

**The Beta Distribution as a Latent Response
Model for Ordinal Data (I): Estimation of
Location and Dispersion Parameters**

Ajit C. Tamhane, Bruce E. Ankenman and Ying Yang

Department of Industrial Engineering & Management Sciences

Northwestern University

Evanston, IL 60208, U.S.A.

Abstract

Ordinal data are often modeled using a continuous latent response distribution, which is partially observed through windows of adjacent intervals defined by cutpoints. In this paper we propose the beta distribution as a model for the latent response. The beta distribution has several advantages over the other common distributions used, e.g., normal and logistic. In particular, it enables separate modeling of location and dispersion effects which is essential in the Taguchi method of robust design.

First, we study the problem of estimating the location and dispersion parameters of a single beta distribution (representing a single treatment) from ordinal data assuming known equispaced cutpoints. Two methods of estimation are compared: the maximum likelihood method and the method of moments. Two methods of treating the data are considered: in raw discrete form and in smoothed continuousized form. A large scale simulation study is carried out to compare the different methods. The mean square errors of the estimates are obtained under a variety of parameter configurations. Comparisons are made based on the ratios of the mean square errors (called the relative efficiencies). No method is universally the best, but the maximum likelihood method using continuousized data is found to perform generally well, especially for estimating the dispersion parameter. This method is also computationally much faster than the other methods and does not experience convergence difficulties in case of sparse or empty cells.

Next, the problem of estimating unknown cutpoints is addressed. Here the multiple treatments setup is considered since in an actual application, cutpoints are common to all treatments, and must be estimated from all the data. A two-step iterative algorithm is proposed for estimating the location and dispersion parameters of the treatments, and the cutpoints. The proposed beta model and McCullagh's (1980) proportional odds model are compared by fitting them to two real data sets.

Keywords: Maximum likelihood method; method of moments; mean square error; latent variable model; proportional odds model; logistic response distribution; incomplete data; ordered categorical data; Taguchi method.

1. Introduction

Ordinal data are commonly used in many areas of application, some examples being the quality of an item or service or performance (poor, fair, good, very good or excellent), seriousness of a defect (minor, major, critical), taste of food (too mild, just right, too spicy) and extent of agreement (strongly disagree, disagree, neutral, agree, strongly agree). Many techniques are available for analyzing stochastic shifts in ordinal data; for a review see Agresti (1984). However, serious difficulties arise when inferences are desired on both location and dispersion effects; see Nair (1986) and Hamada and Wu (1990) and the accompanying discussions. Both these effects are of interest in Taguchi's (1986) approach to quality improvement where the goal is not only to have the average product performance on target, but also to minimize the variation around the target. Taguchi recommends identifying two sets of design factors: control factors that can be used to minimize the dispersion and adjustment factors that can be used to bring the location on target. To implement this strategy it is necessary to be able to measure and analyze the location and dispersion effects separately from each other.

The main cause of difficulty in separating the location effects from dispersion effects when the data are ordinal is that the number of categories is usually small (between 3 and 10). Therefore, when the location parameter is pushed to the limit (either too high or too low), most of the data fall in the extreme category giving a false impression of reduced variance.

A common approach to the analysis of ordinal data is to assume a continuous latent response distribution that is observed through windows of ordered intervals with fixed, but unknown cutpoints. This approach is implicit in the proportional odds model (McCullagh, 1980), which can be derived from an underlying logistic response distribution and in other generalized linear models (McCullagh and Nelder, 1989). These models are typically based on the assumption of a symmetric continuous latent response having an infinite domain. They also do not separate the location and dispersion effects besides failing to take into account the finite nature of the measurement scale used in ordinal data.

To resolve these difficulties, we propose the beta distribution as a model for the latent variable. The following three properties of the beta distribution make it especially suitable

for modeling ordinal data:

1. The beta distribution has a finite domain.
2. The beta distribution can flexibly model a wide variety of shapes including a bell-shape (symmetric or skewed), U-shape and J-shape.
3. For the beta distribution on the interval $[0, 1]$ with the probability density function (p.d.f.):

$$f(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \text{ for } x \in [0, 1], \quad (1.1)$$

where $a, b > 0$, the mean and variance are given by

$$\mu = \frac{a}{a+b} \text{ and } \sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mu(1-\mu)}{a+b+1} = \mu(1-\mu)\eta^2, \quad (1.2)$$

where

$$\eta^2 = \frac{1}{(a+b+1)}. \quad (1.3)$$

Note that μ , which may be taken to be the location parameter, affects σ^2 through $\mu(1-\mu)$. In particular, $\sigma^2 \rightarrow 0$ as $\mu \rightarrow 0$ or 1 , as it should due to the location shift that pushes data into extreme categories. Also, effects on σ^2 of changes in location and scale are separately quantified through two terms, $\mu(1-\mu)$ and η^2 , respectively. Therefore we may regard η^2 as a pure dispersion parameter.

The basic idea here is similar to that in Leon, Shoemaker and Kacker (1987) where the goal was to find a performance measure of dispersion that is independent of adjustment factors (PerMIA), i.e., that is unaffected by changes in location. Box (1988) also made a similar suggestion that the data be transformed to achieve ‘‘separation,’’ i.e., eliminate the dependence of the variance on the mean. We claim that the beta distribution provides a suitable underlying metric for ordinal data and η^2 is an appropriate PerMIA.

Our long term goal in the proposed research program is to develop a comprehensive statistical methodology for analyzing multifactor experiments with ordinal data to identify adjustment factors that have significant effects on the location parameter μ and control factors that have significant effects on the dispersion parameter η^2 . A key step in this methodology is to impute continuous beta scores from the count data. Normalizing transformations will

then be applied to the beta scores, and the resulting data will be analyzed for location and dispersion effects using suitably modified normal theory techniques that are familiar to many engineers and are available in commercial packages. To obtain the beta scores, we first need to develop a good method for estimating μ and η^2 . The primary goal of the present paper is to compare the maximum likelihood method and the method of moments for this purpose using the raw (discrete) data or smoothed (continuousized) data. How the estimates of μ and η^2 will be utilized in the proposed methodology will be the topic of future papers.

One might argue that there is no need to look beyond the maximum likelihood method. However, there are issues other than statistical efficiency, such as numerical stability of estimates and convergence, which are equally important in practice, and it is not *a priori* clear that one method will have an advantage over the other. In fact, even as far as statistical efficiency is concerned, it is not clear that one method will always be superior. Therefore it is necessary to compare the competing methods in detail so that we can determine which method should be used under which conditions. We investigate these questions by first focusing on a single beta distribution corresponding to a single treatment with known equispaced cutpoints. After having determined the best method for estimating μ and η^2 for this setup, we then address the problem of estimating unknown cutpoints. In this latter case we consider the multiple treatments setup since the cutpoints are common to all treatments and must be estimated from all the data, whereas μ and η^2 will be estimated separately for each treatment. Of course, these two estimation problems are interdependent. Therefore we propose a two-step iterative method of estimation. We begin with some starting values for the cutpoints (e.g., equispaced) and estimate μ and η^2 (or equivalently a and b) for each treatment. Using these estimates of (μ, η^2) from each treatment we then estimate the cutpoints for all treatments. These two steps are iterated until convergence.

