J. R. Statist. Soc. B (1996)
58, *No.* 1, *pp.* 253–266

On the Comparison of Multinomial and Poisson Log-linear Models

By JOSEPH B. LANG†

University of Iowa, Iowa City, USA

[Received March 1994. Revised January 1995]

SUMMARY

We introduce a method for comparing multinomial and Poisson log-linear models which affords an explicit description of their equivalences and differences. The method involves specifying the model in terms of constraint equations, rather than the more common freedom equations. The Poisson and multinomial large sample distributions of log-linear model parameter estimators are derived and compared within this constraint equation context; reparameterizations are thereby avoided. As a by-product, the method provides the practitioner with the adjustment that is necessary to make valid inferences about all multinomial log-linear parameters when, as a matter of convenience, the Poisson log-linear model is fitted. This implies that valid large sample inferences about the multinomial cell probabilities can be made directly by using the Poisson log-linear model. To illustrate the utility of this approach, several examples are considered.

Keywords: CATEGORICAL DATA; CONSTRAINT EQUATIONS; FREEDOM EQUATIONS; FREEDOM PARAMETERS; LAGRANGE MULTIPLIERS; RESTRICTED LIKELIHOOD EQUATIONS

1. INTRODUCTION

In practice, the Poisson log-linear model is often assumed as a matter of convenience. In particular, the Poisson log-linear model is simpler to fit using maximum likelihood methods than the multinomial log-linear model is. This follows since

(a) the mean parameters are not required to satisfy sampling constraints and
(b) the components of the random vector are independent.

In fact, the Poisson log-linear model is a univariate generalized linear model (McCullagh and Nelder, 1989) and can be fitted by using the iterative reweighted least squares algorithm.

It has been pointed out by many researchers (e.g. Birch (1963), Bishop *et al.* (1975), Palmgren (1981), McCullagh and Nelder (1989) and Agresti (1990)) that there are several equivalences between the Poisson and multinomial log-linear models. The method that these researchers used to show the equivalences generally, depending on the original parameterization, requires several reparameterizations. When the log-linear model satisfies certain sufficient conditions the Poisson log-likelihood can be partitioned into a sum of two log-likelihoods: one is the reparameterized multinomial log-likelihood; the other is a Poisson log-likelihood. For the proper parameterization, which exists when the sufficient conditions are met, the two log-likelihoods are functions of distinct parameters. From this partitioning it follows that the large sample likelihood inferences about the parameters in the reparameterized multi-

†*Address for correspondence*: Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA.
E-mail: lang@stat.uiowa.edu

 0035–9246/96/58253

nomial log-likelihood are identical for both sampling assumptions (see Palmgren (1981)). These equivalences enable the user to fit the model by using the Poisson assumption even when it is known that the data are product multinomial.

We present a different method for showing these equivalences, a method which also allows for an explicit description of the differences. Our method does not depend on the particular parameterization chosen; it is easily seen to depend only on the spanning set of the design, or basis, matrix. Reparameterization of the log-likelihoods will be avoided, allowing us to derive results about the parameters in the original parameterization. We show that as a by-product of this method of proof we obtain correction matrices that can be used to alter the Poisson variance estimates under the original parameterization so that they are equal to the product multinomial variance estimates. This implies that valid inference about the product multinomial cell probabilities can be made directly by using the Poisson log-linear model. We shall derive the asymptotic distributions of the multinomial and the Poisson maximum likelihood estimators using the approach of Aitchison and Silvey (1958, 1960). Their method involves specifying the model by using constraint equations and then maximizing the log-likelihood subject to these constraints.

In Section 2, we introduce the notation that will be used in the remainder of the paper. The multinomial and Poisson log-linear models are also defined. Two different, but equivalent, specifications of a log-linear model are considered in Section 3. We also present the non-restrictive model assumptions that will be sufficient for the equivalence results to hold. In Section 4, the multinomial log-linear model is described in detail. We derive the restricted likelihood equations (see Aitchison and Silvey (1960)) and describe the large sample behaviour of the solution to these equations. In Section 5, we mirror the discussion of Section 4, this time for Poisson log-linear models. A comparison of the two models is conducted in Section 6. In this section, we give an interesting equivalence result that has several practical implications. Section 7 illustrates these results via examples. A brief discussion of the paper is given in the final section.

