

Some Observations in Likelihood Based Fitting of Longitudinal Models for Binary Data

Subir Ghosh and Arunava Chakravartty
University of California, Riverside

Abstract: Different models are used in practice for describing a binary longitudinal data. In this paper we consider the joint probability models, the marginal models, and the combined models for describing such data the best. The combined model consists of a joint probability model and a marginal model at two different levels. We present some striking empirical observations on the closeness of the estimates and their standard errors for some parameters of the models considered in describing a data from Fitzmaurice and Laird (1993) and consequently giving new insight from this data. We present the data in a complete factorial arrangement with 4 factors at 2 levels. We introduce the concept of “data representing a model completely” and explain “data balance” as well as “chance balance”. We also consider the best model selection problem for describing this data and use the Search Linear Model concepts known in Fractional Factorial Design research (Srivastava (1975)).

Key words: Balance, binary data, data representing a model, deviance statistic, joint probabilities, likelihood, longitudinal data, marginal probabilities, profile explanatory matrix.

1. Introduction

We consider a longitudinal data with a binary response variable Y observed at t time points on each of the n individuals. A set of p explanatory variables X_1, X_2, \dots, X_p is also observed with the response variable Y . We denote the observations on Y and X_1, X_2, \dots, X_p for the i^{th} individual at time point j by $(y_{ij}; x_{ij1}, \dots, x_{ijp})$, $i = 1, \dots, n$, and $j = 1, \dots, t$. The value y_{ij} of Y for the i^{th} individual at the j^{th} time point is equal to 1 with probability μ_{ij} and is equal to 0 with probability $(1 - \mu_{ij})$. We further assume that our data are complete. For describing the data throughout this paper, we consider the observed values of Y at the fixed values of X_1, X_2, \dots, X_p . Consequently, the probability μ_{ij} depends on $(x_{ij1}, \dots, x_{ijp})' = \mathbf{x}_{ij}$. We also consider a bigger probability structure giving

μ_{ij} 's. The profile response of the i^{th} individual is denoted by \mathbf{Y}_i and its observed value is denoted by $\mathbf{y}_i = (y_{i1}, \dots, y_{it})'$. The observed profile explanatory matrix of the i^{th} individual is denoted by $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it})'$. We denote $P(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{x}_i) = \pi(\mathbf{y}_i, \mathbf{x}_i)$. The $\pi(\mathbf{y}_i, \mathbf{x}_i)$'s are the joint probabilities and the μ_{ij} 's are the marginal probabilities. The π 's and μ 's are naturally related to each other [Diggle *et al.* (2002), Dobson (2002), Fitzmaurice *et al.* (2004), Molenberghs and Verbeke (2005)].

The number of distinct vectors for \mathbf{y}_i , $i = 1, \dots, n$, is 2^t since $y_{ij} = 0$ or 1 . We denote one of possible \mathbf{y}_i vectors by $\mathbf{u} = (u_1, \dots, u_t)'$, where $u_j = 0$ or 1 , $j = 1, \dots, t$, and

$$n_{\mathbf{u}} = \{\text{number of } i \in \{1, \dots, n\} \text{ with } \mathbf{y}_i = \mathbf{u}\}. \quad (1.1)$$

The data are represented by the values of $n_{\mathbf{u}}$ for all \mathbf{u} . Table 1 in Section 2 gives an example with the data in terms of the values of $n_{\mathbf{u}}$. Section 3 presents the classes of joint probability models describing the dependence of $\log \pi(\mathbf{y}_i, \mathbf{x}_i)$ on \mathbf{y}_i . Section 4 explains the concepts of "data balance" and "chance balance". Moreover, the concept of "data representing a model completely" is introduced. Section 5 describes the best model selection for fitting the classes of models given in Section 3 to the data in Table 1 using the likelihood method. Section 6 presents the marginal models describing the dependence of $\log \pi(\mathbf{y}_i, \mathbf{x}_i)$ on \mathbf{x}_i . Section 7 describes the combined model

(Fitzmaurice and Laird (1993)) consisting of a joint probability model in Section 3 and a marginal model in Section 6. Section 8 presents some striking observations on the closeness of the estimates and their standard errors for some parameters of the joint probability models, the marginal models and the combined models. We have posted Tables 5 - 12 on our website¹.

