Journal of Statistical Planning and Inference 14 (1986) 187-202
North-Holland

COMPARISON OF DICHOTOMOUS AND POLYTOMOUS RESPONSE MODELS

Siu-Keung TSE

University of New Hampshire, Durham, NH 03824, USA

C.F. Jeff WU

University of Wisconsin-Madison, Madison, WI 53706, USA

Received 18 March 1985; revised manuscript received 2 October 1985
Recommended by R. Bartoszynski

Abstract: In some experimental situations, it is feasible to classify the outcome of the experiment into more than two ordered categories. To compensate for the extra effort required in the more refined classification, it is important to know whether it will lead to more efficient estimates than the binary classification. It is shown that the Fisher information for the common parameters in the more elaborate model is greater, which is further supported in a numerical study. Based on a simulation study, a similar conclusion is reached regarding the extent to which such gains are achieved in small sample situations. Some empirical results are given theoretical justification.

AMS Subject Classification: 62K99, 62N05, 62B15.

Key words: Dichotomous quantal response; Polytomous response; Fisher information; Missing information principle; LD50; Logistic curve.

1. Introduction

In binary response experiments, the observed response is a nomial variable y, which takes on two possible outcomes; response O_1 $(y=1)$ or non-response O_2 $(y=2)$. The experiment is characterized by assuming a response curve

$$F(x) = \Pr\{y=1 \mid x\}$$

where $F(x)$ is the probability of response of an experimental subject under stimulus level x. This model is common to many areas of research. In engineering research, the stimulus levels may be the force applied to equipment or the temperature increases in a system. Then the number of brittle failures or unsuccessful performances can be recorded. In material testing, glass containers may be submitted to a drop height test. The response is 'broken' or 'not broken'. In a drug test that involves houseflies under exposure to a particular insecticide, the number of houseflies that are either dead or alive are observed.

Our main interest is not in the whole response curve F but instead in estimating some of the percentiles. The $100p$-th percentile LD_p is defined by

$$F(\mathrm{LD}_p)=p, \tag{1}$$

i.e., LD_p is the stimulus level at which the experimental subjects would respond on the average $100p\%$ of the time. The median of F, $\mathrm{LD}_{0.5}$, has received much attention in the literature. But the extreme percentiles are often more relevant and of intrinsic interest in practical situations.

In the aforementioned examples, the outcome of the experiment can be classified into $s>2$ ordered categories, for example, according to the varying degrees of damage or moribundity. The response is then called *polytomous*. Our purpose is to show that the polytomous response model is more efficient in estimating percentiles, and also to identify the situations in which such gains are substantial.

Throughout the paper, we consider experiments with only one stimulus variable, since the percentile LD_p is well defined only in the single variable case. In Section 2, we prove that there is a definite gain in Fisher information by using the polytomous response model for estimating the parameters and percentiles associated with the dichotomous response curve. An extensive numerical study is reported in Section 3. With the response curves taking the logistic form where $F(x\mid\beta,\alpha_1)=[1+\exp(-\beta x-\alpha_1)]^{-1}$, and under different design schemes, we compare the determinants of the Fisher information matrices for the common parameters (β,α_1) in the dichotomous and trichotomous response models. The gain in using the latter model is more significant if the probabilities of the categories O_2 and O_3 are not negligible (Theorem 3). We also report an extensive simulation study of the small sample properties of the estimators of (β,α_1) and of the percentiles under the two models. It is found that the trichotomous response model provides more efficient estimators, especially in the case of extreme percentiles, and that the percent reduction in variance of estimating the slope is bigger than that of estimating the intercept (Theorem 4). This model deserves serious consideration if one is interested in the estimation of the extreme percentiles with a relatively small sample and the extra cost required in clearly defining and classifying the outcomes is not large. In Section 4, an example on the degree of disturbed dreams among boys is given to illustrate the application of the model.

2. Comparison of information in large samples

2.1. Dichotomous response model

Assume that k groups consisting of $N_1,\dots,N_k$ subjects are tested at stimulus levels $x_1,x_2,\dots,x_k$. The response Y can take one of two values. In a material testing situation, the response of the material may be 'broken' or 'not broken'. Denote these two outcomes by O_1 and O_2. Assume the outcome O_1 has a response curve given by the distribution function $F(x \mid \beta, \alpha_1) = \Pr\{\text{outcome is } O_1 \mid x\}$ and F satisfies:

(i) the first and second partial derivatives of F with respect to β, α_1 exist and are not identically zero,

(ii) $F(x \mid \beta, \alpha)$ is a strictly monotone function of α and is monotonically increasing in x,

(iii) $\lim_{\alpha \to \infty} F(x \mid \beta, \alpha) = 1$ and $\lim_{\alpha \to -\infty} F(x \mid \beta, \alpha) = 0$.

Note that in this model, $\Pr\{\text{outcome is } O_2 \mid x\} = 1 - F(x \mid \beta, \alpha_1)$. Usually the response curve F assumes the form of a probit or logit. Other models include the angular or the rectangular distribution. Finney (1978) compares these four models and concludes that they are indistinguishable for response rates between 0.05 and 0.95. Thus, in our numerical and simulation study, we use the logistic function for the response curves although some of the theoretical results hold more generally.

For notational convenience, define $P_{i1} = \Pr\{\text{outcome is } O_1 \mid x_i\} = F(x_i \mid \beta, \alpha_1)$. In particular, the $100p$-th percentile LD_p is defined by $F(\text{LD}_p) = p$. Suppose, at level x_i, r_{i1} of the N_i subjects have outcome O_1. The log-likelihood function of this model is

$$\log L_2 = \sum_{i=1}^{k} \{r_{i1} \log P_{i1} + (N_i - r_{i1}) \log(1 - P_{i1})\}.$$

The Fisher information matrix for (β, α_1) under this model is

$$J_2 = E \begin{bmatrix} \dfrac{-\partial^2 \log L_2}{\partial \beta^2} & \dfrac{-\partial^2 \log L_2}{\partial \beta \, \partial \alpha_1} \\ \dfrac{-\partial^2 \log L_2}{\partial \beta \, \partial \alpha_1} & \dfrac{-\partial^2 \log L_2}{\partial \alpha_1^2} \end{bmatrix}. \tag{2}$$

2.2. *Polytomous response model*

For the dichotomous response model described above, suppose the outcome O_2 can be further classified into a number, say $s-1$, of ordered categories. For example, the category 'not broken' in the material testing example can be classified into more ordered categories according to the different degrees of damage. For $s=3$, we may define O_1 = broken, O_2 = partially damaged, O_3 = no damage. In these cases, the response variable is an ordered categorical variable which takes s possible values. These s values are considered to be qualitative rather than quantitative in nature. Some references can be found in Aitchison and Silvey (1957), Ashford (1959) and Gurland, Lee and Dahm (1960). A general class of regression models for ordinal data is developed in McCullagh (1980). More details can be found in McCullagh and Nelder (1983).