2. NOTATION

Throughout this paper we shall assume that the random vector of counts $\mathbf{Y}$ satisfies either $\mathcal{M}$—$\mathbf{Y} = \text{vec}(\mathbf{Y}_1, \ldots, \mathbf{Y}_K)$, where $\mathbf{Y}_k = (Y_{k1}, \ldots, Y_{kr})' \sim$ indep mult$(n_k, \boldsymbol{\pi}_k)$ and $\boldsymbol{\pi}_k = (\pi_{k1}, \ldots, \pi_{kr})'$, $\Sigma_{j=1}^r \pi_{kj} = 1$, $k = 1, \ldots, K$, or $\mathcal{P}$—$\mathbf{Y} = \text{vec}(\mathbf{Y}_1, \ldots, \mathbf{Y}_K)$, where $Y_{kj} \sim$ indep Poisson (μ_{kj}), $k = 1, \ldots, K$, $j = 1, \ldots, r$. That is, the $s \times 1$ vector $\mathbf{Y}$, where $s = rK$, is either product multinomial or product Poisson.

For the product multinomial sampling scheme $\mathcal{M}$, let $\mu_{kj} = n_k \pi_{kj}$ represent the expected cell counts. For both sampling schemes, let the $s \times 1$ vector $\boldsymbol{\mu} = \text{vec}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$, where $\boldsymbol{\mu}_k = \text{vec}(\mu_{kj}; j = 1, \ldots, r)$, and let $\boldsymbol{\xi} = \log \boldsymbol{\mu}$.

Consider the log-linear models $[\omega^{(M)}]$ and $[\omega^{(P)}]$ with corresponding model spaces

$$\omega^{(M)} = \{\boldsymbol{\xi}: \boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta}, \text{samp}(\boldsymbol{\xi}) = \mathbf{0}\}, \tag{2.1}$$

$$\omega^{(P)} = \{\boldsymbol{\xi}: \boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta}\}. \tag{2.2}$$

Here the constraint $\text{samp}(\boldsymbol{\xi}) = \mathbf{0}$ is the multinomial sampling constraint; it is

$\text{samp}(\boldsymbol{\xi}) = (\oplus_{k=1}^{K} \mathbf{1}_r')\mathrm{e}^{\boldsymbol{\xi}} - (n_1, \ldots, n_K)' = \mathbf{0}$. The symbol $\oplus_{k=1}^{K} \mathbf{1}_r'$ represents the direct sum of K components, in this case K identical components. Specifically,

$$\oplus_{k=1}^{K} \mathbf{1}_r' = \begin{pmatrix} \mathbf{1}_r' & \mathbf{0}_r' & \mathbf{0}_r' & \ldots & \mathbf{0}_r' \\ \mathbf{0}_r' & \mathbf{1}_r' & \mathbf{0}_r' & \ldots & \mathbf{0}_r' \\ & \vdots & & \ddots & \\ \mathbf{0}_r' & \mathbf{0}_r' & \mathbf{0}_r' & \ldots & \mathbf{1}_r' \end{pmatrix}$$

is a $K \times s$ matrix. The multinomial sampling constraints simply state that within each of the K levels, say level k, the sum of the expected counts is n_k.

The model $[\omega^{(M)}]$ along with the distributional assumption $\mathcal{M}$ is called a product multinomial log-linear model. The model $[\omega^{(P)}]$ along with the distributional assumption $\mathcal{P}$ is called a Poisson log-linear model.

3. CONSTRAINT *VERSUS* FREEDOM SPECIFICATION OF THE MODEL

Aitchison and Silvey (1958) introduced a terminology that is useful for describing model specification methods. The model spaces $\omega^{(M)}$ and $\omega^{(P)}$ of equations (2.1) and (2.2) are said to be specified using freedom equations. A motivation for this terminology is as follows. The space $\omega^{(P)}$ comprises all those vectors $\boldsymbol{\xi}$ that can be written as $\boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta}$ where the parameter $\boldsymbol{\beta}$ is completely unrestricted in the sense that each component of $\boldsymbol{\beta}$ is free to take on any value on the real line. It is for this reason that the parameters in $\boldsymbol{\beta}$ are called freedom parameters and the model specification is called the freedom equation specification. For distinctness, we shall refer to the parameters in $\boldsymbol{\xi}$ as model parameters.

There is an equivalent way to specify these models. It is known as the constraint equation specification. Requiring that $\boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta}$ for some unrestricted $\boldsymbol{\beta}$ is tantamount to requiring $\boldsymbol{\xi}$ to fall in the range space of $\mathbf{X}$, denoted $R(\mathbf{X})$. This in turn is equivalent to requiring $\boldsymbol{\xi}$ to fall in the null space of a matrix $\mathbf{U}'$ where the columns of $\mathbf{U}$ span the space that is orthogonal to $R(\mathbf{X})$, namely $R(\mathbf{X})^{\perp}$. It follows that the model spaces $\omega^{(M)}$ and $\omega^{(P)}$ can be specified equivalently as

$$\omega^{(M)} = \{\boldsymbol{\xi}\colon \mathbf{U}'\boldsymbol{\xi} = \mathbf{0},\ \text{samp}(\boldsymbol{\xi}) = \mathbf{0}\}, \tag{3.1}$$

$$\omega^{(P)} = \{\boldsymbol{\xi}\colon \mathbf{U}'\boldsymbol{\xi} = \mathbf{0}\}, \tag{3.2}$$

where $\mathbf{U}$ satisfies $\mathbf{U}'\mathbf{X} = \mathbf{0}$ and $R(\mathbf{U}) = R(\mathbf{X})^{\perp}$.

