J. Statist. Comput. Simul., 2002, Vol. 72(7), pp. 549–564

Taylor & Francis
Taylor & Francis Group

A GENERALIZATION OF THE LOGISTIC LINEAR MODEL

JOHN S. J. HSU*

Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA

(Received 6 October 2000; In final form 9 January 2002)

Consider the logistic linear model, with some explanatory variables overlooked. Those explanatory variables may be quantitative or qualitative. In either case, the resulting true response variable is not a binomial or a beta-binomial but a sum of binomials. Hence, standard computer packages for logistic regression can be inappropriate even if an overdispersion factor is incorporated. Therefore, a discrete exponential family assumption is considered to broaden the class of sampling models. Likelihood and Bayesian analyses are discussed. Bayesian computation techniques such as Laplacian approximations and Markov chain simulations are used to compute posterior densities and moments. Approximate conditional distributions are derived and are shown to be accurate. The Markov chain simulations are performed effectively to calculate posterior moments by using the approximate conditional distributions. The methodology is applied to Keeler's hardness of winter wheat data for checking binomial assumptions and to Matsumura's Accounting exams data for detailed likelihood and Bayesian analyses.

Keywords: Gibbs sampling; Laplacian approximation; Metropolis-Hastings algorithm; Beta-binomial model; Discrete exponential family model; Logistic regression

1 INTRODUCTION

Let $Y_1, Y_2, \ldots, Y_n$ be n independent random variables corresponding to successes out of $m_1, m_2, \ldots, m_n$ trials in n different groups, and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote the corresponding $(q+1) \times 1$ design vectors. The standard logistic linear model assumes that, given $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{iq})^T$, the random variable Y_i possesses a binomial distribution with parameters m_i and p_i, where p_i denotes the probability of success for that group. The standard logistic linear model is defined as

$$\operatorname{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_q x_{iq},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_q)^T$ is a vector of $q+1$ unknown parameters. Statistical inferences such as estimation, hypothesis testing, prediction can be performed using standard computer packages, such as SAS and S-PLUS. However, it is quite often that some explanatory variables are overlooked in the analysis. In such cases, the variable Y_i is no longer

* Tel.: 805-893-4055; Fax: 805-893-2334; E-mail: hsu@pstat.ucsb.edu

ISSN 0094-9655 print; ISSN 1563-5163 online
DOI: 10.1080/009496502900144218

a binomial but a sum of binomials. For example, we consider a dose-response model, where the response is the number of insects killed and the true model consists of two explanatory variables, dosage levels and gender. Suppose that the experimenter overlooked the factor of gender. Then the resulting true response is not a binomial but the sum of two binomials, one for male and the other for female. The sum of binomials may be well approximated by a single binomial in some cases but not in general. Hsu *et al.* (1991) suggested a discrete p-parameter exponential family model to approximate the sum of binomials. They assume that the observations $y_1, y_2, \ldots, y_n$ are independent and the probability $p(y_i = j)$ satisfies

$$p(y_i = j) = \frac{e^{\lambda_i(j)}}{\sum_{h=0}^{m_i} e^{\lambda_i(h)}} \qquad (j = 0, 1, \ldots, m_i), \tag{1}$$

where the multivariate logits $\lambda_i(j)$ satisfy

$$\lambda_i(j) = \log {}^{m_i}C_j + \gamma_1 j + \gamma_2 j^2 + \cdots + \gamma_p j^p \tag{2}$$

with

$${}^{m_i}C_j = \frac{m_i!}{j!(m_i - j)!}.$$

A polynomial model is used in Eq. (2) for the logits, in the spirit of Bock (1972) and others. The model defined in Eqs. (1) and (2) is flexible and the fit to the data can be improved by increasing the number of parameters in the model. Therefore, the model provides an effectively nonparametric fit to a discrete distribution. The explicit parameters in Eq. (2) are not meaningful on their own, since the model is motivated toward a reasonable fit to the data. However, many parameters of interest can be expressed as a function of the parameters in Eq. (2). The probability $p(y_i = j)$ in Eq. (1) is an example of such parameters.

Extending the model in (1) and (2) to the cases where the covariates $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are considered, we assume that the $y_1, y_2, \ldots, y_n$ are independent, and that given $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)^T$ and $\gamma = (\gamma_2, \gamma_2, \ldots, \gamma_p)^T$, the ith group response y_i possesses probability mass function

$$p(y_i = j \mid \boldsymbol{\beta}, \gamma, \mathbf{x}_i) = \frac{e^{\lambda_i(j)}}{\sum_{h=0}^{m_i} e^{\lambda_i(h)}} \quad (j = 0, 1, \ldots, m_i), \tag{3}$$

where the multivariate logits $\lambda_i(j)$ satisfy

$$\lambda_i(j) = \log {}^{m_i}C_j + j\mathbf{x}_i^T\boldsymbol{\beta} + \gamma_2 j^2 + \gamma_3 j^3 + \cdots + \gamma_p j^p. \tag{4}$$

The model defined in (3) and (4) consists of two parts: If the parameters $\gamma_2, \gamma_3, \ldots, \gamma_p$ are set equal to zero, then (3) and (4) provide the logistic linear model, under the binomial sampling assumption of y_i, with sample size m_i and probability of success $e^{\mathbf{x}_i^T\boldsymbol{\beta}}/(1 + e^{\mathbf{x}_i^T\boldsymbol{\beta}})$; if the parameters $\beta_1, \beta_2, \ldots, \beta_q$ are set equal to zero, then we have a polynomial model for the logits, and the model provides the discrete p-parameter exponential family model as described in Eqs. (1) and (2).

Therefore, Eq. (4) could be used to investigate the deviations from the logistic linear model. The model in (3) and (4) is now referred to as the p-parameter discrete exponential family logistic linear model. The model provides an alternative to the beta-binomial approach (Williams, 1975, 1982; Leonard and Novick, 1986; Prentice and Barlow, 1988), which adds a single extra parameter to the logistic linear model. The latter handles overdispersion, but does not address other deviations from the binomial assumption.

