Bayesian auxiliary variable models for binary and multinomial regression

(Bayesian Analysis, 2006)

Authors: Chris Holmes
Leonhard Held

As interpreted by: Rebecca Ferrell

UW Statistics 572, Talk #1

April 10, 2014
Categorical data setup

Classical framework with binary responses:

\[ y_i \sim \text{Bernoulli}(p_i) \]
\[ p_i = g^{-1}(\eta_i), \ g^{-1} : \mathbb{R} \rightarrow (0, 1) \]
\[ \eta_i = x_i \beta, \ i = 1, \ldots, n \]
\[ x_i = ( x_{i1} \ldots x_{ip} ) \]
\[ \beta = ( \beta_1 \ldots \beta_p )^T \]

Put a prior on the unknown coefficients:

\[ \beta \sim \pi(\beta) \]

Inferential goal: compute posterior \( \pi(\beta \mid y) \propto p(y \mid \beta)\pi(\beta) \)
Holmes & Held (H&H) set out to take regression models for categorical outcomes and ...
Holmes & Held (H&H) set out to take regression models for categorical outcomes and ...
Why is logistic regression hard to Bayesify?

- Maximum likelihood not that easy either!
  - Fit using iterative methods
  - Asymptotics sidestep unknown finite sample distributions
- No conjugate priors 😞
- Most previous approaches involve Metropolis-Hastings and need tuning, or otherwise rely on accept-reject steps (e.g. Gamerman, 1997; Chen & Dey, 1998)
- Adaptive-rejection sampling (Dellaportas & Smith, 1993) only updates individual coefficients, resulting in poor mixing when coefficients are correlated

What we would like: **automatic and efficient Bayesian inference**
Mixing demonstration
Mixing demonstration

Gibbs steps

\( z \)

\( \beta \)
Mixing demonstration

Gibbs steps

$\beta$

$z$

$-6 -4 -2 0 2 4 6$

$-3 -2 -1 0 1 2 3$

$-3 -2 -1 0 1 2 3$

$-6 -4 -2 0 2 4 6$
Mixing demonstration
Mixing demonstration

Gibbs steps

\[ \begin{align*}
\beta & \quad \text{versus} \quad z \\
-6 & \quad -4 \quad -2 \quad 0 \quad 2 \quad 4 \quad 6 \\
-3 & \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3
\end{align*} \]
Mixing demonstration

Gibbs steps

\[ z \]

\[ \beta \]

\[ \bullet \]

\[ \bullet \]
Mixing demonstration
Mixing demonstration
Mixing demonstration
Mixing demonstration

Joint steps

\( z \)

\( \beta \)
Mixing demonstration
Mixing demonstration
Mixing demonstration

Joint steps

$z$

$\beta$
Mixing demonstration

Joint steps

$\beta$

$z$
H&H goals

H&H address four aspects of Bayesian inference for categorical data regression models:

1. **Probit link**: use auxiliary variable method from Albert & Chib (A&C, 1993) to run MCMC automatically with Gibbs sampling, but with efficient joint updates

2. **Logit link**: make auxiliary variable method and joint updating work with logistic regression

3. **Model uncertainty**: extend methods to situations with uncertain covariate sets (e.g. Bayesian model averaging)

4. **Polychotomous data**: extend methods to data with more than two outcomes
Probit regression

A&C auxiliary variable approach: introduce unobserved auxiliary variables \( z_i \) and re-write the probit model as

\[
y_i = 1_{[z_i > 0]}
\]

\[
z_i = x_i \beta + \epsilon_i
\]

\[
\epsilon_i \sim N(0, 1)
\]

\[
\beta \sim \pi(\beta)
\]

Equivalent to probit model in standard framework:

\[
p_i = P(z_i > 0 \mid \beta) = P(x_i \beta + \epsilon_i > 0 \mid \beta)
\]

\[
= 1 - \Phi(-x_i \beta) = \Phi(x_i \beta) = g^{-1}(x_i \beta)
\]
Probit regression

From joint posterior, obtain nice conditional distributions of the parameters to simulate from in Gibbs steps:

\[ \pi(\beta, z \mid y) \propto p(y \mid \beta, z) p(z \mid \beta) \pi(\beta), \text{ so:} \]

\[ \pi(y \mid z) \]

\[ \pi(\beta \mid z, y) \propto p(z \mid \beta) \pi(\beta) = \pi(\beta) \prod_{i=1}^{n} p(z_i \mid \beta) \]

\[ N(x_i \beta, 1) \]