Suppose the response of the polytomous response model is one of the s ordered outcomes, say, $O_1, O_2, \ldots, O_s$. Let Y be the ordered categorical variable with assigned value i if the response is O_j, $j = 1, 2, \ldots, s$. Assume $F(x \mid \beta, \alpha)$ satisfies (i), (ii), (iii) in Section 2.1. An ordered polytomous quantal response model is given by

$$\Pr\{Y\le j \mid x\} = F(x \mid \beta, \alpha_j), \quad j=1,\dots,s-1, \tag{3}$$

where $\alpha_1<\alpha_2<\dots<\alpha_{s-1}$. There are $s-1$ response curves in this more elaborate model. We have assumed that all the response curves take the same functional form with different α_j while β is assumed to be the same. With assumption (ii) in Section 2.1, these response curves would not cross each other so that the probability of each category O_j is well defined, a fact pointed out by Gurland, Lee and Dahm (1961). Also, under this assumption, the interpretation of β would not depend on the choice of categories. The conclusion would not be affected even when adjacent categories are pooled (McCullagh, 1980).

For notational convenience, denote $P_{ij}=F(x_i \mid \beta,\alpha_j)$, $\alpha_0=-\infty$ and $\alpha_s=+\infty$. From (3), it follows that

$$\Pr\{Y=j \mid x_i\} = P_{ij}-P_{i,j-1}, \quad j=1,2,\dots,s.$$

The parameters in this model are $\beta,\alpha_1,\dots,\alpha_{s-1}$. Define the LD_p to be the stimulus level at which on the average $100p$ percent of the test subjects will have outcome O_1. Then, we have

$$F(\mathrm{LD}_p \mid \beta,\alpha_1)=p.$$

Therefore, the estimation of LD_p depends only on the parameters β and α_1.

For level x_i, let $r_{i1},\dots,r_{is}$ be the numbers of subjects with outcomes $O_1,O_2,\dots,O_s$ respectively. The log-likelihood function of this model takes the form

$$\log L_s = \sum_{i=1}^{k} \{r_{i1}\log P_{i1} + r_{i2}\log(P_{i2}-P_{i1}) + \cdots \\ + (N_i - r_{i1} - \cdots - r_{i,s-1})\log(1-P_{i,s-1})\}.$$

Denote $\theta=(\beta,\alpha_1,\alpha_2,\dots,\alpha_{s-1})^{\mathrm{T}}$. The Fisher information matrix is

$$J_s=(J_{ij}(s)), \quad 1\le i,j\le s,$$

where

$$J_{ij}(s)=E\left(\frac{\partial \log L_s}{\partial\theta_i}\right)\left(\frac{\partial \log L_s}{\partial\theta_j}\right) = -E\left(\frac{\partial^2 \log L_s}{\partial\theta_i\,\partial\theta_j}\right).$$

Under some regularity conditions on $F(x \mid \beta,\alpha)$, the matrix J_s is well defined and is positive definite. Furthermore, the asymptotic distribution of the MLE $\hat\theta$ of θ is $\mathrm{N}(\theta, J_s^{-1})$ (Fahrmeir and Kaufmann, 1985).

One can partition J_s into

$$\begin{bmatrix} J_{11}^{(m)}(s) & J_{12}^{(m)}(s) \\ J_{21}^{(m)}(s) & J_{22}^{(m)}(s) \end{bmatrix}$$

for any $m=1,2,\dots,s-1$ where $J_{11}^{(m)}(s)$ is an $m\times m$ matrix, and $J_{22}^{(m)}(s)$ an $(s-m)\times$

$(s-m)$ matrix. In particular,

$$J_{11}^{(2)}(s)=E\left(\frac{\partial \log L_s}{\partial \beta}, \frac{\partial \log L_s}{\partial \alpha_1}\right)^{\mathrm{T}}\left(\frac{\partial \log L_s}{\partial \beta}, \frac{\partial \log L_s}{\partial \alpha_1}\right), \tag{4}$$

$$J_{22}^{(2)}(s)=E\left[\frac{-\partial^2 \log L_s}{\partial \alpha_i\, \partial \alpha_j}\right]_{2\le i,j\le s-1}, \tag{5}$$

$$J_{12}^{(2)}(s)=J_{21}^{(2)'}(s)=E\left(\frac{\partial \log L_s}{\partial \beta}, \frac{\partial \log L_s}{\partial \alpha_1}\right)^{\mathrm{T}}\left(\frac{\partial \log L_s}{\partial \alpha_j}\right)_{2\le j\le s-1} \tag{6}$$

Then the Fisher information matrix for (β,α_1) is $J_{11}^{(2)}(s)-J_{12}^{(2)}(s)J_{22}^{(2)\ -1}(s)J_{21}^{(2)}(s)$.

2.3. Comparison of the Fisher information matrices

In an $(s+1)$-response model, the numbers of subjects with outcomes $O_1, O_2, \dots, O_{s+1}$ at level x_i are $r_{i1}, r_{i2}, \dots, r_{i,s+1}$ respectively. Let $\boldsymbol{R}_i=(N_i, r_{i1}, \dots, r_{i,s+1})$, $i=1,2,\dots,k$, be the vectors of responses for the polytomous response model. If the outcomes O_s, O_{s+1} are collapsed into one class, then the number of response subjects in this combined category is the sum of the corresponding r_{ij}'s. Equivalently, O_s and O_{s+1} can be viewed as missing data with only the total O_s+O_{s+1} available. Therefore, the information available in the latter s-response model can be summarized by the vectors $\boldsymbol{R}_i'=(N_i, r_{i1}, \dots, r_{is}+r_{i,s+1})$, $i=1,2,\dots,k$, i.e. each $\boldsymbol{R}_i'$ is a function of $\boldsymbol{R}_i$.

