Statistika, Vol. 8 No. 1, 61 – 68
Mei 2008

The Beta-Binomial Multivariate Model for Correlated Categorical Data

Nusar Hajarisman[1)], and **Asep Saefuddin**[2)]

[1)] Department of Statistics, Faculty of Mathematics and Natural Sciences, Bandung Islamic University,
Jalan Tamansari 1, Bandung 40116, West Java, Indonesia
e-mail: nrisman@yahoo.co.uk
[2)] Departement of Statistics, Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University, West Java, Indonesia
e-mail: asaefuddin@indo.net.id

Abstract

Over the past year, a significant amount of research has explored the logistic regression models for analyzing correlated categorical data. In these models, it is assumed that the data occur in clusters, where individuals within each cluster are correlated, but individuals from different clusters are assumed independent. A commonly used in modeling correlated categorical univariate data is to assume that individual counts are generated from a Binomial distribution, with probabilities vary between individuals according to a Beta distribution. The marginal distribution of the counts is then Beta-Binomial. In this paper, a generalization of the model is made allowing the number of respondent m, to be random. Thus both the number units m, and the underlying probability vector are allowed to vary. We proposed the model for correlated categorical data, which is generalized to account for extra variation by allowing the vectors of proportions to vary according to a Dirichlet distribution. The model is a mixture distribution of multinomial and Dirichlet distribution, and we call the model as the beta-binomial multivariate model.

Keywords: Beta-Binomial distribution; Correlated outcomes; Dirichlet distribution; Dirichlet-Multinomial model; Multinomial distribution; Overdispersion.

1. Introduction

The independence assumption of standard logistic regression (LR) for binary data may not hold, and this non-independence will lead to a variance that is greater than binomial variability. The phenomenon would be indicative of overdispersion in binary data. The consequence of this overdispersion is that the confidence interval of the estimate is narrowing, so statistical hypothesis will always be rejected. The beta-binomial (BB) regression model is an alternative model to handle overdispersion in binary data analysis. The BB model is an extension of the traditional LR model by including the correlation parameter between binary responses. However, beta-binomial model is appropriate for overdispersion modeling from cluster of univariate binary responses. Fortunately, in multivariate case, where binary responses from each subunit within clusters can be classified into 1, 2, …, C categories, the beta-binomial model is not appropriate anymore. A significant amount of research has explored the use of regression models for analyzing correlated categorical data. In these models, it is assumed that the data occur in clusters, where elements within each cluster (or subunit) are correlated, but elements from different clusters are independent.

Unlike for continuous multivariate data, where the multivariate normal plays a central role, no convenient multivariate distribution for correlated discrete data is readily available. Hence researchers have been pursuing ways to define regression models that build joint probability from either marginal and/or conditional probabilities (Gange, 1995). Much of this work has been developed specifically for clustered multivariate binary responses – see Pendergast, *et al.* (1993). Tallis (1962) proposed the use of the generalized-multinomial model for dependent multinomial. This model is generalized to account for extra variation by allowing the vectors of proportions to vary according to Dirichlet distribution. Wilson and Koehler (1988) used the generalized-Dirichlet multinomial model to account for extra variation. The model allows for a second order of pairwise correlation among subunit, a type of assumption found reasonable in some biological data, Kuper and Haseman (1978).

In this paper a generalization of the model is made allowing the number of respondents *m*, to be random. Thus both the number of units *m*, and the underlying probability vector are allowed to vary. The model presented here are an extension of beta-binomial model, where the binomial distribution from that model is replaced by multinomial distribution, and the beta distribution is replaced by Dirichlet distribution. This model is known as the Dirichlet-multinomial model, and more specific, we call this model as the beta-binomial multivariate model. Section 2 outlines the beta-binomial model. Section 3 discusses the properties of the beta-binomial multivariate model. An application of the beta-binomial multivariate model will be discussed in Section 4.

2. The Beta-Binomial Model

The beta-binomial model, proposed originally by William (1975) and later applied by Crowder (1978) in a similar situation, assume that (1) individual level of interest reflects the outcome of a series of independent responses and can be characterized by the binomial distribution with parameter p, which we denote by $P(X = x \mid n, p)$; and (2) the values of p are distributed across the population according to a beta distribution, denoted by $g(p)$.