If we use a normal prior for \( \pi(\beta) \), then \( \pi(\beta \mid z, y) \) is also normal

\[ \pi(z \mid \beta, y) \propto p(y \mid z) p(z \mid \beta) \]

\[ = \prod_{i=1}^{n} \left( 1[z_i > 0]1[y_i = 1] + 1[z_i \leq 0]1[y_i = 0] \right) \phi(z_i - x_i \beta) \]

\[ \pi(z_i \mid \beta, y_i) \approx \text{truncated normal} \]
Smarter probit Gibbs

H&H improve mixing by updating \((\beta, z)\) jointly: simulate from \(\pi(z \mid y)\), then from \(\pi(\beta \mid z, y)\). Assuming \(\pi(\beta)\) normal:

\[
\pi(\beta, z \mid y) = \pi(\beta \mid z, y) \pi(z \mid y) \text{ implies (known form) normal}
\]

\[
\pi(z \mid y) \sim \text{truncated multivariate normal}
\]

Truncated multivariate normal hard to sample from, but univariate conditionals can be Gibbsed:

\[
\pi(z_i \mid z_{-i}, y) \succsim \begin{cases} 
    N(m_i, v_i) 1_{[z_i > 0]} & \text{if } y_i = 1 \\
    N(m_i, v_i) 1_{[z_i \leq 0]} & \text{if } y_i = 0
\end{cases}
\]

where \(m_i\) and \(v_i\) are known (ugly) functions of \(z\), data, and prior
Logistic regression

So far: sampling the posterior for a Bayesian probit model can be done automatically and efficiently! 😊

Probit is a reasonable model for binary valued data, so why bother with a logit extension?

- Coefficients correspond to change in log odds
- Logit link has heavier tails than probit
- Probit link is not analytic and observations corresponding to extreme predicted probabilities can have numerical issues
From probit to logit

How to extend auxiliary variables to logistic regression?

\[ y_i = 1_{[z_i > 0]} \]
\[ z_i = x_i \beta + \epsilon_i \]
\[ \epsilon_i \sim N(0, \lambda_i) \]
\[ \lambda_i = (2 \psi_i)^2, \quad \psi_i \sim KS \]
\[ \beta \sim \pi(\beta) \]

Equivalent to logit model because \( \epsilon_i \) has a logistic distribution (Andrews & Mallows, 1974) and CDF of logistic is expit function:

\[ p_i = P(z_i > 0 | \beta) = P(\epsilon_i > -x_i \beta | \beta) \]
\[ = 1 - \expit(-x_i \beta) = \expit(x_i \beta) = g^{-1}(x_i \beta) \]
Logistic Gibbs

In similar fashion to probit model, simulate from posterior conditionals:

\[
\pi(\beta, z, \lambda \mid y) \propto p(y \mid \beta, z, \lambda) p(z \mid \beta, \lambda)p(\lambda) \pi(\beta)
\]

\[
= p(y \mid z)
\]

\[
\pi(\beta \mid z, \lambda, y) \propto p(z \mid \beta, \lambda) \pi(\beta) \approx \text{normal if } \pi(\beta) \text{ normal}
\]

\[
\pi(z \mid \beta, \lambda, y) \propto p(y \mid z)p(z \mid \beta, \lambda) \approx \text{indep. truncated normals}
\]

\[
\pi(\lambda \mid \beta, z, y) \propto p(z \mid \beta, \lambda)p(\lambda) \approx \text{indep. normal } \times \text{ KS}^2
\]

This last conditional distribution is non-standard, but easy to simulate from (no tuning needed)
Smarter logistic Gibbs

Joint updates for logistic to speed up mixing? A couple of possibilities:

\( (A) \quad \pi(z, \lambda \mid \beta, y) = \underbrace{\pi(z \mid \beta, y)}_{\text{truncated logistic}} \underbrace{\pi(\lambda \mid \beta, z)}_{\text{rejection}} \) followed by

\( \pi(\beta \mid z, \lambda) \) normal

\( (B) \quad \pi(\beta, z \mid \lambda, y) = \underbrace{\pi(z \mid \lambda, y)}_{\text{truncated normal}} \underbrace{\pi(\beta \mid z, \lambda)}_{\text{normal}} \) followed by

\( \pi(\lambda \mid \beta, z) \) rejection
Next time, aspirationally

- Performance of joint updating scheme for probit regression
- Performance of two joint updating schemes for logistic regression
- Auxiliary variable approaches under model uncertainty
- Auxiliary variable approaches with polychotomous outcomes
- What’s happened since H&H 2006? (Go to the James Scott seminar next Thursday!)