The following non-restrictive model assumption $\mathcal{A}1$ will be sufficient for the equivalences between Poisson and multinomial log-linear models to hold:

$$R(\mathbf{X}) \supseteq R(\oplus_{k=1}^{K} \mathbf{1}_r).$$

For convenience, we shall also assume that $\mathbf{X}$ is $s \times p$ of full column rank p (assumption $\mathcal{A}2$) and $\mathbf{U}$ is $s \times (s-p) \equiv s \times u$ of full column rank u (assumption $\mathcal{A}3$).

Assumpion $\mathcal{A}1$ is satisfied whenever a parameter is included for each of the K independent multinomials, i.e. the fixed-by-design parameters are included. This assumption implies that the sufficient conditions for equality of point estimates under the two sampling schemes are met (Birch, 1963).

Requiring $\mathbf{X}$ and $\mathbf{U}$ to be of full column rank ($\mathcal{A}2$ and $\mathcal{A}3$) is not necessary.

However, it is convenient to assume this for expository reasons, as generalized inverses can be avoided. When all three assumptions $\mathcal{A}1$, $\mathcal{A}2$ and $\mathcal{A}3$ hold, we shall simply say that assumption $\mathcal{A}$ holds.

4. MULTINOMIAL LOG-LINEAR MODELS

4.1. *Restricted Likelihood Equations*

Our first objective is to find the product multinomial restricted maximum likelihood estimate of $\boldsymbol{\xi}$. The product multinomial log-likelihood is

$$\sum_k \sum_j y_{kj} \log \pi_{kj} + \sum_k \log \begin{pmatrix} & n_k & \\ y_{k1} & \ldots & y_{kr} \end{pmatrix} = \sum_k \sum_j y_{kj} \log \mu_{kj} + c = \sum_k \sum_j y_{kj} \xi_{kj} + c,$$

where c is a constant with respect to $\boldsymbol{\pi}$. Thus, the kernel of the log-likelihood is, using vector notation, $l^{(M)}(\boldsymbol{\xi}; \mathbf{y}) = \mathbf{y}'\boldsymbol{\xi}$. Hence, the maximum likelihood estimator is the solution $\hat{\boldsymbol{\xi}}$ to

$$\sup_{\boldsymbol{\xi} \in \omega^{(M)}} \{ l^{(M)}(\boldsymbol{\xi}; \mathbf{y}) \} = \sup_{\boldsymbol{\xi} \in \omega^{(M)}} (\mathbf{y}'\boldsymbol{\xi}) = \mathbf{y}'\hat{\boldsymbol{\xi}}. \tag{4.1}$$

We are to maximize this function $l^{(M)}$ subject to the constraints as specified in equation (3.1). As did Aitchison and Silvey (1958), we shall use Lagrange's method of undetermined multipliers to solve equation (4.1). In particular, let $\boldsymbol{\tau}$ be the $K \times 1$ vector of undetermined multipliers corresponding to samp$(\boldsymbol{\xi}) = \mathbf{0}$ and let $\boldsymbol{\lambda}$ be the $u \times 1$ vector of undetermined multipliers corresponding to the constraint $\mathbf{U}'\boldsymbol{\xi} = \mathbf{0}$.

The solution vec$(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\tau}})$ to the restricted likelihood equations

$$\begin{pmatrix} \mathbf{y} + \mathbf{D}_{\mathrm{e}^{\hat{\xi}}}(\oplus_{k=1}^{K} \mathbf{1}_r)\hat{\boldsymbol{\tau}} + \mathbf{U}\hat{\boldsymbol{\lambda}} \\ \mathbf{U}'\hat{\boldsymbol{\xi}} \\ (\oplus_{k=1}^{K} \mathbf{1}'_r)\mathrm{e}^{\hat{\xi}} - (n_1, \ldots, n_K)' \end{pmatrix} = \mathbf{0}$$

is the maximum likelihood estimate of $(\boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\tau})$. Here, $\mathbf{D}_\mathbf{x}$ is the diagonal matrix with elements in $\mathbf{x}$ on the diagonal. Assuming that assumption $\mathcal{A}1$ holds, we can solve for $\hat{\boldsymbol{\tau}}$ explicitly. Premultiplying the first equation by $(\oplus_{k=1}^{K} \mathbf{1}'_r)$, we obtain