The mixture distribution of the behavior of interest, denoted $P(X = x \mid n)$, is obtained by weighting each $P(X = x \mid n, p)$ by the likelihood of that value of p occurring, $g(p)$. This is formally denoted by

$$\begin{aligned} P(X = x \mid n) &= \int_0^1 P(X = x \mid n, p) g(p) dp \\ &= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \end{aligned} \tag{1}$$

where B(·) is a beta function. The mean and variance of the beta binomial distribution are given by $E(X) = n\alpha/(\alpha + \beta)$ and $\text{var}(X) = n\alpha\beta(\alpha + \beta + n)/(\alpha + \beta)^2(\alpha + \beta + 1)$, respectively.

The widespread use of the beta-binomial model is ascribed to the ubiquitous nature of heterogeneous count data, its ability to account for binomial overdispersion, and its predictive ability. Although all attributions are true, the beta-binomial model has gained widespread use mainly because integration of the binomial frequency function with respect to the beta density yields a closed-form marginal distribution.

3. Extension of the Beta-Binomial Model

The beta-binomial multivariate model is an extension of the beta-binomial model. This model results when we assume that (1) the individual-level behavior of interest reflects of outcomes of a series of independents responses, and can be characterized by the multinomial distribution with parameter vector $\mathbf{p}$, which we denote by $P(\mathbf{X} = \mathbf{x} \mid n, \mathbf{p})$, and (2) the values of $\mathbf{p}$ are distributed across the population according to a Dirichlet distribution, denoted by $g(\mathbf{p})$.

The mixture distribution of the behavior of interest, denoted by $P(\mathbf{X} = \mathbf{x} \mid n)$, is obtained by weighting each $P(\mathbf{X} = \mathbf{x} \mid n, \mathbf{p})$, by the likelihood of that value of the vector $\mathbf{p}$ occurring (i.e., $g(\mathbf{p})$). This is denoted by

$$P(\mathbf{X} = \mathbf{x} \mid n) = \int P(\mathbf{X} = \mathbf{x} \mid n, \mathbf{p}) g(\mathbf{p}) d\mathbf{p} \tag{2}$$

More formally, we should note that since the elements of any $\mathbf{p}$, of length k, sum to 1, the integration is actually performed with respect to the $k - 1$ variables $p_1, p_2, \ldots, p_{k-1}$, where the integration limits are [0, 1], [0, $1 - p_1$], …, [0, $1 - \sum_{j=1}^{k-2} p_j$], respectively.

In order to derive the mixture distribution associated with a multinomial response process at the individual level and Dirichlet heterogeneity, we must solve the following integral:

$$\begin{aligned} P(\mathbf{X}=\mathbf{x}) &= \int_0^1\int_0^{1-p_1}\cdots\int_0^{1-\sum_{j=1}^{k-2}p_j}\binom{n}{x_1,\ldots,x_k}\left(\prod_{j=1}^{k-1}p_j^{x_j}\right)\left(1-\sum_{j=1}^{k-1}p_j\right)^{n-\sum_{j=1}^{k-1}x_j} \\ &\times\frac{\Gamma(S)}{\prod_{j=1}^{k}\Gamma(a_j)}\left(\prod_{j=1}^{k-1}p_j^{a_j-1}\right)\left(1-\sum_{j=1}^{k-1}p_j\right)^{a_k-1}dp_{k-1}\ldots dp_2dp_1 \end{aligned} \tag{3}$$

This is done in the following manner:

(1) Combine terms and move all non-p_j terms to the left of the integral sign. This gives us

$$\begin{aligned} P(\mathbf{X}=\mathbf{x}) &= \binom{n}{x_1,\ldots,x_k}\frac{\Gamma(S)}{\prod_{j=1}^{k}\Gamma(a_j)}\times \\ &\int_0^1\int_0^{1-p_1}\cdots\int_0^{1-\sum_{j=1}^{k-2}p_j}\left(\prod_{j=1}^{k-1}p_j^{a_j+x_j-1}\right)\left(1-\sum_{j=1}^{k-1}p_j\right)^{a_k+n-\sum_{j=1}^{k-1}x_j-1}dp_{k-1}\ldots dp_2dp_1 \end{aligned} \tag{4}$$

(2) We therefore have to solve the definite integral

$$\int_0^1\int_0^{1-p_1}\cdots\int_0^{1-\sum_{j=1}^{k-2}p_j}\left(\prod_{j=1}^{k-1}p_j^{a_j+x_j-1}\right)\left(1-\sum_{j=1}^{k-1}p_j\right)^{a_k+n-\sum_{j=1}^{k-1}x_j-1}dp_{k-1}\ldots dp_2dp_1$$

The trick is to transform the terms to the right of the integral sign into a known probability density function.