When p is specified, the likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ given $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, under assumptions (3) and (4), is

$$l(\boldsymbol{\theta} \mid \mathbf{y}) = l(\boldsymbol{\beta}, \gamma \mid \mathbf{y}) = \exp\left\{ s_0 + \mathbf{s}_1^T \boldsymbol{\beta} + \mathbf{s}_2^T \gamma - \sum_{i=1}^{n} D_i(\boldsymbol{\beta}, \gamma) \right\}, \tag{5}$$

where,

$$s_0 = \sum_{i=1}^{n} \log {}^{m_i}C_{y_i}, \tag{6}$$

$$\mathbf{s}_1 = \sum_{i=1}^{n} y_i \mathbf{x}_i, \tag{7}$$

$$\mathbf{s}_2 = \left(\sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} y_i^3, \ldots, \sum_{i=1}^{n} y_i^p \right)^T, \tag{8}$$

and

$$D_i(\boldsymbol{\beta}, \gamma) = \log \sum_{h=0}^{m_i} {}^{m_i}C_h \exp(h \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{u}_h^T \gamma), \tag{9}$$

with

$$\mathbf{u}_h = (h^2, h^3, \ldots, h^p)^T. \tag{10}$$

Since the ith response y_i can be reinterpreted as polychotomous, we have a generalized linear model for the logits of a multinomial distribution with $m_i + 1$ cells and unit sample size. Consequently, the parameters in (4) can be estimated and the model can be analyzed using standard computer packages. Furthermore, strong consistency and asymptotic normality will hold for our maximum likelihood estimates, as $n \to \infty$, with m_i, p and q fixed, with $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ concentrated on a bounded region and $\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$ remaining positive definite as $n \to \infty$. See Chiu *et al.* (1996) for a related asymptotic development.

When p is not specified, we may choose p to maximize the generalized information criterion (GIC)

$$\text{GIC} = L_{p+q} - \frac{1}{2}\alpha(p+q),$$

where L_{p+q} is the logarithm of the likelihood (5), evaluated at the $p + q$ maximum likelihood estimates, and α represents a penalty per parameter included in the model. Commonly used penalties are $\alpha = 2$, which leads to Akaike's information criterion (Akaike, 1978):

$$\text{AIC} = L_{p+q} - (p+q),$$

and $\alpha = \log_e n$, which leads to Schwarz's information criterion (Schwarz, 1978):

$$\mathrm{BIC} = L_{p+q} - \frac{1}{2}(p+q)\log_e n.$$

Information criteria have been discussed and compared in many papers. Please see Akaike (1978), Schwarz (1978), Stone (1977, 1979), Atilgan (1983), Shibata (1981), Thompson (1978a, 1978b) and Nishi (1984) for details.

The model in (3) and (4) may be checked via a chi-square statistic. Let $e_i(\boldsymbol{\beta}, \gamma)$ and $v_i(\boldsymbol{\beta}, \gamma)$ respectively denote the mean and variance of the distribution in (3), and let $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$ denote the maximum likelihood vectors of $\boldsymbol{\beta}$ and γ. Then, with $\hat{e}_i = e_i(\hat{\boldsymbol{\beta}}, \hat{\gamma})$ and $\hat{v}_i = v(\hat{\boldsymbol{\beta}}, \hat{\gamma})$, the model in (3) and (4) may be tested by referring the statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{e}_i)^2}{\hat{v}_i} \tag{11}$$

to the tables of the chi-square distribution with $n - q - p - 1$ degrees of freedom.

2 THE KEELER DATA, AN ILLUSTRATIVE EXAMPLE

The data of Table I are a subset of an experiment conducted by Keeler (1985). They performed an experiment to determine the hardness of two strains of winter wheat, Norstar and Frederick to thermal stress. Plants were cooled to a predetermined temperature and were then removed to a growth room to determine survival by regrowth. The predetermined temperatures were reported in column 1. The number of dead plants and the number of plants on test for varieties Norstar and Frederick are in columns 2–3 and 4–5, respectively. We fit the data using the logistic linear model:

$$\mathrm{logit}(p_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \tag{12}$$

where, p_i is the proportion of dead plants, $x_{i1} = \log(-Temp)$ and, $x_{2i} = 1$ if the variety Norstar was used and $x_{2i} = 0$ otherwise, for the ith temperature-variety combination group. The maximum likelihood estimate $\hat{\beta}_2$ of β_2 is 2.3808, with a standard error of 0.2676. This indicates that variety is an important factor in the analysis since $\hat{\beta}_2$ is more than eight standard errors from zero. For illustrative purpose, we suppose that the important factor, variety, was wrongly ignored. In such case, only the numbers combining the two varieties would be used and are reported in Table I (column 6 for number of dead plants and 7 for total plants on test). In such case, the dead plants in each temperature group is

TABLE I The Keeler Data.

	Norstar		*Frederick*		*Combined*	
Temperature	*Dead*	*Plants on Test*	*Dead*	*Plants on Test*	*Dead*	*Plants on Test*
−10 C	1	41	1	40	2	81
−12 C	1	41	15	41	16	82
−14 C	2	41	36	43	38	84
−16 C	7	41	40	40	47	81
−18 C	27	41	40	40	67	81
−20 C	39	42	40	40	79	82

the sum of two binomials, where one for Norstar and the other for Frederick. Nevertheless, we fit the standard logistic linear model

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{i1}. \tag{13}$$

The likelihood ratio chi-square value for goodness of fit was $\chi^2 = 7.04$, with 4 degrees of freedom, and the p-value was 0.1340. The goodness-of-fit chi-square test does not suggest the inadequacy of using model (13), while an important factor, variety, was wrongly ignored. We then fit data to the p-parameter discrete exponential family logistic linear model in (3) and (4) with the multivariate logits $\lambda_i(j)$ satisfy

$$\lambda_i(j) = \log {}^{m_i}C_j + (\beta_0 + \beta_1 x_{i1})j + \gamma_2 j^2 + \gamma_3 j^3 + \cdots + \gamma_p j^p.$$

The maximized log-likelihoods are -15.4945, -13.6520, and -12.7583 for $p = 1, 2$, and 3, respectively. Therefore, both AIC and BIC pick $p = 2$ and prefer a 2-parameter discrete exponential family logistic linear model to a simple logistic linear model ($p = 1$). Furthermore, the maximum likelihood estimate $\hat{\gamma}_2$ of γ_2 was -0.0568 with a standard error of 0.0160. This suggests that γ_2 is different from zero, since $\hat{\gamma}_2$ is more than three standard errors from zero, hence refuting the logistic linear model (13). This example shows that model in (3) and (4) can be used as a useful tool for testing the adequacy of the logistic linear model assumptions.

