Bayesian analysis of longitudinal binary data using Markov regression models with skewed links

Seongho Song

Department of Mathematical Sciences, University of Cincinnati

Dec 10, 2013

Joint work with

Younshik Chung
Department of Statistics, Pusan National University

and

Dipak K. Dey
Department of Statistics, University of Connecticut

- Special thanks to CCTST
- This project was supported by an Institutional Clinical and Translational Science Award, NIH/NCATS Grant Number 8UL1TR000077-04. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

- Introduction
  - Bayesian Variable Selection
    George and McCulloch (1993, 1997)
    Kuo and Mallick (1998)
    Green (1995); Richardson and Green (1997)
  - Skewed link for binary response
    Chen, Dey and Shao (1999)
  - Markov Regression for longitudinal data
    - Non-Bayesian approach
      Cox (1970), Bishop (1975), Diggle et al (1994) etc
    - Bayesian approach
      Erkanli, Soyer and Angold (2001)

Outline

Model

- Bayesian Analysis of Longitudinal binary data using Markov regression models with skewed link
  - Model
  - Likelihood
  - Informative Prior
  - Reversible Jump
- Application of Indonesian children's health study (ICHS)
- Discussion

- Bayes Factor - Berger (1995)
- DIC - Spiegelhalter et al (2002)
- L-measure - Gelfand and Ghosh (1998)

- George and McCulloch (1993) - Using Hierarchical Model

  Consider the linear model.

$$Y|\beta, \sigma^2 \sim N_n(X\beta, \sigma^2 I)$$

  where Y is $n \times 1$, $X = [X_1, \ldots, X_p]$ is $n \times p$, $\beta = (\beta_1, \ldots, \beta_p)^T$, and σ^2 is a scalar.

- George and McCulloch (1993) Continued

  Consider normal mixture prior of β with the latent variable $\gamma_i = 0$ or 1.

$$\beta_i | \gamma_i \sim (1 - \gamma_i) N(0, \tau_i^2) + \gamma_i N(0, c_i^2 \tau_i^2)$$

  and

$$P(\gamma_i = 1) = 1 - P(\gamma_i = 0) = p_i$$

  - Set $\tau_i (> 0)$ small so that if $\gamma_i = 0$, then β_i would be small.
  - Set c_i large ($c_i > 1$ always) so that $\gamma_i = 0$, then β_i should be included in the model.

- George and McCulloch (1993) Continued

By obtaining the marginal posterior distribution of γ,

$$f(\gamma|Y) \propto f(Y|\gamma)f(\gamma) \ ,$$

the variable selection can be done.

- Kuo and Mallick (1998) - Using Indicators

Consider the linear model with indicator variables.

$$y_i = \sum_{j=1}^{p} \beta_j \gamma_j x_{ij} + \epsilon_i$$

When $\gamma_j = 1$, include the j-th covariate in the model.
When $\gamma_j = 0$, omit the j-th covariate from the model.
Again, by obtaining the marginal posterior distribution of γ,

$$f(\gamma|Y) \propto f(Y|\gamma) f(\gamma) \ ,$$

the variable selection can be done.

- Reversible Jump - Green (1995), Ricahdson and Green (1997)
    - Consider the dimension of parameter space as unknown random.
- It is useful for the following model:
    - variable selection in regression
    - non-nested regression model
    - Bayesian choice between models with different numers of parameters
    - Change point problems

      etc

- Skewed Link model - Chen, Dey and Shao (1999)

  Let $w = (w_1, \ldots, w_n)^T$ be a vector of independent latent variables. The Skewed Link model is formulated as

$$Y_{it} = \begin{cases} 1, & \text{if } w_i < 0 \\ 0, & \text{if } w_i \geq 0 \end{cases},$$

  where

$$w_i = x_i^T \beta + \delta z_i + \epsilon_i \ , \quad z_i \sim G$$

  and $\epsilon_i \sim F$, z_i and ϵ_i are independent, G is the cdf of skewed dist., and F is the cdf of a symmetric dist. δ is a skewness parameter.

Then,

$$p_i = P(y_i = 1) = \int_{-\infty}^{\infty} F(x_i^T \beta + \delta z_i) g(z_i)\ dz_i\ ,$$

and

$$1 - p_i = P(y_i = 0) = \int_{-\infty}^{\infty} [1 - F(x_i^T \beta + \delta z_i)] g(z_i)\ dz_i\ ,$$

where $g(z_i)$ is the pdf of z_i.