$$\mathbf{0} = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_K \end{pmatrix} + \mathrm{diag}(n_1, \ldots, n_K)\hat{\boldsymbol{\tau}} + (\oplus_{k=1}^{K} \mathbf{1}'_r)\mathbf{U}\hat{\boldsymbol{\lambda}} = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_K \end{pmatrix} + \mathrm{diag}(n_1, \ldots, n_K)\hat{\boldsymbol{\tau}},$$

since, by assumption $\mathcal{A}1$, $\oplus_{k=1}^{K} \mathbf{1}_r = \mathbf{XB}$ for some matrix $\mathbf{B}$ and $\mathbf{U}'\mathbf{X} = \mathbf{0}$. It is this orthogonality of the Jacobians that allows for simple Poisson–multinomial comparisons. It follows that $\hat{\boldsymbol{\tau}} = -\mathbf{1}_K$. Thus, we can find vec$(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}})$ by solving the reduced set of equations

$$\begin{pmatrix} \mathbf{y} - \mathrm{e}^{\hat{\xi}} + \mathbf{U}\hat{\boldsymbol{\lambda}} \\ \mathbf{U}'\hat{\boldsymbol{\xi}} \end{pmatrix} = \mathbf{0}. \tag{4.2}$$

Lang and Agresti (1994) outlined a simple iterative scheme for solving equations (4.2). The algorithm can be used for a more general class of models as well.

4.2. *Asymptotic Behaviour of Multinomial Estimators*

We shall explore the large sample behaviour of $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\beta}}$ under the assumption that the model $[\omega^{(M)}]$ truly does hold. Let the symbol $n = \Sigma_{k=1}^{K} n_k$ represent the total sample size and the matrix $\mathbf{N} = \oplus_{k=1}^{K} n_k \mathbf{I}_r$, where $\mathbf{I}_r$ is the $r \times r$ identity matrix. All the asymptotics will hold as $n \to \infty$ in such a way that $n^{-1}\mathbf{N} \to \mathbf{W}$, where the matrix $\mathbf{W} = \oplus_{k=1}^{K} w_k \mathbf{I}_r$, and $0 < w_k \leqslant 1$, $k = 1, \ldots, K$, i.e. the sample sizes are assumed to grow large at approximately the same rate.

To describe the asymptotic behaviour of $\mathrm{vec}(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}})$, we begin by noting that $\mathrm{vec}(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}})$ is the solution to equations (4.2) if and only if it is the solution to

$$\begin{pmatrix} n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\hat{\boldsymbol{\xi}}}) + n^{-1/2}\mathbf{U}\hat{\boldsymbol{\lambda}} \\ n^{1/2}\mathbf{U}'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \end{pmatrix} = \mathbf{0}, \tag{4.3}$$

since $\mathbf{U}'\boldsymbol{\xi} = \mathbf{0}$ for $\boldsymbol{\xi} \in \omega^{(M)}$.

Now, expanding the restricted likelihood equations (4.3) about $\mathrm{vec}(\boldsymbol{\xi}, \mathbf{0})$, we show in Appendix A that

$$\begin{pmatrix} n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ n^{-1/2}\hat{\boldsymbol{\lambda}} \end{pmatrix} \to \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Gamma}) \qquad \text{in distribution,}$$

where $\mathrm{MVN}(\mathbf{0}, \boldsymbol{\Gamma})$ represents a multivariate normal random vector with zero mean and variance $\boldsymbol{\Gamma}$, which can be written as

$$\begin{pmatrix} \mathbf{W}^{-1}\mathbf{D}_\pi^{-1} - \mathbf{W}^{-1}\mathbf{D}_\pi^{-1}\mathbf{U}(\mathbf{U}'\mathbf{D}_\pi^{-1}\mathbf{W}^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D}_\pi^{-1}\mathbf{W}^{-1} - (\oplus_{k=1}^{K}\mathbf{1}_r\mathbf{1}_r')\mathbf{W}^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{U}'\mathbf{D}_\pi^{-1}\mathbf{W}^{-1}\mathbf{U})^{-1} \end{pmatrix}.$$

The upper left block of $\boldsymbol{\Gamma}$ is the asymptotic variance of $n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})$. It can be rewritten as

$$\lim_{n\to\infty} [n\{\mathbf{D}_\mu^{-1} - \mathbf{D}_\mu^{-1}\mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D}_\mu^{-1} - \oplus_{k=1}^{K}\mathbf{1}_r\mathbf{1}_r'/n_k\}],$$

since $\lim_{n\to\infty}(n\mathbf{D}_\mu^{-1}) = \mathbf{W}^{-1}\mathbf{D}_\pi^{-1}$ and $n\mathbf{N}^{-1} = \mathbf{W}^{-1} + o(1)$. The following result summarizes the behaviour of several multinomial estimators.