The above reformulation paves the way for proving the following theorem by using the Missing Information Principle (Orchard and Woodbury, 1972). This theorem states that the Fisher information matrix

$$J_{11}^{(s)}(s+1)-J_{12}^{(s)}(s+1)J_{22}^{(s)\ -1}(s+1)J_{21}^{(s)}(s+1)$$

associated with the first s parameters, namely, $\beta, \alpha_1, \dots, \alpha_{s-1}$ in the $(s+1)$-response model is always greater than or equal to the Fisher information matrix J_s for the parameters $\beta, \alpha_1, \dots, \alpha_{s-1}$ in the s-response model. The second model is obtained by combining the last two outcomes of the first model into one group. Define $A\ge B$ if $A-B$ is non-negative definite.

Theorem 1. *For* $s\ge 2$,

$$J_{11}^{(s)}(s+1)-J_{12}^{(s)}(s+1)J_{22}^{(s)\ -1}(s+1)J_{21}^{(s)}(s+1)\ge J_s.$$

In fact,

$$J_{11}^{(s)}(s+1)-J_{12}^{(s)}(s+1)J_{22}^{(s)\ -1}(s+1)J_{21}^{(s)}(s+1)-J_s$$

$$=\begin{bmatrix} a_{11} & 0\cdots & a_{21} \\ 0 & 0 & 0 \\ \vdots & & \vdots \\ a_{21} & 0\cdots & a_{22} \end{bmatrix}-\frac{1}{a_{33}}\begin{bmatrix} a_{31} \\ 0 \\ \vdots \\ a_{32} \end{bmatrix}[a_{31}, 0, \dots, a_{32}]$$

where

$$a_{11}=\sum_{i=1}^{k} N_i\left[\frac{-1}{1-P_{i,s-1}}\left(\frac{\partial P_{i,s-1}}{\partial\beta}\right)^2 + \frac{1}{P_{is}-P_{i,s-1}}\left(\frac{\partial P_{is}}{\partial\beta}-\frac{\partial P_{i,s-1}}{\partial\beta}\right)^2+\frac{1}{1-P_{is}}\left(\frac{\partial P_{is}}{\partial\beta}\right)^2\right],$$

$$a_{21}=\sum_{i=1}^{k} N_i\left[\frac{1-P_{is}}{(1-P_{i,s-1})(P_{is}-P_{i,s-1})}\frac{\partial P_{i,s-1}}{\partial\beta}-\frac{1}{P_{is}-P_{i,s-1}}\frac{\partial P_{i,s}}{\partial\beta}\right]\frac{\partial P_{i,s-1}}{\partial\alpha_{s-1}},$$

$$a_{22}=\sum_{i=1}^{k} N_i\left[\frac{1-P_{is}}{(1-P_{i,s-1})(P_{is}-P_{i,s-1})}\right]\left(\frac{\partial P_{i,s-1}}{\partial\alpha_{s-1}}\right)^2,$$

$$a_{31}=\sum_{i=1}^{k} N_i\left[\frac{1-P_{i,s-1}}{(P_{is}-P_{i,s-1})(1-P_{is})}\frac{\partial P_{is}}{\partial\beta}-\frac{1}{P_{is}-P_{i,s-1}}\frac{\partial P_{i,s-1}}{\partial\beta}\right]\frac{\partial P_{is}}{\partial\alpha_s},$$

$$a_{32}=\sum_{i=1}^{k}\left[\frac{-N_i}{(P_{is}-P_{i,s-1})}\frac{\partial P_{is}}{\partial\alpha_s}\frac{\partial P_{i,s-1}}{\partial\alpha_{s-1}}\right],$$

$$a_{33}=\sum_{i=1}^{k} N_i\frac{1-P_{i,s-1}}{(P_{is}-P_{i,s-1})(1-P_{is})}\left(\frac{\partial P_{is}}{\partial P_{is}}\right)^2.$$

All the proofs in this paper are given in the Appendix.

By a similar argument, we can show that more information is gained by having a more refined classification of other ordered responses. Therefore, although there are additional parameters to be estimated in the augmented model, there is still a gain in information as a result of a more refined classification of the response.

A result similar to Theorem 1 may also hold for other information criteria. In particular, the next result compares the Fisher information matrix associated with the parameters (β,α_1) of any polytomous response model with that of the corresponding dichotomous response model. Based on this result, we can conclude that LD_p can be estimated more efficiently by using the polytomous response model.

Theorem 2. *Consider the Fisher information matrix J_s of an s-response model and partition J_s into*

$$J_s=\begin{bmatrix} J_{11}^{(2)}(s) & J_{12}^{(2)}(s)\\ J_{21}^{(2)}(s) & J_{22}^{(2)}(s)\end{bmatrix}$$

where $J_{11}^{(2)}(s)$ is the 2×2 *upper-left sub-matrix. Then $J_{11}^{(2)}(s)-J_{12}^{(2)}(s)J_{22}^{(2)\,-1}(s)J_{21}^{(2)}(s)\geq J_2$ for any $s\geq 3$.*

This theorem states that the Fisher information matrix associated with the parameters (β,α_1) in the polytomous model is always greater than or equal to the Fisher information matrix for the corresponding dichotomous response model. It is concluded that more information is gained in using the polytomous response model for estimating LD_p, or any smooth function of (β,α_1) than the dichotomous response model in large samples.

3. Numerical study

3.1. Evaluation of the Fisher information matrices

In this section, we perform a numerical comparison of the Fisher information matrices associated with the dichotomous response model and the polytomous response model. In particular, we assume the response curves to be logistic and $s=3$. Therefore, for given stimulus level x, it follows that

$$\begin{aligned}\Pr\{y=1 \mid x\} &= [1+\exp(-\beta x-\alpha_1)]^{-1},\\ \Pr\{y=2 \mid x\} &= [1+\exp(-\beta x-\alpha_2)]^{-1}-[1+\exp(-\beta x-\alpha_1)]^{-1},\\ \Pr\{y=3 \mid x\} &= 1-[1+\exp(-\beta x-\alpha_2)]^{-1}, \quad \alpha_1<\alpha_2.\end{aligned} \tag{7}$$

The dichotomous response model is obtained by combining the second and the third categories into one group. Thus the common parameters of interest are β and α_1.

Without loss of generality, we assume the first response curve is given by the standard logistic curve ($\beta=1, \alpha_1=0$). Two different design patterns are chosen. Scheme I chooses the 10, 40, 60, 90 percentiles of the first response curve as design points while the 5, 45, 55, 95 percentiles are used in Scheme II. Different values of α_2 are used. We fix the total sample size to be 40. Various numbers of observations are allocated to each percentile, the allocations varying from the balanced to the unbalanced, either skewed to the left or right. The purpose is to study the gain in information under different combinations of design scheme, allocation pattern of the subjects and the relative distance of the two response curves.