2. An Example

We now consider a data from Fitzmaurice and Laird (1993) in Table 1 on page 145. This is a subset of Six Cities data from Ware *et al.*, 1984 on a longitudinal study of the health effects of air pollution. The n individuals for this study are 537 children from Steubenville, Ohio and thus $n = 537$. Each child was examined annually at ages 7 through 10. Therefore the time point j is the age $(6 + j)$, $j = 1, \dots, 4$ and $t = 4$. The response variable Y is the wheezing status : yes or no. The $y_{ij} = 1$ for yes and $y_{ij} = 0$ for no. Two explanatory variables ($p = 2$) are $Age - 9 = X_1$, Maternal Smoking (MS) = X_2 . The $x_{ij2} = 1$ for the i^{th} child and for all j if the mother smoked regularly since the first year of the study (time point 1 or equivalently $j = 1$) and $x_{ij2} = 0$ otherwise. The vector $\mathbf{u}' = (u_1, u_2, u_3, u_4)$

¹<http://www.statistics.ucr.edu/papers/ghosh.html>

and the $n_{\mathbf{u}}$ values are given in Table 1.

If the mother of the i^{th} child smoked regularly since the first year of the study, then the matrix \mathbf{x}_i is

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ \mathbf{x}'_{i3} \\ \mathbf{x}'_{i4} \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \quad (2.1)$$

If the mother of the i^{th} individual did not smoke since the first year of the study, then the matrix \mathbf{x}_i is

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ \mathbf{x}'_{i3} \\ \mathbf{x}'_{i4} \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ -1 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (2.2)$$

The observed profile response vector \mathbf{y}_i of the i^{th} individual is one of the 16 vectors \mathbf{u} listed in Table 1, with probability $P[\mathbf{Y}_i = \mathbf{y}_i | \mathbf{x}_i] = \pi(\mathbf{y}_i, \mathbf{x}_i)$ where \mathbf{x}_i is either in (2.1) and (2.2).

Table 1: The values of $n_{\mathbf{u}}$ when $x_{ij2} = 0$ and $x_{ij2} = 1$

| \mathbf{u}' | | | | $n_{\mathbf{u}}$ | | |
|---------------|-------|-------|--------|------------------|---------------|-------|
| $(u_1$ | u_2 | u_3 | $u_4)$ | $x_{ij2} = 0$ | $x_{ij2} = 1$ | Total |
| 0 | 0 | 0 | 0 | 237 | 118 | 355 |
| 0 | 0 | 0 | 1 | 10 | 6 | 16 |
| 0 | 0 | 1 | 0 | 15 | 8 | 23 |
| 0 | 0 | 1 | 1 | 4 | 2 | 6 |
| 0 | 1 | 0 | 0 | 16 | 11 | 27 |
| 0 | 1 | 0 | 1 | 2 | 1 | 3 |
| 0 | 1 | 1 | 0 | 7 | 6 | 13 |
| 0 | 1 | 1 | 1 | 3 | 4 | 7 |
| 1 | 0 | 0 | 0 | 24 | 7 | 31 |
| 1 | 0 | 0 | 1 | 3 | 3 | 6 |
| 1 | 0 | 1 | 0 | 3 | 3 | 6 |
| 1 | 0 | 1 | 1 | 2 | 1 | 3 |
| 1 | 1 | 0 | 0 | 6 | 4 | 10 |
| 1 | 1 | 0 | 1 | 2 | 2 | 4 |
| 1 | 1 | 1 | 0 | 5 | 4 | 9 |
| 1 | 1 | 1 | 1 | 11 | 7 | 18 |

3. Joint Probability Models

We first present a joint probability model introduced in Fitzmaurice and Liard(1993). We define for $h = 2, \dots, t$ and $t \geq 2$,

$$\begin{aligned}\Psi &= (\psi_1, \dots, \psi_t)', \\ \Omega_h &= (\omega_{12}, \dots; \omega_{123}, \dots; \omega_{12\dots h}, \dots, \omega_{(t-h+1)\dots t})', \\ \mathbf{y}_h^{(i)} &= (y_{i1}y_{i2}, \dots; y_{i1}y_{i2}y_{i3}, \dots; y_{i1}y_{i2}\dots y_{ih}, \dots, y_{i(t-h+1)}\dots y_{it})',\end{aligned}\quad (3.1)$$

where Ψ is a $(t \times 1)$ vector, Ω_h and $\mathbf{y}_h^{(i)}$ are $(l_h \times 1)$ vectors, $l_h = \binom{t}{2} + \dots + \binom{t}{h}$. The model is

