Revista Colombiana de Estadística
Volumen 28 N° 1. pp. 63 a 76. Junio 2005

Estimating multilevel models for categorical data via generalized least squares

Minerva Montero Díaz*
Valia Guerra Ones**

Resumen

Montero, Castell & Ojeda (2002) propusieron una estrategia para formular modelos multinivel para tablas de contingencia basada en la aplicación del modelo lineal general a datos categóricos jerárquicos. Aplicando el método a un modelo de regresión logística multinivel con datos simulados, encontramos que las estimaciones de los parámetros aleatorios son inadmisibles en ciertas situaciones, con sesgos grandes y estimaciones negativas de la varianza cuando los conjuntos de datos son desbalanceados. Para corregir los estimadores proponemos una técnica basada en descomposición de valores singulares truncados en la solución de mínimos cuadrados generalizados para estimar los parámetros aleatorios. Mediante simulación mostramos la efectividad de la técnica en cuanto a la reducción del sesgo de los estimadores.

Palabras claves: *Modelos multinivel, mínimos cuadrados generalizados, valores singulares truncados.*

Abstract

Montero et al. (2002) proposed a strategy to formulate multilevel models related to a contingency table sample. This methodology is based on the application of the general linear model to hierarchical categorical data. In this paper we applied the method to a multilevel logistic regression model using simulated data. We find that the estimates of the random parameters are inadmissible in some circumstances; large bias and negative estimates of the variance are expected for unbalanced data sets. In order to correct the estimates we propose to use a numerical technique based on the Truncated Singular Value Decomposition (TSVD) in the solution of the problem of generalized least squares associated to the estimation of the random parameters. Finally a simulation study is presented to shows the effectiveness of this technique for reducing the bias of the estimates.

Keywords: *Multilevel models, Generalized least squares, Truncated Singular Value.*

*Instituto de Cibernética, Matemática y Física. Ciudad Habana. Cuba. E-mail: minerva@icmf.inf.cu
**Instituto de Cibernética, Matemática y Física. Ciudad Habana. Cuba

1. Introduction

The analysis of a sample of contingency tables plays an important role in many fields of science. Non-normal generalized linear models with random effects have become increasingly accepted for the analysis of such data (Lee & Nelder (1996, 2001), Breslow & Clayton (1993)). In making inferences from this class of models, a marginal-likelihood analysis is often troubled by intractable integration. To avoid this, during recent years, various approximate methods of inference to fit multilevel models for binary or count data have been proposed (Longford 1994, Goldstein 1991).

Montero et al. (2002) consider the GSK approach (Grizzle, Starmerc & Koch 1969), as a tool to formulate multilevel models for analyses a sample of contingency tables and introduce an estimation procedure that may be applied to fit these models. This procedure relegates the analysis of a sample of contingency tables to a class of problem that can be handled by Generalized Least Squares (GLS). One of the main advantages of the procedure is the similarity with the case of the multilevel linear model; hence it can be used in situations where other methods impose the solution of complicated mathematical expressions.

In this paper the validity of this procedure for the analysis of a sample of contingency tables is explored by means of a multilevel logistic regression model with random slope. When the data are balanced (Montero, Castell & Ojeda 2003) the procedure can obtain estimates of the random parameters at accepted levels of bias and precision; however, the estimations can be inadmissible when the data are unbalanced. In this paper, we analyze the theoretical and numerical reasons that justify the disturbing estimations for the random parameters. The analysis is based on the effect of the smallest singular values of the design matrix on the random parameter estimation. We propose the Truncated Singular Value Decomposition (TSVD) as a technique for avoiding the inadmissible solutions and the L-curve criterion is suggested for calculating the truncation index. Simulations varying the degree of imbalance and the variance size of the random effects are presented to illustrate the effectiveness of the proposed technique.

2. The Model and Estimation Procedure

We consider a 2-level hierarchical data structure. Suppose a sample of J contingency tables (level-2 units) where the rows of each table, called subpopulations, represent I levels (level-1 units) of an explanatory variable or combinations of levels of several explanatory variables. Random samples of size n_{ij} $(i = 1, 2, \ldots, I;\ j = 1, 2, \ldots, J)$ are selected from rows. The responses are classified according to r categories with $n_{ilj}, (l = 1, 2, \ldots, r)$ denoting the number of elements classified in the lth response category for the ith subpopulation of the jth table.