The paper is organized as follows. Section 2 gives the formulation of the problem and the necessary notation. Sections 3 through 6 focus on the estimation of μ and η^2 for a single beta distribution. Section 3 discusses the maximum likelihood and the method of moment estimation using discrete data. Section 4 extends these methods to continuousized data, which are obtained from the raw data by spreading the observations in each cell uniformly. When there are empty or sparse cells, all except one of these methods face

convergence problems. Section 5 proposes to avoid this difficulty by adding 1/2 (called a pseudo-observation) to each cell. Section 6 describes a simulation study to compare the different methods of estimation. Section 7 addresses the problem of estimation of cutpoints. Section 8 applies the proposed method to a real data example; McCullagh's proportional odds model is also fitted to the same data for comparison purposes. Finally, Section 9 gives conclusions and recommendations.

2. Problem Formulation and Notation

Let $I \geq 2$ denote the number of treatments (combinations of different factor levels). We first consider the problem of finding the best method of estimating μ and η^2 for each treatment. For this purpose we begin by assuming known cutpoints and focus on the single treatment case through Section 6. Suppose that the data from the treatment consist of a random sample of n items, which are classified into $J \geq 2$ ordered categories (cells) with n_j items in the j th cell and $\sum_{j=1}^J n_j = n$. As in McCullagh (1980), we assume that these categorized data are a manifestation of latent continuous random variables (r.v.'s) X_1, X_2, \dots, X_n such that the i th item is observed to fall in the j th cell if and only if $c_{j-1} \leq X_i < c_j$ where $c_0 < c_1 < c_2 < \dots < c_{J-1} < c_J$ are known cutpoints; a special case of interest is the equispaced cutpoints: $c_j = j/J$, $j = 0, 1, 2, \dots, J$. We assume that the X_i 's are i.i.d. with the beta distribution given by (1.1). The goal is to estimate (μ, η^2) from the vector of observed cell counts $\mathbf{n} = (n_1, n_2, \dots, n_J)$.

Let

$$F(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x u^{a-1}(1-u)^{b-1} du,$$

denote the cumulative distribution function (c.d.f.) of the beta distribution. For brevity, we will use the notation $F_j = F(c_j|a, b)$, $j = 0, 1, 2, \dots, J$. Then the cell counts vector $\mathbf{n} = (n_1, n_2, \dots, n_J)$ is a realization of a multinomial random vector $\mathbf{N} = (N_1, N_2, \dots, N_J)$ with cell probabilities $p_j = F_j - F_{j-1}$, $j = 1, 2, \dots, J$ and $\sum N_j = \sum n_j = n$.

Because the beta distribution is parameterized in terms of (a, b) , it is more convenient to find the estimates (\hat{a}, \hat{b}) first and then calculate

$$\hat{\mu} = \frac{\hat{a}}{\hat{a} + \hat{b}} \text{ and } \hat{\eta}^2 = \frac{1}{\hat{a} + \hat{b} + 1}. \quad (2.1)$$

3. Estimation for Discrete Data

3.1 Maximum Likelihood Method

The standard method for estimation of parameters from incomplete data is the maximum likelihood method; see, e.g., Sundberg (1974, 1976) and Dempster, Laird and Rubin (1976). The maximum likelihood estimate of (a, b) maximizes the kernel of the log-likelihood function:

$$\ln L = \sum_{j=1}^J n_j \ln p_j. \quad (3.1)$$

We refer to the resulting estimates as MLE-D (maximum likelihood estimates for discrete data).

Various algorithms can be used to find (\hat{a}, \hat{b}) . One possibility is to find (\hat{a}, \hat{b}) directly using a suitable nonlinear programming algorithm. The second possibility is to solve the equations

$$\frac{\partial \ln L}{\partial a} = \sum_{j=1}^J \frac{n_j}{p_j} \frac{\partial p_j}{\partial a} = 0 \text{ and } \frac{\partial \ln L}{\partial b} = \sum_{j=1}^J \frac{n_j}{p_j} \frac{\partial p_j}{\partial b} = 0, \quad (3.2)$$

using either the Newton-Raphson or the scores algorithm. The third possibility is to use the EM-algorithm of Dempster, Laird and Rubin (1976). Since the beta distribution is a member of the exponential family, closed form expressions can be obtained for the iterative steps of the EM-algorithm; the details are omitted for brevity.

In our empirical study all three algorithms for computing the MLE-D of (a, b) experienced convergence problems when μ was close to 0 or 1, and/or when η^2 was small. In these cases most of the data are concentrated in a few cells. The most satisfactory method was the nonlinear programming algorithm.

3.2 Method of Moments

To apply the method of moments we begin by assuming that all observations in each cell are concentrated at the midpoint of that cell. Let m_j be the midpoint of cell j :

$$m_j = \begin{cases} \frac{1}{2}(c_{j-1} + c_j) & \text{for general cutpoints} \\ \frac{1}{2J}(2j - 1) & \text{for equispaced cutpoints.} \end{cases} \quad (3.3)$$

Then denoting the sample mean and the sample variance of these data by \bar{x} and s^2 , respectively, the method of moments estimates of μ and η^2 can be easily computed using

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^J n_j m_j \text{ and } \hat{\eta}^2 = \frac{s^2}{\hat{\mu}(1 - \hat{\mu})} = \frac{\sum_{j=1}^J n_j m_j^2 - (\sum_{j=1}^J n_j m_j)^2/n}{(n-1)\bar{x}(1 - \bar{x})}. \quad (3.4)$$

We will refer to the resulting estimates as MME-D (method of moments estimates for discrete data).

We performed preliminary simulations to evaluate these estimates which showed that $\hat{\mu}$ and $\hat{\eta}^2$ are highly biased. In general, $\hat{\mu} = \bar{x}$ is positively biased for $\mu < 1/2$ and negatively biased for $\mu > 1/2$. Due to the assumption that all observations in each cell are concentrated at the midpoint of that cell, s^2 often underestimates σ^2 , especially when observations are clumped into a few cells. The simulation study also showed that in general $\bar{x}(1 - \bar{x})$ overestimates $\mu(1 - \mu)$. Therefore $\hat{\eta}^2$ generally underestimates η^2 . Even if a modified method of moments for bias reduction due to MacKinnon and Smith (1998) and described in Section 4.2 is used, a substantial negative bias remains in $\hat{\eta}^2$. This suggests that an improved method of moments estimates can be obtained by spreading the data in each cell. We refer to the resulting data as *continuousized data*. Estimation methods for continuousized data are given in the next section. Henceforth we will not consider the MME-D method.