(3) Looking closely at this, we see that its structure mirrors the density of the Dirichlet distribution with parameter $a_j + z_j$ (j = 1, …, k); all that is missing is a $\Gamma(S+n)/\prod_{j=1}^{k}\Gamma(a_j+x_j)$ term. We can therefore write our expression for $P(\mathbf{X}=\mathbf{x})$ as

$$\begin{aligned} P(\mathbf{X}=\mathbf{x}) &= \binom{n}{x_1,\ldots,x_k}\frac{\Gamma(S)}{\prod_{j=1}^{k}\Gamma(a_j)}\frac{\prod_{j=1}^{k}\Gamma(a_j+x_j)}{\Gamma(S+n)}\times \\ &\int_0^1\int_0^{1-p_1}\cdots\int_0^{1-\sum_{j=1}^{k-2}p_j}\frac{\Gamma(S+n)}{\prod_{j=1}^{k}\Gamma(a_j+x_j)}\left(\prod_{j=1}^{k-1}p_j^{a_j+x_j-1}\right)\times \\ &\left(1-\sum_{j=1}^{k-1}p_j\right)^{a_k+n-\sum_{j=1}^{k-1}x_j-1}dp_{k-1}\ldots dp_2dp_1 \end{aligned} \tag{5}$$

(4) As the integrand is a Dirichlet pdf, the definite integral, by definition, equals 1, and we therefore write the equation as

$$P(\mathbf{X}=\mathbf{x}) = \binom{n}{x_1,\ldots,x_k}\frac{\Gamma(S)}{\prod_{j=1}^{k}\Gamma(a_j)}\frac{\prod_{j=1}^{k}\Gamma(a_j+x_j)}{\Gamma(S+n)} \tag{6}$$

This is called the Dirichlet-multinomial model or we call this model as the beta-binomial multivariate (BBM) model.

3.1. Mean and Variance of the Beta-Binomial Multivariate

The mean of the BBM can easily be derived through conditional distribution. To do so, we evaluate

$$E(\mathbf{X}) = E_{\mathbf{p}}\left[E(\mathbf{X}\mid\mathbf{p})\right]$$

where $E_{\mathbf{p}}[\cdot]$ denotes expectation with respect to the distribution of the vector **p**. Conditional on **p**, **X** is distributed multinomial, and the mean of the multinomial distribution is n**p**; therefore E(**X**) = E(n**p**). Since n is scalar constant, this is equivalent to E(**X**) = nE(**p**). As the latent vector **p** has a Dirichlet distribution, and we know that the mean of the Dirichlet distribution is E(**X**) = **a**/S, with E(X_j) = a_j/S. It follows that the mean of the beta-binomial multivariate is

$$E(\mathbf{X}) = \frac{n}{S}\mathbf{a},\ \text{with } E(X_j) = \frac{na_j}{S} \tag{7}$$

The derivation of the variance-covariance of the beta-binomial multivariate is more complex and we therefore present the result without derivation:

$$\operatorname{var}\left(X_j\right)=\frac{na_j\left(S-a_j\right)\left(S+n\right)}{S^2(S+1)}$$

$$\operatorname{cov}\left(X_j,X_{j`}\right)=\frac{-na_ja_{j`}(S+n)}{S^2(S+1)} \tag{8}$$

This can be re-written as:

$$\operatorname{cov}\left(X_j,X_{j`}\right)=n\frac{a_j}{S}\left(\delta_{j=j`}-\frac{a_{j`}}{S}\right)\left(\frac{S+n}{S+1}\right) \tag{9}$$

where $\delta_{j=j`}$ is the Kronecker delta, defined as

$$\delta_{j=j`}=\begin{cases}1 & \text{if } j=j` \\ 0 & \text{otherwise}\end{cases}$$

Let $\bar{\mathbf{p}}$ be the mean vector of the Dirichlet distribution with j-th element $\bar{p}=a_j/S$. We therefore have

$$\operatorname{cov}\left(X_j,X_{j`}\right)=n\bar{p}_j\left(\delta_{j=j`}-\bar{p}_{j`}\right)\left(\frac{S+n}{S+1}\right),$$

and can therefore express the variance-covariance of the beta-binomial multivariate in matrix form as

$$\operatorname{var}(\mathbf{X})=\left(\frac{S+n}{S+1}\right)n\left[\operatorname{Diag}\left(\bar{\mathbf{p}}\right)-\bar{\mathbf{p}}\bar{\mathbf{p}}`\right] \tag{10}$$