Let $\pi_j = \left(\pi'_{1j}, \pi'_{2j}, \ldots, \pi'_{ij}\right)$, $\pi_{ij} = (\pi_{i1j}, \pi_{i2j}, \ldots, \pi_{irj})'$ with $\sum\limits_{i} \pi_{ilj} = 1$, represents a vector of probabilities for the jth table. Each set of probabilities has $r - 1$

linearly independent elements.

Let $F(\pi_j) = [F_1(\pi_j), F_2(\pi_j), \ldots, F(\pi_j)]'$ be a vector of $a < I(r-1)$ functions of π_j.

Different types of functions may be represented in a relatively simple manner using matrix notation ()(Forthoper and Koch, 1973). The function of the probabilities can be simple (e.g., the same probability) or complex (e.g., a rank correlation coefficient between two response variables, etc).

By analyzing tables where the I samples of the J tables are independent, the GSK approach establish that once the function has been specified it can be used as dependent variable in a linear model. However, when analyzing a sample of contingency tables, the lack of independence between subpopulations results in distortion of variance estimates and this can result in problems with type I error for ordinary test statistic.

The procedure presented in this paper explicitly takes into account dependence across tables as well as within tables. The values of the functions of the probabilities become realizations of the dependent variable in a multilevel linear model. Dependencies between the observations are modeled via random effects. Once the model is formulated, it is possible to apply the asymptotic theory of estimation in the framework of the general linear model. The estimation procedure is based on iterative generalized least squares.

2.1. Multilevel Model for Proportions

In this paper we are mainly concerned with the logit function. We consider a sample of contingency tables with a set of two proportions, p_{ij}, $1-p_{ij}$ as outcomes, for the individuals classified in the ith row from the jth table. The following 2-level logit model with a single dichotomy explanatory variable is considered:

$$f_{ij} = \mathrm{logit}(p_{ij}) = \log(p_{ij}) - \log(1-p_{ij}) = \gamma_{00} + \gamma_{10}x_{ij} + u_j z_{ij} + e_{ij} \qquad (1)$$

where x_{ij} is a covariate having fixed effect γ_{10}, z_{ij} is a covariate having random effects u_{ij} at the 2-level and e_{ij} are independent level-1 random errors. The situation studied in this paper correspond to the case where $z_{ij} = x_{ij}$

We assumed that the observed proportions follow a binomial distribution, but a simplification was introduced. As suggested by Goldstein (1987) we can simply required the variances to be inversely proportional to n_{ij}, then, the level-1 variance of f_{ij} is also inversely proportional to n_{ij}. If we further assume a simple random variation of the f_{ij} across tables, then the between tables variation is assumed to be the same for each of the I subpopulations.

We assume that:

$$E(u_{ij}) = E(e_{ij}) = 0$$

$$Var(u_j) = \sigma_u^2, \quad Var(e_{ij}) = \frac{\sigma_{e_i}^2}{n_{ij}} \ \text{ and } \ \mathrm{Cov}(u_j, e_{ij}) = 0$$

An expression for the total variance of f_{ij} in the model (1) can be written as:

$$\sigma_u^2 z_{ij} + \frac{\sigma_{e_i}^2}{n_{ij}}$$

for the ith subpopulation. The model requires then the estimation of three random parameters, σ_u^2, $\sigma_{e_0}^2$ and $\sigma_{e_1}^2$.

Let p_j be the vector of observed proportions, given in the same way as π_j. Note that model (1) can be written as a special case of the general linear mixed model:

$$F(p) = X\Gamma + Zu + e \tag{2}$$

where $F(p) = A\log(p)$ is the logit function for the observed proportions, whit A denoting the matrix of the coefficients of the natural logarithms of the vector $p = (p_1, p_2, \ldots, p_J)$; Γ is a vector of fixed coefficients with design matrix X; u is a vector of random effects whit design matrix Z and e is a vector of random errors.

