x) = \log(1 + e^x)$

The Variational Approximation Trick

Write $-b(x)$ variationally:

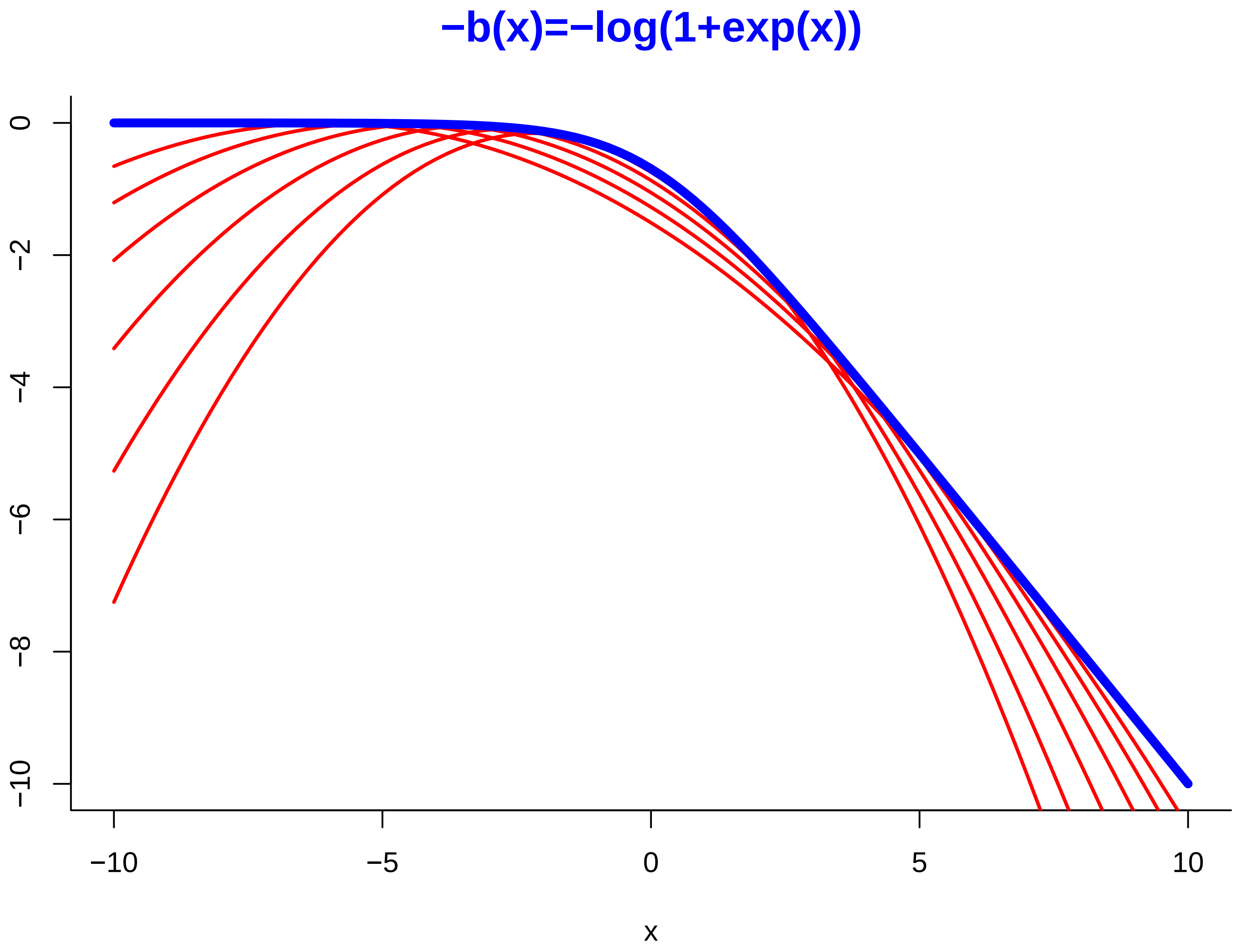
$$-b(x) = -\log(1 + e^x) = \max_{\xi}\{A(\xi)x^2 + B(\xi)x + C(\xi)\}$$

The Variational Approximation Trick

Write $-b(x)$ variationally:

$$-b(x) = -\log(1+e^x) = \max_{\xi}\{A(\xi)x^2 + B(\xi)x + C(\xi)\}$$

$$\begin{aligned} A(\xi) &= -\tanh(\xi/2)/(4\xi) \\ B(\xi) &= -1/2 \\ C(\xi) &= \xi/2 - \log(1+e^{\xi}) + \xi\tanh(\xi/2)/4 \end{aligned}$$



−b(x)=−log(1+exp(x))
0
−2
−4
−6
−8
−10
−10
−5
0
5
10
x

Family of Variational (Approximate) Solutions

$$[\beta_1|y;\xi] \sim N(\mu(\xi), \sigma^2(\xi))$$

$$\mu(\xi) = \frac{(2n\overline{\lambda}(\xi) + 10^{-8})(x^T y - \overline{x}/2)}{(2n\overline{\lambda}(\xi) + 10^{-8})\{2(x^2)^T \lambda(\xi) + 10^{-8}\} - 4\{\lambda(\xi)^T x\}}$$

$$\sigma^2(\xi) = [2(x^2)^T \lambda(\xi) + 10^{-8} - 4\{\lambda(\xi)^T x\}^2/\{2n\overline{\lambda}(\xi) + 10^{-8}\}]^{-1}$$

where $\quad \lambda(\xi) = \tanh(\xi/2)/(4\xi).$

Choice of Variational Parameters

Choice of

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$$

can be made via an

EM argument.

Reference: Jaakkola & Jordan, Statistics and Computing, 2000.

Full Algorithm

Let $[\boldsymbol{\beta_0}, \boldsymbol{\beta_1} | \boldsymbol{y}; \boldsymbol{\xi}] \sim \boldsymbol{N}(\boldsymbol{\mu}(\boldsymbol{\xi}), \boldsymbol{\Sigma}(\boldsymbol{\xi}))$ be var. approx. to $[\boldsymbol{\beta_0}, \boldsymbol{\beta_1} | \boldsymbol{y}]$.

CYCLE:

1. $\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1} \leftarrow \mathbf{10}^{-8}\boldsymbol{I} + \mathbf{2}\boldsymbol{X}^T \text{diag}\{\boldsymbol{\lambda}(\boldsymbol{\xi})\}\boldsymbol{X}$
2. $\boldsymbol{\mu}(\boldsymbol{\xi}) \leftarrow \boldsymbol{\Sigma}(\boldsymbol{\xi})\boldsymbol{X}^T(\boldsymbol{y} - \tfrac{\mathbf{1}}{\mathbf{2}}\mathbf{1})$
3. $\boldsymbol{\xi} \leftarrow \sqrt{\text{diagonal}[\boldsymbol{X}\{\boldsymbol{\Sigma}(\boldsymbol{\xi}) + \boldsymbol{\mu}(\boldsymbol{\xi})\boldsymbol{\mu}(\boldsymbol{\xi})^T\}\boldsymbol{X}^T]}$

We have recently developed

alternative variational approximation methods

that give promising results.