$$\begin{aligned}M_h : \quad \log \pi(\mathbf{y}_i, \mathbf{x}_i) &= \psi_0 + \Psi' \mathbf{y}_i + \Omega_h' \mathbf{y}_h^{(i)} \\ &= \psi_0 + (\psi_1 y_{i1} + \dots + \psi_t y_{it}) \\ &\quad + (w_{12} y_{i1} y_{i2} + \dots + w_{123} y_{i1} y_{i2} y_{i3} + \dots \\ &\quad + w_{12\dots h} y_{i1} y_{i2} \dots y_{ih} + \dots \\ &\quad + w_{(t-h+1)\dots t} y_{i(t-h+1)} \dots y_{it}),\end{aligned}\quad (3.2)$$

where the constant ψ_0 depends on Ψ and Ω_h so that

$$\sum_{i=1}^n \pi(\mathbf{y}_i, \mathbf{x}_i) = 1. \quad (3.3)$$

The right hand side of (5) depends on i through \mathbf{y}_i and $\mathbf{y}_h^{(i)}$ but does not depend on \mathbf{x}_i . We present some special cases to explain the model clearly. For $t = 2$

$$M_2 : \log \pi(\mathbf{y}_i, \mathbf{x}_i) = \psi_0 + (\psi_1 y_{i1} + \psi_2 y_{i2}) + w_{12} y_{i1} y_{i2}.$$

For $t = 3$

$$\begin{aligned}M_2 : \log \pi(\mathbf{y}_i, \mathbf{x}_i) &= \psi_0 + (\psi_1 y_{i1} + \psi_2 y_{i2} + \psi_3 y_{i3}) \\ &\quad + (w_{12} y_{i1} y_{i2} + w_{13} y_{i3} + w_{23} y_{i2} y_{i3}), \\ M_3 : \log \pi(\mathbf{y}_i, \mathbf{x}_i) &= \psi_0 + (\psi_1 y_{i1} + \psi_2 y_{i2} + \psi_3 y_{i3}) \\ &\quad + (w_{12} y_{i1} y_{i2} + w_{13} y_{i3} + w_{23} y_{i2} y_{i3}) + w_{123} y_{i1} y_{i2} y_{i3}.\end{aligned}$$

We now generate models that are nested within M_h . For $t = 3$ and $h = 2$, the nested models within M_2 are

$$\begin{aligned}M_{121} : \log \pi(\mathbf{y}_i, \mathbf{x}_i) &= \psi_0 + (\psi_1 y_{i1} + \psi_2 y_{i2} + \psi_3 y_{i3}) + (w_{12} y_{i1} y_{i2}), \\ M_{122} : \log \pi(\mathbf{y}_i, \mathbf{x}_i) &= \psi_0 + (\psi_1 y_{i1} + \psi_2 y_{i2} + \psi_3 y_{i3}) + (w_{13} y_{i1} y_{i3}),\end{aligned}$$

$$M_{123} : \log \pi(\mathbf{y}_i, \mathbf{x}_i) = \psi_0 + (\psi_1 y_{i1} + \psi_2 y_{i2} + \psi_3 y_{i3}) + (w_{23} y_{i2} y_{i3}).$$

In describing the data by a model, we can never be sure whether one model describes the data better over the other possible models. For this reason, we now consider a class of models for describing the data. Our goal is to identify a model describing the data best (Srivastava (1975)).

Let $\boldsymbol{\Omega}_{khs}$ be a $(k \times 1)$ column vector whose elements are the s^{th} subset of k elements of $\boldsymbol{\Omega}_h$, $s = 1, \dots, \binom{l_h}{k}$. Let $\mathbf{y}_{khs}^{(i)}$ be a $(k \times 1)$ column vector whose elements are the s^{th} subset of the k elements of $\mathbf{y}_h^{(i)}$, $s = 1, \dots, \binom{l_h}{k}$. We now consider a class of $\binom{l_h}{k}$ models given below.

$$M_{khs} : \log \pi(\mathbf{y}_i, \mathbf{x}_i) = \psi_0 + \boldsymbol{\Psi}' \mathbf{y}_i + \boldsymbol{\Omega}'_{khs} \mathbf{y}_{khs}^{(i)}, \quad (3.4)$$

where $s = 1, \dots, \binom{l_h}{k}$ and ψ_0 depends on $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}_{khs}$.