In Table 1, we give the ratio of the determinants of the Fisher information matrices for (β, α_1) under the two models. The formulae are given by (1), (4), (5), (6).

Table 1
Ratio of the determinants of the Fisher information matrices for the 3-response and 2-response model ($\beta=1$, $\alpha_1=0$)

Design scheme		α_2 = 0.1	0.5	1.0	1.5	2.0	3.0	4.0
(15,5,5,15)	I	1.04	1.18	1.29	1.32	1.28	1.15	1.06
	II	1.05	1.23	1.43	1.53	1.50	1.29	1.12
(10,10,10,10)	I	1.04	1.19	1.36	1.44	1.43	1.27	1.12
	II	1.05	1.25	1.53	1.77	1.87	1.62	1.29
(15,15,5,5)	I	1.04	1.21	1.39	1.48	1.47	1.29	1.13
	II	1.06	1.32	1.67	1.94	2.01	1.67	1.31
(5,5,15,15)	I	1.04	1.18	1.33	1.43	1.44	1.29	1.14
	II	1.03	1.17	1.38	1.58	1.70	1.58	1.29
(5,15,15,5)	I	1.04	1.19	1.37	1.50	1.54	1.39	1.19
	II	1.05	1.25	1.57	1.92	2.18	2.11	1.60

Note: (15,5,5,15) I means that 15,5,5,15 observations are respectively assigned to the four levels of design scheme I.

The two response curves are related in the following manners:

(i) If α_2 is close to α_1, the two curves are close to each other, i.e. the probability that $y=2$ is small and most of the responses observed are classified either as $y=1$ or $y=3$.

(ii) If α_2 is too large, the curve corresponding to O_1+O_2 is close to 1, i.e. the probability that $y=3$ is small and the responses observed are mostly $y=1$ or $y=2$.

In either case, one of the classes O_2, O_3 has few subjects. Therefore, it is not worth the additional labor in refining the response O_2+O_3 into O_2 and O_3. Thus, the gain in using the trichotomous response model in these cases would not be very significant. As we can see from the table, the ratio increases as α_2 increases to 2.0 and then decreases. When α_2 is sufficiently large, the corresponding ratio of the determinants of the information matrices is close to 1. These facts can be justified by the following results.

Define $\Delta I(\beta,\alpha_1)=J_{11}^{(2)}(3)-J_{12}^{(2)}(3)J_{22}^{(2)\,-1}(3)J_{21}^{(2)}(3)-J_2$ to be the gain in information in using the 3-response model, instead of the 2-response model, for estimating (β,α_1).

Theorem 3. *If the probability of each category is given by* (7), *then for fixed* α_1, $\Delta I(\beta,\alpha_1)\to 0$ *as* (i) $\alpha_2\to\alpha_1$ *or* (ii) $\alpha_2\to\infty$.

The percentage gains shown in the table are at least 5% with most of them more than 15%, some up to 100%. This gives an asymptotic justification for choosing the 3-response model to estimate the parameters. We have also computed such ratios for other design schemes. They exhibit the same pattern, and are therefore omitted.

In the previous comparison we use the determinant of the Fisher information matrix as a convenient summary criterion. A practical question is to what extent does this criterion reflect the performance of estimates of different parameters in the finite sample situations? It can only be answered by Monte Carlo simulation. Such results, reported in the next section, support the general conclusion of this section in using the 3-response model, instead of the 2-response model, for estimating (β,α_1).

3.2. *A simulation study*

We reiterate that our main interest is in estimating the percentiles associated with the outcome O_1. In this section, we present an extensive simulation study to compare the 3-response model and the 2-response model for estimating β, α_1, $LD_{0.1}$, $LD_{0.3}$, $LD_{0.5}$, $LD_{0.7}$ and $LD_{0.9}$ under different design schemes. The Fisher information in Section 3.1 is an asymptotic measure while Table 2 presents simulation results for sample size 40. The latter results provide a more realistic indication of the performance of these estimators in small samples under the two models.

Again, we assume that the underlying response curve $F(x\mid\beta,\alpha)$ is logistic. The $100p$-th percentile of the first response curve is given by

Table 2

Comparison of MSE of the MLE of parameters α_1, β and the percentiles in standard logistic model with 1000 simulation samples ($\alpha_1=0,\ \beta=1$)

Parameter	α_2 in 3-response model							2-response model
	0.1	0.5	1.0	1.5	2.0	3.0	4.0	
Design scheme (15,5,5,15) I								
α_1	0.26	0.26	0.24	0.23	0.25	0.26	0.27	0.33
β	0.15	0.15	0.11	0.10	0.11	0.15	0.18	0.64
$LD_{0.1}$	0.65	0.62	0.52	0.51	0.53	0.64	0.72	0.74
$LD_{0.3}$	0.26	0.25	0.22	0.22	0.23	0.26	0.28	0.28
$LD_{0.5}$	0.20	0.20	0.19	0.18	0.19	0.20	0.20	0.19
$LD_{0.7}$	0.28	0.28	0.27	0.25	0.26	0.26	0.27	0.27
$LD_{0.9}$	0.70	0.69	0.63	0.58	0.59	0.65	0.70	0.72
Design scheme (10,10,10,10) I								
α_1	0.18	0.18	0.17	0.17	0.17	0.17	0.18	0.19
β	0.21	0.21	0.14	0.11	0.13	0.17	0.23	0.28
$LD_{0.1}$	0.96	0.81	0.62	0.63	0.64	0.76	0.86	0.99
$LD_{0.3}$	0.28	0.25	0.21	0.22	0.22	0.24	0.26	0.29
$LD_{0.5}$	0.17	0.15	0.15	0.16	0.15	0.16	0.16	0.16
$LD_{0.7}$	0.29	0.26	0.24	0.23	0.23	0.26	0.27	0.28
$LD_{0.9}$	0.99	0.83	0.70	0.66	0.67	0.80	0.88	0.97
Design scheme (15,15,5,5) I								
α_1	0.19	0.19	0.20	0.19	0.20	0.20	0.20	0.39
β	0.23	0.19	0.16	0.13	0.14	0.16	0.27	1.72
$LD_{0.1}$	0.80	0.61	0.47	0.44	0.47	0.58	0.72	0.92
$LD_{0.3}$	0.18	0.18	0.16	0.17	0.17	0.19	0.19	0.21
$LD_{0.5}$	0.24	0.21	0.20	0.19	0.19	0.18	0.21	0.23
$LD_{0.7}$	0.64	0.46	0.41	0.36	0.36	0.37	0.50	0.63
$LD_{0.9}$	1.98	1.33	1.12	0.94	0.94	1.06	1.51	2.02
Design scheme (5,5,15,15) I								
α_1	0.20	0.19	0.19	0.18	0.19	0.20	0.19	0.22
β	0.22	0.20	0.19	0.14	0.14	0.18	0.23	0.25
$LD_{0.1}$	1.74	2.77	3.60	2.52	3.71	3.93	3.89	4.41
$LD_{0.3}$	0.54	0.89	1.15	0.81	1.19	1.24	1.24	1.38
$LD_{0.5}$	0.21	0.31	0.38	0.29	0.39	0.40	0.40	0.43
$LD_{0.7}$	0.21	0.19	0.21	0.20	0.20	0.21	0.19	0.22
$LD_{0.9}$	0.86	0.94	1.17	0.93	1.16	1.27	1.16	1.40
Design scheme (5,15,15,5) I								
α_1	0.14	0.15	0.14	0.14	0.14	0.14	0.14	0.15
β	0.32	0.26	0.25	0.21	0.19	0.19	0.29	0.39
$LD_{0.1}$	2.79	3.58	2.84	2.47	3.18	2.73	2.60	5.29
$LD_{0.3}$	0.57	0.83	0.67	0.60	0.76	0.63	0.60	1.20
$LD_{0.5}$	0.21	0.20	0.20	0.19	0.20	0.19	0.18	0.25
$LD_{0.7}$	0.65	0.42	0.43	0.37	0.38	0.42	0.42	0.55
$LD_{0.9}$	3.00	2.53	2.24	1.89	2.22	2.18	2.14	3.58