We have recently developed

alternative variational approximation methods

that give promising results.

I will call these

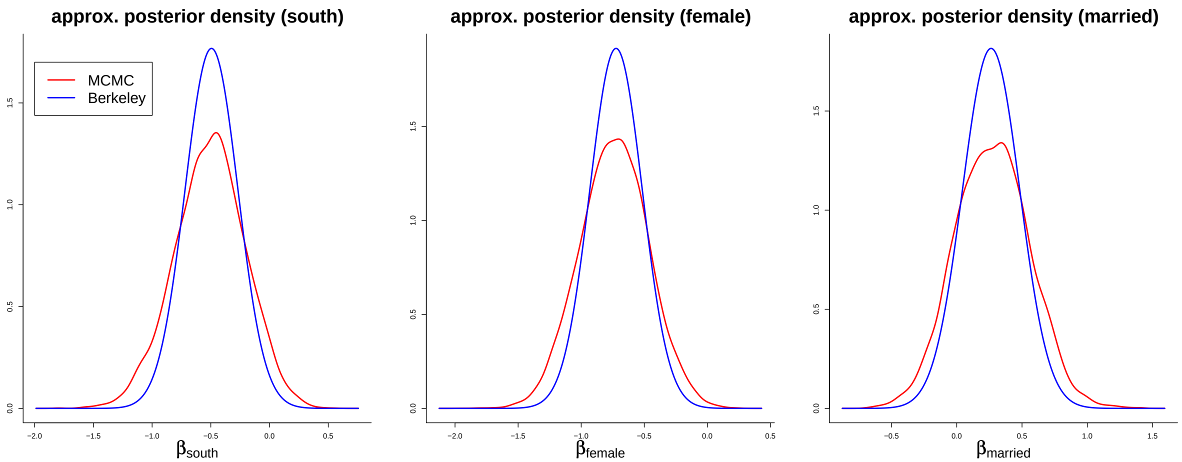
Wollongong I

and

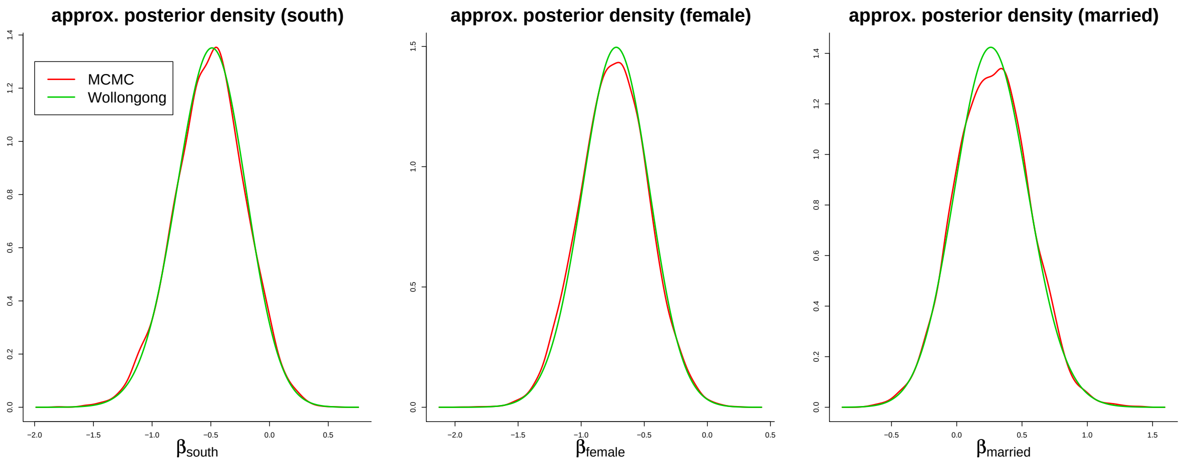
Wollongong II

Berkeley versus Wollongong I

Berkeley Variational Approximation Answer

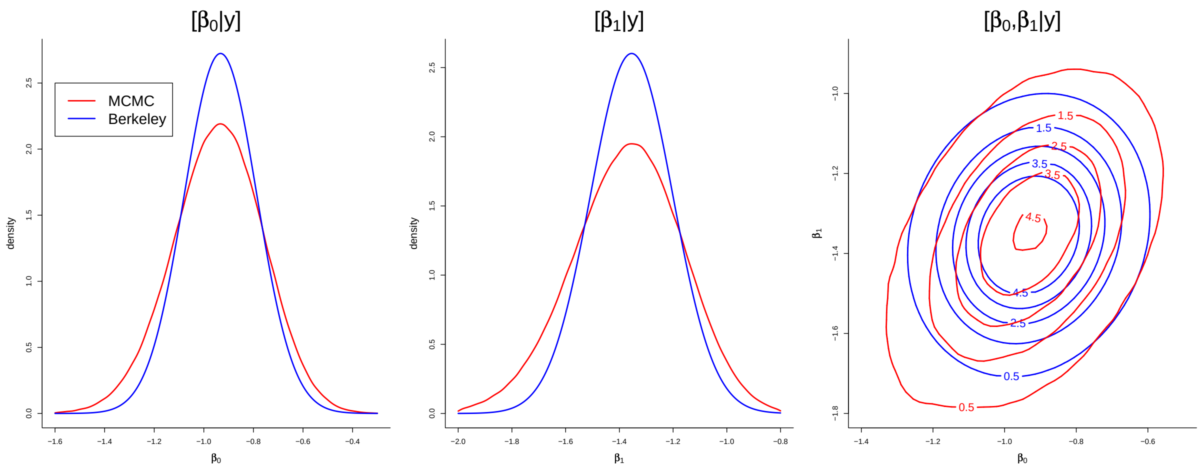


Wollongong I Variational Approximation Answer

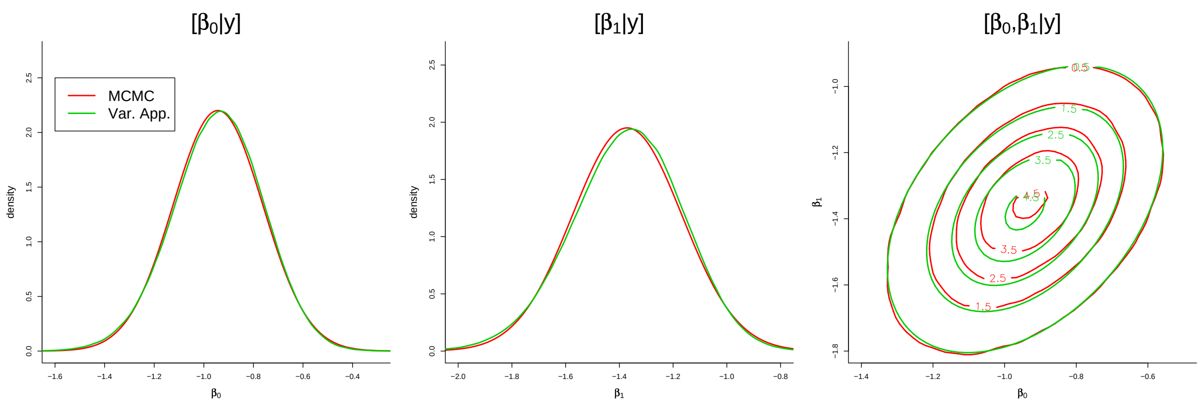


Berkeley versus Wollongong II

Berkeley Variational Approximation Answer



Wollongong II Variational Approximation Answer



Details of Wollongong I

$$\text{posterior of slope} = p(\beta_1|y) = \frac{p(\beta_1, y)}{p(y)} \propto p(\beta_1, y).$$