3.2. Estimating Model Parameters

In order to apply the BBM model, we must first develop estimates of its parameter vector **a**, from the given sample data. There are two methods that can be used to estimate the parameter of the model: maximum likelihood and method of moments. Let $\mathbf{x}_i$ be the vector of response ith (i = 1, 2, …, N) across the k groups, and n_i the number of category responses ($n_i=\sum_{j=1}^{k}x_{ij}$); x_{ij} denotes the number of times outcome j occurs in n_i independent trials.

For maximum likelihood method, by definition likelihood function is the joint density of the observed data. Assuming the observations are independent, this is the product of the BBM probabilities for each $\mathbf{x}_i$. The log-likelihood function is therefore

$$LL(\mathbf{a}\,|\,\text{data})=\sum_{i=1}^{N}\ln\left[P\left(\mathbf{X}=\mathbf{x}_i\,|\,n_i,\mathbf{a}\right)\right] \tag{11}$$

Using the standard numerical optimization software, we find the value of the parameter vector **a** that maximizes this log-likelihood function; this is the maximum likelihood estimate of **a**.

For the case of $n_i = n \;\forall\; i$, another approach to estimating the parameters of the BBM model is the method of moments (Johnson and Kotz, 1969). Let us denote the sample mean vector by $\bar{\mathbf{x}}$, the jth element of which is

$$\bar{x}_j=\frac{1}{N}\sum_{i=1}^{N}x_{ij}\,,\quad \text{for } j=1,\,\ldots,\,k$$

Equating the sample mean vector with its population counterpart, we have

$$\bar{x}_j=\frac{na_j}{S} \tag{12}$$

Given an estimate of S, it follows that

$$\hat{a}_j = \frac{\hat{S}\overline{x}_j}{n}, \text{ for } j = 1, \ldots, k$$

We therefore need an estimate of *S*. Let us denote the sample variance of $\mathbf{X}_j$ by s_j^2. Equating this with its population counterpart, we have

$$s_j^2 = \frac{na_j\left(S - a_j\right)\left(S + n\right)}{S^2\left(S + 1\right)} \tag{13}$$

From (12) we have $\overline{x}_j / n = a_j / S$. Substituting this into (13), we have

$$s_j^2 = \frac{n\overline{x}_j\left(n - \overline{x}_j\right)(S + n)}{n(S + 1)}$$

Solving this for *S*, we get

$$\hat{S} = \frac{n\left[\overline{x}_j\left(n - \overline{x}_j\right) - s_j^2\right]}{ns_j^2 - \overline{x}_j\left(n - \overline{x}_j\right)}$$

Hardie and Fader (2001) showed the easier way to estimating the parameters of the BBM model through the variance-covariance of the BBM model (see (10)). Looking closely at that expression, we see that it is $(S + n)/(S + 1) \times$ the variance-covariance matrix of the multinomial distribution compute with $\overline{\mathbf{p}}$. This leads to the following procedure for developing an estimate of *S*.

4. Numerical Example

To illustrate the use of BBM model, in this case we used the data in estimating behavior of succeeding of college students in Bandung Islamic University at 1998 (Hajarisman, *et al.*, 1998, and Hajarisman (1998)). The sample was a cluster sample, where classes are clusters or subunit. The population of interest will be restricted to student from Department Statistics, Bandung Islamic University, and the responding number of cluster is 15. In this case, we just look at two of the clusters, where the first cluster is a student who takes a math class and the second cluster is a student who takes an English class. Table 1 shows the data of succeeding of college students, where one margin is sex, and the other margin is succeeding of the student when he/she take the math class and English class.

For illustrative purpose, in our paper present is a 2 × 2 categorical table with one margin being sex of the student and the other margin response to a succeeding of the student who take some classes, success or fail. This response was a series of classes who it was taken by student. And it is of interest to know if these succeeding differ across categories of student, sex being one category. We have then four cross categories with cell numbers respectively 1, 2, 3, 4 corresponding to pair (female, fail), (male, fail), (female, success), and (male, success). Thus the first cell corresponding to female who have been failed in math class (for the first panel data), cell 4 are male who have been succeed in math class, etc.