Result 1. The asymptotic variances of the multinomial estimators, denoted by $\boldsymbol{\Gamma}$, are

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\xi}}) = \mathbf{D}_\mu^{-1} - \mathbf{D}_\mu^{-1}\mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D}_\mu^{-1} - \oplus_{k=1}^{K}\mathbf{1}_r\mathbf{1}_r'/n_k, \tag{4.4}$$

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}(\hat{\boldsymbol{\xi}})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{4.5}$$

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\mu}}) = \mathbf{D}_\mu - \mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}' - \oplus_{k=1}^{K}\boldsymbol{\mu}_k\boldsymbol{\mu}_k'/n_k, \tag{4.6}$$

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\pi}}) = \mathbf{N}^{-1}\boldsymbol{\Gamma}(\hat{\boldsymbol{\mu}})\mathbf{N}^{-1}. \tag{4.7}$$

Notice that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\mu}} = \mathrm{e}^{\hat{\boldsymbol{\xi}}}$ and $\hat{\boldsymbol{\pi}} = \mathbf{N}^{-1}\hat{\boldsymbol{\mu}}$. Therefore, the last three results can be shown by using the delta method.

The asymptotic variances of result 1 can be used to understand better what effect parsimonious modelling has on variance estimates. For instance, in a 2×2 table, suppose that there is full multinomial sampling ($K = 1$). For the multinomial log-linear model of independence, the matrix $\mathbf{U}' = (1, -1, -1, 1)$. By equations (4.6) and (4.7), the asymptotic variance of $\hat{\boldsymbol{\pi}}$ can be written as

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\pi}}) = \frac{1}{n}(\mathbf{D}_\pi - \boldsymbol{\pi}\boldsymbol{\pi}') - \frac{1}{n^2\,\mathrm{avar}(\log\bar{\theta})}\begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix},$$

where $\bar{\theta}$ is the sample odds ratio and

$$\mathrm{avar}(\log\bar{\theta}) = \sum_{i=1}^{2}\sum_{j=1}^{2} \mu_{ij}^{-1}$$

is the asymptotic variance of $\log\bar{\theta}$. Evidently, the asymptotic variance of the independence model estimator $\hat{\boldsymbol{\pi}}$ can be obtained by subtracting the matrix involving $\mathrm{avar}(\bar{\theta})$ from the saturated model variance estimate. More generally, since the matrix $\mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}'$ in equation (4.6) is non-negative definite, we immediately see that good fitting parsimonious models, those models with many constraints, produce estimators with smaller variances.

5. POISSON LOG-LINEAR MODELS

5.1. *Restricted Likelihood Equations*

The Poisson maximum likelihood estimate of $\boldsymbol{\xi}$ is the solution $\tilde{\boldsymbol{\xi}}$ to

$$\sup_{\boldsymbol{\xi}\in\omega^{(P)}} \{l^{(P)}(\boldsymbol{\xi}; \mathbf{y})\} = \sup_{\boldsymbol{\xi}\in\omega^{(P)}} (\mathbf{y}'\boldsymbol{\xi} - \mathrm{e}^{\boldsymbol{\xi}'}\mathbf{1}_s) = \mathbf{y}'\tilde{\boldsymbol{\xi}} - \mathrm{e}^{\tilde{\boldsymbol{\xi}}'}\mathbf{1}_s. \tag{5.1}$$

As in the multinomial setting, we solve equation (5.1) by using Lagrange's method of undetermined multipliers. For Poisson log-linear models, we need only to introduce the $u \times 1$ vector of undetermined multipliers $\boldsymbol{\lambda}$ corresponding to the constraint $\mathbf{U}'\boldsymbol{\xi} = \mathbf{0}$; there are no sampling constraints.

The solution $\mathrm{vec}(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\lambda}})$ to the restricted likelihood equations

$$\begin{pmatrix} \mathbf{y} - \mathrm{e}^{\tilde{\boldsymbol{\xi}}} + \mathbf{U}\tilde{\boldsymbol{\lambda}} \\ \mathbf{U}'\tilde{\boldsymbol{\xi}} \end{pmatrix} = \mathbf{0} \tag{5.2}$$

is the maximum likelihood estimate.

5.2. *Asymptotic Behaviour of Poisson Estimators*

We explore the large sample behaviour of $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\beta}}$ under the assumption that the model $[\omega^{(P)}]$ truly does hold. Let the symbol $\mu_+ = \Sigma_k \Sigma_j \mu_{kj}$. The asymptotics will hold as $\mu_+ \to \infty$ in such a way that $\mu_+^{-1}\mathbf{D}_\mu \to \mathbf{V}$, where the matrix $\mathbf{V} = \mathrm{diag}(v_{11}, \ldots, v_{Kr})$ and $0 < v_{kj} < 1$, $k = 1, \ldots, K$, $j = 1, \ldots, r$. Thus, all the expected cell counts are assumed to grow large at approximately the same rate.