Details of Wollongong I

$$\text{posterior of slope} = p(\beta_1|y) = \frac{p(\beta_1, y)}{p(y)} \propto p(\beta_1, y).$$

$$p(\beta_1, y) = \int_{-\infty}^{\infty} [\beta_1, \beta_0, y] \, d\beta_0$$

Details of Wollongong I

$$\text{posterior of slope} = p(\beta_1|y) = \frac{p(\beta_1, y)}{p(y)} \propto p(\beta_1, y).$$

$$\begin{aligned} p(\beta_1, y) &= \int_{-\infty}^{\infty} [\beta_1, \beta_0, y] \, d\beta_0 \\ &= e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\beta_0 1^T y - 1^T b(\beta_0 1 + \beta_1 x) - \beta_0^2/(2 \times 10^8)\} \, d\beta_0 \end{aligned}$$

Details of Wollongong I

$$\text{posterior of slope} = p(\boldsymbol{\beta}_1|\boldsymbol{y}) = \frac{p(\boldsymbol{\beta}_1, \boldsymbol{y})}{p(\boldsymbol{y})} \propto p(\boldsymbol{\beta}_1, \boldsymbol{y}).$$

$$\begin{aligned}
p(\boldsymbol{\beta}_1, \boldsymbol{y}) &= \int_{-\infty}^{\infty} [\boldsymbol{\beta}_1, \boldsymbol{\beta}_0, \boldsymbol{y}]\, d\boldsymbol{\beta}_0 \\
&= e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\boldsymbol{\beta}_0 \mathbf{1}^T \boldsymbol{y} - \mathbf{1}^T \boldsymbol{b}(\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x}) - \boldsymbol{\beta}_0^2/(2 \times 10^8)\}\, d\boldsymbol{\beta}_0 \\
&\geq e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\boldsymbol{\beta}_0 \mathbf{1}^T \boldsymbol{y} - (\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x})^T \text{diag}\{\boldsymbol{A}(\boldsymbol{\xi})\}(\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x}) \\
&\qquad - \boldsymbol{B}(\boldsymbol{\xi})^T (\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x}) - \mathbf{1}^T \boldsymbol{C}(\boldsymbol{\xi}) - \boldsymbol{\beta}_0^2/(2 \times 10^8)\}\, d\boldsymbol{\beta}_0
\end{aligned}$$

Details of Wollongong I

$$\text{posterior of slope} = p(\beta_1|y) = \frac{p(\beta_1, y)}{p(y)} \propto p(\beta_1, y).$$

$$\begin{aligned}
p(\beta_1, y) &= \int_{-\infty}^{\infty} [\beta_1, \beta_0, y]\, d\beta_0 \\
&= e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\beta_0 1^T y - 1^T b(\beta_0 1 + \beta_1 x) - \beta_0^2/(2 \times 10^8)\}\, d\beta_0 \\
&\geq e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\beta_0 1^T y - (\beta_0 1 + \beta_1 x)^T \text{diag}\{A(\xi)\}(\beta_0 1 + \beta_1 x) \\
&\quad -B(\xi)^T(\beta_0 1 + \beta_1\, x) - 1^T C(\xi) - \beta_0^2/(2 \times 10^8)\}\, d\beta_0 \\
&= \text{explicit function of } \beta_1 \text{ and } \xi
\end{aligned}$$

Details of Wollongong I

posterior of slope $= p(\boldsymbol{\beta}_1|\boldsymbol{y}) = \dfrac{p(\boldsymbol{\beta}_1, \boldsymbol{y})}{p(\boldsymbol{y})} \propto p(\boldsymbol{\beta}_1, \boldsymbol{y})$.

$$\begin{aligned}
p(\boldsymbol{\beta}_1, \boldsymbol{y}) &= \int_{-\infty}^{\infty} [\boldsymbol{\beta}_1, \boldsymbol{\beta}_0, \boldsymbol{y}]\, d\boldsymbol{\beta}_0 \\
&= e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\boldsymbol{\beta}_0 \mathbf{1}^T \boldsymbol{y} - \mathbf{1}^T \boldsymbol{b}(\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x}) - \boldsymbol{\beta}_0^2/(2 \times 10^8)\}\, d\boldsymbol{\beta}_0 \\
&\geq e^{-\beta_1^2/(2\times 10^8)} \int_{-\infty}^{\infty} \exp\{\boldsymbol{\beta}_0 \mathbf{1}^T \boldsymbol{y} - (\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x})^T \text{diag}\{\boldsymbol{A}(\boldsymbol{\xi})\}(\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1 \boldsymbol{x}) \\
&\qquad - \boldsymbol{B}(\boldsymbol{\xi})^T (\boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{\beta}_1\, \boldsymbol{x}) - \mathbf{1}^T \boldsymbol{C}(\boldsymbol{\xi}) - \boldsymbol{\beta}_0^2/(2 \times 10^8)\}\, d\boldsymbol{\beta}_0 \\
&= \text{explicit function of } \boldsymbol{\beta}_1 \text{ and } \boldsymbol{\xi} \\
&= \text{explicit}(\boldsymbol{\beta}_1; \boldsymbol{\xi}), \quad \text{(say)}
\end{aligned}$$

Details of Wollongong I (continued)

Set up a grid: $\boldsymbol{\beta}_1^{[1]}, \ldots, \boldsymbol{\beta}_1^{[G]}$ over domain of $\boldsymbol{p}(\boldsymbol{\beta}_1|\boldsymbol{y})$.

Details of Wollongong I (continued)

Set up a grid: $\boldsymbol{\beta}_1^{[1]}, \ldots, \boldsymbol{\beta}_1^{[G]}$ over domain of $p(\boldsymbol{\beta}_1|\boldsymbol{y})$.

For $1 \leq g \leq G$ choose $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}^{[g]}$ to maximise explicit$(\boldsymbol{\beta}_1^{[g]}, \boldsymbol{\xi})$.

Details of Wollongong I (continued)

Set up a grid: $\boldsymbol{\beta}_1^{[1]}, \ldots, \boldsymbol{\beta}_1^{[G]}$ over domain of $p(\boldsymbol{\beta}_1|\boldsymbol{y})$.

For $1 \leq g \leq G$ choose $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}^{[g]}$ to maximise explicit$(\boldsymbol{\beta}_1^{[g]}, \boldsymbol{\xi})$.