When $h = 2$, we get $l_2 = \binom{2}{2}$. For $k = 1$, we have l_2 models M_{12s} ($s = 1, \dots, l_2$) in the class of our models. For $k = 2$, we have $\binom{l_2}{2}$ models M_{22s} ($s = 1, \dots, \binom{l_2}{2}$) in the class. It can be seen from (3.4) that for $h = 2$ and $k = 1, \dots, l_2$,

$$\begin{aligned} \omega_{12} = & \log \pi((0, 0, u_3, \dots, u_t), \mathbf{x}_i) + \log \pi((1, 1, u_3, \dots, u_t), \mathbf{x}_i) \\ & - \log \pi((0, 1, u_3, \dots, u_t), \mathbf{x}_i) - \log \pi((1, 0, u_3, \dots, u_t), \mathbf{x}_i), \end{aligned} \quad (3.5)$$

where the $(u_1, u_2, u_3, \dots, u_t)$'s are the distinct \mathbf{y}_i 's. Thus $\omega_{12} = 0$ if and only if

$$\begin{aligned} & \pi((0, 0, u_3, \dots, u_t), \mathbf{x}_i) \pi((1, 1, u_3, \dots, u_t), \mathbf{x}_i) \\ & = \pi((0, 1, u_3, \dots, u_t), \mathbf{x}_i) \pi((1, 0, u_3, \dots, u_t), \mathbf{x}_i), \end{aligned} \quad (3.6)$$

for all the 2^{t-2} vectors (u_3, \dots, u_t) . The (3.5) and (3.6) provide the relationship between an individual model parameter in ω_{khs} with the probabilities $\pi(\mathbf{y}_i, \mathbf{x}_i)$'s in (3.4). Similarly for $\omega_{jj'}$; $j, j' = 1, \dots, t, j < j'$.

A special case of (3.2) assuming $\boldsymbol{\Omega}_h = \mathbf{0}$ gives the model,

$$M_0 : \log \pi(\mathbf{y}_i, \mathbf{x}_i) = \psi_0 + \boldsymbol{\Psi}' \mathbf{y}_i, \quad (3.7)$$

where ψ_0 depends on $\boldsymbol{\Psi}$. We compare the model M_0 with the models M_{khs} in (3.4) as a first step in our model comparisons.

The log-likelihood for M_{khs} is given by

$$\begin{aligned} l_{khs} &= \sum_{i=1}^n \log \pi(\mathbf{y}_i, \mathbf{x}_i) \\ &= \sum_{\mathbf{u}} n_{\mathbf{u}} \log \pi(\mathbf{u}, \mathbf{x}), \end{aligned} \quad (3.8)$$

where \mathbf{u} is one of the 2^t distinct \mathbf{y}_i 's; $n_{\mathbf{u}}$ defined in (1.1) is the number of times the response vector \mathbf{u} occurs among the n profile response vectors \mathbf{y}_i , $i = 1, \dots, n$; and \mathbf{x} is the \mathbf{x}_i corresponding to \mathbf{y}_i . The maximum value of l_{khs} in the method of maximum likelihood is denoted by \hat{l}_{khs} . Similarly, the log-likelihood for M_0 and its maximum value are denoted by l_0 and \hat{l}_0 , respectively. The deviance statistic is given by

$$D_{khs} = 2 [\hat{l}_{khs} - \hat{l}_0], \quad (3.9)$$

which is approximately χ^2 distributed with k degrees of freedom (df). The higher the value of D_{khs} ; $h = 2, \dots, t$, $s = 1, \dots, \binom{h}{k}$, $k = 1, \dots, l_h$; means better the description of the data by the model M_{khs} . We will consider only the D_{khs} values that are greater than $\chi_{\alpha; k}^2$, the upper 100α percentage point of the χ^2 distribution with k df . When one model $M_{k_1 h_1 s_1}$ is nested within another $M_{k_2 h_2 s_2}$, where $k_2 > k_1$, we consider the difference,

$$D_{k_2 h_2 s_2} - D_{k_1 h_1 s_1} = 2 [\hat{l}_{k_2 h_2 s_2} - \hat{l}_{k_1 h_1 s_1}], \quad (3.10)$$

which is approximately χ^2 distributed with $(k_2 - k_1)$ df . When $D_{k_2 h_2 s_2} - D_{k_1 h_1 s_1}$ is greater than $\chi_{\alpha; (k_2 - k_1)}^2$, the model $M_{k_2 h_2 s_2}$ is considered significantly different from $M_{k_1 h_1 s_1}$ in describing the data at the α level of significance.