4. Estimation for Continuousized Data

Recall that the observed data are discrete cell counts $\mathbf{n} = (n_1, n_2, \dots, n_J)$. Let $\{x_{jk}, k = 1, 2, \dots, n_j; j = 1, 2, \dots, J\}$ be the continuousized data obtained by uniform spreading in each cell:

$$x_{jk} = \begin{cases} c_{j-1} + \frac{kd_j}{(n_j+1)} & \text{for general cutpoints} \\ \frac{j-1}{J} + \frac{k}{(n_j+1)J} & \text{for equispaced cutpoints,} \end{cases} \quad (4.1)$$

where $d_j = c_j - c_{j-1}$.

4.1 Maximum Likelihood Method

Taking $\{x_{jk}, k = 1, 2, \dots, n_j; j = 1, 2, \dots, J\}$ as data from a beta distribution, the MLE's of a and b are solutions to the equations (see equations (21.1) and (21.2) in Johnson and

Kotz (1970)):

$$\begin{aligned}\psi(a) - \psi(a+b) &= \frac{1}{n} \sum_{j=1}^J \sum_{k=1}^{n_j} \ln x_{jk} \\ \psi(b) - \psi(a+b) &= \frac{1}{n} \sum_{j=1}^J \sum_{k=1}^{n_j} \ln(1 - x_{jk}),\end{aligned}\quad (4.2)$$

where $\psi(\cdot)$ is the digamma function. These equations can be solved iteratively as suggested in Johnson and Kotz (1970). We refer to the resulting estimates as MLE-C (maximum likelihood estimates for continuousized data).

4.2 Method of Moments

In this section we develop a modified method of moments due to MacKinnon and Smith (1998) that approximately eliminates the bias. First calculate the sample mean of the x_{jk} :

$$\bar{x} = \frac{1}{n} \sum_{j=1}^J \sum_{k=1}^{n_j} x_{jk} = \frac{1}{n} \sum_{j=1}^J n_j m_j \quad (4.3)$$

and the sample variance of the x_{jk} :

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{j=1}^J \sum_{k=1}^{n_j} (x_{jk} - \bar{x})^2 \\ &= \frac{1}{n-1} \left[\sum_{j=1}^J \sum_{k=1}^{n_j} (x_{jk} - m_j)^2 + \sum_{j=1}^J n_j (m_j - \bar{x})^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{j=1}^J \frac{n_j(n_j^2 - 1)}{12d_j^2(n_j + 1)^2} + \sum_{j=1}^J n_j m_j^2 - n\bar{x}^2 \right] \\ &= \frac{1}{n-1} \left[\frac{1}{12} \sum_{j=1}^J \frac{n_j(n_j - 1)}{d_j^2(n_j + 1)} + \sum_{j=1}^J n_j m_j^2 - n\bar{x}^2 \right].\end{aligned}\quad (4.4)$$

The basic method of moments uses estimating equations (3.4) with the above values of \bar{x} and s^2 . However, the resulting estimates are biased because $E(\bar{X}) \neq \mu$ and $E(S^2) \neq \sigma^2$, where \bar{X} and S^2 are the r.v.'s corresponding to the observed quantities \bar{x} and s^2 given by (4.3) and (4.4), respectively. The estimating equations for the modified method of moments are

$$E(\bar{X}) = \bar{x} \text{ and } E(S^2) = s^2. \quad (4.5)$$

We solve these equations for \hat{a} and \hat{b} , and then calculate $\hat{\mu}$ and $\hat{\eta}^2$ using (2.1). We refer to the resulting estimates as MME-C (method of moments estimates for continuousized data).

To compute the expectations in equations (4.5), we use the expressions

$$E(N_j) = np_j, \text{Var}(N_j) = np_j(1 - p_j) \text{ and } \text{Cov}(N_j, N_k) = -np_jp_k,$$

where the N_j are the r.v.'s corresponding to the observed cell counts n_j . The resulting expectations are functions of a and b through $p_j = F(c_j|a, b) - F(c_{j-1}|a, b)$. First,

$$E(\bar{X}) = \frac{1}{n} \sum_{j=1}^J E(N_j)m_j = \sum_{j=1}^J p_jm_j. \quad (4.6)$$

Next,

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left[\frac{1}{12} E \left(\sum_{j=1}^J \frac{N_j(N_j-1)}{d_j^2(N_j+1)} \right) + E \left(\sum_{j=1}^J N_jm_j^2 \right) - nE(\bar{X})^2 \right] \\ &= \frac{1}{n-1} \left[\frac{1}{12} E_1 + E_2 - nE_3 \right], \end{aligned} \quad (4.7)$$

where the three expectations are computed as follows.

To compute E_1 we use the following lemma.

Lemma: Let X be a binomial r.v. with sample size n and success probability p (denoted as $X \sim \text{Bin}(n, p)$). Then

$$E \left[\frac{X(X-1)}{X+1} \right] = np - 2 + \frac{2(1 - q^{n+1})}{p(n+1)},$$

where $q = 1 - p$.

Proof:

$$\begin{aligned} E \left[\frac{X(X-1)}{X+1} \right] &= \sum_{x=0}^n \left[\frac{x(x-1)}{x+1} \right] \binom{n}{x} p^x q^{n-x} \\ &= \frac{1}{(n+1)p} \sum_{x=0}^n x(x-1) \binom{n+1}{x+1} p^{x+1} q^{(n+1)-(x+1)} \\ &= \frac{1}{(n+1)p} \sum_{y=1}^n (y-1)(y-2) \binom{n+1}{y} p^y q^{(n+1)-y} \quad (\text{by putting } y = x+1) \\ &= \frac{1}{(n+1)p} \left[E\{(Y-1)(Y-2)\} - 2q^{n+1} \right] \quad (\text{where } Y \sim \text{Bin}(n+1, p)) \\ &= \frac{1}{(n+1)p} \left[E(Y^2 - 3Y + 2) - 2q^{n+1} \right] \\ &= \frac{1}{(n+1)p} \left[(n+1)pq + (n+1)^2p^2 - 3(n+1)p + 2 - 2q^{n+1} \right], \end{aligned}$$

which upon algebraic simplification gives the desired result. \square

Using the above lemma, we have

$$\begin{aligned} E_1 &= E \left(\sum_{j=1}^J \frac{N_j(N_j - 1)}{d_j^2(N_j + 1)} \right) \\ &= \sum_{j=1}^J \frac{1}{d_j^2} \left[np_j - 2 + \frac{2(1 - q_j^{n+1})}{p_j(n + 1)} \right], \end{aligned} \quad (4.8)$$

where $q_j = 1 - p_j$. Note that if $p_j \rightarrow 0$, the term inside the square brackets approaches 0 using L'Hôpital's rule. This approximation is used for cell probabilities close to 0 to avoid numerical errors.