It should been now be noted that:

$$E(F(p)) = X\Gamma, \quad \text{Let } \; Var(e) = \Omega_e, \;\; Var(u) = \Omega_u \quad \text{and} \quad \operatorname{Cov}(e,u) = 0$$

We can then say that:

$$Var(F(p)) = V_F = Z\Omega_u Z' + \Omega_e$$

The model (2) is a special case of the general linear model:

$$F(p) = X\Gamma + e^*$$

where $e^* = Zu + e$, $E(e^*) = 0$ and $\operatorname{Cov}(e^*, e^*) = V_F$.

If the covariance matrix is known, the parameter vector Γ, is estimated by generalized least squares:

$$\Gamma = \left(X' V_F^{-1} X\right) X' V_F^{-1} F(p) \tag{3}$$

When V_F is unknown a common practice is to substitute $\widehat{V}_F$ for an estimate V_F in the expression (3). We carry out an iterative procedure analogous to the described in Goldstein (1995) which alternates between estimates of fixed and random parameters until convergence. We estimated the fixed parameters from a generalized least squares (GLS) fit for categorical data ignoring the random errors at level 2 (see appendix A).

Once suitable starting values for the fixed parameters are obtained we form the "raw" residuals $\widetilde{F} = F(p) - \widehat{\Gamma}A$ which can be used to estimate the random parameters in the model. Then form the cross-product matrix $E(F^*) = E\left(\widetilde{F}\widetilde{F}\right) = V_F$. We vectorize the cross-product matrix $F^{**} = \operatorname{vec}(F^*)$, and similarly we construct the vector $\operatorname{vec}(V_F)$. The relationship between these vectors can be expressed as the following linear model involving the random parameter vector θ, so that:

$$F^{**} = Z^*\theta + R \tag{4}$$

where Ω_u and Ω_e are the elements of θ, Z^* is the design matrix for the random parameters and R is a residual vector. In order to estimate the random parameter vector θ, we carry out a generalized least squares analysis, so that:

$$\widehat{\theta} = \left(Z^{*'} V^{*^{-1}} Z^*\right)^{-1} Z^{*'} V^{*^{-1}} F^{**}$$

where V^* is the Kronecker square product of V_F, namely $V^* = V_F \otimes V_F$. The estimated covariance matrix for the fixed parameters is:

$$\mathrm{Cov}\left(\widehat{\Gamma}\right) = \left(X' V^{-1} X\right)^{-1}$$

and for the random parameters:

$$\mathrm{Cov}\left(\widehat{\theta}\right) = \left(Z^{*'} V^{*^{-1}} Z^*\right)^{-1} Z^{*^t} V^{*^{-1}} \mathrm{Cov}\left(F^{**}\right) V^{*^{-1}} Z^{*^t} \left(Z^{*^t} V^{*^{-1}} Z^*\right)^{-1}$$

Assuming multivariate normality, Goldstein & Rasbash (1992) show that:

$$\mathrm{Cov}\left(\widehat{\theta}\right) = 2\left(Z^{*'} V^{*-1} Z^{*'}\right)^{-1}$$

We observed that in some circumstances the estimation procedure can produce inadmissible estimates of the random parameters. We consider the case where the quality of estimations is affected by imbalance among the subpopulation sample sizes.

3. Analysis of Simulated Data

Simulation studies (Montero et al. 2003) based on the model (1), used to investigate how the effects of sample size may affect the estimation of the parameters, show that, the proposed estimation procedure seems to perform adequately for balanced data, in the situations taken into account.

To explore the properties of estimators for unbalanced data we use the same hierarchical structure as in the balanced case; i.e., the values of parameters γ_{00} and γ_{10} in the model (1) were set to 0.5 and 1.0, respectively. The level-2 random effects u_j are generated from independent normal distribution with zero mean and finite variance. Logit(π_{ij}) is obtained adding the fixed part and level-2 random effects. Finally, the values of the variable n_{ij} (used for obtain p_{ij}) are generated from a binomial distribution with parameter π_{ij}.

The number of contingency tables is fixed at 50. Several different uniforms distributions were used to generate the sample sizes of each row in a set of tables. The designs are classified in four different types of designs depending on the degree of imbalance of the tables, that is:

Type B: Design Balance, $n_{ij} = 200$ for all i, j.
Type S: Design Slightly unbalanced, $n_{ij} \sim U\,(199, 200)$.
Type M: Design Moderately unbalanced, $n_{ij} \sim U\,(150, 200)$.
Type L: Design Largely unbalanced, $n_{ij} \sim U\,(100, 200)$.