4. Data Representation

In the log-likelihood l_{khs} in (3.8) for M_{khs} , we assume,

$$\max_{\mathbf{u}} \pi(\mathbf{u}, \mathbf{x}) = \pi(\mathbf{w}, \mathbf{x}). \quad (4.1)$$

Note that $\pi(\mathbf{u}, \mathbf{x})$ in (3.4) depends on \mathbf{u} but not on \mathbf{x} . Also \mathbf{w} is one of the 2^t possible \mathbf{u} 's. For brevity, we write $\pi(\mathbf{u}, \mathbf{x}) = \pi_{\mathbf{u}}$. We have

$$n_{\mathbf{w}} = n - \sum_{\mathbf{u} \neq \mathbf{w}} n_{\mathbf{u}}. \quad (4.2)$$

The $n_{\mathbf{u}}$ values are given in the column "Total" of Table 1. It can be checked from (3.8) that

$$l_{khs} = n \log \pi_{\mathbf{w}} - \sum_{\mathbf{u} \neq \mathbf{w}} n_{\mathbf{u}} (\log \pi_{\mathbf{w}} - \log \pi_{\mathbf{u}}). \quad (4.3)$$

From (4.1) we have $(\log \pi_{\mathbf{w}} - \log \pi_{\mathbf{u}}) \geq 0$ for all \mathbf{u} and hence

$$l_{khs} \leq n \log \pi_{\mathbf{w}}. \quad (4.4)$$

In (4.4), the equality holds for all values of $n_{\mathbf{u}}$'s satisfying $\sum_{\mathbf{u}} n_{\mathbf{u}} = n$ if and only if the $\pi_{\mathbf{u}}$'s are all equal. We now assume

$$\max_{\mathbf{u}} n_{\mathbf{u}} = n_{\mathbf{v}}. \quad (4.5)$$

Then

$$n_{\mathbf{v}} = n - \sum_{\mathbf{u} \neq \mathbf{v}} n_{\mathbf{u}}, l_{khs} = n \log \pi_{\mathbf{v}} - \sum_{\mathbf{u} \neq \mathbf{v}} n_{\mathbf{u}} (\log \pi_{\mathbf{w}} - \log \pi_{\mathbf{u}}) - \sum_{\mathbf{u} \neq \mathbf{v}} n_{\mathbf{u}} (\log \pi_{\mathbf{v}} - \log \pi_{\mathbf{w}}). \quad (4.6)$$

The last term on the right hand side of (4.6) vanishes when $\mathbf{v} = \mathbf{w}$. The situation ' $\mathbf{v} = \mathbf{w}$ ' is naturally of interest because the data represents the model in terms of the maximum $\pi_{\mathbf{u}}$'s. The "data represents the model more completely" if the \mathbf{u} 's give the same ordering of $n_{\mathbf{u}}$'s and $\pi_{\mathbf{u}}$'s from their smallest value to their largest value.

In most of our analyses, we assume that the data represents a model completely and we draw inferences under that particular model. In reality, we do not know this model but we are able to postulate a class of models so that one of them is better represented by the data than the others. The problem is to search and identify a model within the postulated class which is best represented by the data (Srivastava (1975), Shirakura, Takahashi, and Srivastava (1996), Ghosh and Teschmacher (2002)). The postulated class of models in this paper consists of the models M_{khs} defined in (3.4) and the data are the $n_{\mathbf{u}}$ values. We draw inferences on the elements of Ψ and Ω_h from the $n_{\mathbf{u}}$ values.

When the values of $n_{\mathbf{u}}$'s are all equal, we describe the situation as a "data balance" and otherwise a "data imbalance". A measure of the data imbalance is

$$s_n^2 = \frac{1}{2^t - 1} \sum_{\mathbf{u}} (n_{\mathbf{u}} - \bar{n})^2, \quad (4.7)$$

where $\bar{n} = \frac{\sum_{\mathbf{u}} n_{\mathbf{u}}}{2^t}$. A higher value of s_n^2 indicates a greater data imbalance. We calculate the numerical value of s_n^2 for measuring the degree of imbalance in the data. When the $\pi_{\mathbf{u}}$'s are all equal, we describe the situation as a "chance balance" or otherwise a "chance imbalance". A measure of chance imbalance is