Next,

$$E_2 = \sum_{j=1}^J E(N_j)m_j^2 = n \sum_{j=1}^J p_j m_j^2. \quad (4.9)$$

Finally,

$$\begin{aligned} E_3 &= E(\bar{X}^2) \\ &= \frac{1}{n^2} E \left(\sum_{j=1}^J N_j m_j \right)^2 \\ &= \frac{1}{n^2} \left\{ \sum_{j=1}^J m_j^2 E(N_j^2) + 2 \sum_{k=j+1}^J \sum_{j=1}^{J-1} m_j m_k E(N_j N_k) \right\} \\ &= \frac{1}{n^2} \left\{ \sum_{j=1}^J m_j^2 [np_j(1 - p_j) + n^2 p_j^2] + 2 \sum_{k=j+1}^J \sum_{j=1}^{J-1} m_j m_k [-np_j p_k + n^2 p_j p_k] \right\} \\ &= \frac{1}{n} \left\{ \sum_{j=1}^J m_j^2 p_j + (n - 1) \sum_{j=1}^J m_j^2 p_j^2 + 2(n - 1) \sum_{k=j+1}^J \sum_{j=1}^{J-1} m_j m_k p_j p_k \right\} \\ &= \frac{1}{n} \sum_{j=1}^J m_j^2 p_j + \frac{n - 1}{n} \left(\sum_{j=1}^J m_j p_j \right)^2. \end{aligned} \quad (4.10)$$

Substituting E_1, E_2 and E_3 from (4.8), (4.9) and (4.10) in (4.7) we get the final expression for $E(S^2)$.

5. Use of Pseudo-Observations

Both the MLE-D and the MME-C methods may fail to converge when there are sparse or empty cells. The MLE-C method does not face this problem because once the data are spread uniformly within each cell, no further use is made of the cell boundaries. A practical

solution to the lack of convergence is to add $1/2$ (referred to as a pseudo-observation) to each cell count.

Denote the new cell counts by $n'_j = n_j + 1/2$ and the resulting total sample size by $n' = \sum_{j=1}^J n'_j = n + J/2$. We assume that the pseudo-observations are placed at the midpoints m_j . The actual observations are spread uniformly in each cell as described in Section 4. Let \bar{x}' and s'^2 be the sample mean and the sample variance of the continuousized data including the pseudo-observations. We can express \bar{x}' and s'^2 in terms of \bar{x} and s^2 as follows:

$$\bar{x}' = \frac{1}{n'} \left(\sum_{j=1}^J n_j m_j + \frac{1}{2} \sum_{j=1}^J m_j \right) = \frac{n\bar{x} + J\bar{m}/2}{n'}, \quad (5.1)$$

where $\bar{m} = \sum m_j / J$ (for equispaced cutpoints, $\bar{m} = J/2$) and

$$s'^2 = \frac{1}{n' - 1} \left[(n - 1)s^2 + \frac{nJ^2}{4n'^2} (\bar{x} - \bar{m})^2 + \frac{1}{2} \sum_{j=1}^J \left\{ \left(m_j - \frac{J\bar{m}}{2n'} \right) - \frac{n\bar{x}}{n'} \right\}^2 \right]. \quad (5.2)$$

The algorithm for calculating the MLE-D estimates does not need any modification if pseudo-observations are added. To calculate the MME-C estimates, it would appear that the estimating equations should be changed to

$$E(\bar{X}') = \bar{x}' \text{ and } E(S'^2) = s'^2, \quad (5.3)$$

where \bar{X}' and S'^2 are the random variables corresponding to the observed \bar{x}' and s'^2 , respectively. It is not difficult to derive expressions for $E(\bar{X}')$ and $E(S'^2)$ using (4.3) and (4.4). However, from (5.1) we see that \bar{x}' is a linear function of \bar{x} and hence $E(\bar{X}')$ is the same linear function of $E(\bar{X})$. Therefore the equation $E(\bar{X}') = \bar{x}'$ is identical to the equation $E(\bar{X}) = \bar{x}$. This is the equation that causes convergence problems. To obviate the convergence difficulty with equations (5.3), we use the following equations:

$$E(\bar{X}) = \bar{x}' \text{ and } E(S^2) = s'^2. \quad (5.4)$$

These equations may result in a slight bias in the estimates, but their solutions are found to always converge.

6. Empirical Comparison of the Estimation Methods

6.1 Design of the Simulation Experiment

Three methods of estimation were compared in the simulation experiment: MLE-D, MLE-C and MME-C. Pseudo-observations were used with MLE-D and MME-C, but not with MLE-C. Two values of the number of cells were studied: $J = 5$ and 10 . For each J , equispaced cutpoints $c_j = j/J$, $j = 0, 1, \dots, J$ (assumed to be known) were chosen to generate the data. Thus, e.g., for $J = 5$, we have $c_0 = 0, c_1 = 0.2, c_2 = 0.4, c_3 = 0.6, c_4 = 0.8, c_5 = 1$. Two values of the sample size, $n = 30$ and 60 , were examined for each value of J .

For each J and the associated cutpoints, the μ values were chosen at the midpoints of the cells. Thus, for $J = 5$, we chose $\mu = 0.5, 0.7$ and 0.9 , and for $J = 10$, we chose $\mu = 0.55, 0.75$ and 0.95 . These are in a sense the “worst” cases when η^2 is small because then most of the data fall in a single cell; the “best” cases would be to choose μ equal to one of the cutpoints because then the data would be distributed in the two adjacent cells even when η^2 is small. The values $\mu < 0.5$ were not studied because the results are similar to those for $\mu > 0.5$ due to symmetry.

For each value of μ , three values of η^2 were chosen: $\eta^2 = 0.3, 0.1$ and 0.03 . Note that $\eta^2 < 1/3$ implies that $a + b > 2$ and hence at least one of a or b is > 1 . This means that none of the simulated beta distributions were U-shaped; all were either J-shaped or bell-shaped. This choice was made based on the types of data encountered in practice which generally do not cluster at both ends of the scale. The values of (a, b) were obtained from (μ, η^2) by solving (1.2) and (1.3), resulting in

$$a = \frac{\mu(1 - \eta^2)}{\eta^2} \text{ and } b = \frac{(1 - \mu)(1 - \eta^2)}{\eta^2}. \quad (6.1)$$

6.2 Computational Details and Simulation Results

For each configuration of (μ, η^2) or equivalently (a, b) , we performed 10,000 simulation runs. First, the cell probabilities $p_j = F(c_j|a, b) - F(c_{j-1}|a, b)$, $j = 1, 2, \dots, J$ were calculated. Each simulation run consisted of generating the multinomial cell count vector

(n_1, n_2, \dots, n_J) with the cell probability vector (p_1, p_2, \dots, p_J) and total sample size n .

All three methods of estimation were applied to each generated cell count vector, and estimates (\hat{a}, \hat{b}) and $(\hat{\mu}, \hat{\eta}^2)$ were calculated. The penalty-barrier method of nonlinear programming (Bazaraa, Sherali and Shetty, 1993) was used to find the MLE-D (by maximizing the log-likelihood function (3.1) subject to $a, b > 0$) as well as to solve the equations (4.2) for MLE-C and equations (4.5) for MME-C. The results were averaged over 10,000 runs to find the bias and variance estimates. The mean square error (MSE) was calculated as $\text{MSE} = \text{Bias}^2 + \text{Variance}$. The simulation estimates of bias, variance and MSE for the three methods are summarized in Tables 2-5.