One small and one large level-2 variance $\left(\sigma_u^2 = 0.5 \text{ and } \sigma_u^2 = 1.0\right)$ were assumed. Thus, there are $4 \times 2 = 8$ different design conditions and for each condition 100 simulated data sets were generated.

The estimations of the fixed and random parameters were obtained for simulations under the different conditions of the designs. The procedure produced reasonably unbiased estimates for the fixed parameters γ_{00} and γ_{10}, but it exhibits big difficulties in the estimates of the random parameters for unbalanced samples. We focus our attention on the estimates of the variance of the random effects, that is, $\widehat{\sigma}_u^2$. Because of similar behaviors of the estimates, in this section we only show the case where the random parameter is sufficiently large to be interesting $\left(\sigma_u^2 = 1\right)$.

Figure 1 shows plots of the distributions of 100 estimates of the random parameters for each design considered in the study. Note that large bias and negative estimates of the variance are expected for the three unbalanced data sets. The situation is particularly bad when the tables are slightly unbalanced. In contrast, the estimates for tables more unbalanced appear to be less biased, but are still inadmissible. Paradoxically, the biggest differences are between the balanced set data and the slightly balanced.

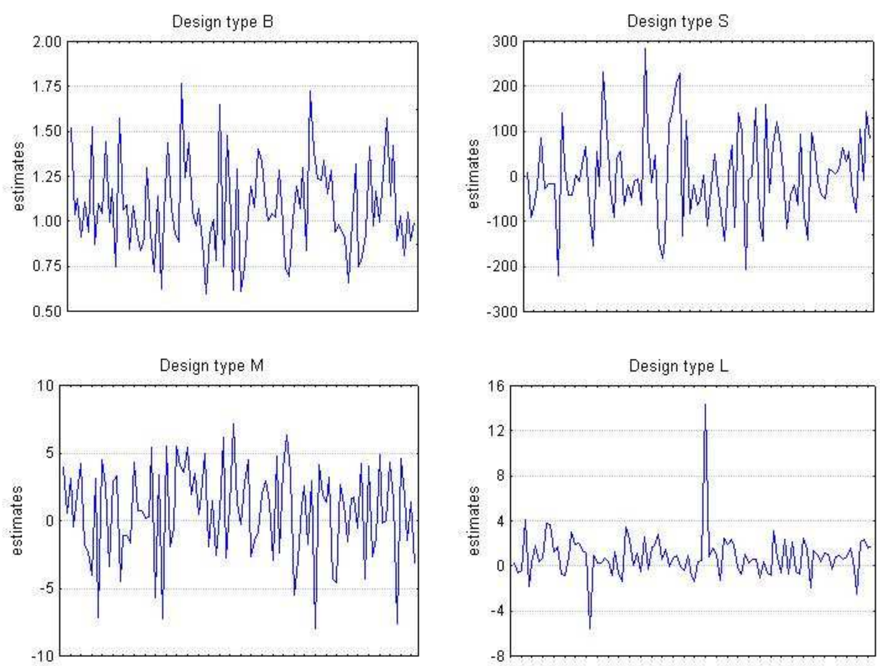

Figure 1: Line plot of the distributions of 100 estimates of the random parameters for the four designs considered.

4. Understanding and Solving the Numerical Difficulties

The origin of the inadmissible estimates of the random parameters for unbalanced data is related to the numerical solution of the general linear model (4).

Consider the Cholesky decomposition of the symmetric positive definite covariance matrix $\tau^2 V^* = BB^t$. Then, the solution vector θ in (4) can be calculated solving the least square problem:

$$\min \left\| B^{-1} \left(Z^* \theta - F^{**} \right) \right\|_2 \tag{5}$$

This problem should be solved using a stable algorithm suggested by Paige (1979), where the pseudoinverse of B is not calculated implicitly. However, if B is a well-conditioned matrix an obvious computational approach to this problem is to apply any standard technique to minimize $\left\| \left(B^{-1} Z^* \right) \theta - B^{-1} F^{**} \right\|$.