Note: The 2-response model corresponds to $\alpha_2=\infty$ in the 3-response model.

$$\mathrm{LD}_p = -[\alpha_1 + \log(p^{-1} - 1)]/\beta. \tag{8}$$

In the simulation study, we use $\alpha_1 = 0.0$, $\beta = 1.0$. Therefore, we have $\mathrm{LD}_{0.1} = -2.1972 = -\mathrm{LD}_{0.9}$, $\mathrm{LD}_{0.3} = -0.8473 = -\mathrm{LD}_{0.7}$, $\mathrm{LD}_{0.5} = 0$. The design schemes are the *same* as those considered in Table 1. Only results for scheme I are reported in Table 2. Results of scheme II and others exhibit similar patterns, and are therefore omitted. The simulation was performed on a VAX 11/750 at the University of Wisconsin-Madison. The uniform random numbers were generated by the IMSL subroutine GGUBS. For each of the design schemes considered, we generated 1000 random samples. The method of maximum likelihood was used in estimating the parameters. The MLE's were found by using the iterative Newton–Raphson method. Therefore, if $\hat{\alpha}_1, \hat{\beta}$ are the MLE of α_1, β, the MLE of LD_p is given by

$$\hat{\mathrm{L}}\mathrm{D}_p = -[\hat{\alpha}_1 + \log(p^{-1} - 1)]/\hat{\beta}. \tag{9}$$

Note that the MLE $\hat{\alpha}_1$ and $\hat{\mathrm{L}}\mathrm{D}_{0.5}$ need not take the same value although $\alpha_1 = \mathrm{LD}_{0.5} = 0$.

Summary of Table 2

1. For nearly all the design schemes considered, the estimates of (β, α_1) from the trichotomous response model have smaller MSE than those from the dichotomous response model. The reduction is about 50% in the best case. Therefore, even for small samples, the results of Section 2 appear to hold.

2. The MSE of the estimates $\hat{\beta}$ and $\hat{\alpha}_1$ are minimized at $\alpha_2 = 1.5$ or 2. But, the MSE increases when α_2 is too large or small. Therefore, the trichotomous response model has an advantage only if the two curves for O_1 and $O_1 + O_2$ are not too close to each other, or if the curve for $O_1 + O_2$ is not too close to 1. See Theorem 3 for a theoretical justification.

3. The simulation study shows that there is a significant improvement in estimating β by using the trichotomous response model while the gain is not so significant in estimating α_1. This can be explained by observing that the slope parameter β is common to the two curves in the 3-response model and can be estimated efficiently by 'borrowing strength' from data in all categories. An attempt to justify this is made in Theorem 4.

4. For estimation of the extreme percentiles, $\mathrm{LD}_{0.1}$ or $\mathrm{LD}_{0.9}$, there is a large reduction in MSE by using the 3-response model instead of the 2-response model. On the other hand, the reduction is not so remarkable for the median $\mathrm{LD}_{0.5}$.

5. When α_2 is too large or small, i.e. $\alpha_2 = 0.1$ or 4, the MSE from the trichotomous response model may produce slightly larger MSE than those from the dichotomous response model. However, the differences in all cases are less than 5%. This may be due to the sampling fluctuation in the Monte Carlo simulation.

6. For the estimation of low (high) percentiles, the designs that are skew to the right (left), i.e. allocating more subjects at the low (high) stimulus levels, perform better than the others. The MSE associated with the extreme percentiles $\mathrm{LD}_{0.1}$ and

$LD_{0.9}$ are higher than the others. This illustrates the fact that the estimates of these extreme percentiles are very unstable. In particular, the V-shaped design (i.e. more observations on the extremes) gives the smallest MSE for the extreme percentiles while those associated with the Λ-shaped design (i.e. fewer observations on the extremes) are the largest. However, the balanced design gives overall good performance in estimating the percentiles.

Point 3 will be further justified as follows. Let $\theta=(\beta,\alpha_1)$ and $\delta=\alpha_2-\alpha_1$. Under the 3-response model, asymptotically,

$$\mathrm{Cov}(\hat{\theta}) \cong [J_{11}^{(2)}(3)-J_{12}^{(2)}(3)J_{22}^{(2)\,-1}(3)J_{21}^{(2)}(3)]^{-1} = \begin{pmatrix} \lambda^{\beta\beta} & \lambda^{\alpha\beta} \\ \lambda^{\alpha\beta} & \lambda^{\alpha\alpha} \end{pmatrix}.$$

Similarly, under the 2-response model,

$$\mathrm{Cov}(\hat{\theta}) \cong J_2^{-1} = \begin{pmatrix} I^{\beta\beta} & I^{\alpha\beta} \\ I^{\alpha\beta} & I^{\alpha\alpha} \end{pmatrix}.$$

Theorem 4. *Assume the response curves are logistic. If the design scheme is symmetric about the median* $LD_{0.5}$ *of the first response curve, then for small* $\delta>0$,

$$\frac{I^{\alpha\alpha}-\lambda^{\alpha\alpha}}{I^{\alpha\alpha}} < \frac{I^{\beta\beta}-\lambda^{\beta\beta}}{I^{\beta\beta}},$$

i.e. the percent-reduction in variance of the estimate $\hat{\beta}$ *is larger than that of* $\hat{\alpha}_1$.