3 SIMULATION RESULTS

Three cases were considered for the study and one thousand simulations were performed in each case. For all cases, we considered a logistic linear regression with two explanatory variables. We assume that one of the explanatory variables was binary and was wrongly omitted from the study. The resulting response y_i for the ith group is in fact a sum of two binomials, where each of them corresponds to the response of one of the two subgroups classified according to the binary explanatory variable. Let m_1 and m_2 be the subgroup sizes and p_1 and p_2 be the probabilities of success for the two groups, respectively, and

$$\text{logit}(p_1) = \alpha_0 + \alpha_1 x, \tag{14}$$

and

$$\text{logit}(p_2) = \beta_0 + \beta_1 x. \tag{15}$$

In each simulation, $n = 100$ pairs of (y_1, y_2) were simulated, where y_1, and y_2 were simulated according to logistic regression functions (14) and (15), respectively, with a common x value and a common subsample size $m_1 = m_2 = 10$. The x values were $0.1, 0.2, \ldots, 0.9, 1.0$ and were repeated ten times for each simulation. In the absence of the binary explanatory variable which appeared in the true model, only the total $y = y_1 + y_2$ was recorded.

Three cases (C1) $\alpha_0 = 2, \alpha_1 = 1, \beta_0 = 2, \beta_1 = 1$; (C2) $\alpha_0 = -2, \alpha_1 = 1, \beta_0 = 2, \beta_1 = 1$; (C3) $\alpha_0 = -2, \alpha_1 = 1, \beta_0 = 4, \beta_1 = 1$, and four models (M1) logistic linear regression model; (M2) beta-binomial regression model; (M3) 2-parameter discrete exponential family logistic linear model; (M4) 3-parameter discrete exponential family logistic linear model were studied and compared. Table II presents the average of the maximized log-likelihoods

TABLE II Simulation Study.

Model	*Parameters used*	*Cases chosen by AIC*	*Average maximum log-likelihood*
Case C1: $\alpha_0 = 2, \alpha_1 = 1, \beta_0 = 2, \beta_1 = 1$			
M1	2	837	−148.2156
M2	3	34	−148.0394
M3	3	116	−147.7164
M4	4	13	−147.6520
Case C2: $\alpha_0 = -2, \alpha_1 = 1, \beta_0 = 2, \beta_1 = 1$			
M1	2	0	−194.0016
M2	3	0	−194.0022
M3	3	759	−180.1360
M4	4	241	−179.4593
Case C3: $\alpha_0 = -2, \alpha_1 = 1, \beta_0 = 4, \beta_1 = 1$			
M1	2	0	−188.0875
M2	3	0	−188.0882
M3	3	337	−165.6953
M4	4	663	−165.7182

for each model, and the number of times that model was chosen according to AIC. For case C1, when $\alpha_0 = 2, \alpha_1 = 1, \beta_0 = 2$, and $\beta_1 = 1$, the two regression functions are identical and the true model is in fact a logistic linear model, our study shows that among the 1000 simulations, 837 correctly selected the true model. For cases C2 and C3, the parameters α_0 and β_0 are different, hence the true model is no longer a logistic linear model. Our study shows that none of the simulations selected the usual logistic linear model or the beta-binomial regression model, and the average maximized log-likelihoods were substantially larger for the two discrete exponential family logistic linear models than the logistic regression model and the beta-binomial regression model. This simulation study shows that model in (3) and (4) together with the information criterion provides a useful tool for checking the logistic linear regression assumptions.

4 THE MATSUMURA DATA – LIKELIHOOD ANALYSIS

The data in the Appendix provide the observed exam scores for $n = 145$ University of Wisconsin students in Professor Matsumura's Accounting class. Each student completed four multiple choice tests, the first two containing 25 each, the third one containing 22, and the last one containing 29 dissimilar items. We first fit the data using the logistic regression model (including an intercept term), with the last exam scores as the response variable and the proportions correct on the first three exams as explanatory variables. Our chi-square value (11) for the logistic regression model was $\chi^2 = 236.455$, with 140 degrees of freedom. The corresponding p-value was 0.0000007. Obviously, the logistic regression model does not fit the data well. We then fit the data using the p-parameter discrete exponential family logistic linear model and to the beta-binomial model. Both models fit the data well. For the beta-binomial model, the chi-square value was $\chi^2 = 145.899$, with 139 degrees of freedom, and the corresponding p-value was 0.327. For the p-parameter discrete exponential family logistic linear model, AIC and BIC both were maximized when $p = 2$. The corresponding chi-square value was 150.595, with 139 degrees of freedom, and the p-value was 0.237. The maximum likelihood estimates of $\beta_0, \beta_1, \beta_2, \beta_3$ and γ_2, together with their standard errors, are reported in Table III. Note that $\hat{\gamma}_2$ is more than five standard errors from zero,

TABLE III Maximum Likelihood Analysis for Matsumura Data.

Parameter	β_0	β_1	β_2	β_3	γ_2
MLE	−2.169	0.343	1.210	0.823	0.033
Standard Error	0.233	0.349	0.311	0.221	0.006

refuting a standard logistic regression model ($\gamma_2 = 0$). One possible interpretation of this phenomenon is that an important explanatory variable was overlooked, that is, the degree of difficulty of each item. Some items are easier and some items are more difficult in most tests. However, it is not easy to quantify the level of difficulty of each item and therefore, it is seldom recorded in practice. Please note that only the total scores for each test were reported and the individual responses to each item were unfortunately not reported. Therefore, the standard item response models (Van der Linden and Hambleton, 1997) will not be able to be directly applied in this example.

In addition to the parameters specified in the model, many parameters of interest can be represented as functions of the parameters in the model. For instance, the predicted score for an individual, given $\mathbf{x}$ may be of interest. It is essential to predict a student's score when this student missed the test and his previous test scores are available. In this case, the parameter of interest is the expected value of y given the observed $\mathbf{x}$, that is,

$$\begin{aligned}\eta = E(y \mid \boldsymbol{\beta}, \gamma, \mathbf{x}) &= \sum_{j=0}^{m} j p(y = j \mid \boldsymbol{\beta}, \gamma, \mathbf{x}) \\ &= \frac{\sum_{j=0}^{m} j \exp(\log {}^{m}C_j + j\mathbf{x}^T\boldsymbol{\beta} + \mathbf{u}_j\gamma)}{\sum_{h=0}^{m} \exp(\log {}^{m}C_h + h\mathbf{x}^T\boldsymbol{\beta} + \mathbf{u}_h\gamma)}. \end{aligned} \tag{16}$$