The Singular Value Decomposition (SVD) is a useful tool to solve (5) and to understand the numerical results shown in Figure 1. Given the matrix $W = B^{-1} Z^*$, it always exists orthonormal matrices U and V and a diagonal matrix S such that:

$$W = USV^{'}$$

the diagonal elements S_i of S are called singular values of W.

Using the SVD of W, the random parameter vector θ in (4) can be written as:

$$\theta = \sum_{i=1}^{\mathrm{rank}(W)} \frac{U_i^{'} F^{**}}{S_i} V_i \tag{6}$$

where U_i and V_i are the columns of the matrices U and V, respectively, and rank(W) denotes the rank of W.

Expression (6) permits to understand the numerical results shown in Figure 1. Note that if the matrix W has very small singular values S_i; then the corresponding coefficients $(U_i F^{**} |\ S_i)$ can increase drastically the magnitude of the solution θ. Likewise, the presence of small singular values in the matrix W can produce huge changes when the coefficients of W are slightly perturbed.

In the simulation study of section 3, we have observed that in the case of balanced data, the singular values of matrix W are not small, except one of them, that is smaller than the computer precision. It means the matrix is rank one deficient and the summand corresponding to the smallest singular value is not considered in (6). It explains the acceptable estimates obtained for the random parameters when the data are balanced. However, in the unbalanced cases, where large bias and negative estimates of the variance are obtained, we observe the presence of very small singular values S_i in the matrix W that are not considered as zero by the computer and then the summands corresponding to these singular values are included for calculating θ in the expression (6).

A possible way to obtain acceptable values for the random parameter vector θ is truncating the expression (6) to include only the k summands corresponding to singular values greater than a given tolerance. In other words, the random parameter vector θ is approximated by:

$$\theta = \sum_{i=1}^{k} \frac{U_i^{!} F^{**}}{S_i} V_i \tag{7}$$

This technique is known as Truncated Singular Value Decomposition (TSVD), (Golub & Loan 1996).

The determination of the tolerance parameter can be a difficult task. When there is a well-determined gap between large and small singular values, the parameter k is chosen equal to the number of the large singular values. However, when all singular values decay gradually to zero, and there is no gap in the singular value spectrum, the parameter k should be calculated using a numerical technique, for example the L-curve criterion, (Hansen 1998).

This criterion is based on the determination of the corner of a discrete parametric plot of the norm of the solution θ_k versus the norm of the corresponding residual $\left\| \left(B^{-1}Z^*\right) \theta_k - B^{-1}F^{**} \right\|$, see details in Hansen (1998).

It is important to say that other approximations for θ can be considered for avoiding the overestimation and underestimation of the random parameters. The main idea is to filter the contribution of each summand of the expression (5) to the calculated vector. This aspect will be analyzed in future studies. Next section illustrates the numerical results obtained using the expression (7) and taking the tolerance parameter as 10^{-5}.

5. Simulation Study

In order to study the performance of the correction introduced, we now simulate data under the same conditions as one of the preceding simulation study of section 3 and fit the multilevel model (1) by the modified procedure. For every model specification, 500 data sets were generated. The estimation procedure converged in all 3000 simulated data sets.

To analyze the parameter estimates two criterions, bias and efficiency, are used. Tables 1 and 2 display for each parameter the true value and the values of the estimated fixed and random parameters averaged over the 500 simulations conducted every design. The mean of the correspondent Mean Squared Errors (MSE), and the mean of estimated standard errors are also given. First we discuss the case where the variance of random effects is large $\left(\sigma_u^2 = 1.0\right)$.

As we can see from Table 1, it is evident that the application of the Truncated Singular Value Decomposition improves substantially the random parameter estimates. The procedure gives good estimates for the fixed parameters and reasonably biased estimates for the random parameters at level 2.