- Non-Baysesian Approaches - MLE, GEE approach

  Cox (1970)
  Bishop (1975)
  Diggle et al (1994)

- Markov Regression Model
  - Baysesian Approaches - Erkanli, Soyer and Angold (2001)
  - Consider the binary observations $(Y_{i1}, \ldots, Y_{iT})$,
    where $Y_{it} = \begin{cases} 1, & \text{the } i^{th} \text{ patient has an event at time t} \\ 0, & \text{otherwise} \end{cases}$
    for $i = 1, \ldots, n$ and $t = 1, \ldots, T$.
  - $H_{it} = \{y_{i1}, \ldots, y_{i,t-1}\}$ is the history vector for i^{th} subject available up to time t.
  - $p_{it} = Pr(Y_{it} = 1|H_{it})$ is the probability of a subject having an event at time t, which depends on the subject's past events through vector H_{it}.
  - $x'_{it} = (x_{it1}, \ldots, x_{itp})$ is the corresponding covariates of y_{it}.

- Markov Regression model with logit link

$$p(y_{i1}, \ldots, y_{iT}) = \prod_{t=1}^{T} p(y_{it}|H_{it}),$$

where

$$p(y_{it}|H_{it}) = p_{it}^{y_{it}}(1 - p_{it})^{(1-y_{it})} \ ,$$
$$logit(p_{it}) = \mu + x_{it}'\delta + g(H_{it}),$$

where

$$g(\mathcal{H}_{it}) = \sum_{k=1}^{q} \gamma_k y_{i,t-k}$$

describes a qth order Markov logistic regression model. Consider an extended version for the variable selection as

$$g(\mathcal{H}_{it}) = \sum_{k=1}^{q} \gamma_k \beta_k y_{i,t-k}$$

Markov Regression model with Skewed link

Outline

- Model with skewed link
- Observed and Complete Likelihood Functions
- Informative Prior
- Reversible Jump

- Consider the binary observations $(Y_{i1}, \ldots, Y_{iT})$, $i = 1, \ldots, n$ where $Y_{it} = \begin{cases} 1, & \text{the } i^{th} \text{ patient has an event at time t} \\ 0, & \text{otherwise} \end{cases}$ for $t = 1, \ldots, T$.
- $H_{it} = \{y_{i1}, \ldots, y_{i,t-1}\}$ is the history vector for i^{th} subject available up to time t.
- $p_{it} = Pr(Y_{it} = 1 | H_{it})$ is the probability of a subject having an event at time t, which depends on the subject's past events through vector H_{it}.
- $x'_{it} = (x_{it1}, \ldots, x_{itp})$ is the corresponding covariates of y_{it}.

Model

$$p(y_{i1}, \ldots, y_{iT}) = \prod_{t=1}^{T} p(y_{it}|H_{it}),$$

where

$$p(y_{it}|H_{it}) = p_{it}^{y_{it}}(1 - p_{it})^{(1-y_{it})},$$

$$link(p_{it}) = \mu + x'_{it}\beta + h(H_{it}),$$

and $\beta' = (\beta_1, \ldots, \beta_p)$.

Choice of $h(\cdot)$

- $h(\cdot)$ is depending on the different belief about the subjects' transition behaviors.
- Consider the q^{th} order additive function,

$$h(H_{it}) = \sum_{k=1}^{q} \gamma_k y_{it-k}$$

for $q \in \{1, \ldots, q_{max}\}$, $1 \leq q_{max} \leq T - 1$.

- Assume that q is unknown random order of $h(\cdot)$.

Link Function
Consider the skewed links, discussed by Chen, Dey and Shao(1999).
For any cdf F, for example, $F(t) = \Phi(t)$ and $F(t) = \frac{e^t}{1+e^t}$,

$$p_{it} = Pr(Y_{it} = 1) = \int_{-\infty}^{\infty} F(\mu + x_{it}'\beta + h(\mathcal{H}_{it}) + \delta z_{it}) g(z_{it}) dz_{it}$$

and

$$1-p_{it} = Pr(Y_{it} = 0) = \int_{-\infty}^{\infty} \left[1 - F(\mu + x_{it}'\beta + h(\mathcal{H}_{it}) + \delta z_{it})\right] g(z_{it}) dz_{it},$$

where $g(\cdot)$ is the pdf of z_{it}.