Table 1. Data of succeeding of college students

Cluster / Sex	Math Class			English Class		
	Fail	Success	Total	Fail	Success	Total
Female	27	211	238	60	178	238
Male	25	145	170	37	133	170
	52	356	408	97	311	408

For the data in first cluster (math class), the probability is

$$P(\mathbf{X}=\mathbf{x})=\binom{408}{27 \quad 25 \quad 211 \quad 145}\frac{\Gamma(S)}{\prod_{j=1}^{4}\Gamma(a_j)}$$
$$=\frac{\Gamma(a_1+27)\Gamma(a_2+25)\Gamma(a_3+211)\Gamma(a_4+145)}{\Gamma(S+408)}$$

and for in second cluster (English class), the probability is

$$P(\mathbf{X}=\mathbf{x})=\binom{408}{60 \quad 37 \quad 178 \quad 133}\frac{\Gamma(S)}{\prod_{j=1}^{4}\Gamma(a_j)}$$
$$=\frac{\Gamma(a_1+60)\Gamma(a_2+37)\Gamma(a_3+178)\Gamma(a_4+133)}{\Gamma(S+408)}$$

The model of interest is the independence model, where sex and succeeding of college students are independent. To find the estimator of the parameter, we used the maximum likelihood methods so we need the score function and higher-order derivatives of the log-likelihood function with respect to S. For this model the maximum likelihood estimator for S is 78.9, with estimator for the cell probabilities of success and the expectation of Xj (E(Xj)), respectively, can be seen on Table 2.

Table 2. Estimator for the cell probabilities of success and the expectation of X_j ($E(X_j)$)

Cluster / Sex	Math Class		English Class	
	Fail	Success	Fail	Success
Female	0.082 (33.456)	0.472 (192.576)	0.127 (51.816)	0.512 (208.576)
Male	0.069 (28.152)	0.377 (153.816)	0.075 (30.600)	0.286 (116.688)

Note: the expectation value for X_j is in the bracket.

From the estimator for cell probabilities of success we can see that the ratio of 1.18 of female versus male, and a ratio of 5.76 of female that have been succeed in math class, and a ratio of 5.46 of male who have been succeed in math class. For English class we can see that the ratio of 1.69 of female versus male, and a ratio of 4.03 of female that have been succeed in math class, and a ratio of 5.46 of male who have been succeed in math class 3.81. Based on these results we can see that the chance of male student succeeding in math class is almost the same as female student. However, the chance of male student in English class is smaller than female student.

The variance-covariance matrix of the beta-binomial multivariate model for first cluster (math class) and second cluster (English class) is given by

$$\operatorname{var}\left(\mathbf{X}^{(1)}\right)=\begin{bmatrix} 178.805 & -13.381 & -91.536 & -73.112 \\ -13.381 & 159.717 & -80.974 & -64.676 \\ -91.536 & -80.974 & 619.625 & -442.422 \\ -73.112 & -64.676 & -442.422 & 583.959 \end{bmatrix}$$

and

$$\operatorname{var}\left(\mathbf{X}^{(2)}\right)=\begin{bmatrix} 275.658 & -23.682 & -163.248 & -90.307 \\ -23.682 & 172.487 & -96.406 & -53.331 \\ -163.248 & -96.406 & 620.856 & -367.629 \\ -90.307 & -53.331 & -367.629 & 507.712 \end{bmatrix}$$

5. Discussion and Extension

The Dirichlet distribution provides a convenient model for describing variation among vectors of proportions since it has relatively simple mathematical properties. The Dirichlet-Multinomial model has been studied by Mosimann (1962). Brier (1980) used the model to analyze sample proportions obtained from a single two-stage cluster sample. Koehler and Wilson (1986) extended some of Brier's results to analyze vectors of proportions obtained from several two-stage cluster samples. In arguing for this methodology there is a question to be discussed: with this model why used maximum likelihood as an estimation procedure?

Brier (1980) in the original paper used method of moments. However Brier does not cover either differing cluster sizes or non-response. Method of moments is very messy when one tries to generalize to these situations. However with maximum likelihood estimation procedure in the most general situation is fairly straightforward, and assuming we have no sparse cells or disproportionately large clusters the asymptotic of the direct maximum likelihood estimators can be established.