It is clear that the fixed parameter estimates are close to their true value; that is, the bias of the estimates is small. For the fixed parameters the approach performs excellently with a bias of 3.7% at the most. Table 1 shows that the

Table 1: Mean values of multilevel logit estimates for 500 simulated data sets for model (1) assuming $\sigma_u^2 = 1.0$

	Parameters	True value	Estimate	MSE	e.s
Design type S					
	γ_{00}	0.5	0.505	0.001	0.024
	γ_{10}	1	1.031	0.025	0.150
	σ_u^2	1	1.114	0.073	0.124
Design type M					
	γ_{00}	0.5	0.503	0.000	0.022
	γ_{10}	1	1.026	0.023	0.148
	σ_u^2	1	1.082	0.065	0.123
Design type L					
	γ_{00}	0.5	0.504	0.000	0.020
	γ_{10}	1	1.014	0.020	0.148
	σ_u^2	1	1.088	0.065	0.124

estimation procedure results in very small MSE for the fixed parameters, especially for γ_{00}.

The random parameter estimates represent a considerable improvement, but are still subject to a small bias. The estimates for the three unbalanced data sets are 11.4, 8.2 and 8.8 percent upward bias respectively. The standard deviation of estimates is small and none of these biases are statistically different from zero. The MSE values reported in Table 1 show that the procedure is less efficient in estimating the random parameters. Table 2 shows that when the variance of the random effects is small $\left(\sigma_u^2 = 0.5\right)$ none of the estimates is significantly biased. The estimates of the parameter σ_u^2 are 14.4, 12.8 and 9.4 percent upward biases.

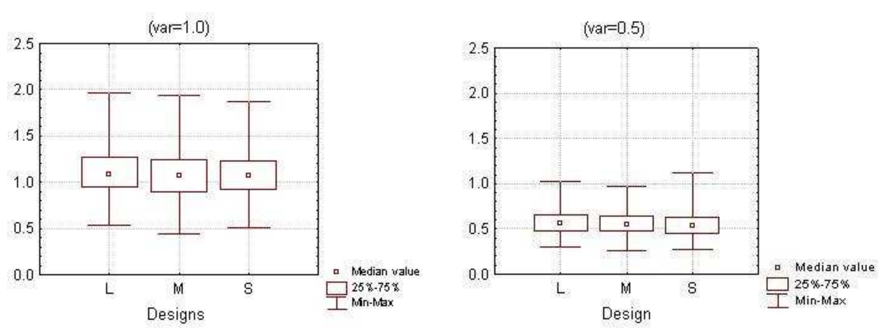

Figure 2: Boxplots of estimates of $\widehat{\sigma}_u^2$.

Table 2: Mean values of multilevel logit estimates for 500 simulated data sets for model (1) assuming $\sigma_u^2 = 0.5$

	Parameters	True value	Estimate	MSE	e.s
Design type S					
	γ_{00}	0.5	0.502	0.001	0.024
	γ_{10}	1	1.022	0.012	0.109
	σ_u^2	0.5	0.572	0.023	0.083
Design type M					
	γ_{00}	0.5	0.504	0.000	0.022
	γ_{10}	1	1.023	0.013	0.108
	σ_u^2	0.5	0.564	0.019	0.083
Design type L					
	γ_{00}	0.5	0.502	0.000	0.021
	γ_{10}	1	1.010	0.012	0.106
	σ_u^2	0.5	0.547	0.018	0.082

Finally, we consider how the quality of estimation is affected by the imbalance of the data when the TSVD is applied. The values of MSE reported in Table 1 and 2 show that the estimator is equally efficient for the three unbalanced designs. Figure 2 shows graphically the sampling distributions for the estimations of each design. A general suggestion of this figure is that estimation of random parameters is little affected by the imbalance of the tables. Quality of estimation seems fairly insensitive to unbalance.

Figure 3 shows the normal probability plots of the random parameter estimates produced by the proposed method. Except for a few outliers the plots for all the estimates are reasonably consistent with the expected asymptotic normality.

6. Conclusions

Our aim was to examine the behavior of an estimation procedure based on the generalized least squares method for categorical data analysis, in the frame of multilevel models related to a two-level hierarchical data structure coming from a sample of contingency tables. We are particularly interested in the multilevel logistic regression model, but the method can be applied to other models and in situations where other methods impose the solution of complicated mathematical expressions. The main advantage of this approach is the similarity with the case of the linear model.