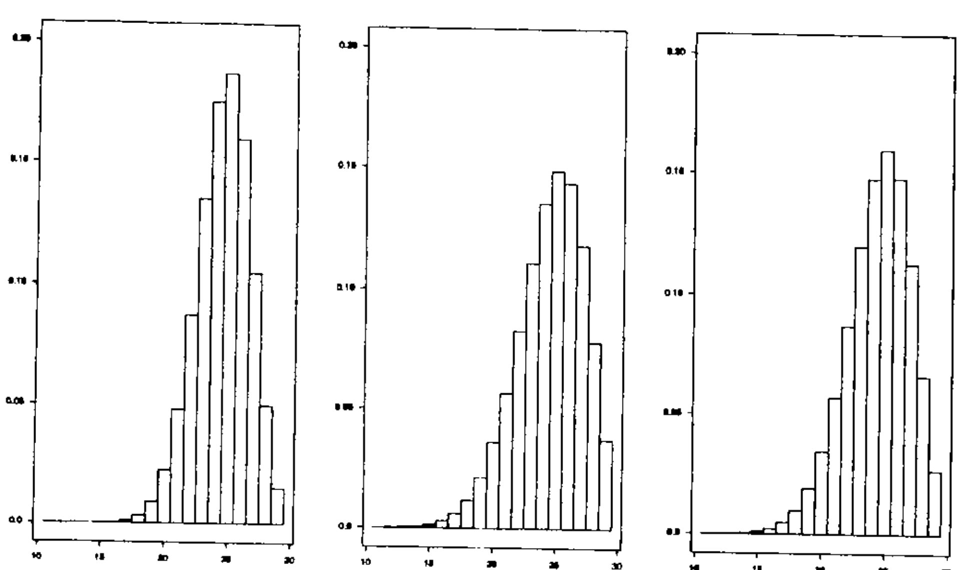

FIGURE 1 Estimated sampling distributions: estimated sampling distribution for a student who correctly answered 23, 21, and 20 items for the first three tests, using: (a) Logistic Regression model; (b) Beta-binomial model; (c) Exponential family logistic linear model.

We now use the scores (23, 21, 20, 27) of the first student in the data set for illustrative purposes. Suppose a student correctly answered 23, 21 and 20 items for the first three tests, his predicted final exam score can be obtained by substituting the maximum likelihood estimates in (16). The predicted scores were 23.491, 23.495, and 23.402 for the logistic regression model, beta-binomial model and 2-parameter discrete exponential family logistic linear model respectively. Please note, although the expected values of y given $\mathbf{x}$ are not very distinct, the predicted probability mass functions of y can be quite different. Hence it may make a significant difference when we perform statistical inferences. Figure 1 presents three histograms which describe the estimated distributions of y given $\mathbf{x} = (1, 23/25, 21/25, 20/22)^T$, for the three different models: Figure 1a for logistic regression model; Figure 1b for beta-binomial model; and Figure 1c for 2-parameter discrete exponential family logistic linear model.

5 BAYESIAN ANALYSIS

In some situations it may be possible to incorporate prior information regarding $\boldsymbol{\beta}$ and γ via independent multivariate normal prior distributions, say with mean $\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\gamma$, and covariance matrices $\mathbf{C}_\beta, \mathbf{C}_\gamma$. The posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma^T)^T$ given $\mathbf{y}$ is proportional to

$$\bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) = \exp\left\{ s_0 + \mathbf{s}_1^T \boldsymbol{\beta} + \mathbf{s}_2^T \gamma - \sum_{i=1}^{n} D_i(\boldsymbol{\beta}, \gamma) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^T \mathbf{C}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta) \right.$$
$$\left. - \frac{1}{2}(\gamma - \boldsymbol{\mu}_\gamma)^T \mathbf{C}_\gamma^{-1} (\gamma - \boldsymbol{\mu}_\gamma) \right\}. \tag{17}$$

where s_0, $\mathbf{s}_1$, $\mathbf{s}_2$ and $D_i(\boldsymbol{\beta}, \gamma)$ are defined in (6)–(9). However, as $|\mathbf{C}_\beta| \to \infty$ and $|\mathbf{C}_\gamma| \to \infty$, the prior information becomes vague, and the two quadratic terms, within the exponential of (17) vanishes, and (17) becomes proportional to the likelihood (5). Let $\eta = g(\boldsymbol{\theta})$ be the parameter of interest. The posterior distribution of η given $\mathbf{y}$ can be (a) closely approximated using Laplacian approximations, (b) approximated using approximate conditional distributions, or (c) simulated using Gibbs sampler/Metropolis-Hastings algorithm.

(a) Laplacian Approximation. The posterior distribution of $\eta = g(\boldsymbol{\theta})$ can be approximated by

$$\pi^*(\eta \mid \mathbf{y}) \propto \bar{\pi}(\boldsymbol{\theta}_\eta \mid \mathbf{y}) |\mathbf{R}_\eta|^{-1/2} f(\eta \mid \boldsymbol{\theta}_\eta, \mathbf{R}_\eta^{-1}), \tag{18}$$

where $\boldsymbol{\theta}_\eta$ conditionally maximizes (17) for each fixed $\eta = g(\boldsymbol{\theta})$, and satisfies

$$\left[\frac{\partial \log \bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y})}{\partial \boldsymbol{\theta}} - \lambda_\eta \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta} = \boldsymbol{\theta}_\eta} = 0,$$

λ_η is a Lagrange multiplier,

$$\mathbf{R}_\eta = -\left[\frac{\partial^2 \log \bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} - \lambda_\eta \frac{\partial^2 g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]_{\boldsymbol{\theta} = \boldsymbol{\theta}_\eta},$$

and $f(\eta \mid \boldsymbol{\mu}, \mathbf{C})$ denotes the density of $\eta = g(\boldsymbol{\theta})$ while $\boldsymbol{\theta}$ possesses a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. For details of Laplacian approximations, see Leonard (1982), Leonard *et al.* (1989), Hsu (1995), and Leonard and Hsu (1999).