As in the multinomial setting, we expand a properly standardized version of the restricted likelihood equations (5.2) about $\mathrm{vec}(\boldsymbol{\xi}, \mathbf{0})$ in a Taylor expansion. Using arguments that are analogous to those used to describe the asymptotic behaviour of product multinomial estimators, it is straightforward to show that

$$\begin{pmatrix} \mu_+^{1/2}(\tilde{\xi}-\xi) \\ \mu_+^{-1/2}\tilde{\lambda} \end{pmatrix} \rightarrow \mathrm{MVN}(\mathbf{0}, \mathbf{\Gamma}_\mathrm{P}) \qquad \text{in distribution,}$$

where

$$\mathbf{\Gamma}_\mathrm{P} = \begin{pmatrix} \mathbf{V} & -\mathbf{U} \\ -\mathbf{U}' & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V} & -\mathbf{U} \\ -\mathbf{U}' & \mathbf{0} \end{pmatrix}^{-1},$$

which simplifies to

$$\mathbf{\Gamma}_\mathrm{P} = \begin{pmatrix} \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{U}(\mathbf{U}'\mathbf{V}^{-1}\mathbf{U})^{-1}\mathbf{U}\mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{U}'\mathbf{V}^{-1}\mathbf{U})^{-1} \end{pmatrix}.$$

The upper left-hand block in this expression is the asymptotic variance of $\mu_+^{1/2}(\tilde{\xi}-\xi)$. It can be rewritten as

$$\lim_{\mu_+\rightarrow\infty} [\mu_+\{\mathbf{D}_\mu^{-1} - \mathbf{D}_\mu^{-1}\mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D}_\mu^{-1}\}],$$

since $\lim_{\mu_+\rightarrow\infty}(\mu_+\mathbf{D}_\mu^{-1}) = \mathbf{V}^{-1}$. The following result summarizes the asymptotic behaviour of several Poisson estimators.

Result 2. The asymptotic variances of the Poisson estimators, denoted by $\mathbf{\Gamma}$, are

$$\mathbf{\Gamma}(\tilde{\xi}) = \mathbf{D}_\mu^{-1} - \mathbf{D}_\mu^{-1}\mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D}_\mu^{-1}, \tag{5.3}$$

$$\mathbf{\Gamma}(\tilde{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}(\tilde{\xi})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{5.4}$$

$$\mathbf{\Gamma}(\tilde{\mu}) = \mathbf{D}_\mu - \mathbf{U}(\mathbf{U}'\mathbf{D}_\mu^{-1}\mathbf{U})^{-1}\mathbf{U}'. \tag{5.5}$$

Similarly to result 1, the last two results can be obtained by using the delta method.

6. COMPARISON: MULTINOMIAL *VERSUS* POISSON LOG-LINEAR MODELS

We begin by comparing point estimates—multinomial *versus* Poisson. Equations (4.2) and (5.2) evidently give rise to the same solutions, i.e. $\hat{\xi} = \tilde{\xi}$ and $\hat{\lambda} = \tilde{\lambda}$. These numerical equivalences occur because assumption $\mathcal{A}1$ implies the sufficient conditions of Birch (1963), namely, as long as the Poisson fitted values $\mu_{kj}(\tilde{\beta})$ satisfy the multinomial sampling constraints, the multinomial and Poisson estimates are identical. It also follows that the freedom parameter estimates are identical (i.e. $\hat{\beta} = \tilde{\beta}$). This follows since $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\xi} = \tilde{\beta}$.

Comparing results 1 and 2, it is immediately obvious that the following result holds.

Result 3. Assuming that model assumption $\mathcal{A}$ holds, the asymptotic variances for the multinomial and Poisson estimators are related according to

$$\Gamma(\hat{\xi}) = \Gamma(\tilde{\xi}) - \oplus_{k=1}^{K} \mathbf{1}_r\mathbf{1}_r'/\, n_k, \tag{6.1}$$

$$\Gamma(\hat{\mu}) = \Gamma(\tilde{\mu}) - \oplus_{k=1}^{K} \mu_k\mu_k'/n_k, \tag{6.2}$$

$$\Gamma(\hat{\pi}) = \mathbf{N}^{-1}\Gamma(\tilde{\mu})\mathbf{N}^{-1} - \mathbf{N}^{-1}(\oplus_{k=1}^{K} \mu_k\mu_k'/n_k)\mathbf{N}^{-1}, \tag{6.3}$$

$$\Gamma(\hat{\beta}) = \Gamma(\tilde{\beta}) - \Lambda, \tag{6.4}$$

where the $p \times p$ matrix Λ is non-negative definite and has the form

$$\Lambda = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\oplus_{k=1}^{K} \mathbf{1}_r\mathbf{1}_r'/n_k)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{6.5}$$