This gives

$$\text{explicit}(\boldsymbol{\beta}_1^{[1]}, \widehat{\boldsymbol{\xi}}^{[1]}), \ldots, \text{explicit}(\boldsymbol{\beta}_1^{[G]}, \widehat{\boldsymbol{\xi}}^{[G]})$$

as an approximation to

$$p(\boldsymbol{\beta}_1^{[1]}, \boldsymbol{y}), \ldots, p(\boldsymbol{\beta}_1^{[G]}, \boldsymbol{y}).$$

Details of Wollongong I (continued)

Set up a grid: $\boldsymbol{\beta}_1^{[1]}, \ldots, \boldsymbol{\beta}_1^{[G]}$ over domain of $p(\boldsymbol{\beta}_1|\boldsymbol{y})$.

For $1 \leq g \leq G$ choose $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}^{[g]}$ to maximise explicit$(\boldsymbol{\beta}_1^{[g]}, \boldsymbol{\xi})$.

This gives

$$\text{explicit}(\boldsymbol{\beta}_1^{[1]}, \widehat{\boldsymbol{\xi}}^{[1]}), \ldots, \text{explicit}(\boldsymbol{\beta}_1^{[G]}, \widehat{\boldsymbol{\xi}}^{[G]})$$

as an approximation to

$$p(\boldsymbol{\beta}_1^{[1]}, \boldsymbol{y}), \ldots, p(\boldsymbol{\beta}_1^{[G]}, \boldsymbol{y}).$$

Final step: Normalise using (one-dimensional) quadrature to approximate $p(\boldsymbol{\beta}_1|\boldsymbol{y})$.

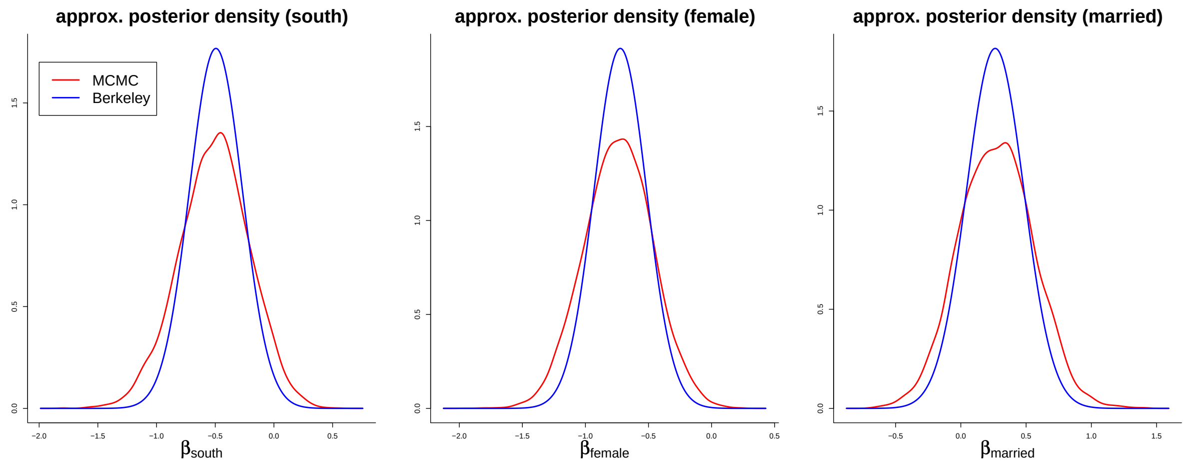
Jaakkola & Jordan idea applied

grid-wise

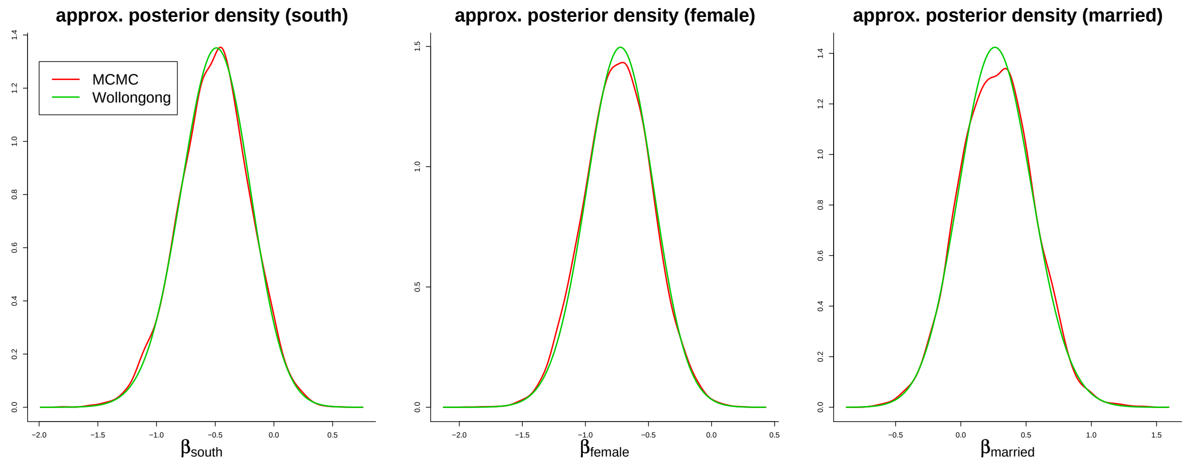
rather than globally.

Berkeley versus Wollongong I

Berkeley Variational Approximation Answer



Wollongong I Variational Approximation Answer



Details of Wollongong II

Consider the

Bayesian Poisson regression model

$$\boldsymbol{p(y|\beta) = \exp\{y^T X\beta - 1^T \log(1 + e^{X\beta}) - 1^T \log(y!)\}}$$

$$\boldsymbol{\beta_{p\times 1} \sim N(0, F)}$$

The log marginal likelihood is (ignoring constants):

$$\log p(y) = \log \int_{\mathbb{R}^p} p(y|\beta)p(\beta)\, d\beta$$

The log marginal likelihood is (ignoring constants):

$$\begin{aligned}\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\ &= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}}\end{aligned}$$

The log marginal likelihood is (ignoring constants):

$$\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}}
\end{aligned}$$

The log marginal likelihood is (ignoring constants):

$$\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\Sigma)}\left[\frac{\exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]
\end{aligned}$$

The log marginal likelihood is (ignoring constants):

$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{\exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right] \\
&\geq \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left(\log\left[\frac{\exp\{\boldsymbol{y}^T \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \mathbf{1}^T \exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T \boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\boldsymbol{\pi})^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]\right)
\end{aligned}
$$

The log marginal likelihood is (ignoring constants):

$$
\begin{aligned}
\log p(\boldsymbol{y}) &= \log \int_{\mathbb{R}^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \int_{\mathbb{R}^p} \exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \tfrac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\} \\
&\qquad \times \frac{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{\beta}} \\
&= \log \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\Sigma)}\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right] \\
&\geq \boldsymbol{E}_{\widetilde{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\Sigma)}\left(\log\left[\frac{\exp\{\boldsymbol{y}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{1}^T\exp(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \frac{1}{2}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{F}^{-1}\widetilde{\boldsymbol{\beta}}\}}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})^T\Sigma^{-1}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\mu})\}}\right]\right) \\
&= \boldsymbol{y}^T\boldsymbol{X}\boldsymbol{\mu} - \boldsymbol{1}^T\exp\{\boldsymbol{X}\boldsymbol{\mu} + \tfrac{1}{2}\text{diagonal}(\boldsymbol{X}\Sigma\boldsymbol{X}^T)\} - \tfrac{1}{2}\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} \\
&\qquad -\tfrac{1}{2}\{\text{tr}(\boldsymbol{F}^{-1}\Sigma) + \log|\Sigma|\} \\
&= \log \underline{p(\boldsymbol{y},\boldsymbol{\mu},\Sigma)} \quad = \text{variational lower bound on } \log p(\boldsymbol{y})
\end{aligned}
$$

Variational Approximation of Poisson Regression Bayes Factor

We have just shown $p(y) \geq \underline{p(y; \mu, \Sigma)}$ for all $\mu_{p\times 1}$ and $\Sigma_{p\times p}$.