(b) Approximate conditional distributions. The posterior density of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma^T)^T$ given $\mathbf{y}$ can be decided by the conditional distributions of $\boldsymbol{\beta}$ given γ and $\mathbf{y}$, and of γ given $\boldsymbol{\beta}$ and $\mathbf{y}$. Each of the two conditional distributions can be approximated by a normal distribution. Therefore, we can approximate the posterior density of $\boldsymbol{\theta}$ given $\mathbf{y}$ by the above two approximated conditional distributions. Then the Gibbs sampling method may be used to calculate the posterior moments of $\eta = g(\boldsymbol{\beta}, \gamma)$, given $\mathbf{y}$, by simulating $\boldsymbol{\beta}$ and γ from the two approximated conditional distributions, respectively. The two approximated conditional distributions are derived below:
For a given γ, expanding $\log \bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y})$ in a Taylor series about $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_\gamma$, where $\hat{\boldsymbol{\beta}}_\gamma$ maximizes the posterior density $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ and $\bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y})$ is defined in (17), gives

$$\begin{aligned}\log \bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) = \log \bar{\pi}(\hat{\boldsymbol{\beta}}_\gamma, \gamma) &+ \left[\mathbf{s}_1^T - \sum_{i=1}^{n} e_i^* \mathbf{x}_i^T - (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)^T \mathbf{C}_\beta^{-1}\right](\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_\gamma) \\ &- \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_\gamma)^T (v_i^* \mathbf{x}_i \mathbf{x}_i^T + \mathbf{C}_\beta^{-1})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_\gamma) \\ &+ \text{ cubic and higher order terms,}\end{aligned} \tag{19}$$

where,

$$e_i^* = e_i^*(\gamma) = E(y_i \mid \hat{\boldsymbol{\beta}}_\gamma, \gamma, \mathbf{x}_i) = \sum_{h=0}^{m_i} h p(y_i = h \mid \hat{\boldsymbol{\beta}}_\gamma, \gamma, \mathbf{x}_i), \tag{20}$$

and

$$\begin{aligned} v_i^* = v_i^*(\gamma) &= \mathrm{Var}(y_i \mid \hat{\boldsymbol{\beta}}_\gamma, \gamma, \mathbf{x}_i) \\ &= \sum_{h=0}^{m_i} h^2 p(y_i = h \mid \hat{\boldsymbol{\beta}}_\gamma, \gamma, \mathbf{x}_i) - \left[\sum_{h=0}^{m_i} h p(y_i = h \mid \hat{\boldsymbol{\beta}}_\gamma, \gamma, \mathbf{x}_i)\right]^2, \end{aligned} \tag{21}$$

with $p(y_i = h \mid \boldsymbol{\beta}, \gamma, \mathbf{x}_i)$ defined in (3). Neglecting cubic and higher order terms and completing the square in (19), we find that the log-posterior can be approximated by

$$\log \bar{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \approx \log \bar{\pi}(\hat{\boldsymbol{\beta}}_\gamma, \gamma) + \frac{1}{2}\mathbf{d}_\gamma^T \mathbf{Q}_\gamma \mathbf{d}_\gamma - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_\gamma^*)^T \mathbf{Q}_\gamma (\boldsymbol{\beta} - \boldsymbol{\beta}_\gamma^*) \tag{22}$$

where

$$\mathbf{d}_\gamma = \sum_{i=1}^{n}(y_i - e_i^*)\mathbf{x}_i - \mathbf{C}_\beta^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta), \tag{23}$$

$$\mathbf{Q}_\gamma = \sum_{i=1}^{n} v_i^* \mathbf{x}_i \mathbf{x}_i^T + \mathbf{C}_\beta^{-1}, \tag{24}$$

and

$$\boldsymbol{\beta}_\gamma^* = \hat{\boldsymbol{\beta}}_\gamma + \mathbf{Q}_\gamma^{-1}\mathbf{d}_\gamma. \tag{25}$$

Equation (22) tells us that the conditional posterior distribution of $\boldsymbol{\beta}$, given γ, is approximately multivariate normal, with mean vector $\boldsymbol{\beta}_\gamma^*$ and covariance matrix $\mathbf{Q}_\gamma^{-1}$. Analogous to the above derivation, the approximate conditional distribution of γ, given β, may be derived by expanding $\log\bar{\pi}(\boldsymbol{\theta}|\mathbf{y})$ in a Taylor series up to the second order term, about $\gamma = \hat{\gamma}_\beta$, where $\hat{\gamma}_\beta$ maximizes the posterior density in (17) for a given $\boldsymbol{\beta}$. Then the conditional posterior distribution of γ given $\boldsymbol{\beta}$ is approximated by a normal distribution with mean vector γ_β^* and covariance matrix $\mathbf{Q}_\beta^{-1}$, where

$$\mathbf{d}_\beta = \sum_{i=1}^{n}[\mathbf{u}_{y_i} - E(\mathbf{u}_{y_i}|\boldsymbol{\beta}, \hat{\gamma}_\beta, \mathbf{x}_i)] - \mathbf{C}_\gamma^{-1}(\hat{\gamma} - \boldsymbol{\mu}_\gamma). \tag{26}$$

$$\begin{aligned}\mathbf{Q}_\beta &= \sum_{i=1}^{n}\mathrm{Cov}(\mathbf{u}_{y_i}|\boldsymbol{\beta}, \hat{\gamma}_\beta, \mathbf{x}_i) + \mathbf{C}_\gamma^{-1} = \sum_{i=1}^{n}[E(\mathbf{u}_{y_i}\mathbf{u}_{y_i}^T|\boldsymbol{\beta}, \hat{\gamma}_\beta, \mathbf{x}_i) \\ &\quad - E(\mathbf{u}_{y_i}|\boldsymbol{\beta}, \hat{\gamma}_\beta, \mathbf{x}_i)E(\mathbf{u}_{y_i}^T|\boldsymbol{\beta}, \hat{\gamma}_\beta, \mathbf{x}_i)] + \mathbf{C}_\gamma^{-1}, \end{aligned} \tag{27}$$

$$\gamma_\beta^* = \hat{\gamma}_\beta + \mathbf{Q}_\beta^{-1}\mathbf{d}_\beta, \tag{28}$$

and the vector $\mathbf{u}$ is defined in (10). Note that the expectations, variance, and covariance in (20), (21), (26) and (27) are with respect to probability mass function (3).

Let $\eta = g(\boldsymbol{\beta}, \gamma)$ be any parameter of interest. The approximate posterior mean may be calculated, using Gibbs sampling method, as follows: Given γ, a $\boldsymbol{\beta}$ is generated from the multivariate normal distribution with mean vector $\boldsymbol{\beta}_\gamma^*$ and covariance matrix $\mathbf{Q}_\gamma^{-1}$. Given $\boldsymbol{\beta}$, a γ is generated from the multivariate normal distribution with mean vector γ_β^* and covariance matrix $\mathbf{Q}_\beta^{-1}$. The quantity $\eta = g(\boldsymbol{\beta}, \gamma)$ is then calculated. The posterior mean of η given $\mathbf{y}$ is approximated by the long-term average of the calculated η. The simulation process continues until the average converges. For details of the Gibbs sampling method, see for example Gelfand and Smith (1990), Leonard *et al.* (1994), Gelman *et al.* (1995), Leonard and Hsu (1999).