A heuristic justification of point 4 follows from the above result. Let $\boldsymbol{a}=(-LD_p/\beta,-1/\beta)$, then $\mathrm{var}(\hat{LD}_p)=\boldsymbol{a}\,\mathrm{cov}(\hat{\theta})\boldsymbol{a}^{\mathrm{T}}$. For extreme percentiles, $(LD_p/\beta)^2\,\mathrm{var}(\hat{\beta})$ becomes the dominant term of $\mathrm{var}(\hat{LD}_p)$. A large percent-reduction in variance of $\hat{B}$ would help reduce the variance of $\hat{LD}_p$.

4. An example

The data in Table 3 are taken from Maxwell (1961, p. 70) and concern the degree of disturbed dreams among 223 boys aged 5-15. Originally, there were four categories. We combine the data so that the degree of severity of disturbed dreams of each boy is classified into three ordered and mutually exclusive classes.

Define $x=-$(mid-point of each age-category). A preliminary plot of the transformed variables

$$\log\left(\frac{r_{i1}+\frac{1}{2}}{N_i-r_{i1}+\frac{1}{2}}\right) \quad \text{and} \quad \log\left(\frac{r_{i1}+r_{i2}+\frac{1}{2}}{N_i-r_{i1}-r_{i2}+\frac{1}{2}}\right)$$

against x reveals approximately linear and parallel relationships. Therefore, we fit the data to the logistic model (7). The estimates are $\hat{\beta}=0.18$, $\hat{\alpha}_1=0.38$, $\hat{\alpha}=1.38$. The estimated covariance matrix for $(\hat{\beta},\hat{\alpha}_1)$ is

$$\hat{\text{cov}}(\hat{\beta},\hat{\alpha}_1)=\begin{pmatrix}0.0029 & 0.03\\ 0.03 & 0.34\end{pmatrix}.$$

The Pearson χ^2-statistic is 5.01 with 7 degrees of freedom.

Suppose we are interested in the minimum age that a boy would suffer from severely disturbed dream for only 10% of the time, i.e. to estimate the 10 percentile of the response curve related to category 1. Based on the above result, we have $\hat{\text{LD}}_{0.10}=-14.32$ with $\hat{\text{S}}.\text{E}.=2.29$, i.e. the age is 14.32.

Suppose we only rate the disturbed dreams as 'severe' or 'not severe'. Therefore, the data are obtained by combining categories 2 and 3 into one group. We fit a binary logistic curve to the combined data. The corresponding estimates are $\tilde{\beta}=0.19$ and $\tilde{\alpha}_1=0.54$. The estimated covariance matrix for $(\tilde{\beta},\tilde{\alpha}_1)$ is

$$\hat{\text{cov}}(\tilde{\beta},\tilde{\alpha}_1)=\begin{pmatrix}0.0045 & 0.047\\ 0.047 & 0.51\end{pmatrix}.$$

The Pearson χ^2-statistics is 1.79 with 3 degrees of freedom. The estimate of the 10 percentile is $\hat{\text{LD}}_{0.10}=-14.41$ with $\hat{\text{S}}.\text{E}.=2.71$.

Comparing the estimates from the two models, we can see that there is about 20% reduction in standard errors of the two estimates of β and α_1. For estimation of the $\text{LD}_{0.10}$, the reduction is about 15%.

5. Conclusion

Generally, our main concern is with the estimation of percentiles. The simulation study shows that in most cases considered, the MSE of $\hat{\text{LD}}_p$ from the trichotomous response model is less than that of the dichotomous response model. Although the reductions are negligible in some cases, the majority of them are from 5% to 10%, some up to 20%. Therefore, it is worthwhile to use the trichotomous response model if the classification of outcomes can be defined clearly and implemented easily. Especially if one is interested in estimating the extreme percentiles and the test subjects are expensive, it would be very costly to use a relatively large sample size in order to have reliable estimates. The use of the trichotomous response model is a viable alternative for providing better estimates of these extreme percentiles. Therefore it should be seriously considered in practice.

Appendix

Proof of Theorem 1. Define $\boldsymbol{\theta}=(\beta,\alpha_1,\ldots,\alpha_s)$, $\boldsymbol{R}=(\boldsymbol{R}_1,\boldsymbol{R}_2,\ldots,\boldsymbol{R}_k)$, $\boldsymbol{R}'=(\boldsymbol{R}_1',\boldsymbol{R}_2',\ldots,\boldsymbol{R}_k')$ and let $J(\boldsymbol{\theta}\mid\boldsymbol{R}), J(\boldsymbol{\theta}\mid\boldsymbol{R}')$ be the Fisher information matrices associated with the vectors $\boldsymbol{R},\boldsymbol{R}'$ respectively, i.e. information from the $(s+1)$-response and s-response model.

By the result of the general Missing Information Principle (Orchard and Woodbury, 1972), we have

$$J(\theta \mid \boldsymbol{R}) = J(\theta \mid \boldsymbol{R}') + J(\theta, \boldsymbol{R}' \mid \boldsymbol{R}) \tag{A1}$$

and $J(\theta; \boldsymbol{R}' \mid \boldsymbol{R}) \geq 0$, which is the information lost from combining O_s and O_{s+1} into one category. In particular, $J(\theta \mid \boldsymbol{R})$ is the expectation of the symmetric matrix

$$\begin{bmatrix} -\partial^2 \log L_{s+1}/\partial\beta^2 & \cdots & \\ -\partial^2 \log L_{s+1}/\partial\beta\,\partial\alpha_1 & -\partial^2 \log L_{s+1}/\partial\alpha_1^2 & \\ \vdots & & \ddots \\ -\partial^2 \log L_{s+1}/\partial\beta\,\partial\alpha_s & & & -\partial^2 \log L_{s+1}/\partial\alpha_s^2 \end{bmatrix}$$

$$= \begin{bmatrix} J_{11}^{(s+1)} & J_{12}^{(s+1)} \\ J_{21}^{(s+1)} & J_{22}^{(s+1)} \end{bmatrix}$$

and $J(\theta \mid \boldsymbol{R}')$ is obtained by replacing the last row and column of the above matrix by zero, i.e.