Variational Approximation of Poisson Regression Bayes Factor

We have just shown $p(y) \geq \underline{p(y; \mu, \Sigma)}$ for all $\mu_{p \times 1}$ and $\Sigma_{p \times p}$.

$\underline{p(y; \mu, \Sigma)}$ is relatively easy to compute (1D numerical integration; nice integrands).

Variational Approximation of Poisson Regression Bayes Factor

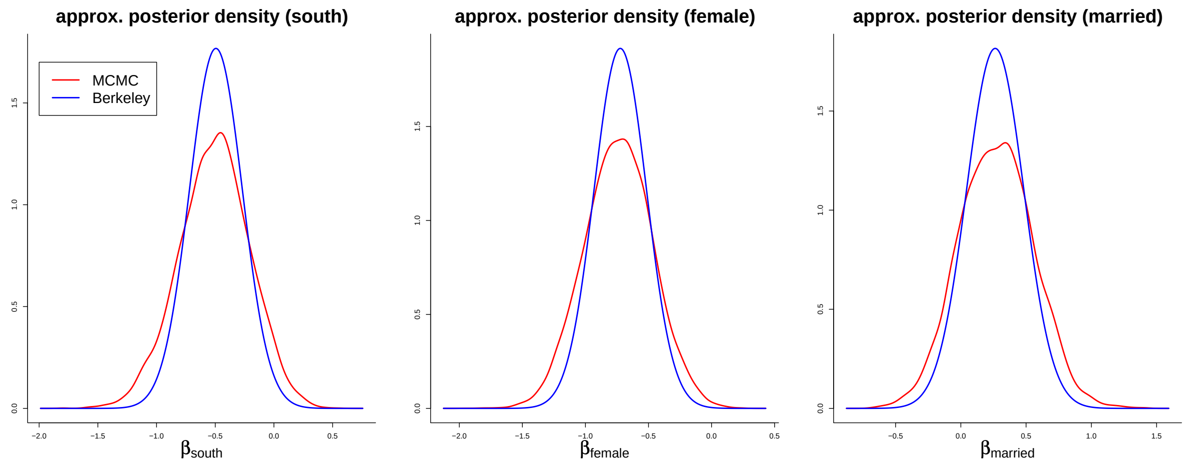
We have just shown $\boldsymbol{p(y)} \geq \underline{\boldsymbol{p(y;\mu,\Sigma)}}$ for all $\boldsymbol{\mu_{p\times 1}}$ and $\boldsymbol{\Sigma_{p\times p}}$.

$\underline{\boldsymbol{p(y;\mu,\Sigma)}}$ is relatively easy to compute (1D numerical integration; nice integrands).

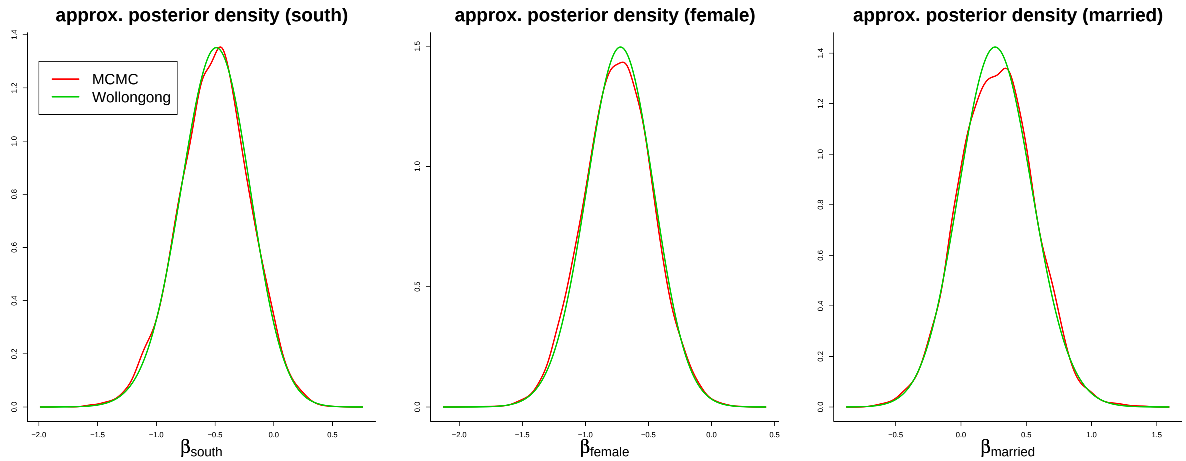
We choose these variational parameters $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ to maximise the right-hand-side (i.e. make bound as tight as we can).