(c) Exact distribution and moments. The exact posterior distribution and moments of any parameter of interest can be simulated using Gibbs sampler/Metropolis-Hastings algorithm via the approximate conditional posterior densities derived in part (*b*). Let $\pi^*(\boldsymbol{\beta}|\gamma, \mathbf{y})$ and $\pi^*(\gamma|\boldsymbol{\beta}, \mathbf{y})$ denote the multivariate normal densities with means $\boldsymbol{\beta}_\gamma^*, \gamma_\beta^*$, and covariance matrices $\mathbf{Q}_\gamma^{-1}, \mathbf{Q}_\beta^{-1}$, respectively, where $\boldsymbol{\beta}_\gamma^*, \gamma_\beta^*, \mathbf{Q}_\gamma$, and $\mathbf{Q}_\beta$ are defined in (25), (28), (24), and (27). Note that the densities $\pi^*(\boldsymbol{\beta}|\gamma, \mathbf{y})$ and $\pi^*(\gamma|\boldsymbol{\beta}, \mathbf{y})$ approximate the conditional posterior densities $\pi(\boldsymbol{\beta}|\gamma, \mathbf{y})$ and $\pi(\gamma|\boldsymbol{\beta}, \mathbf{y})$, respectively. The posterior mean of $\eta = g(\boldsymbol{\beta}, \gamma)$ given $\mathbf{y}$, can be obtained by simulating $\boldsymbol{\beta}$ and γ from the normal distributions with densities $\pi^*(\boldsymbol{\beta}|\gamma, \mathbf{y})$ and $\pi^*(\gamma|\boldsymbol{\beta}, \mathbf{y})$. In the tth simulation, let $\eta^{(t)} = g(\boldsymbol{\beta}^{(t)}, \gamma^{(t)})$ be the simulated η. To simulate $\eta^{(t)}$, we sample a candidate point $\boldsymbol{\beta}^*$ from $\pi^*(\boldsymbol{\beta}|\gamma^{(t-1)}, \mathbf{y})$ and set

$$\boldsymbol{\beta}^{(t)} = \begin{cases} \boldsymbol{\beta}^* & \text{with probability} \min(P_\beta, 1) \\ \boldsymbol{\beta}^{(t-1)} & \text{otherwise} \end{cases}$$

where

$$P_\beta = \frac{\pi(\boldsymbol{\beta}^* \mid \gamma^{(t-1)}, \mathbf{y})/\pi^*(\boldsymbol{\beta}^* \mid \gamma^{(t-1)}, \mathbf{y})}{\pi(\boldsymbol{\beta}^{(t-1)} \mid \gamma^{(t-1)}, \mathbf{y})/\pi^*(\boldsymbol{\beta}^{(t-1)} \mid \gamma^{(t-1)}, \mathbf{y})}.$$

Then, sample a candidate point γ^* from $\pi^*(\gamma \mid \boldsymbol{\beta}^{(t)}, \mathbf{y})$ and set

$$\gamma^{(t)} = \begin{cases} \gamma^* & \text{with probability} \min(P_\gamma, 1) \\ \gamma^{(t-1)} & \text{otherwise} \end{cases}$$

where

$$P_\gamma = \frac{\pi(\gamma^* \mid \boldsymbol{\beta}^{(t)}, \mathbf{y})/\pi^*(\gamma^* \mid \boldsymbol{\beta}^{(t)}, \mathbf{y})}{\pi(\gamma^{(t-1)} \mid \boldsymbol{\beta}^{(t)}, \mathbf{y})/\pi^*(\gamma^{(t-1)} \mid \boldsymbol{\beta}^{(t)}, \mathbf{y})}.$$

The exact posterior mean of η is the long term average of the simulated $\eta^{(t)} = g(\boldsymbol{\beta}^{(t)}, \gamma^{(t)})$. See Gelman *et al.* (1995) for details of the above simulation procedure.

6 THE MATSUMURA DATA – BAYESIAN INFERENCE

The Bayesian marginalization techniques discussed in Section 5, for the discrete exponential family logistic linear model defined in (3) and (4), are applied to the Matsumura data.

Following the discussions is Section 4, we again use the scores (23, 21, 20, 27) of the first student in the data set as an example. The following four parameters are of interest and will be discussed, under the vague prior for $\boldsymbol{\beta}$ and γ by letting $|\mathbf{C}_\beta| \to \infty$ and $|\mathbf{C}_\gamma| \to \infty$, in (17).

(A) $\eta_A = \gamma_2$. The standard logistic regression model will be refuted when the posterior distribution of γ_2 is not concentrated about zero.

(B) $\eta_B = p(y = 27 \mid \boldsymbol{\beta}, \gamma, \mathbf{x})$; the probability that a student correctly answered 27 items in the final test given the fact that he correctly answered 23, 21 and 20 items for the first three tests.

(C) $\eta_C = E(y \mid \boldsymbol{\beta}, \gamma, \mathbf{x})$; the predicted score for an individual who correctly answered 23, 21 and 20 items for the first three tests.

Table IV presents the simulated posterior means for the above four parameters utilizing approximate conditional (normal) distributions and Gibbs sample/Metropolis-Hastings algorithm, which were discussed in Section 5(b) and 5(c), respectively.

TABLE IV Posterior Means.

Parameter of interest	η_A	η_B	η_C
Conditional normal approximations	0.03131	0.06591	23.4063
Gibbs sampler/Metropolis-Hastings algorithm	0.03083	0.06585	23.4119

The newly derived approximation, based on conditional (normal) approximations, produced surprisingly close approximates to the simulated posterior means, which are obtained via Gibbs sampler/Metropolis-Hastings algorithm.

Figures 2 to 4 present the posterior densities of the parameters η_A, η_B and η_C described above. In each figure, histogram (a) used 100,000 simulations for the Gibbs sampler/Metropolis-Hastings algorithm, and curve (b) was obtained by using the Laplacian approximation. Please note the close correspondence between the approximate smooth curve (b) and simulated histogram (a) for these figures.

Figure 2 describes the posterior density of $\eta_A = \gamma_2$. As the posterior density of $\eta_A = \gamma_2$ is concentrated on the region (0.01, 0.05), this confirms that a standard logistic regression is not quite adequate.

Figure 3 describes the posterior density of $\eta_B = p(y = 27 \mid \boldsymbol{\beta}, \gamma, \mathbf{x})$. The figure shows that the probability, for a student who correctly answered 23, 21 and 20 items for the first three tests respectively, to correctly answer 27 items on the final test is about 0.03 to 0.11.

Figure 4 describes the posterior density of $\eta_C = E(y \mid \boldsymbol{\beta}, \gamma, \mathbf{x})$. The figure tells us that for a student who correctly answered 23, 21 and 20 items for the first three test, is likely to answer 22 to 24.5 items correctly on the last test.