Remark 1. By relationship (6.1), the asymptotic variance of any linear contrast of log-means within a covariate level is the same under both the multinomial and the Poisson models. This follows since a linear contrast of log-means within level k of the covariate has form $\mathbf{c}_k'\xi$ where $\mathbf{c}_k' = (0, 0, \ldots, 0, c_{k1}, \ldots, c_{kr}, 0, \ldots, 0)$ with c_{k1} in the $(k-1)r+1$ position and $\Sigma_{j=1}^{r} c_{kj} = 0$. Thus, by relationship (6.1),

$$\begin{aligned} \operatorname{avar}(\mathbf{c}_k'\hat{\xi}) &= \mathbf{c}_k'\Gamma(\hat{\xi})\mathbf{c}_k \\ &= \mathbf{c}_k'\Gamma(\tilde{\xi})\mathbf{c}_k - \mathbf{c}_k'(\oplus_{k=1}^{K} \mathbf{1}_r\mathbf{1}_r'/n_k)\mathbf{c}_k \\ &= \mathbf{c}_k'\Gamma(\tilde{\xi})\mathbf{c}_k - 0 \\ &= \operatorname{avar}(\mathbf{c}_k'\tilde{\xi}). \end{aligned}$$

Therefore, inferences about measures like odds and odds ratios are the same for both models. In contrast, inferences about such measures as relative risk and differences between probabilities will be different for these two models.

Remark 2. To illustrate how important these differences between the models can be, we consider the following: suppose that there is full multinomial sampling ($K=1$). Our goal is to estimate $\mu'\mathbf{1}_s$, the sum of the expected counts. For full multinomial sampling, we know that $\hat{\mu}'\mathbf{1}_s = n$ with probability 1. Since n is considered fixed for multinomial sampling, our variance estimate of $\hat{\mu}'\mathbf{1}_s$ should reflect this; it should be 0. With Poisson sampling our point estimate is the same, but the variance of our estimate is significantly inflated. In fact, $\widehat{\operatorname{avar}}(\tilde{\mu}'\mathbf{1}_s) = \tilde{\mu}'\mathbf{1}_s = n$. Notice that by relationship (6.2) we can easily adjust the Poisson variance estimate so that it equals the multinomial variance estimate, i.e.

$$\widehat{\operatorname{avar}}(\hat{\mu}'\mathbf{1}_s) = \mathbf{1}_s'\tilde{\Gamma}(\tilde{\mu})\mathbf{1}_s - \mathbf{1}_s'\,\frac{\hat{\mu}\hat{\mu}'}{n}\,\mathbf{1}_s = n - n = 0.$$

Remark 3. Relationship (6.3) is of practical importance. By that relationship, we can draw valid inference about the multinomial cell probabilities, even when a Poisson log-linear model is fitted. We simply subtract the quantity $\mathbf{N}^{-1}(\oplus_{k=1}^{K} \mu_k\mu_k'/n_k)\mathbf{N}^{-1}$, or an estimate thereof, from the Poisson variance estimator $\mathbf{N}^{-1}\Gamma(\tilde{\mu})\mathbf{N}^{-1}$, an estimator that can easily be obtained by using standard generalized linear model fitting methods such as iterative reweighted least squares (McCullagh and Nelder, 1989; Aitkin *et al.*, 1989).

Remark 4. With regard to relationship (6.4), it is of practical importance to know which of the components in Λ are 0. For, if they are 0, inference about the corresponding freedom parameters will be the same for both sampling schemes. Palmgren (1981) investigated this relationship by using an alternative approach. We end this section with a more thorough investigation of the form of Λ.

Notice that relationship (6.5) can be written as

$$\Lambda = \{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\oplus_{k=1}^{K} \mathbf{1}_r/\sqrt{n_k})\}\{(\oplus_{k=1}^{K}\mathbf{1}_r'/\sqrt{n_k})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\},$$

and that, since $R(\oplus_{k=1}^{K} \mathbf{1}_r) \subseteq R(\mathbf{X})$ by assumption $\mathcal{A}1$, we can write $(\oplus_{k=1}^{K} \mathbf{1}_r/\sqrt{n_k}) = \mathbf{XB}$ for some matrix $\mathbf{B}$. By assumption $\mathcal{A}1$, there must be a collection of T ($T \leqslant p$) columns in $\mathbf{X}$, say $\mathbf{X}^* = (\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_T})$, that spans a space containing $R(\oplus_{k=1}^{K} \mathbf{1}_r/\sqrt{n_k})$. Without loss of generality, we assume that this collection is a minimal spanning set in the sense that any smaller collection has span that does not contain $R(\oplus_{k=1}^{K} \mathbf{1}_r/\sqrt{n_k})$. It follows that the matrix $\mathbf{B}'$ can be written as $\mathbf{B}' = (\mathbf{b}_1 \ \mathbf{b}_2 \ \ldots \ \mathbf{b}_p)$, where $\mathbf{b}_i = \mathbf{0}_K$ if $i \notin \{i_1, \ldots, i_T\}$. This is so because $R(X^*)$ contains $R(\oplus_{k=1}^{K} \mathbf{1}_r/\sqrt{n_k})$, which implies that $(\oplus_{k=1}^{K} \mathbf{1}_r/\sqrt{n_k})$ can be written as a linear combination of the columns in $\mathbf{X}^*$ alone. Therefore, with this choice of $\mathbf{B}$, we have that $\Lambda = \mathbf{BB}'$ and the following result.