Berkeley versus Wollongong I

Berkeley Variational Approximation Answer

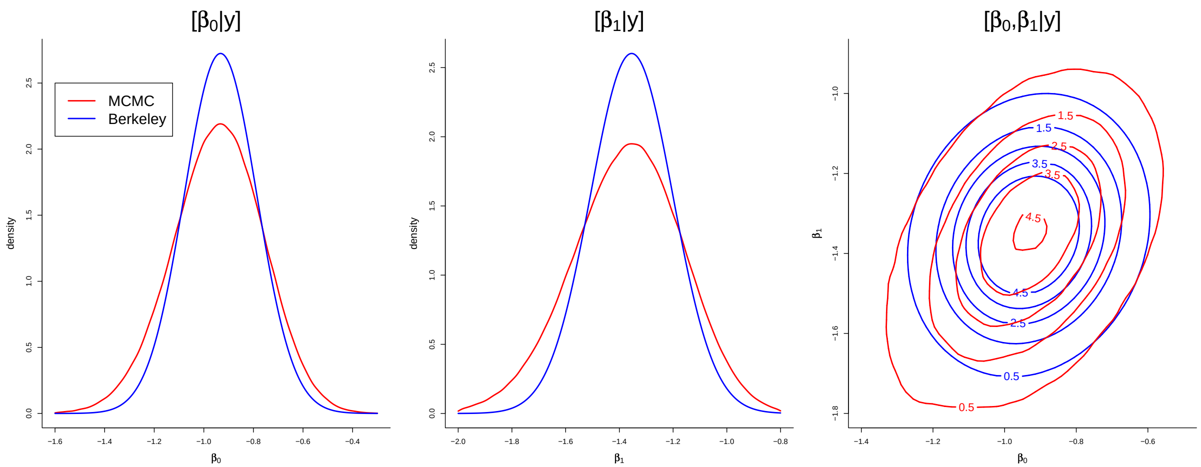


Wollongong I Variational Approximation Answer

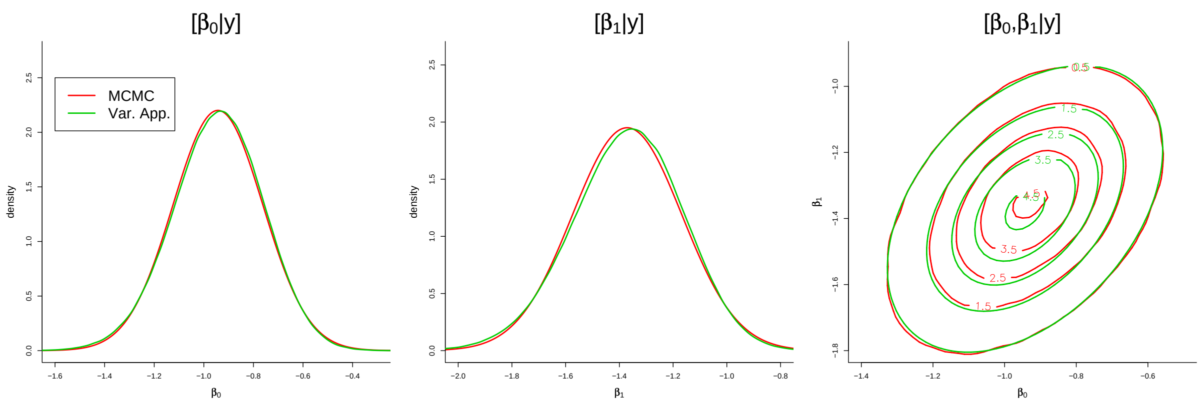


Berkeley versus Wollongong II

Berkeley Variational Approximation Answer



Wollongong II Variational Approximation Answer



Conclusions Thus Far

Conclusions Thus Far

- Hierarchical Bayes models and DAGs are useful structure for semiparametric regression.

Conclusions Thus Far

- Hierarchical Bayes models and DAGs are useful structure for semiparametric regression.
- We can always fall back on Markov chain Monte Carlo (MCMC) and BUGS.

Conclusions Thus Far

- Hierarchical Bayes models and DAGs are useful structure for semiparametric regression.
- We can always fall back on Markov chain Monte Carlo (MCMC) and BUGS.
- But MCMC slow, requires convergence assessment.

Conclusions Thus Far

- Hierarchical Bayes models and DAGs are useful structure for semiparametric regression.
- We can always fall back on Markov chain Monte Carlo (MCMC) and BUGS.
- But MCMC slow, requires convergence assessment.
- Variational inference engines emerging as a (better?) alternative.

Conclusions Thus Far

- Hierarchical Bayes models and DAGs are useful structure for semiparametric regression.
- We can always fall back on Markov chain Monte Carlo (MCMC) and BUGS.
- But MCMC slow, requires convergence assessment.
- Variational inference engines emerging as a (better?) alternative.
- Wollongong variational inference ‘school’ showing early promising results.

Conclusions Thus Far

- Hierarchical Bayes models and DAGs are useful structure for semiparametric regression.
- We can always fall back on Markov chain Monte Carlo (MCMC) and BUGS.
- But MCMC slow, requires convergence assessment.
- Variational inference engines emerging as a (better?) alternative.
- Wollongong variational inference 'school' showing early promising results.
- But bugger all ('diddly-squat' in US) in the way of theory.

Start of...

NEW THEORETICAL RESULTS FOR

Generalised Linear Mixed Models (GLMMs)

(Note - we now switch to being frequentists!)

Some Simple GLMMs

Logistic Response

$$\text{logit}\{P(y_{ij} = 1 | U_i)\} = \beta_0 + \beta_1 x_i + U_i$$

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2)$$

Poisson Response

$$y_{ij} = 1 | U_i \sim \text{Poisson}\{\exp(\beta_0 + \beta_1 x_i + U_i)\}$$

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2)$$

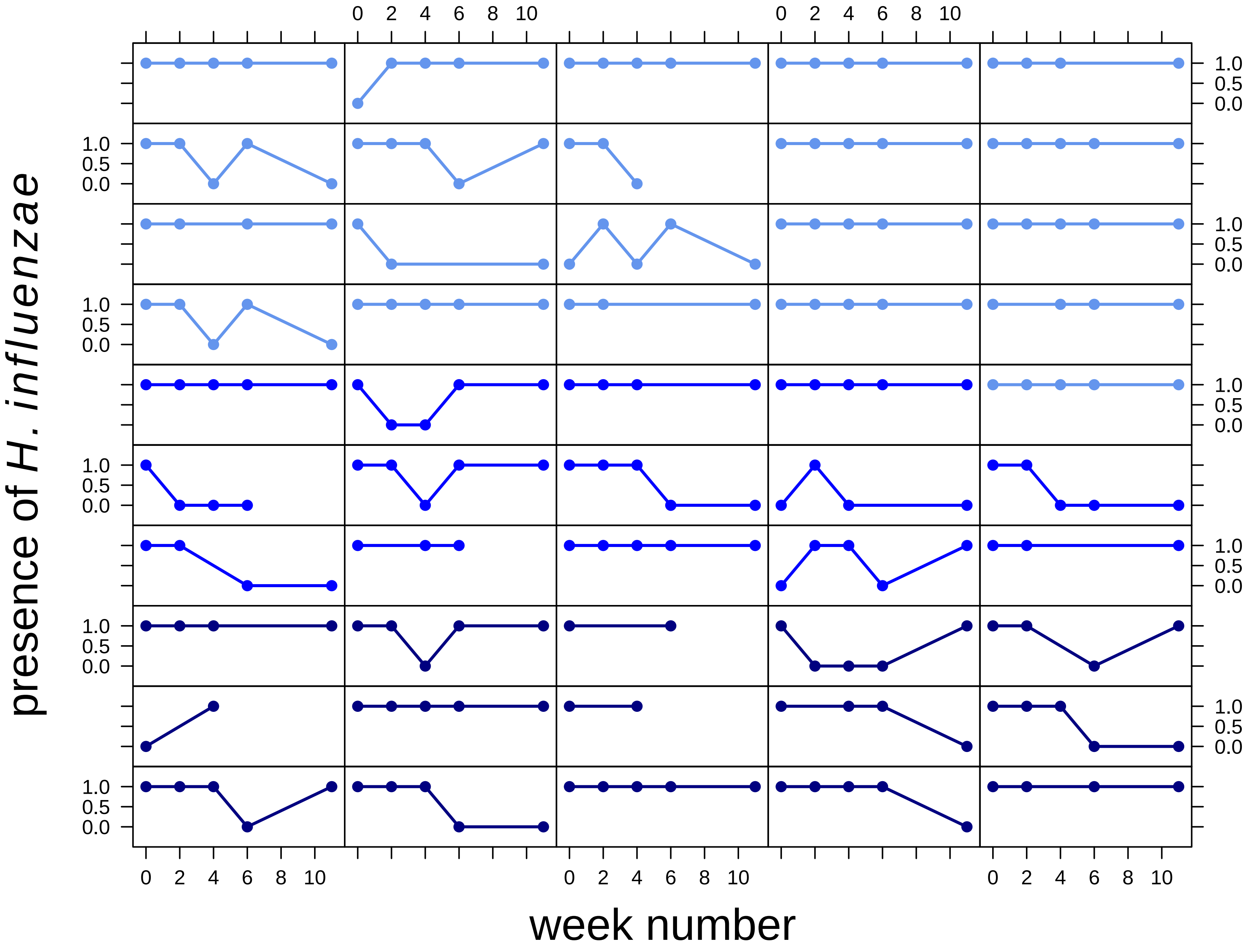
Relevance Check

GLMMs are

really, really important.