$$s_{\pi}^2 = \frac{1}{2^t - 1} \sum_{\mathbf{u}} (\pi_{\mathbf{u}} - \bar{\pi})^2, \quad (4.8)$$

where $\bar{\pi} = \frac{\sum_{\mathbf{u}} \pi_{\mathbf{u}}}{2^t}$. A higher value of s_{π}^2 indicates a greater chance imbalance. Since the $\pi_{\mathbf{u}}$'s are all unknown, we cannot calculate the numerical value of s_{π}^2 , but we can estimate it based on the values of $n_{\mathbf{u}}$. The effects "chance imbalance" and "data imbalance" can be seen in (4.3) and (4.4) for maximizing the

log-likelihood $l_{kh,s}$.

5. Best Models

In this section, we fit the class of models M_{k2s} to the data given in Table 1. We have $t = 4$ and $l_2 = \binom{4}{2} = 6$. For $k = 1$, there are $\binom{l_2}{1} = 6$ models M_{12s} , $s = 1, \dots, 6$. We compare all these 6 models with M_0 to obtain the numerical values of the deviance statistic D_{12s} , $s = 1, \dots, 6$ given in Table 2. All the values of D_{12s} in Table 2 are greater than the critical value $\chi_{0.05;1}^2$ which is equal to 3.84. The model that gives the highest value of D_{12s} is chosen. The subset $s = 4$ with elements (2, 3) gives the highest value of D_{12s} . Hence the model M_{124} provides the best description for the data within the class of models M_{12s} , $s = 1, \dots, 6$. In Table 3, we present the model providing the best description of the data within the class of models M_{k2s} , $s = 1, \dots, \binom{l_2}{k}$, for $k = 1, \dots, l_2$. The best model for k is nested within the best model for k' ; $k, k' = 1, \dots, 6$, $k < k'$.

Table 4 presents the numerical values of $D_{(k+1)2s} - D_{k2s}$, $k = 1, \dots, (l_2 - 1)$. Three models M_{42s} with the elements in the subset (23, 34, 12, 14), M_{52s} with (23, 34, 12, 14, 24), and M_{62s} with (23, 34, 12, 14, 24, 13) are definitely the top candidates to be considered. The model M_{42s} with the elements in the subset (23, 34, 12, 14) provides an adequate description of the data within the class of models M_{k2s} , $s = 1, \dots, \binom{l_2}{k}$ for $k = 1, \dots, l_2$ and $l_2 = 6$. Table 5 (Ghosh and Chakravartty (2007)) presents the estimate (E), standard error (SE), observed χ^2 ($(E/SE)^2$), and p-value for the parameters of the best models in Table 3. The best models within the k classes of models M_{k2s} , $s = 1, \dots, \binom{6}{k}$ for $k = 1, 2, 3, 4$ have all ψ parameters and ω parameters highly significant at $\alpha = 0.05$ and $\alpha = 0.01$. The best models within the k classes of models M_{k2s} , $s = 1, \dots, \binom{6}{k}$ for $k = 5$ and 6 have all ψ parameters and ($\omega_{23}, \omega_{34}, \omega_{12}, \omega_{14}$) parameters significant at both $\alpha = 0.05$ and $\alpha = 0.01$. Moreover, the parameters ω_{24} for $k = 5$, ω_{24} and ω_{13} for $k = 6$ are significant at $\alpha = 0.05$ but not significant at $\alpha = 0.01$.

Table 2: The values of D_{12s} , $s = 1, \dots, 6$

| Subset s | Elements of subset s | D_{12s} |
|------------|------------------------|-----------|
| 1 | 1 2 | 54.42 |
| 2 | 1 3 | 41.36 |
| 3 | 1 4 | 44.12 |
| 4 | 2 3 | 83.26 |
| 5 | 2 4 | 45.18 |
| 6 | 3 4 | 58.50 |

Table 3: Best models within M_{k2s} , $s = 1, \dots, \binom{l_2}{k}$, for $k = 1, \dots, l_2$

| k | Elements of subset s for the best model | D_{12s} | $\chi_{0.05,k}^2$ |
|-----|--|-----------|-------------------|
| 1 | 2 3 | 83.26 | 3.84 |
| 2 | 2 3, 3 4 | 141.76 | 5.99 |
| 3 | 2 3, 3 4, 1 2 | 196.18 | 7.81 |
| 4 | 2 3, 3 4, 1 2, 1 4 | 225.64 | 9.49 |
| 5 | 2 3, 3 4, 1 2, 1 4, 2 4 | 230.14 | 11.07 |
| 6 | 2 3, 3 4, 1 2, 1 4, 2 4, 1 3 | 235.34 | 12.59 |