$$J(\theta, \boldsymbol{R}') = \begin{pmatrix} J_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

By (A1), we have $J(\theta \mid \boldsymbol{R}) \geq J(\theta \mid \boldsymbol{R}')$, which is equivalent to

$$\begin{bmatrix} J_{11}^{(s)}(s+1) - J_s & J_{12}^{(s)}(s+1) \\ J_{21}^{(s)}(s+1) & J_{22}^{(s)}(s+1) \end{bmatrix} \geq 0$$

which in turn implies $J_{11}^{(s)}(s+1) - J_s - J_{12}^{(s)}(s+1) J_{22}^{(s)\,'}(s+1) J_{21}^{(s)}(s+1) \geq 0$, thus proving Theorem 1. The formulae for a_{ij} are easily obtained from computing

$$E\left(\frac{-\partial^2 \log L_{s+1}}{\partial\theta_i\,\partial\theta_j}\right) \quad \text{and} \quad E\left(\frac{-\partial^2 \log L_s}{\partial\theta_i\,\partial\theta_j}\right).$$

Proof of Theorem 2. Since

$$J(\theta \mid \boldsymbol{R}') = E \begin{bmatrix} -\partial^2 \log L_2/\partial\beta^2 & -\partial^2 \log L_2/\partial\beta\,\partial\alpha_1 & 0 & \cdots & 0 \\ -\partial^2 \log L_2/\partial\beta\,\partial\alpha_1 & -\partial^2 \log L_2/\partial\alpha_1^2 & 0 & & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix},$$

the result follows by applying an argument similar to the above.

Proof of Theorem 3. Let $P_{ij} = [1 + \exp(-\beta x_i - \alpha_j)]^{-1}$ and $Q_{ij} = 1 - P_{ij}$, $j = 1, 2$, $i = 1, 2, \ldots, k$. The a_{ij} defined in Theorem 1 are simplified as follows:

$$a_{11}=\sum_{i=1}^{k} n_i x_i^2(P_{i2}-P_{i1})Q_{i1}Q_{i2}, \qquad a_{12}=\sum_{i=1}^{k} n_i x_i P_{i1}Q_{i1}Q_{i2},$$

$$a_{22}=\sum_{i=1}^{k}\frac{n_i P_{i1}^2 Q_{i1}Q_{i2}}{P_{i2}-P_{i1}}, \qquad a_{13}=\sum_{i=1}^{k} n_i x_i P_{i2}Q_{i1}Q_{i2},$$

$$a_{23}=-\sum_{i=1}^{k}\frac{n_i P_{i1}P_{i2}Q_{i1}Q_{i2}}{P_{i2}-P_{i1}}, \qquad a_{33}=\sum_{i=1}^{k}\frac{n_i P_{i2}^2 Q_{i1}Q_{i2}}{P_{i2}-P_{i1}}.$$

(i) $\delta=\alpha_2-\alpha_1\to\infty$:

The positive definiteness of J_3 implies $J_{21}^{(2)}(3)J_{22}^{(2)\,'}(3)J_{21}^{(2)}(3)\geq 0$. By Theorem 1, we have $0\leq \Delta I(\beta,\alpha_1)\leq J_{11}^{(2)}(3)-J_2$. As $\delta\to\infty$, $P_{i2}\to 1$ and $Q_{i2}\to 0$. But

$$J_{11}^{(2)}(3)-J_2=\begin{pmatrix} a_{11} & a_{12}\\ a_{12} & a_{22}\end{pmatrix}\to 0$$

because Q_{i2} in a_{ij} converges to 0. This implies $\Delta I(\beta,\alpha_1)\to 0$.

(ii) $\delta=\alpha_2-\alpha_1\to 0$:

This implies $P_{i2}\to P_{i1}$ and $Q_{i2}\to Q_{i1}$, $i=1,2,\ldots,k$. Using Taylor series expansion of P_{i2} in α_2 around α_1, i.e. expansion about $\alpha=0$, for $i=1,2,\ldots,k$, we have $P_{i2}-P_{i1}=\delta P_{i1}Q_{i1}+h_i(\delta)$ where $h_i(\delta)=\mathrm{O}(\delta^2)$ which implies

$$\frac{P_{i2}-P_{i1}}{\delta P_{i1}Q_{i1}}=1+\mathrm{O}(\delta) \tag{A2}$$

where $\mathrm{O}(\delta)$ is a term of order δ. Consider

$$\begin{aligned}
&a_{11}-a_{13}^2/a_{33}\\
&=\sum_{i=1}^{k} n_i x_i^2(P_{i2}-P_{i1})Q_{i1}Q_{i2}-\left(\sum_{i=1}^{k} n_i x_i P_{i2}Q_{i1}Q_{i2}\right)^2\left(\sum_{i=1}^{k}\frac{n_1 P_{i2}^2 Q_{i1}Q_{i2}}{P_{i2}-P_{i1}}\right)^{-1}\\
&=\sum_{i=1}^{k} n_i x_i^2(P_{i2}-P_{i1})Q_{i1}Q_{i2}-\delta\left(\sum_{i=1}^{k} n_i x_i P_{i2}Q_{i1}Q_{i2}\right)^2\left(\sum_{i=1}^{k}\frac{n_i P_{i2}^2 Q_{i1}Q_{i2}}{P_{i1}Q_{i1}[1+\mathrm{O}(\delta)]}\right)^{-1}\\
&\to 0 \quad \text{as } \delta\to 0, \text{ where } n_i, x_i \text{ are fixed.}
\end{aligned}$$

By using a similar argument, every entry of

$$\Delta I(\beta,\alpha_1)=\begin{bmatrix} a_{11}-a_{13}^2 a_{33}^{-1} & a_{12}-a_{13}a_{23}a_{33}^{-1}\\ a_{12}-a_{13}a_{23}a_{33}^{-1} & a_{22}-a_{23}^2 a_{33}^{-1}\end{bmatrix}$$

tends to zero, thus $\Delta I(\beta,\alpha_1)\to 0$.