Observed Likelihood Function

Let $D_{obs} = (n, Y, X)$ denote the observed data. Then, the likelihood function for the skewed link model is given by

$$\begin{aligned} L(\mu, \beta, \delta, \gamma | D_{obs}) &= \prod_{i=1}^{n} \prod_{t=1}^{T} p(y_{it} | \mathcal{H}_{it}, \mu, \beta, \delta) \\ &= \prod_{i=1}^{n} \prod_{t=1}^{T} \int_{-\infty}^{\infty} [F(x_{it}'\beta + h(\mathcal{H}_{it}) + \delta z_{it})]^{y_{it}} \\ &\quad [1 - F(x_{it}'\beta + h(\mathcal{H}_{it}) + \delta z_{it})]^{1-y_{it}} g(z_{it}) dz_{it}. \\ &= \prod_{i=1}^{n} \prod_{t=1}^{T} \int \int \{1_{(} y_{it} = 0) 1_{(} w_{it} \leq 0) + 1_{(} y_{it} = 1) 1_{(} w_{it} > 0)\} \\ &\qquad f(w_{it} - \mu - x_{it}'\beta - h(\mathcal{H}_{it}) - \delta z_{it}) g(z_{it}) \;\; dw_{it} \; dz_{it}. \end{aligned}$$

where $f(\cdot)$ is the pdf of $F(\cdot)$.

Complete Likelihood Function

Let $z_i = (z_{i1}, \ldots, z_{iT})'$, $w_i = (w_{i1}, \ldots, w_{iT})'$ be the augmented data.
Let $D = (n, y, X, z, w)$ denote the complete data.
Then the complete-data likelihood function of the parameter $(\mu, \beta, \delta, \gamma, z, w)$ can be written as

$$\begin{aligned} L^*(\mu, \beta, \delta, \gamma, w, z \quad | \quad & n, y, X) \\ = & \prod_{i=1}^{n} \prod_{t=1}^{T} \left[\{ 1_{(} y_{it} = 0) 1_{(} w_{it} \leq 0) + 1_{(} y_{it} = 1) 1_{(} w_{it} > 0) \} \right. \\ & \left. \times f(w_{it} - \mu - x_{it}'\beta - h(\mathcal{H}_{it}) - \delta z_{it}) g(z_{it}) \right] \end{aligned}$$

Informative Priors
Assume that μ, δ, γ and β are independently distributed with

$$\begin{aligned} \mu &\sim N(0, \sigma_\mu^2), \\ \beta &\sim N_p(0, \sigma_\beta^2), \\ \gamma &\sim N_q(\gamma_0, \sigma_\gamma^2) \end{aligned}$$

and

$$\delta \sim N(0, \sigma_\delta^2),$$

Reversible Jumps

Assume that the random order of $h(\cdot)$ follows Poisson distribution with parameter λ,

$$p(q) = \frac{e^{-\lambda}\lambda^q}{q!}, \quad q = 0, 1, \ldots .$$

A Poisson distribution truncate to $q < n$ or to $q < q_{max}$ for a suitable choice of q_{max} is more sensible here.

Three move types for our problem

- Updating all the parameters given the value q
- Birth step w.p.

$$b_q = c \cdot \min\left\{1, \frac{p(q+1)}{p(q)}\right\}$$

- Death step w.p.

$$d_{q+1} = c \cdot \min\left\{1, \frac{p(q)}{p(q+1)}\right\},$$

where the constant c is chosen as large as possible subject to

$$b_q + d_q \leq .9, \text{ for all } q = 0, 1, \ldots, q_{max},$$

$d_0 = 0$ and $b_{q_{max}} = 0$

Joint posterior distributions of parameters, for a given q,

$$
\begin{aligned}
&p(\mu, \beta, \delta, \gamma, w, z|n, y, X) \\
&\quad = \Big[\prod_{i=1}^{n}\prod_{t=1}^{T}\{1_(y_{it} = 0)1_(w_{it} \leq 0) + 1_(y_{it} = 1)1_(w_{it} > 0)\} \\
&\qquad \times f(w_{it} - \mu - x'_{it}\beta - h(\mathcal{H}_{it}) - \delta z_{it})g(z_{it})\Big] \\
&\qquad \times N(\mu|0, \sigma^2_\mu) \\
&\qquad \times N_p(\beta|0, \sigma^2_\beta I_p) \\
&\qquad \times N(\delta|0, \sigma^2_\delta) \\
&\qquad \times N_q(\gamma|\gamma_0, \sigma^2_{gamma} I_q))
\end{aligned}
$$

For Full Conditional Distributions, assume that $f = \phi$ and $g = \phi^{+}$.

$$[\mu|\cdots\cdots] = N\left(\mu|\left(\sigma_\mu^{-2} + nT\right)^{-1}[w_{it} - x_{it}^T\beta - h(\mathcal{H}_{it}) - \delta * z_{it}], \left(\sigma_\mu^{-2} + nT\right)^{-1}\right)$$

