Bayesian Methods

LABORATORY

Lesson 7: March 3 2002

Software: **BUGS**

Miscellanea in BUGS: a probit model via latent variable and Stochastic Search Variable Selection via a hierarchical normal mixture model

Probit Model via Latent Variable. Vasoconstriction data

Example taken from Finney, D. J. (1947) *The estimation from individual records of the relationship between dose and quantal response*, *Biometrika*, **34**, Issue 3/4, 320-334

On the course web page):

Dataset: **vasoconstriction.dat**;

BUGS program: **latentv.b** *Bernoulli data: via probit/logit or via latent variable*

Reference code: *Program 4.12 SATM scores*, example 4.12 in Congdon's book

The data

Measurements: 39 on 3 subjects

Dose: V, R

Response: vasoconstriction (in the skin of the fingers)

Visual inspection of the data: contour curves:

$$V^{\beta_1} R^{\beta_2} = constant$$

$$Y = \beta_0 + \beta_1 log(10V) + \beta_2 log(10R)$$

Biometric problem

We are required to study the relationship between **dose** (magnitude of a stimulus applied to certain test subjects) and response (measure of the effect) from individual records.

In some classes of data, the response is 'all-or-nothing' or **quantal**, and cannot be measured quantitatively.
An additional difficulty sometimes is that the intensity of the stimulus cannot be specified in advance of a test, but can only be measured after the test has taken place. Then, the records consist in a list of doses with, for each, either 0 or 100% responding.

In the present application, moreover, the dose is expressed in terms of two measurements.

Theoretical setting

Suppose we have a collection of binary responses with associated predictor variables:
$y_i \in \{0,1\}$, $i = 1, \ldots, n$,
x_i, k-dimensional predictors.
De£ne the **latent variables** z_i as:

$$z_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n,$$

where

$$\epsilon_i \overset{iid}{\sim} F(\cdot) \quad \text{with } F(\cdot) \text{ a cdf}$$

Consider the model:

$$Y_i = \begin{cases} 0 & if\ Z_i \leq 0 \\ 1 & if\ Z_i > 0 \end{cases}.$$

If $F = \Phi$, the standard normal distribution, the **latent variable model is equivalent to the probit model** for $p_i = P(Y_i = 1)$:

$$y_i \sim Bern(p_i)$$

$$p_i = \Phi(\mathbf{x}_i^T \boldsymbol{\beta}) \quad or \quad \Phi^{-1}(p_i) = probit(p_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

From the latent variable formulation we have that:

$$p_i = P(Y_i = 1) = P(Z_i > 0) = P(\epsilon_i > -\mathbf{x}_i^T \boldsymbol{\beta}) = 1 - F(-\mathbf{x}_i^T \boldsymbol{\beta}) \overset{F\ simmetric}{=} F(\mathbf{x}_i^T \boldsymbol{\beta})$$

If $F = e^\eta/(1+e^\eta)$, the logistic cdf, then $F^{-1} = logit$, the alternative frequently used link function for Bernoulli data.

Bayesian computation

Under a prior for $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu_0}, \Sigma_0),$$

how has Gibbs sampling to be implemented for a Bayesian analysis of the two equivalent models?

Probit/logit link model:

$$\boldsymbol{\beta} \Longrightarrow p_i \longrightarrow y_i$$

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto \prod_i p_i^{y_i}(1-p_i)^{1-y_i} N(\boldsymbol{\mu_0}, \Sigma_0)$$

where $p_i = F(\mathbf{x}_i^T \boldsymbol{\beta})$.
$p(\boldsymbol{\beta}|\mathbf{y})$ is the only full conditional. It is non-conjugate, but it is log-concave.
BUGS uses the free-derivative Adaptive Rejection sampling to update the node.

Latent variable model:

$$\boldsymbol{\beta} \Longrightarrow \mu_i \longrightarrow z_i \Longrightarrow y_i$$

$$p(\boldsymbol{\beta}, \mathbf{z}|\mathbf{y}) \propto \prod_i N(z_i|\mu_i, 1) I[low, high]\, N(\boldsymbol{\mu_0}, \Sigma_0)$$

where $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$,
$low = I[1 - y_i > 0] * c$, $high = I[y_i > 0] * c$.

That is, the sampling of Z may be based on draws from a truncated Normal: truncation is to the right (0=ceiling value) if $y_i = 0$, to the left (by 0) if $y_i = 1$.

The full conditionals for $\boldsymbol{\beta}$ and $\mathbf{Z}$ are trivial.
Again $p(\boldsymbol{\beta}|\mathbf{z}, \mathbf{y})$ is (non-conjugate)log-concave.

SSVS is a procedure to select 'promising' subsets of $X_1, \ldots, X_p$ for further considerations.

SSVS is based on **embedding the entire regression** setup in a hierarchical Bayes normal **mixture** model, where **latent variables** are used to identify subset choices.

Gibbs Sampler updates the initial probabilities assigned to the different subset choices which may be not necessarily2^p.

Stochastic Search Variable Selection. Simulated data

Methodology presented in George, E. I. and McCulloch R. E. (1993) *Variable Selection Via Gibbs Sampling*, *JASA*, **88** Issue 423, 881-889.

BUGS program (on the course web page):

Stochastic Search Variable Selection. Simulated data example in **SSVSsim.b**.

Reference code: *Program 4.25 Two-stage variable selection with simulated data*, Example 4.25 in Congdon's book, p 141.

Two-stage variable selection model

- **I stage**

$$y_i \,|\, \boldsymbol{\beta}, \sigma^2 \overset{ind}{\sim} N(\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)$$

- **II stage**: finite Normal mixture

$$\beta_j \,|\, \gamma_i \sim (1-\gamma_i)N(0, \sigma^2_{\beta j}) + \gamma_j N(0, c_j^2\sigma^2_{\beta j})$$

$$P(\gamma_j = 1) = 1 - P(\gamma_j = 0) = p_i,$$

marginal of any discrete distribution

$$f(\boldsymbol{\gamma})$$

$$\sigma^2 \sim IG(\nu_\gamma/2, \nu_\gamma\lambda_\gamma/2)$$

Variable selection problem

A crucial problem in building a multiple regression model is the **selection of predictors to include**.

That is, given a dependent variable Y and a set of potentia predictors $X_1, \ldots, X_p$, the problem is to find the 'best' model of the form

$$Y = X_1^*\beta_1^* + \ldots + X_q^*\beta_q^* + \epsilon,$$

where $X_1^*, \ldots, X_q^*$ is a 'selected' subset of $X_1, \ldots, X_p$.

There are 2^p potential regression models. A wide variety o selection procedures are based on a comparison *of all* 2^p *possible submodels*.

Simulated data example

- Uniform prior on N=12 possible regression options:

$$f(k) = 1/12$$

- $\nu = 0$: $1/\sigma^2 \sim G(.001, .001)$
- small $\sigma_{\beta j}$, large c_j: $\sigma_{\beta j} \equiv .33$, $c_j \equiv 10$
  If inclusion of β_j is not supported by the data then the prior with the default variance $\sigma^2_{\beta j}$ will tend to be selected more often. That is, the data provide little support for a nonzero β_j, then SSVS model does not select the correspondent predictor x_j.