Proof of Theorem 4. Recall $\Delta I(\beta,\alpha_1)=J_{11}^{(2)}(3)-J_2-J_{12}^{(2)}(3)J_{22}^{(2)\,'}(3)J_{21}^{(2)}(3)\to 0$ as $\delta\to 0$. For small δ, the asymptotic covariance matrix of $(\hat\beta-\beta,\hat\alpha_1-\alpha_1)$ under the 3-response model is

$$\begin{aligned}
[J_{11}^{(2)}(3)-J_{12}^{(2)}(3)J_{22}^{(2)\,'}(3)J_{21}^{(2)}(3)]^{-1} &= [J_2+\Delta I(\beta,\alpha_1)]^{-1}\\
&\simeq J_2^{-1}-J_2^{-1}\Delta I(\beta,\alpha_1)J_2^{-1}
\end{aligned}$$

where

$$J_2 = \begin{bmatrix} \sum_{i=1}^{k} n_i x_i^2 P_{i1} Q_{i1} & \sum_{i=1}^{k} n_i x_i P_{i1} Q_{i1} \\ \sum_{i=1}^{k} n_i x_i P_{i1} Q_{i1} & \sum_{i=1}^{k} n_i P_{i1} Q_{i1} \end{bmatrix} = \begin{bmatrix} l_{\beta\beta} & l_{\alpha\beta} \\ l_{\alpha\beta} & l_{\alpha\alpha} \end{bmatrix}.$$

From the symmetry of design, $l_{\alpha\beta}=0$. Then $J_2^{-1}=\operatorname{diag}(l_{\beta\beta}^{-1}, l_{\alpha\alpha}^{-1})$ so $l^{\beta\beta}=l_{\beta\beta}^{-1}$, $l^{\alpha\alpha}=l_{\alpha\alpha}^{-1}$ and $l^{\alpha\beta}=0$. Also

$$J_2^{-1}\Delta I(\beta,\alpha_1)J_2^{-1} = \begin{bmatrix} l_{\beta\beta}^{-2}(a_{11}-a_{13}^2 a_{33}^{-1}) & l_{\alpha\alpha}^{-1} l_{\beta\beta}^{-1}(a_{12}-a_{13}a_{23}a_{33}^{-1}) \\ l_{\alpha\alpha}^{-1} l_{\beta\beta}^{-1}(a_{12}-a_{13}a_{23}a_{33}^{-1}) & l_{\alpha\alpha}^{-2}(a_{22}-a_{23}^2 a_{33}^{-1}) \end{bmatrix}$$

Using Taylor series expansion of P_{i2} in α_2 around α_1, and neglecting terms of order higher than δ^2,

$$P_{i2} = P_{i1} + P_{i1}Q_{i1}\delta + \tfrac{1}{2}\delta^2 P_{i1}Q_{i1}(Q_{i1}-P_{i1}) + O(\delta^3). \tag{A3}$$

Substitute (A3) for P_{i2} and Q_{i2} in a_{ij}. With the symmetry condition, after some algebra, we have

$$a_{11} \simeq \delta \sum_{i=1}^{k} n_i x_i^2 P_{i1} Q_{i1}^3, \qquad a_{13} \simeq \sum_{i=1}^{k} n_i x_i P_{i1}^2 Q_{i1},$$

$$a_{22} \simeq \left(\frac{1}{\delta}-\frac{1}{2}\right) l_{\alpha\alpha}, \qquad a_{23} \simeq \left(\frac{1}{\delta}-\frac{\delta}{4}\right) l_{\alpha\alpha}, \qquad a_{33} \simeq \left(\frac{1}{\delta}+\frac{1}{2}-\frac{\delta}{4}\right) l_{\alpha\alpha}.$$

Therefore

$$\frac{l^{\alpha\alpha}-\lambda^{\alpha\alpha}}{l^{\alpha\alpha}} = \frac{l^{\alpha\alpha}-[l^{\alpha\alpha}-l_{\alpha\alpha}^{-2}(a_{22}-a_{23}^2 a_{33}^{-1})]}{l^{\alpha\alpha}}$$

$$= \left(\frac{1}{\delta}-\frac{1}{2}\right) - \frac{\left(\frac{1}{\delta}-\frac{\delta}{4}\right)^2}{\left(\frac{1}{\delta}+\frac{1}{2}-\frac{\delta}{4}\right)} \simeq \frac{\delta^2}{8}\left(1-\frac{\delta}{2}\right)^2 = O(\delta^2),$$

$$\frac{l^{\beta\beta}-\lambda^{\beta\beta}}{l^{\beta\beta}} = \frac{l^{\beta\beta}-[l^{\beta\beta}-l_{\beta\beta}^{-2}(a_{11}-a_{13}^2 a_{33}^{-1})]}{l^{\beta\beta}}$$

$$= \frac{\delta}{l_{\alpha\alpha} l_{\beta\beta}} \left\{ \left(\sum_{i=1}^{k} n_i x_i^2 P_{i1} Q_{i1}^3\right)\left(\sum_{i=1}^{k} n_i P_{i1} Q_{i1}\right) - \left(\sum_{i=1}^{k} n_i x_i P_{i1} Q_{i1}^2\right)^2 \right\}.$$

The expression inside the bracket is independent of δ and is positive by the Cauchy-Schwarz Inequality, thus $(l^{\beta\beta}-\lambda^{\beta\beta})/l^{\beta\beta}=O(\delta)$. Therefore, for small δ, $(l^{\alpha\alpha}-\lambda^{\alpha\alpha})/l^{\alpha\alpha} \le (l^{\beta\beta}-\lambda^{\beta\beta})/l^{\beta\beta}$.

Acknowledgements

We thank Paul Rosenbaum for suggesting the reference Ochard and Woodbury (1972), which simplifies the original proof of Theorem 1. Thanks are also due to Mike Meyer for useful comments.

This research is sponsored by the United States Army under contract No. DAAG29-80-C-0041 and by the National Foundation under grant No. MCS-8300140.

References

Aitchison, J. and S.D. Silvey (1957). The generalization of probit analysis to the case of multiple response. *Biometrika* 44, 131–140.

Ashford, J.R. (1959). The approach to the analysis of data for semi-quantal responses in biological assays. *Biometrics* 15, 573–581.

Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* 13, 342–368.

Finney, D.J. (1978). *Statistical Method in Biological Assay.* Hafner, New York.

Gurland, J., I. Lee and P.A. Dahm (1960). Polychotomous quantal response in biological assay. *Biometrics* 16, 382–398.

Maxwell, A.E. (1961). *Analysing Qualitative Data.* Methuen, London.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J. Roy. Statist. Soc. Ser. B* 42, 109–142.

McCullagh, P. and J.A. Nelder (1983). *Generalized Linear Models.* Chapman and Hall, London.

Orchard, T. and M.A. Woodbury (1972). A missing information principle: Theory and applications. *6th Berkeley Symp. Math. Statist. Prob.* 1, 697–715.