For $l = 1, \ldots, p$,

$$[\beta_l | \beta_{(-l)}, \cdots\cdots] = N\left(\beta_l | \mu_{\beta_l}, \left(\sigma_\beta^{-2} + \sum_i \sum_t x_{itl}^2\right)^{-1}\right)$$

where

$$\begin{aligned}\mu_{\beta_l} &= \left(\sigma_\beta^{-2} + \sum_i \sum_t x_{itl}^2\right)^{-1} \\ &\times \left[\sum_i \sum_t x_{itl}(w_{it} - x_{it}^T \beta - h(\mathcal{H}_{it}) - \delta * z_{it} - x_{it}^T \beta_{(-l)})\right]\end{aligned}$$

and

$$x_{it}^T \beta_{(-l)} = x_{it1}^T \beta_1 + x_{it2}^T \beta_2 + \cdots + x_{it,l-1}^T \beta_{l-1} + x_{it,l+1}^T \beta_{l+1} + \cdots + x_{itp}^T \beta_p \ .$$

$$[\delta|\cdots\cdots] = N\left(\delta|\mu_\delta, \left(\sigma_\delta^{-2} + \sum_i \sum_t x_{itl}^2\right)^{-1}\right)$$

where

$$\mu_\delta = \left(\sigma_\delta^{-2} + \sum_i \sum_t z_{it}^2\right)^{-1} \left[\sum_i \sum_t z_{it}(w_{it} - x_{it}^T \beta - h(\mathcal{H}_{it}))\right]$$

For γ_k, $k = 1, \ldots, q$, $q \in \{1, 2, \ldots, q_{max}\}$, $1 \leq q_{max} \leq T - 1$,

$$[\gamma_k | \cdots\cdots] = N\left(\gamma_k | \mu_{\gamma_k}, \left(\sigma_\gamma^{-2} + \sum_i \sum_t y_{i,t-k}^2\right)^{-1}\right)$$

where

$$\begin{aligned} \mu_{\gamma_k} &= \left(\sigma_\gamma^{-2} + \sum_i \sum_t y_{i,t-k}^2\right)^{-1} \\ &\quad \times \left[\frac{\gamma_{0k}}{\sigma_\gamma^2} \sum_i \sum_t y_{i,t-k}(w_{it} - \mu - x_{it}^T \beta - h(\mathcal{H}_{it})_{(-l)} - \delta z_{it})\right] \end{aligned}$$

and

$$h(\mathcal{H}_{it})_{(-l)} = \gamma_1 y_{i,t-1} + \cdots + \gamma_{k-1} y_{i,t-k+1} + \gamma_{k+1} y_{i,t-k-1} + \cdots + \gamma_q y_{i,t-q} \ .$$

For $i = 1, \ldots, n$ and $t = 1, \ldots, T$,

$$[w_{it} | \cdots\cdots] = N^{+} \left(w_{it} | \mu + x_{it}^T \beta + h(\mathcal{H}_{it}) - \delta z_{it}, 1 \right) , \quad w_{it} > 0 \quad \text{if} \quad y_{it} = 1$$

$$N^{-} \left(w_{it} | \mu + x_{it}^T \beta + h(\mathcal{H}_{it}) - \delta z_{it}, 1 \right) , \quad w_{it} < 0 \quad \text{if} \quad y_{it} = 0$$

For $i = 1, \ldots, n$ and $t = 1, \ldots, T$,

$$[z_{it} | \cdots\cdots] = N^{+} \left(z_{it} | \delta(1 + \delta^2)^{-1} (w_{it} - \mu - x_{it}^T \beta - h(\mathcal{H}_{it}), (1 + \delta^2)^{-1} \right) \quad z_{it} > 0$$

- Acceptance Probability for the birth step

$$min \left\{ \text{Likelihood ratio} \times \text{prior ratio} \times \frac{d_{q+1}}{b_q} \times \frac{1}{\gamma_{q+1}}, 1 \right\}$$

- Acceptance Probability for death step is the same form but with re-labelling of the variables and the ratio term inverted.

Indonesian Children's Health Study

- Consider the data on respiratory infection in Indonesian preschool children.
- $n = 122$ preschool children in Indonesian were examined for up to $T = 6$ consecutive quarters for the respiratory infection (Sommer, 1982).
- Consider gender, height for age, seasonal cosine and sine, presence of Xerophthalmia(Vitamin A deficiency), age as covariates.