On the basis of a number of simulations the results revealed that the degree of imbalance of the data has an important impact on the estimation of the random parameters. For unbalanced data, the proposed procedure produces inadmissible estimates of the random parameters. We showed that the TSVD, used to solve the

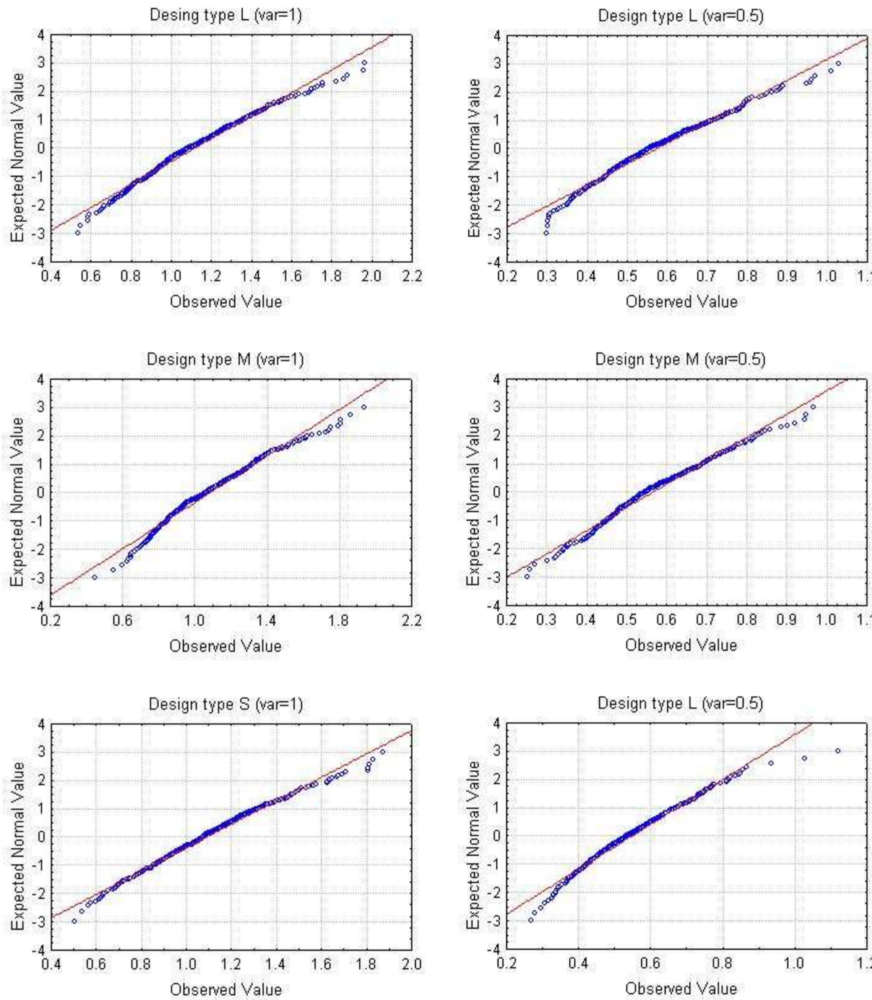

Figure 3: Normal Probability Plot of $\widehat{\sigma}_u^2$.

least squares problem associated to the estimation of the random parameters, can considerably improve the estimates. The study was carried out via a simulations study. Random parameters are estimated at accepted levels of bias and precision after the modification is applied. In summary, TSVD is effective in reducing the bias of random parameters.

For the specifications considered, the comparison between the designs shows that the degree of imbalance seems to have neither a systematic nor a significantly different effect on bias and efficiency of the estimates if a modification, such as the TSVD, is applied. When variance is small , the estimator was found to be slightly more efficient that when the variance is large.

Although it is not appropriate to draw general conclusions from a single simulation study, the results suggest the described procedure should be used as an efficient method to handle multilevel models for hierarchical categorical data. A further analysis of more complex models and extreme data sets is necessary to recommend this approach as a unified approach for modeling a sample of contingency tables.

Acknowledgements

We would like to thank Dr. Jesús E. Sánchez for his careful reading and suggestions on the first version of this paper.