Table 4: The numerical values of $D_{(k+1)2s} - D_{k2s}$, $k = 1, \dots, (l_2 - 1)$

| k | $D_{(k+1)2s} - D_{k2s}$ | p -value |
|-----|-------------------------|------------|
| 1 | 58.50 | 0.0000 |
| 2 | 54.42 | 0.0000 |
| 3 | 29.46 | 0.0000 |
| 4 | 4.50 | 0.0339 |
| 5 | 5.20 | 0.0226 |

6. Marginal Models

We now consider three marginal models with the presence of x_{ij1} for describing the data. The first marginal model MM_1 is

$$MM_1 : \log \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_{12} x_{ij1} x_{ij2}, \quad (6.1)$$

where the β 's are unknown parameters; $(x_{i11}, x_{i21}, x_{i31}, x_{i41})' = (2, -1, 0, 1)'$; $x_{ij2} = 1$ if the mother smoked and $x_{ij2} = 0$ otherwise. The maximum log-likelihood for MM_1 is denoted by \hat{l}_{MM_1} where $-2\hat{l}_{MM_1} = 1819.4800$.

The second marginal model MM_2 is

$$MM_2 : \log \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} \quad (6.2)$$

The model MM_2 is nested within the model MM_1 . The maximum log-likelihood for MM_2 is denoted by \hat{l}_{MM_2} where $-2\hat{l}_{MM_2} = 1819.8894$. The deviance statistic

$$D_{MM12} = 2 \left[\hat{l}_{MM_1} - \hat{l}_{MM_2} \right], \quad (6.3)$$

which is approximately χ^2 distributed with 1 df. The observed $D_{MM12} = 0.4094$ and $\chi_{0.05;1}^2 = 3.84$. Hence there is a strong evidence from the data that the

parameter β_{12} in MM_1 is not significantly different from zero at $\alpha = 0.05$. Consequently, the model MM_2 provides an adequate description of the data when we compare it with the model MM_1 .

The third marginal model MM_3 is

$$MM_3 : \log \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 x_{ij1} \quad (6.4)$$

The maximum log-likelihood for MM_3 is denoted by \hat{l}_{MM_3} where $-2\hat{l}_{MM_3} = 1824.6820$. The deviance statistic for comparing MM_2 with MM_3 is given by

$$D_{MM23} = 2 \left[\hat{l}_{MM_2} - \hat{l}_{MM_3} \right], \quad (6.5)$$

which is approximately χ^2 distributed with 1 df. The observed D_{MM23} is 4.7926 and the p-value is 0.0285. The β_2 in MM_2 is significant at $\alpha = 0.05$ and not significant at $\alpha = 0.01$, by the likelihood ratio test. The model MM_2 provides a better description of the data when we compare it with the model MM_3 at $\alpha = 0.05$. The model MM_3 provides an adequate description of the data when we compare it with the model MM_2 at the $\alpha = 0.01$.

Table 6 presents the E , SE , observed $\chi^2 ((E/SE)^2)$, and p-value for the β parameters in MM_1 , MM_2 and MM_3 . For the model MM_1 , the parameters β_0 and β_1 are significant at $\alpha = 0.05$ and only β_0 is significant at $\alpha = 0.01$. However, the parameter β_1 is not barely significant at $\alpha = 0.01$ for MM_1 . For both the models MM_2 and MM_3 , the parameters β_0 and β_1 are significant at $\alpha = 0.01$. However, the parameter β_1 is barely significant at $\alpha = 0.01$ for both MM_2 and MM_3 .

7. Combined Models

Fitzmaurice and Laird (1993) proposed a combined model simultaneously consisting of a joint probability model and a marginal model. We now consider 18 Fitzmaurice-Laird models consisting simultaneously of one of the 6 best models in Table 3 and one of MM_1 , MM_2 and MM_3 . We then apply the general likelihood based approach introduced in Fitzmaurice and Laird (1993) for drawing inferences on the β (marginal mean) parameters and ω (the conditional association) parameters. We present the E , and the corresponding robust SE (rSE) for the β and ω parameters and their corresponding robust χ^2 ($r\chi^2$) and robust p -value ($r p$ -value) described in the Fitzmaurice-Laird approach.