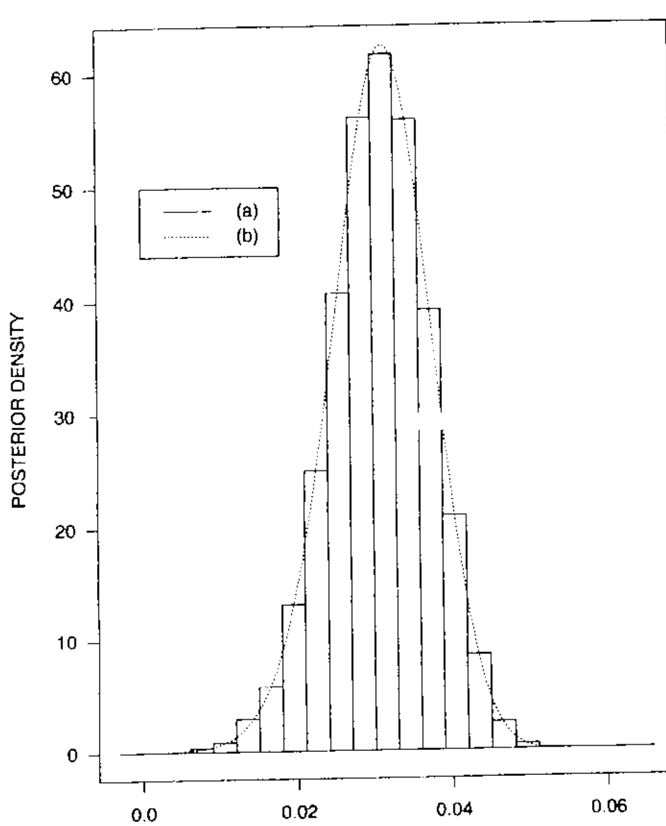

FIGURE 2 Marginal posterior density of η_A: (a) histogram, based on 100,000 simulations for exact posterior density using Gibbs sampler/Metropolis-Hastings algorithm; (b) Laplacian approximation.

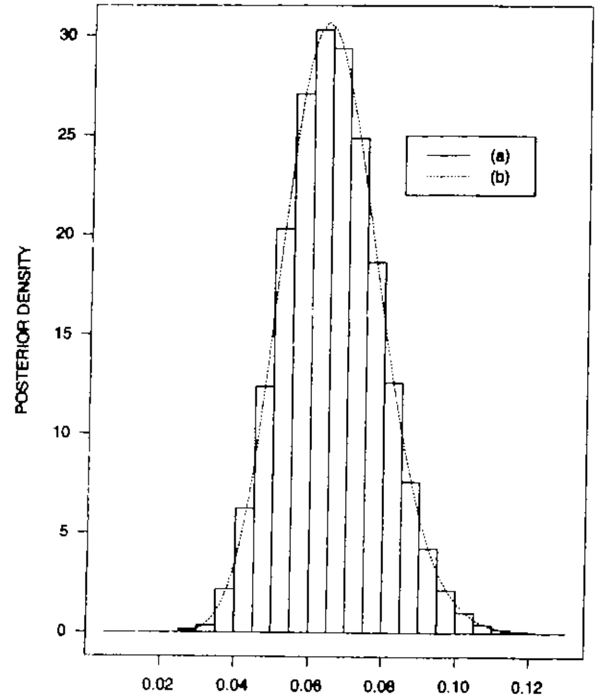

FIGURE 3 Marginal posterior density of η_B: (a) histogram, based on 100,000 simulations for exact posterior density using Gibbs sampler/Metropolis-Hastings algorithm; (b) Laplacian approximation.

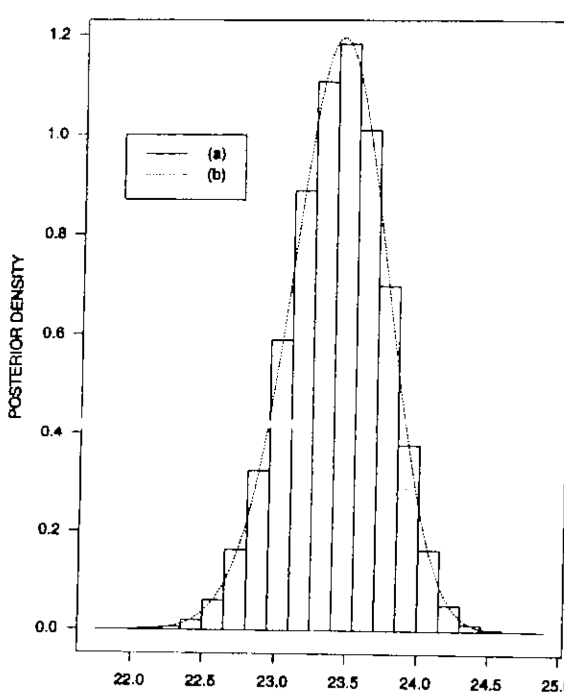

FIGURE 4 Marginal posterior density of η_C: (a) histogram, based on 100,000 simulations for exact posterior density using Gibbs sampler/Metropolis-Hastings algorithm; (b) Laplacian approximation.

7 CONCLUDING REMARKS

The p-parameter discrete exponential family logistic linear model described in Eq. (3) and (4) extends the commonly used logistic linear model, while some explanatory variables were overlooked. The model provides a semiparameter fit to the data. Moreover, the discussions in Sections 2 and 3 showed that the model also provided a useful tool for checking the adequacy of the logistic linear model. Bayesian analyses were addressed in Sections 5 and 6. The computations for the analyses were found to be not straightforward. Bayesian computation methods such as Laplacian approximations and Gibbs sampler/Metropolis-Hastings algorithm were discussed and applied to the Matsumura data. While the simulated posterior means utilized Gibbs sampled/Metropolis-Hastings algorithm are theoretically exact, the approximated posterior means using Laplacian methods were found to be quite accurate. However, it took hours to perform simulations but approximation was performed in seconds, for the Matsumura example.

Acknowledgements

The author wishes to thank Tom Leonard for his helpful discussions and suggestions, Ella Mae Matsumura for her data set and two referees for their valuable comments.

References

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Ann. Inst. Statist. Math.*, **30**, 9–14.

Atilgan, T. (1983). Parameter parsimony model selection, and smooth density estimation. *PhD Thesis*, Department of Statistics, University of Wisconsin-Madison.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, **37**, 29–51.