A. GLS Fit for Categorical Data (or GSK Approach)

Consider the data structure of section 1. If we assume the I subpopulations of the jth table as being uncorrelated with one another a consistent estimator for the covariance matrix of p_j is the matrix:

$$V_j\left(p_j\right) = \operatorname{diag}\left(V_{1j}\left(p_{1j}\right), V_{2j}\left(p_{2j}\right), \ldots, V_{IJ}\left(p_{Ij}\right)\right)$$

with the matrices:

$$V_{ij}\left(p_{ij}\right) = \frac{1}{n_{ij}}\left[D_{p_{ij}} - p_{ij}p_{ij}^{\shortmid}\right], \qquad (i = 1, 2, \ldots, I)$$

where $D_{p_{ij}}$ is a matrix diagonal with elements of the vector p_{ij} on the main diagonal.

Let $F_j \equiv F(p_j)$. We assume that F_j has continuous second order partial derivatives in an open region containing π_j. A consistent estimator for the covariance matrix of F_j is the matrix:

$$\widehat{V}_{F_j} = H_j\left[V_j\left(p_j\right)H_j^{\shortmid}\right]$$

where $H = \left[\partial F\left(\pi_j\right)/\partial\pi_j \mid \pi_j = p_j\right]$ is the $a \times Ic$ matrix of first partial derivatives of the functions F_j evaluated on p_j.

Observations from different tables are mutually independent and, if no function combines probabilities from more than one population, this independence is maintained through the transformation. Thus, the covariance between observations from different tables is zero, and the estimated covariance matrix of $F \equiv F(p)$ has the form:

$$\widehat{V}_F = \operatorname{diag}\left(\widehat{V}_{F_1}, \widehat{V}_{F_2}, \ldots, \widehat{V}_{F_J}\right)$$

The GSK approach applies to linear models for F of the form $F(\pi) = X\Gamma$.

Note: A consistent estimator for the covariance matrix of the function $F(p_j) = B_j \log(p_j)$ (Forthofer & Koch 1973) is the matrix:

$$\widehat{V}_{F_j} = A_j D_j^{-1} \left[\widehat{V_j}(p_j)\right] D_j^{-1} A_j^{-1}$$

where D_j contains the elements of the vector p_j on the main diagonal.

References

Breslow, N. E. & Clayton, D. G. (1993), ‘Approximate inference in generalized linear mixed models’, *American Statistical Association* **88**, 9–25.

Forthofer, R. N. & Koch, G. G. (1973), ‘An analysis for compounded functions of categorical data’, *Biometrics* **29**, 143–159.

Goldstein, H. (1987), *Multilevel Models in Educational and Social Research*, Charles Griffin.

Goldstein, H. (1991), ‘Nonlinear multilevel models, with an application to discrete response data’, *Biometrika* **78**(1), 45–51.

Goldstein, H. (1995), *Multilevel Statistical Models*, 2 edn, Halsted Press.

Goldstein, H. & Rasbash, J. (1992), ‘Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalized least squares’, *Computational Statistics and Data Analysis* **13**, 63–71.

Golub, G. & Loan, C. F. V. (1996), *Matrix Computations*, 3 edn.

Grizzle, J. E., Starmerc, F. & Koch, G. (1969), ‘Analysis of categorical data by linear models’, *Biometrics* **25**, 489–504.

Hansen, P. C. (1998), *Rank-deficient and discrete ill-posed problems: Numerical aspects and linear inversion*, Society for Industrial and Applied Mathematics.

Lee, Y. & Nelder, J. A. (1996), ‘Hierarchical generalized linear models’, *Royal Statistics Society* **B**(58), 619–678.

Lee, Y. & Nelder, J. A. (2001), ‘Hierarchical generalized linear models: a synthesis of generalized linear model and structured dispersion’, *Biometrika* **88**, 987–1006.

Longford, N. (1994), 'Logistic regression with random coefficients', *Computational Statistics and Data Analysis* **97**, 1–15.

Montero, M., Castell, E. & Ojeda, M. M. (2002), Modelos multinivel de una muestra de tablas de contingencia utilizando el enfoque gsk, Technical Report 2002–167, Reporte de investigación del ICIMAF.

Montero, M., Castell, E. & Ojeda, M. M. (2003), Modelos multinivel para una muestra de tablas de contingencia: un estudio por simulación, Technical Report 2003–228, Reporte de investigación del ICIMAF.

Paige, C. C. (1979), 'Fast numerically stable computations for generalizad linear least squares problems', *Society for Industrial and Applied Mathematics* **1**(1), 165–171.