We observe from Table 9 (Ghosh and Chakravartty (2007)) that the 9 Fitzmaurice-Laird models for (k, MM_i) , $k = 4, 5, 6$ and $i = 1, 2, 3$, are comparable to each other. From Table 7 (Ghosh and Chakravartty (2007)), it is clear that the 3 models, (k, MM_3) , $k = 4, 5, 6$, are the best for describing the data. The model for

($k = 4, MM_3$) provides an adequate description of the data in comparison to the other models for ($k = 5, MM_3$) and ($k = 6, MM_3$).

8. Results and Conclusions

In our calculations presented in Tables 5, 6 and 7 (Ghosh and Chakravartty (2007)), we observe some striking empirical results.

Observation 1. In Table 5 (Ghosh and Chakravartty (2007)), the E and SE values of the parameters ψ_1, ψ_2 and ω_{23} when $k = 1$ are exactly the same as the corresponding values when $k = 2$. The E and SE values of ω_{23} are exactly the same for $k = 1, 2$ and 3 . Consequently, the observed χ^2 and p -values values are also the same.

Observation 2. The E and SE values of $\hat{\beta}$ in Table 6 (Ghosh and Chakravartty (2007)) are very close to the corresponding E and rSE values in Table 7 for $MM_u, u = 1, 2, 3$, and $k = 1, \dots, 6$. Table 10 highlights this similarity.

Observation 3. The E and SE values of $\hat{\omega}$ in Table 5 (Ghosh and Chakravartty (2007)) are close to the corresponding E and rSE values in Table 7 (Ghosh and Chakravartty (2007)) for $MM_u, u = 1, 2, 3$, and $k = 1, \dots, 6$. Table 11 highlights this similarity.

Observation 4. The E and rSE values of $\hat{\omega}_{23}$ and $\hat{\omega}_{34}$ in Table 7 (Ghosh and Chakravartty (2007)) for ($k = 2, MM_u$) are very close to the corresponding values for ($k = 3, MM_u$), $u = 1, 2, 3$. Table 12 highlights this similarity.

Fitzmaurice and Laird (1993) demonstrated that $\hat{\beta}$'s and $\hat{\omega}$'s are uncorrelated to each other. Observations 2 and 3 can be explained from this important fact of uncorrelatedness. In other words, $\hat{\beta}$'s and $\hat{\omega}$'s for a combined model are expected to be close to $\hat{\beta}$'s for the marginal model and $\hat{\omega}$'s for the joint probability model separately. Observations 1 and 4 are original in this paper and call for further research.

References

- Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. 2nd ed. Oxford University Press.
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models* 2nd ed. Chapman & Hall

- Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004). *Applied Longitudinal Analysis*. Wiley.
- Fitzmaurice, G. M. and Laird, N. M. (1993). *A likelihood-based method for analysing longitudinal binary responses*. *Biometrika* **80**, 141-151.
- Ghosh, S. and Chakravartty, A. (2007). Tables 5-12: Some observations in Likelihood based fitting of longitudinal models for binary data.²
- Ghosh, S. and Teschmacher, L. (2002). Comparisons of search designs using search probabilities. *Journal of Statistical Planning and Inference* **104**, 439-458.
- Molenberghs, G. and Lesaffre, E. (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine* **18**, 2237-2255.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.
- Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*. Chapman & Hall.
- Shirakura, T., Takahashi, T., and Srivastava, J. N. (1996). Searching probabilities for nonzero effects in search designs for the noisy case. *Annals of Statistics* **24**, 2560-2568.
- Srivastava, J. N. (1975). Designs for searching non-negligible effects. In *a Survey of Statistical Design and Linear Models* (Edited by J. N. Srivastava), 507-520. North-Holland.
- Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **44**, 1049-1060.

Received October 1, 2007; accepted November 30, 2007.

²<http://www.statistics.ucr.edu/papers/ghosh.html>.

Subir Ghosh
Department of Statistics
University of California, Riverside
Riverside, CA 92521, USA
subir.ghosh@ucr.edu

Arunava Chakravartty
Department of Statistics
University of California
Riverside, CA 92521, USA

current address
Genomics Informatics – Statistics
Eli Lilly and Company
Indianapolis, IN 46285, USA
charkrvarty_arunava@lili.com