It is interesting to note that the MME-C and MLE-D methods generally underestimate μ , but always overestimate η^2 . The biases in $\hat{\eta}^2$ become very large when μ is close to 1 and η^2 is close to 0. The MLE-C method does not exhibit such systematic patterns and its biases in $\hat{\eta}^2$ are generally much smaller.

The minimum MSE for each configuration (μ, η^2) is marked with an asterisk. Note that no method is uniformly best for all configurations either for estimating μ or η^2 . However, out of the nine (μ, η^2) configurations for each (J, n) combination, the MLE-C method has the smallest MSE's in the following numbers of cases: for $J = 5, n = 30$, the smallest $\text{MSE}(\hat{\mu})$ in four cases and the smallest $\text{MSE}(\hat{\eta}^2)$ in five cases, for $J = 5, n = 60$, the smallest $\text{MSE}(\hat{\mu})$ in four cases and the smallest $\text{MSE}(\hat{\eta}^2)$ in three cases, for $J = 10, n = 30$, the smallest $\text{MSE}(\hat{\mu})$ in six cases and the smallest $\text{MSE}(\hat{\eta}^2)$ in seven cases, and for $J = 10, n = 60$, the smallest $\text{MSE}(\hat{\mu})$ in three cases and the smallest $\text{MSE}(\hat{\eta}^2)$ in seven cases.

The methods can be compared through the ratios of their MSE's, referred to as the relative efficiencies (RE's). We define the RE's in terms of the MLE-C method. The RE of the MLE-C method with respect to (w.r.t.) another method is the ratio of the MSE of that method to the MSE of the MLE-C method. Thus $\text{RE} > 1$ (or $\log(\text{RE}) > 0$) favors the MLE-C method. The logarithms of the RE of the MLE-C method w.r.t. the MME-C and MLE-D methods for estimating μ and η^2 are plotted in Figures 1 to 4. Each figure is for a different combination (J, n) . The left panel of each figure shows the plots of $\log(\text{RE})$ for estimating μ , while the right panel shows the plots of $\log(\text{RE})$ for estimating η^2 . In both cases the $\log(\text{RE})$ values are plotted against μ and different plots are made for the three

values of η^2 . The plots of $\log(\text{RE})$ w.r.t. the MME-C method are shown with solid lines and solid plotting symbols, while the plots of $\log(\text{RE})$ w.r.t. the MLE-D method are shown with dotted lines and open plotting symbols. The plotting symbols used are diamonds for $\eta^2 = 0.3$, squares for $\eta^2 = 0.1$ and triangles for $\eta^2 = 0.03$.

We see that most of the $\log(\text{RE})$'s are positive, favoring the MLE-C method. Furthermore, the positive values of $\log(\text{RE})$ are much larger than any negative values. In other words, the potential benefits of the MLE-C method are much greater than any potential losses w.r.t. the other two competing methods. The MLE-D method is often competitive, whereas the MME-C method is not preferred in any situation.

7. Maximum Likelihood Estimation of Cutpoints

First let us consider estimation of cutpoints for a single treatment. Taking the partial derivatives of the log-likelihood (3.1) with respect to the c_j and using the fact that $p_j = F(c_j|a, b) - F(c_{j-1}|a, b)$ we get the following likelihood equations

$$\frac{\partial \ln L}{\partial c_j} = f(c_j|a, b) \left\{ \frac{n_j}{p_j} - \frac{n_{j+1}}{p_{j+1}} \right\} = 0 \quad (j = 1, 2, \dots, J-1),$$

where $f(c_j|a, b)$ is the beta density function (1.1). The solutions to these equations are given by

$$p_j = F(c_j|a, b) - F(c_{j-1}|a, b) = \frac{n_j}{n} \quad (j = 1, 2, \dots, J), \quad (7.1)$$

where $F(c_0|a, b) = 0$ and $F(c_J|a, b) = 1$. The MLE-C method (which, as we saw, is the preferred method for estimating (a, b)) equations can be supplemented with these equations, and the two sets of equations can be solved iteratively. For example, we can start with an initial guess of equispaced cutpoints $c_j = j/J$ and solve for (a, b) using (4.2). Then using these estimates of (a, b) , the equations (7.1) can be solved for $(c_1, c_2, \dots, c_{J-1})$. Two remarks are in order concerning the implementation of this recursive algorithm.

1. After having computed the estimates of the c_j , we need to recompute the continuousized data values x_{jk} given by (4.1) in order to iterate with respect to the estimates of (a, b) using (4.2).
2. If some $n_j = 0$ then we get $c_{j-1} = c_j$ from (7.1). One way to address this difficulty is to add a pseudo-observation $= 1/2$ to all cells thus avoiding empty cells. The other

alternative is to merge this cell with an adjacent nonempty cell. We prefer the second alternative, especially for multiple treatments, as explained below. Adding pseudo-observations is less preferable because the MLE-C method for estimating (a, b) does not require pseudo-observations to achieve convergence.

Now consider $I \geq 2$ treatment combinations, each associated with its own beta latent distribution with parameters (a_i, b_i) , $i = 1, 2, \dots, I$. The cutpoints are common to all treatment combinations. Let n_{ij} be the cell count in the j th cell of the i th treatment combination and let $n_i = \sum_{j=1}^J n_{ij}$ be the total number of observations on the i th treatment combination. The log-likelihood function is

$$\ln L = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln p_{ij}, \quad (7.2)$$

where $p_{ij} = F(c_j|a_i, b_i) - F(c_{j-1}|a_i, b_i)$. By taking the partial derivatives of $\ln L$ with respect to (w.r.t.) the c_j ($1 \leq j \leq J - 1$), we get the following equations for the MLE's of the c_j for given (a_i, b_i) , $i = 1, 2, \dots, I$:

$$\sum_{i=1}^I f(c_j|a_i, b_i) \left\{ \frac{n_{ij}}{p_{ij}} - \frac{n_{i,j+1}}{p_{i,j+1}} \right\} = 0 \quad (j = 1, \dots, J_1). \quad (7.3)$$

These equations do not have a simple solution as in (7.1). An alternative method of estimating the cutpoints is by the moments method by equating the expected number of outcomes in each cell to the observed number of outcomes. The corresponding equations are

$$\sum_{i=1}^I n_i p_{ij} = \sum_{i=1}^I n_i [F(c_j|a_i, b_i) - F(c_{j-1}|a_i, b_i)] = \sum_{i=1}^I n_{ij} \quad (j = 1, \dots, J - 1). \quad (7.4)$$

Note that both (7.3) and (7.4) are different generalizations of (7.1).

Both these methods yield $c_{j-1} = c_j$ if the total count in the j th category, $\sum_{i=1}^I n_{ij}$, is zero, i.e., no outcomes are observed in this category for any of the treatments. In this case it seems reasonable to merge this category with an adjacent one since the other option of adding pseudo-observations to all IJ cells is unattractive.