Then, we know that any probability modeling effort is the assumption that the observed individual-level behavior is the realization of a random process with density $f(x|\theta)$, which has unknown parameter(s) θ. By assuming a particular distribution for θ, we are able to derive an aggregate-level model without specific knowledge of any given individual's latent parameter(s).

In many cases, however, we are interested in estimating a given individual's latent, θ. This may be because we wish to rank the individuals on the basis of their true underlying behavioral tendency or because we wish to forecast their behavior in a future period. In either case, the challenge is to make inferences regarding θ, given the individual's observed behavior x. In order to address this problem, Hardie and Fader (2001) make used of Bayes theorem.

References

[1]. Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley and Sons.

[2]. Altham, P. M. E. 1971. The Analysis of Matched Proportions. *Biometrika*, **58**(3), 561-576.

[3]. Brier, S. 1980. Analysis of Contingency Tables Under Cluster Sampling. *Biometrika*, 67, 591-596.

[4]. Collet, D. 1991. *Modeling Binary Data*. London: Chapman and Hall.

[5]. Crowder, M. J. 1978. Beta-Binomial ANOVA for Proportions. Applied Statistics, **27**, 34-37.

[6]. Czado, C. 2000. Multivariate Regression Analysis of Panel Data with Binary Outcomes Applied to Unemployment Data. *Technical Report*, York University, Department of Mathematics and Statistics, North York, Ontario.

[7]. Dobson, A. J. 1983. *Introduction to Statistical Modelling*. London: Chapman and Hall.

[8]. Gange, S. J. 1995. Generating Multivariate Categorical Variates Using the Iterative Proportional Fitting Algorithm, *The American Statistician*, **49**, 134-138.

[9]. Hajarisman, N. 1998. *Kajian Perbandingan Model Regresi Beta-Binomial dengan Model Regesi Logistik dan Penerapannya untuk Menduga Pola Kelulusan Mahasiswa TPB – IPB*. Unpublished Thesis, Bogor: Department of Statistics, Bogor Agriculture University.

[10]. Hajarisman, N., Saefuddin, A., Notodiputro, K. A., and Susetyo, B. 1998. The Beta-Binomial Regression Model for Estimating Behavior of Succeding of College Students. *Forum Statistika dan Komputasi*, Vol

[11]. Hardie, B. G. S., and Fader, P. S. 2000. *Applied Probability Models in Marketing Research: Introduction*. www.petefader.com

[12]. Johnson, N. L., and Kotz, S. 1969. *Multivariate Discrete Distribution*. New York: Houghton Mifflin Company.

[13]. Kupper, L.L., and Haseman, J.K. 1978. The Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments. *Biometrics*, **34**, 69-76.

[14]. McCullagh, P., and Nelder, J. A. 1983. *Generalized Linear Models*. 2nd Ed. New York: Chapman and Hall.

[15]. Moseman, J. E. 1962. On the Compound Multinomial Distribution, the Multivariate β-Distribution, and Correlation among Proportion. *Biometrika*, **49**, 65-82.

[16]. Pendergast, J. F., Gange, S. J., Lindstrom, M. J., Newton, M. A., Palta, M., and Fisher, M. F. 1993. A Survey of Methods for Analyzing Clustered Binary Data. *Technical Report 82*, University of Wisconsin, Department of Biostatistics.

[17]. Prentice, R. L. 1986. Binary Regression Using an Extended Beta-binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors. *Journal of the American Statistical Association*, **81**, 321-327.

[18]. Rizzo, L. 1990. Fitting Categorical Data Models Using Data from Complex Sample Surveys with Item Nonresponse. *Technical Report*, The University of Iowa, Iowa City, Iowa 52242.

[19]. Tallis, G. M. 1962. The Use of A Generalized Multinomial Distribution in the Estimation of Correlation in Discrete Data. *Journal of Royal Statistics Society*, Series B, **24**, 530-534.

[20]. William, D. A. 1975. The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity. *Biometrics*, **31**: 949-952.
[21]. Wilson, J. R. 1987. Logarithmic Series Distribution and Its Use in Analyzing Discrete Data. *Technical Report DIS*, Arizona State University, Tempe, Arizona.
[22]. Wilson, J. R., and Koehler, K. J. 1988. A Generalized Dirichlet Multinomial Model for Categorical Data with Extra Variation. *Technical Report DIS*, Arizona State University, Tempe, Arizona.