Chiu, T. Y. M., Leonard, T. and Tsui, K. (1996). The matrix-logarithmic covariance model. *J. Amer. Statist. Assoc.*, **91**, 198–210.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 393–397.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

Hsu, J. S. J. (1995). Generalized Laplacian approximations in Bayesian inference. *Canadian J. Statist.*, **23**, 399–410.

Hsu, J. S. J., Leonard, T. and Tsui, K. (1991). Statistical inference for multiple choice tests. *Psychometrika*, **56**, 327–348.

Keeler, L. C. (1985). Genotypic and environmental effects on the cold, icing, and flooding tolerance of winter wheat. *MSc Thesis*, University of Guelph, Ontario.

Leonard, T. (1982). Comment on "A simple predictive density function". *J. Amer. Statist. Assoc.*, **77**, 657–658.

Leonard, T. and Hsu, J. S. J. (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge University Press, Cambridge.

Leonard, T., Hsu, J. S. J. and Tsui, K. (1989). Bayesian marginal inference. *J. Amer. Statist. Assoc.*, **84**, 1051–1057.

Leonard, T., Hsu, J. S. J., Tsui, K. and Murray, J. F. (1994). Bayesian and likelihood inference form equally weighted mixtures. *Ann. Inst. Statist. Math.*, **40**, 203–220.

Leonard, T. and Novick, J. B. (1986). Bayesian full rank marginalization for two-way contingency tables. *J. Educ. Statist.*, **11**, 33–56.

Nishi, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.*, **12**, 757–765.

Prentice, R. L. and Barlow, W. E. (1988). Correlated binary regression with covariate specific to each binary observation. *Biometrics*, **44**, 1033–1048.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Math. Statist.*, **6**, 461–464.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika*, **68**, 45–54.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. Roy. Statis. Soc., B*, **39**, 44–47.

Stone, M. (1978). Comments on model selection criteria of Akaike and Schwarz. *J. Roy. Statis. Soc., B*, **41**, 276–278.

Thompson, M. L. (1978a). Selection of variables in multiple regression: Part I. A review and evaluation. *Internat. Statist. Rev.*, **46**, 1–19.

Thompson, M. L. (1978b). Selection of variables in multiple regression: Part II. Chosen procedures, computations and examples. *Internat. Statist. Rev.*, **46**, 126–146.

Van der Linden, W. and Hambleton, R. K. (Eds.) (1997). *Handbook of Modern Item Response Theory.* Springer-Verlag, New York.

Williams, D. A. (1975). The analysis of binary responses from Toxicological experiments involving reproduction and treatogenicity. *Biometrics*, **31**, 949–952.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.*, **31**, 144–148.

APPENDIX: Matsumara's Accounting Exams Data.

	TEST			
No.	*1*	*2*	*3*	*4*
1	23	21	20	27
2	24	21	16	27
3	20	21	17	24
4	22	16	15	19
5	22	18	10	17
6	17	15	18	17
7	23	16	16	20
8	21	16	17	22
9	22	18	18	18
10	25	19	21	26
11	23	23	18	24
12	17	14	11	14
13	22	20	12	13
14	23	16	11	13
15	23	20	16	22
16	18	17	12	25
17	25	22	14	26
18	22	22	17	16
19	18	15	18	22
20	21	22	16	24
21	23	19	12	20
22	23	18	19	20
23	19	20	16	24
24	24	20	15	20
25	19	16	6	18
26	25	20	20	23
27	19	14	14	20
28	20	19	22	26
29	24	18	18	25
30	21	17	16	24
31	20	17	12	19
32	25	20	18	19
33	16	8	13	8
34	20	17	19	23
35	21	17	13	17
36	23	20	16	22
37	19	15	16	21
38	24	22	11	24
39	22	16	13	23
40	23	17	13	23
41	16	16	10	18
42	22	21	16	17
43	21	15	10	18
44	20	20	11	23
45	23	22	13	20
46	22	24	14	27
47	19	14	9	10
48	22	17	18	19
49	22	21	17	24
50	20	15	13	23

	TEST			
No.	*1*	*2*	*3*	*4*
51	25	20	18	25
52	22	19	10	16
53	23	23	21	26
54	25	22	20	20
55	17	14	13	18
56	22	18	13	17
57	22	17	9	16
58	20	19	15	19
59	19	18	15	21
60	25	19	18	24
61	13	9	17	20
62	22	21	18	24
63	21	12	15	16
64	25	20	19	27
65	22	20	21	23
66	23	25	21	26
67	24	22	19	25
68	18	18	12	19
69	25	22	16	25
70	22	18	12	18
71	21	17	14	22
72	23	17	11	25
73	23	20	19	24
74	22	21	14	23
75	23	25	21	23
76	20	17	16	22
77	23	21	15	21
78	19	19	15	14
79	20	18	18	19
80	21	14	11	11
81	20	20	14	17
82	22	18	18	24
83	24	19	19	16
84	21	20	19	19
85	23	17	10	16
86	20	17	17	22
87	25	25	21	29
88	20	10	7	15
89	18	17	11	21
90	20	19	14	19
91	24	24	21	24
92	25	18	18	21
93	22	21	16	22
94	24	20	14	20
95	22	18	16	20
96	23	14	12	16
97	16	12	9	8
98	22	17	10	18
99	24	21	20	24
100	17	17	11	22
101	23	22	20	26

APPENDIX (*Continued*)

	TEST			
No.	*1*	*2*	*3*	*4*
102	23	23	16	22
103	19	12	11	13
104	25	23	18	22
105	19	11	14	14
106	21	20	18	23
107	21	19	18	26
108	16	18	19	15
109	24	21	16	22
110	18	17	15	20
111	20	20	15	19
112	22	21	14	23
113	17	19	9	17
114	22	17	13	22
115	22	17	17	16
116	23	22	14	26
117	25	18	16	26
118	22	19	14	17
119	24	19	16	25
120	21	21	14	20
121	20	15	9	18
122	24	21	15	19
123	24	20	18	24

	TEST			
No.	*1*	*2*	*3*	*4*
124	24	19	16	24
125	18	14	14	18
126	24	21	15	19
127	21	21	12	23
128	22	18	17	20
129	18	16	9	23
130	20	15	17	17
131	23	23	17	18
132	20	18	15	20
133	25	21	18	17
134	21	19	13	19
135	22	12	16	20
136	25	22	21	26
137	19	11	12	16
138	21	23	18	24
139	17	13	12	16
140	20	16	12	24
141	22	18	11	18
142	20	17	13	21
143	16	17	11	15
144	25	21	20	21
145	25	23	18	24