- Results of unknown order q

q	posterior probability
0	0.00000
1	0.97442
2	0.02510
3	0.00044
4	<0.00000

- Results

	mean	sd	2.5%	50%	97.5%
γ	0.0818	0.1608	-0.2358	0.0825	0.3954
δ	-0.0012	0.1728	-0.3109	-0.0003	0.3098
μ	-1.4961	0.1694	-1.8295	-1.4915	-1.1839
$\beta 1$	-0.1166	0.1222	-0.3573	-0.1164	0.1213
$\beta 2$	-0.0127	0.0110	-0.0344	-0.0127	0.0086
$\beta 3$	-0.3695	0.0926	-0.5537	-0.3687	-0.1905
$\beta 4$	-0.1326	0.0970	-0.3253	-0.1326	0.0565
$\beta 5$	0.5811	0.2598	0.0581	0.5865	1.0757
$\beta 6$	-0.0169	0.0038	-0.0244	-0.0169	-0.0097
$\beta 7$	-0.0002	0.0002	-0.0006	-0.0002	0.0001

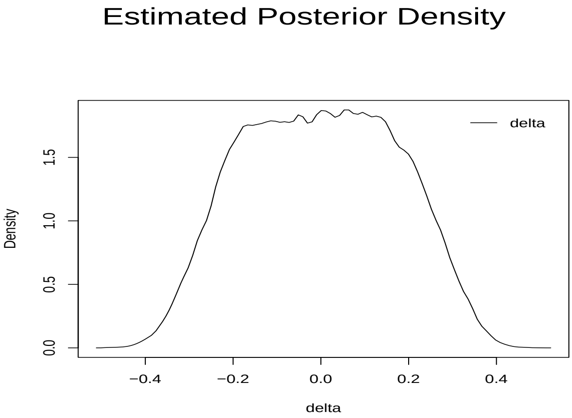

Figure : Plots of δ

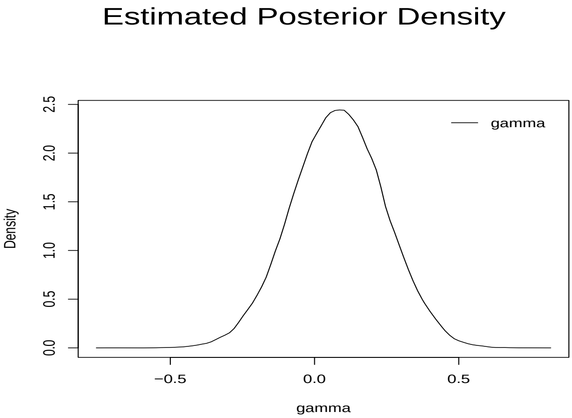

Figure : Plots of γ

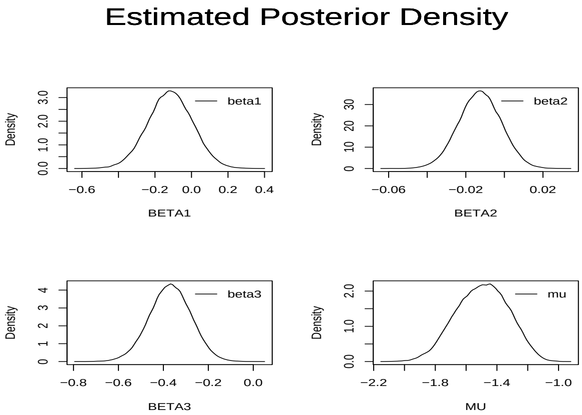

Figure : Plots of μ and β's

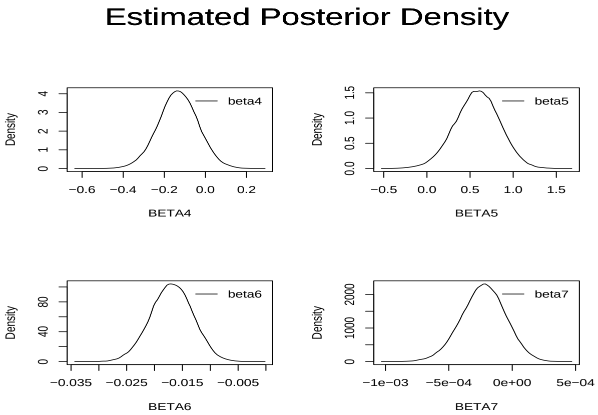

Figure : Plots of β's

- The presence of disease might depend only on the previous response.
- Need not to consider skew parameter of link function for ICHS.
- Consider model comparisons with different link functions
- Simulation work will be done.

Thank you for your attentions.