We found that the method of directly maximizing the log-likelihood function (7.2) w.r.t. the c_j 's for fixed (a_i, b_i) $i = 1, \dots, I$ gives the best numerical performance, and so was adopted. The same penalty barrier method described in Section 6.2 was used for direct

maximization. The constraints $0 < c_1 < c_2 < \dots < c_{J-1} < 1$ were imposed after merging any categories having zero outcomes with adjacent categories.

8. Example

We fitted the proposed beta distribution model to two data sets given in Best and Rayner (1998) (the first data set is from Agresti (1990)) using the MLE-C method. The proportional odds model was also fitted to the same data sets for comparison purposes. This model is given by

$$\ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \frac{c_j - \mu_i}{\sigma_i}, \quad i = 1, 2, \dots, I, j = 1, 2, \dots, J,$$

where $\pi_{ij} = p_{i1} + \dots + p_{ij}$ is the cumulative probability of response $\leq j$ for the i th treatment. Note that both the models have the same number of free parameters, namely $2I$ location and dispersion parameters for the I treatments, and $J - 1$ cutpoints. McCullagh (1986) suggested that the effects of the factors on the μ_i and the $\ln \sigma_i$ can be modeled as linear models.

McCullagh's PLUM program (1988) was used to fit the proportional odds model. The data sets are given in Tables 6 and 7. The first data set gives the counts of people in defense force staff from different regions of U.S. whose likings for black olives were scored on a six-point ordinal scale. The second data set gives the counts of Australian and Japanese consumers from two cities in each country whose likings for Japanese chocolate were scored on a seven-point ordinal scale. Both data sets have treatments with factorial structures; however, our objective here is not to assess the effects of the factors on the ordinal responses, but simply to evaluate the goodness of fits of the two models.

Each table gives the observed cell counts as well as the fitted cell counts using the two models. The last column gives the contribution to the overall chi-square goodness of fit statistic:

$$\sum_{i=1}^I \sum_{j=1}^J \frac{(\text{Observed}_{ij} - \text{Fitted}_{ij})^2}{\text{Fitted}_{ij}}$$

from each row i of the table. This chi-square statistic has $I(J - 1) - 2I - (J - 1) = (I - 1)(J - 1) - 2I$ degrees of freedom (d.f.) for both fitted models.

The overall chi-square statistics along with their d.f. and p -values are shown in the following table.

	Beta Distribution Model	Proportional Odds Model
Data Set 1	$\chi^2 = 14.381$, d.f. = 13, $p = 0.348$	$\chi^2 = 11.554$, d.f. = 13, $p = 0.564$
Data Set 2	$\chi^2 = 12.413$, d.f. = 10, $p = 0.258$	$\chi^2 = 11.755$, d.f. = 10, $p = 0.302$

We see that both models fit reasonably well. The proportional odds model gives somewhat better overall fits, as well as for individual rows of the two tables in all cases except one. Thus it is a distinct competitor to the proposed beta distribution model. However, we feel that the beta distribution model has other advantages to offer as enumerated in Section 1.

9. Conclusions

The simulation results have shown that although there is not a uniformly best method of estimation, in a majority of cases the MLE-C method is superior to the other two. In other cases the MLE-D method is generally the best, but the MLE-C method is not far behind. The MME-C and the MME-D methods are not serious contenders for estimating either μ or η^2 .

Besides its statistical efficiency, the MLE-C method has the following two advantages.

1. It is computationally much faster than the other methods.
2. It does not face any convergence problems even in case of sparse cells. Therefore it does not need data augmentation by pseudo-observations.

Because of these advantages we recommend the MLE-C method for estimation of the location and dispersion parameters of the beta distribution model. This method is then combined with the maximum likelihood method for estimating the cutpoints in an iterative two-step algorithm.

The beta distribution model and the proportional odds model were both fitted to real data sets. Both models fitted reasonably well, but the proportional odds model gave marginally better fits. The beta distribution model has some other advantages, however. In future research both models should be compared in terms of their ability to detect location and dispersion effects in multifactor experiments.

REFERENCES

1. Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, New York: John Wiley.
2. Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley.
3. Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. (1993), *Nonlinear Programming: Theory and Algorithms*, Second Edition, New York: Wiley.
4. Best, D. J. and Rayner, J. C. W. (1998), "Nonparametric analysis of ordinal categorical response data with factorial structure," *Applied Statistics*, **47**, 439 -446.
5. Box, G. E. P. (1988), "Signal-to-noise ratios, performance criteria and transformation," *Technometrics*, **30**, 1 -31.
6. Dempster, A. P., Laird N. M. and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *Journal of Royal Statistical Society*, Ser. B, , 1 -38
7. Hamada, M. and Wu, C. F. J. (1990), "A critical look at acculation analysis and related methods (with discussion)," *Technometrics*, **28**, 302 -306.
8. Leon, R. V., Shoemaker, A. C. and Kacker, R. N. (1987), "Performance measures independent of adjustment: An explanation and extension of Taguchi's signal to noise ratio," *Technometrics*, **29**, 253 -287.
9. McCullagh, P. (1980), "Regression models for ordinal data (with discussion)," *Journal of Royal Statistical Society*, Ser. B, **42**, 109 -142.
10. McCullagh, P. (1988), "PLUM: An interactive computer package for analysis of ordinal data," User's Manual available from pmcc@galton.uchicago.edu.
11. McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models*, Second Edition, London: Chapman and Hall.
12. MacKinnon, J. G. and Smith, A. A., Jr. (1988), "An approximate bias correction in econometrics," *Journal of Econometrics*, **85**, 205 -230.

13. Nair, V. N. (1986), "Testing in industrial experiments with ordered categorical data (with discussion)," *Technometrics*, **28**, 283 -311.
14. Sundberg, R. (1974), "Maximum likelihood theory for incomplete data from an exponential family," *Scandinavian Journal of Statistics*, **1**, 49 -58.
15. Sundberg, R. (1976), "An iterative method for solution of the likelihood equations for incomplete data from an exponential family," *Communications in Statistics*, Ser. B (Simulations and Computations), **5**, 55 -64.
16. Taguchi, G. (1986), *Introduction to Quality Engineering*, White Plains, NY: Kraus International Publications.