No time to explain.

Just take my word for it!


presence of H. influenzae
week number
0 2 4 6 8 10
1.0
0.5
0.0

Exponential Family Models

name	canonical link	$b(\eta)$	$c(y, \phi)$	ϕ
Bernoulli	$\eta = \text{logit}(\mu)$	$\log(1 + e^{\eta})$	0	1
Poisson	$\eta = \ln(\mu)$	e^{η}	$-\ln(y!)$	1
$N(\mu, \sigma^2)$	$\eta = \mu$	$\eta^2/2$	$(y^2/\sigma^2 - \ln(2\pi\sigma^2))/2$	σ^2

Exponential Family GLM

$$\log p(y; \beta, \phi) = \{y^T X\beta - 1^T b(X\beta)\}/\phi + 1^T c(y, \phi)$$

GLMM Extension

$$\log\{p(y|u)\} = \{y^T(X\beta + Zu) - \mathbf{1}^T b(X\beta + Zu)\}/\phi + \mathbf{1}^T c(y, \phi)$$

$$u \sim N(0, G)$$

Maximum Likelihood Estimation

Likelihood is:

$$
\begin{aligned}
\mathcal{L}(\beta, G, \phi) &= p(y; \beta, G) \\
&= \int_{\mathbb{R}^q} p(y, u)\, du \\
&= \int_{\mathbb{R}^q} p(y|u)p(u)\, du \\
&= (2\pi)^{-q/2}|G|^{-1/2} \int_{\mathbb{R}^q} \exp[\{y^T(X\beta + Zu) - 1^T b(X\beta + Zu \\
&\qquad +1^T c(y, \phi) - \tfrac{1}{2} u^T G^{-1} u\}]\, du
\end{aligned}
$$

GLMMs Big Headache

The likelihood involves an

intractable integral

(often high-dimensional).

Poisson Random Intercept Example

Likelihood is:

$$\mathcal{L}(\beta, \sigma^2) = (\sigma^2)^{-m/2} \times \text{const.}$$
$$\times \int_{\mathbb{R}^m} \exp\{y^T(X\beta + Zu) - 1^T \exp(X\beta + Zu) - \tfrac{1}{2\sigma^2} u^T u\}\, du$$

Poisson Random Intercept Example

Likelihood is:

$$\mathcal{L}(\beta, \sigma^2) = (\sigma^2)^{-m/2} \times \text{const.}$$
$$\times \int_{\mathbb{R}^m} \exp\{y^T(X\beta + Zu) - 1^T \exp(X\beta + Zu) - \tfrac{1}{2\sigma^2} u^T u\}\, du$$

Log-likelihood is (ignoring constants):

$$\ell(\beta, \sigma^2) = -(m/2)\log(\sigma^2)$$
$$+ \log \int_{\mathbb{R}^m} \exp\{y^T(X\beta + Z\widetilde{u}) - 1^T \exp(X\beta + Z\widetilde{u}) - \tfrac{1}{2\sigma^2} \widetilde{u}^T \widetilde{u}\}\, d\widetilde{u}$$

Variational Transform of Problem

$$\begin{aligned}\boldsymbol{\ell}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) &= \log \int_{\mathbb{R}^m} \exp\{\boldsymbol{y}^T(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T \exp(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \tfrac{1}{2\boldsymbol{\sigma}^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\} \\ &\quad \times \frac{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})\}} \, d\widetilde{\boldsymbol{u}} - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2)\end{aligned}$$

Variational Transform of Problem

$$
\begin{aligned}
\boldsymbol{\ell}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) &= \log \int_{\mathbb{R}^m} \exp\{\boldsymbol{y}^T(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T \exp(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \tfrac{1}{2\boldsymbol{\sigma}^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\} \\
&\quad \times \frac{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{u}} - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2) \\
&= \log \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)}\left[\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T \exp(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{u}})\} - \frac{1}{2\boldsymbol{\sigma}^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}} - \boldsymbol{\mu})\}}\right.
\end{aligned}
$$

Variational Transform of Problem

$$
\begin{aligned}
\boldsymbol{\ell}(\boldsymbol{\beta},\boldsymbol{\sigma}^2) &= \log \int_{\mathbb{R}^m} \exp\{\boldsymbol{y}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T\exp(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \tfrac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\} \\
&\quad \times \frac{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{u}} - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2) \\
&= \log \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)}\left[\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T\exp(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}})\} - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\right. \\
&\geq \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)}\log\left\{\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T e^{\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\widetilde{u}} - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\right\}
\end{aligned}
$$

Variational Transform of Problem

$$\begin{aligned}
\ell(\beta,\sigma^2) &= \log \int_{\mathbb{R}^m} \exp\{y^T(X\beta+Z\widetilde{u}) - 1^T\exp(X\beta+Z\widetilde{u}) - \tfrac{1}{2\sigma^2}\widetilde{u}^T\widetilde{u}\} \\
&\quad \times \frac{(2\pi)^{-m/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{u}-\mu)^T\Sigma^{-1}(\widetilde{u}-\mu)\}}{(2\pi)^{-m/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{u}-\mu)^T\Sigma^{-1}(\widetilde{u}-\mu)\}}\, d\widetilde{u} - \tfrac{m}{2}\log(\sigma^2) \\
&= \log E_{\widetilde{u}\sim N(\mu,\Sigma)}\left[\frac{\exp\{y^T(X\beta+Z\widetilde{u}) - 1^T\exp(X\beta+Z\widetilde{u})\} - \frac{1}{2\sigma^2}\widetilde{u}^T\widetilde{u}}{(2\pi\sigma^2)^{-m/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{u}-\mu)^T\Sigma^{-1}(\widetilde{u}-\mu)\}}\right. \\
&\geq E_{\widetilde{u}\sim N(\mu,\Sigma)}\log\left\{\frac{\exp\{y^T(X\beta+Z\widetilde{u}) - 1^T e^{X\beta+Z\widetilde{u}} - \frac{1}{2\sigma^2}\widetilde{u}^T\widetilde{u}\}}{(2\pi\sigma^2)^{-m/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{u}-\mu)^T\Sigma^{-1}(\widetilde{u}-\mu)\}}\right\} \\
&= y^T(X\beta+Z\mu) - 1^T\exp\{X\beta+Z\mu+\tfrac{1}{2}\text{diagonal}(Z\Sigma Z^T)\} \\
&\quad -\frac{1}{2\sigma^2}\{\mu^T\mu+\text{tr}(\Sigma)\} + \tfrac{1}{2}\log|\Sigma| - \tfrac{m}{2}\log(\sigma^2)
\end{aligned}$$