Table 1: Values of (a, b) Corresponding to Selected Values of (μ, η^2)

J	μ	η^2	a	b
5	0.5	0.3	1.1667	1.1667
		0.1	4.5000	4.5000
		0.03	16.167	16.167
	0.7	0.3	1.6333	1.6333
		0.1	6.3000	2.7000
		0.03	22.633	9.7000
	0.9	0.3	2.1000	0.2333
		0.1	8.1000	0.9000
		0.03	29.100	3.2333
10	0.55	0.3	1.2833	1.0500
		0.1	4.9500	4.0500
		0.03	17.783	14.550
	0.75	0.3	1.7500	0.5833
		0.1	6.7500	2.2500
		0.03	24.250	8.0833
	0.95	0.3	2.2167	0.1167
		0.1	8.5500	0.4500
		0.03	30.717	1.6167

Table 2: Simulation Estimates of Bias, Variance and Mean Square Error of $\hat{\mu}$ and $\hat{\eta}^2$ for MME-C, MLE-D and MLE-C Methods of Estimation ($J = 5, n = 30$)

μ	η^2	MME-C			MLE-D			MLE-C		
		Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)
0.5	0.3	-0.020	23.12	23.12	-0.018	22.63	22.63	-0.019	21.72	21.72*
	0.1	-0.007	7.94	7.94	-0.006	7.81	7.81*	-0.004	9.77	9.77
	0.03	-0.005	2.89	2.89	-0.005	2.83	2.83*	-0.006	3.48	3.48
0.7	0.3	-1.401	20.44	22.40	-1.472	19.90	22.07*	-2.239	18.67	23.68
	0.1	-1.390	7.53	9.46	-1.571	7.26	9.73	+0.357	8.10	8.22*
	0.03	-1.775	2.33	5.48	-2.041	2.34	6.50	-0.019	3.37	3.37*
0.9	0.3	-2.762	11.41	19.04*	-3.070	11.73	21.15	-6.324	9.03	49.03
	0.1	+0.066	6.88	6.88*	-0.626	7.62	8.01	-3.055	2.37	11.70
	0.03	+4.049	2.20	18.59	+3.671	3.07	16.54	-0.694	0.80	1.28*
μ	η^2	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)
0.5	0.3	+1.214	40.07	41.54	+0.037	35.76	35.76	-1.324	13.63	15.38*
	0.1	+1.618	5.98	8.60	+1.320	5.48	7.22*	+2.251	7.45	12.52
	0.03	+1.949	1.36	5.16*	+2.259	0.94	6.04	+2.182	1.61	6.38
0.7	0.3	+2.204	48.04	52.93	+0.954	43.11	44.02	-4.957	16.59	41.16*
	0.1	+3.697	8.63	22.30	+3.235	7.69	18.16*	+3.616	5.27	18.35
	0.03	+3.189	1.57	11.73	+3.392	1.16	12.66	+2.717	2.90	10.28*
0.9	0.3	+9.875	71.35	168.9	+7.13	86.43	137.2*	+17.13	18.40	311.7
	0.1	+25.56	56.60	709.7	+22.69	54.96	569.7	-2.242	2.68	7.71*
	0.03	+47.03	79.02	2291	+42.56	53.20	1864	+2.250	1.08	6.14*

Multiply the bias by 10^{-2} and the variance and MSE by 10^{-4} .

Asterisk (*) represents the minimum MSE for given (μ, η^2) configuration (row).

Table 3: Simulation Estimates of Bias, Variance and Mean Square Error of $\hat{\mu}$ and $\hat{\eta}^2$ for MME-C, MLE-D and MLE-C Methods of Estimation ($J = 5, n = 60$)

μ	η^2	MME-C			MLE-D			MLE-C		
		Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)
0.5	0.3	-0.036	12.83	12.83	-0.037	12.56	12.56	-0.034	11.05	11.05*
	0.1	-0.010	4.42	4.42	-0.009	4.38	4.38*	-0.008	5.10	5.10
	0.03	-0.015	1.61	1.61	-0.015	1.59	1.59*	-0.016	1.77	1.77
0.7	0.3	-0.756	11.45	12.02	-0.789	11.22	11.84*	-2.311	9.80	15.14
	0.1	-0.741	4.00	4.55	-0.833	3.94	4.63	+0.426	4.13	4.31*
	0.03	-0.968	1.32	2.25	-1.146	1.32	2.63	+0.069	1.81	1.81*
0.9	0.3	-1.448	6.43	8.52*	-1.539	6.24	8.61	-6.381	4.68	45.40
	0.1	+1.168	4.60	5.96	+0.591	4.78	5.13*	-3.053	1.23	10.55
	0.03	+5.792	1.50	35.05	+5.438	2.03	31.60	-0.561	0.46	0.78*
μ	η^2	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)
0.5	0.3	+0.632	20.59	20.99	+0.036	19.36	19.36	-0.204	6.77	6.81*
	0.1	+0.818	3.19	3.86	+0.683	3.06	3.52*	+2.689	4.24	11.47
	0.03	+0.939	0.62	1.50*	+1.210	0.48	1.95	+2.328	0.81	6.23
0.7	0.3	+1.149	25.12	26.44	+0.539	23.68	23.97*	-3.960	8.63	24.31
	0.1	+1.917	4.37	8.05	+1.744	4.07	7.11*	+4.324	2.68	21.38
	0.03	+1.533	0.65	3.00*	+1.838	0.51	3.89	-1.787	1.53	4.73
0.9	0.3	+5.518	49.76	80.21	+4.801	49.25	72.30*	-16.46	9.91	281.0
	0.1	+17.02	25.70	315.3	+15.07	26.96	253.9	-1.787	1.53	4.73*
	0.03	+38.79	64.25	1569	+36.59	74.90	1413	+2.345	0.73	6.23*

Multiply the bias by 10^{-2} and the variance and MSE by 10^{-4} .

Asterisk (*) represents the minimum MSE for given (μ, η^2) configuration (row).

Table 4: Simulation Estimates of Bias, Variance and Mean Square Error of $\hat{\mu}$ and $\hat{\eta}^2$ for MME-C, MLE-D and MLE-C Methods of Estimation ($J = 10, n = 30$)

μ	η^2	MME-C			MLE-D			MLE-C		
		Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)
0.55	0.3	-0.708	18.44	18.94	-0.709	17.73	18.23*	-0.203	22.30	22.34
	0.1	-0.722	6.16	6.68	-0.877	5.70	6.47*	+0.013	8.44	8.44
	0.03	-0.715	2.01	2.52*	-0.914	1.83	2.66	-0.007	2.74	2.74
0.75	0.3	-3.504	14.56	26.83	-3.615	13.49	26.56	-1.512	17.72	20.00*
	0.1	-3.461	5.01	16.99	-4.348	4.42	23.32	+0.112	6.47	6.48*
	0.03	-3.579	1.59	14.40	-4.698	1.38	23.45	+0.001	2.18	2.18*
0.95	0.3	-6.412	4.21	45.32	-6.396	4.02	44.93	-4.227	5.39	23.26*
	0.1	-5.263	1.60	29.30	-6.145	2.50	40.26	-2.472	1.50	7.61*
	0.03	-3.843	0.55	15.31	-4.534	1.70	22.26	-0.949	0.46	1.36*
μ	η^2	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)
0.55	0.3	+1.128	24.15	25.42	+0.114	20.96	20.97	-1.622	17.52	20.16*
	0.1	+3.561	4.17	16.86	+3.488	3.57	15.73	+0.293	5.26	5.35*
	0.03	+4.539	0.53	21.13	+5.210	0.42	27.56	+0.597	0.71	1.07*
0.75	0.3	+4.038	27.26	43.56	+2.927	23.40	31.96*	-5.592	20.24	51.51
	0.1	+8.085	5.39	70.76	+7.115	4.90	55.52	+0.937	5.55	6.42*
	0.03	+9.168	0.94	84.98	+8.140	0.59	66.85	+0.823	0.85	1.53*
0.95	0.3	+20.12	28.29	433.3	+18.40	26.77	365.5*	-20.93	16.42	454.5
	0.1	+38.14	25.16	1479	+33.66	31.29	1164	-4.687	3.30	25.27*
	0.03	+50.93	18.99	2613	+45.48	32.66	2101	+0.104	0.79	0.80*

Multiply the bias by 10^{-2} and the variance and MSE by 10^{-4} .

Asterisk (*) represents the minimum MSE for given (μ, η^2) configuration (row).

Table 5: Simulation Estimates of Bias, Variance and Mean Square Error of $\hat{\mu}$ and $\hat{\eta}^2$ for MME-C, MLE-D and MLE-C Methods of Estimation ($J = 10, n = 60$)

μ	η^2	MME-C			MLE-D			MLE-C		
		Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)	Bias($\hat{\mu}$)	Var($\hat{\mu}$)	MSE($\hat{\mu}$)
0.55	0.3	-0.399	11.03	11.19	-0.402	10.55	10.71*	-0.217	11.45	11.49
	0.1	-0.405	3.70	3.87	-0.495	3.54	3.79*	+0.110	4.37	4.37
	0.03	-0.388	1.20	1.35*	-0.501	1.15	1.40	-0.005	1.41	1.41
0.75	0.3	-1.901	8.71	12.33	-1.966	8.17	12.04	-1.547	9.13	11.52*
	0.1	-1.884	2.93	6.48	-2.404	2.71	8.49	+0.130	3.33	3.34*
	0.03	-1.943	0.93	4.71	-2.576	0.88	7.51	+0.025	1.13	1.13*
0.95	0.3	-3.396	2.55	14.08*	-3.507	2.33	14.63	-4.263	2.59	20.76
	0.1	-2.189	1.06	5.85*	-3.138	1.36	11.21	-2.502	0.81	7.07
	0.03	-0.569	0.36	0.69*	-1.462	0.89	3.02	-0.873	0.25	1.01
μ	η^2	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)	Bias($\hat{\eta}^2$)	Var($\hat{\eta}^2$)	MSE($\hat{\eta}^2$)
0.55	0.3	+0.608	13.68	14.04	+0.065	12.27	12.28	-0.789	9.03	9.66*
	0.1	+1.893	2.42	6.01	+1.954	2.16	5.98	+0.526	2.84	3.12*
	0.03	+2.414	0.31	6.31	+2.886	0.26	8.59	+0.635	0.36	0.76*
0.75	0.3	+2.303	16.84	22.14	+1.742	15.00	18.04*	-4.821	10.29	33.53
	0.1	+4.549	2.82	23.52	+4.097	2.69	19.47	+1.379	2.96	4.87*
	0.03	+5.160	0.38	27.00	+4.723	0.30	22.61	+0.880	0.44	1.22*
0.95	0.3	+15.08	25.20	252.7	+14.11	25.23	224.4*	-20.54	8.16	430.1
	0.1	+30.84	19.54	970.6	+26.28	23.35	714.0	-4.355	1.97	20.93*
	0.03	+45.29	20.51	2072	+38.79	33.65	1538	+0.163	0.46	0.49*

Multiply the bias by 10^{-2} and the variance and MSE by 10^{-4} .

Asterisk (*) represents the minimum MSE for given (μ, η^2) configuration (row).

Table 6: Preference for Black Olives (Olive Data)

Row	Type of Count	Response						Chi-Square
		–	–	0	+	++	+++	
1	Observed	20	15	12	17	16	28	
	Fitted (Beta)	18.224	14.049	18.848	17.695	12.782	26.402	3.660
	Fitted (Prop. Odds)	19.503	13.036	17.185	16.399	12.395	29.482	3.018
2	Observed	18	17	18	18	6	25	
	Fitted (Beta)	18.531	14.163	18.647	16.994	11.838	21.827	4.006
	Fitted (Prop. Odds)	19.931	14.028	17.996	16.154	11.334	22.557	3.802
3	Observed	12	9	23	21	19	30	
	Fitted (Beta)	11.511	13.399	20.990	21.537	16.019	30.545	2.235
	Fitted (Prop. Odds)	10.750	12.088	20.431	22.598	17.418	30.715	1.531
4	Observed	30	22	21	17	8	12	
	Fitted (Beta)	29.314	19.199	22.282	17.259	10.038	11.907	0.917
	Fitted (Prop. Odds)	30.258	20.986	22.486	15.749	8.740	11.781	0.316
5	Observed	23	18	20	18	10	15	
	Fitted (Beta)	22.847	16.701	20.749	17.371	10.944	15.389	0.243
	Fitted (Prop. Odds)	22.928	17.652	21.275	16.797	10.189	15.158	0.175
6	Observed	11	9	26	19	17	24	
	Fitted (Beta)	11.296	13.300	20.621	20.638	14.822	25.324	3.320
	Fitted (Prop. Odds)	10.224	12.390	21.028	22.291	15.985	24.082	2.713

Row 1: Urban Midwest, Row 2: Urban Northeast, Row 3: Urban Southwest

Row 4: Rural Midwest, Row 5: Rural Northeast, Row 6: Rural Southwest

Table 7: Liking for Japanese Chocolate (Chocolate Data)

Row	Type of Count	Response							Chi-Square
		1	2	3	4	5	6	7	
1	Observed	2	1	6	1	8	9	6	
	Fitted (Beta)	1.625	2.755	3.293	2.202	10.042	7.061	6.023	5.035
	Fitted (Prop. Odds)	1.529	2.693	3.221	2.083	9.400	7.565	6.511	4.692
2	Observed	1	6	2	2	10	5	5	
	Fitted (Beta)	2.044	3.136	3.536	2.276	9.658	6.064	4.287	4.168
	Fitted (Prop. Odds)	2.036	3.404	3.777	2.287	9.026	6.009	4.462	3.719
3	Observed	0	1	3	4	15	7	1	
	Fitted (Beta)	0.136	1.223	3.159	2.923	15.503	6.785	1.271	0.662
	Fitted (Prop. Odds)	0.148	1.091	3.016	2.978	16.327	6.355	1.084	0.686
4	Observed	1	1	2	3	16	6	2	
	Fitted (Beta)	0.487	2.014	3.507	2.729	13.053	6.870	2.340	2.549
	Fitted (Prop. Odds)	0.300	1.502	3.228	2.781	14.292	7.046	1.851	2.657

Row 1: Sydney, Australia, Row 2: Melbourne, Australia, Row 3: Tokyo, Japan, Row 4: Osaka, Japan