Variational Transform of Problem

$$
\begin{aligned}
\boldsymbol{\ell}(\boldsymbol{\beta},\boldsymbol{\sigma}^2) &= \log \int_{\mathbb{R}^m} \exp\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T\exp(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \tfrac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\} \\
&\quad \times \frac{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{u}} - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2) \\
&= \log \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)}\left[\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T\exp(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}})\} - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\right. \\
&\geq \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)}\log\left\{\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T e^{\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}} - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\right\} \\
&= \boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z\mu}) - \mathbf{1}^T\exp\{\boldsymbol{X\beta}+\boldsymbol{Z\mu}+\tfrac{1}{2}\text{diagonal}(\boldsymbol{Z\Sigma Z}^T)\} \\
&\quad -\frac{1}{2\boldsymbol{\sigma}^2}\{\boldsymbol{\mu}^T\boldsymbol{\mu}+\text{tr}(\boldsymbol{\Sigma})\} + \tfrac{1}{2}\log|\boldsymbol{\Sigma}| - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2) \\
&\equiv \underline{\boldsymbol{\ell}(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\mu},\boldsymbol{\Sigma})}
\end{aligned}
$$

Variational Transform of Problem

$$
\begin{aligned}
\ell(\boldsymbol{\beta},\boldsymbol{\sigma}^2) &= \log \int_{\mathbb{R}^m} \exp\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T\exp(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \tfrac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\} \\
&\quad \times \frac{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}{(2\boldsymbol{\pi})^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\, d\widetilde{\boldsymbol{u}} - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2) \\
&= \log \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)}\left[\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T\exp(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}})\} - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\right. \\
&\geq \boldsymbol{E}_{\widetilde{u}\sim N(\mu,\Sigma)} \log\left\{\frac{\exp\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}) - \mathbf{1}^T e^{\boldsymbol{X\beta}+\boldsymbol{Z}\widetilde{\boldsymbol{u}}} - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{u}}^T\widetilde{\boldsymbol{u}}\}}{(2\boldsymbol{\pi}\boldsymbol{\sigma}^2)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{u}}-\boldsymbol{\mu})\}}\right\} \\
&= \boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Z\mu}) - \mathbf{1}^T\exp\{\boldsymbol{X\beta}+\boldsymbol{Z\mu}+\tfrac{1}{2}\text{diagonal}(\boldsymbol{Z\Sigma Z}^T)\} \\
&\quad -\frac{1}{2\boldsymbol{\sigma}^2}\{\boldsymbol{\mu}^T\boldsymbol{\mu}+\text{tr}(\boldsymbol{\Sigma})\} + \tfrac{1}{2}\log|\boldsymbol{\Sigma}| - \tfrac{m}{2}\log(\boldsymbol{\sigma}^2) \\
&\equiv \underline{\ell(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\mu},\boldsymbol{\Sigma})} \\
&= \text{variational lower bound on } \ell(\boldsymbol{\beta},\boldsymbol{\sigma}^2).
\end{aligned}
$$

The variational approx. max. lik. est. is:

$$(\underline{\widehat{\boldsymbol{\beta}}}, \underline{\widehat{\boldsymbol{\sigma}}^{\mathbf{2}}}),$$

the $(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}})$ component of

$$\underset{\boldsymbol{\beta},\boldsymbol{\sigma}^{\mathbf{2}},\boldsymbol{\mu},\boldsymbol{\Sigma}}{\text{argmax}}\, \underline{\boldsymbol{\ell}(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}.$$

Variational Approximate Fisher Information

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}^2) =$ parameters of interest

$\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) =$ variational parameters

Variational Approximate Fisher Information

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \text{parameters of interest}$$

$$\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{variational parameters}$$

Pretending that $\underline{\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\Sigma})} = \underline{\ell(\boldsymbol{\theta}, \boldsymbol{\eta})}$ is a log-likelihood then

the Fisher information is

$$\underline{\boldsymbol{I}_{(\boldsymbol{\theta},\boldsymbol{\eta})}} = -\boldsymbol{E}\{\mathsf{H}\underline{\ell(\boldsymbol{\theta}, \boldsymbol{\eta})}\} = \begin{bmatrix} \underline{\boldsymbol{I}_{\boldsymbol{\theta\theta}}} & \underline{\boldsymbol{I}_{\boldsymbol{\theta\eta}}}^T \\ \underline{\boldsymbol{I}_{\boldsymbol{\theta\eta}}} & \underline{\boldsymbol{I}_{\boldsymbol{\eta\eta}}} \end{bmatrix}$$

Asymptotic covariance matrix is $(\underline{\boldsymbol{I}_{\boldsymbol{\theta\theta}}} - \underline{\boldsymbol{I}_{\boldsymbol{\theta\eta}}}^T \underline{\boldsymbol{I}_{\boldsymbol{\eta\eta}}}^{-1} \underline{\boldsymbol{I}_{\boldsymbol{\theta\eta}}})^{-1}$.

LATE BREAKING NEWS!

LATE BREAKING NEWS!

CONSISTENCY RESULTS ESTABLISHED FOR GROUPED DATA GLMMS!!!

LATE BREAKING NEWS!

CONSISTENCY RESULTS ESTABLISHED FOR GROUPED DATA GLMMS!!!

Peter Hall spotted leaving the scene.

References

1. Ormerod, J.T. and Wand, M.P. (2008). Variational approximations for logistic mixed models. *Proceedings of the Ninth Iranian Statistics Conference, Isfahan, Iran*, pp. 450–467.

2. Wand, M.P. and Ormerod, J.T. (2009). Comment on paper by Rue, Martino & Chopin. *Journal of the Royal Statistical Society, Series B*, in press.

3. Wand, M.P. (2009). Semiparametric regression and graphical models. *Australian and New Zealand Journal of Statistics*, in press.

4. Ormerod, J.T., Hall, P. and Wand, M.P. (2009). Gaussian variational approximation for generalized linear mixed models. In progress.

5. Ormerod, J.T. and Wand, M.P. (2009). Understanding variational approximations. In progress. (a la George Casella!)

Second and third of these are on Wand papers web-site.

Final (Three-Point!) Summary

- Variational approximations have great potential in semiparametric regression.
- Early Ormerod/Wand (mainly Ormerod PhD thesis) work showing good practical perfomance.
- Some interesting statistical theory emerging.

Parting Words

It is too early to tell if

Variational Approximation

will become a major player the future of

semiparametric regression analysis.

But if it does then you can say that you heard about it first at the:

11th UF Dept Statistics Winter Workshop!

Papers, Contact etc.

www.uow.edu.au/∼mwand