Result 4. Assuming that model assumption $\mathcal{A}$ holds, the asymptotic variances of the multinomial estimator $\hat{\beta}$ and the Poisson estimator $\tilde{\beta}$ are related according to

$$\Gamma(\hat{\beta}) = \Gamma(\tilde{\beta}) - \Lambda,$$

where Λ is non-negative definite with components satisfying

$$\Lambda_{ij} = 0, \qquad \text{if } (i, j) \notin \{i_1, \ldots, i_T\} \times \{i_1, \ldots, i_T\}.$$

The collection of subscripts $\{i_1, \ldots, i_T\}$ indexes the T columns of the design matrix $\mathbf{X}$ that span a set that contains $R(\oplus_{k=1}^{K} \mathbf{1}_r)$.

This result follows immediately on noting that if either i or j is not in the set $\{i_1, \ldots, i_T\}$ then either row $\mathbf{b}_i'$ or $\mathbf{b}_j'$ is the zero vector. Thus, the (i, j)th component of Λ, which has form $\mathbf{b}_i'\mathbf{b}_j$, must be 0.

Result 4 allows us to determine which parameter estimators, for the original parameterization, have large sample distributions which are invariant to sampling scheme, Poisson or product multinomial.

7. EXAMPLES

7.1. *Example 1*

Consider a prospective study whereby $n_1 = 50$ subjects are assigned treatment ($T = 1$) and $n_2 = 75$ subjects are assigned control ($T = 2$). After a certain length of time, the subjects are observed and their disease status determined: $D = 1$ means that the disease is present and $D = 2$ means that the disease is absent. The resulting data and corresponding distributions are shown in Table 1.

For this example, the number of independent multinomials, or covariate levels, is $K = 2$ and the number of response levels within each covariate is $r = 2$. The product multinomial log-linear model space corresponding to homogeneity (or, loosely, independence of T and D) can be specified as

$$\omega^{(M)} = \{\boldsymbol{\xi}\colon \xi_{kj} = \alpha + \alpha_k^T + \alpha_j^D,\ \alpha_1^T = \alpha_1^D = 0,\ \mathrm{samp}(\boldsymbol{\xi}) = \mathbf{0}\}$$
$$= \{\boldsymbol{\xi}\colon \mathbf{U}'\boldsymbol{\xi} = \mathbf{0},\ \mathrm{samp}(\boldsymbol{\xi}) = \mathbf{0}\},$$

where $\xi_{kj} = \log(n_k \pi_{kj})$, $k = 1, 2, j = 1, 2$. Here, the freedom parameter vector is $\boldsymbol{\beta}' = (\alpha, \alpha_2^T, \alpha_2^D)$, the matrix $\mathbf{U}' = (1, -1, -1, 1)$ and the design matrix $\mathbf{X}$ corresponding to the freedom parameter identifiability constraints ($\alpha_1^T = \alpha_1^D = 0$) is

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

The multinomial sampling constraint is

$$\mathrm{samp}(\boldsymbol{\xi}) = (\oplus_{k=1}^2 \mathbf{1}_2')\mathrm{e}^{\boldsymbol{\xi}} - (n_1, n_2)' = \begin{pmatrix} \mu_{1+} - 50 \\ \mu_{2+} - 75 \end{pmatrix} = \mathbf{0}.$$

Notice that the first two columns of $\mathbf{X}$ span a space that contains $R(\oplus_{k=1}^2 \mathbf{1}_2)$. In general, to check whether $R(\mathbf{A}) \subseteq R(\mathbf{B})$ use the fact that $R(\mathbf{A}) \subseteq R(\mathbf{B})$ if and only if $\{\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\}\mathbf{A} = \mathbf{0}$. Evidently, assumption $\mathcal{A}$ is satisfied and so by result 4

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Gamma}(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}_{ij} = 0$, if $(i, j) \notin \{1, 2\} \times \{1, 2\}$. Hence, $\boldsymbol{\Lambda}$ has the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} x & y & 0 \\ y & z & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where x, y and z are generally non-zero. We conclude that, if we had fitted the Poisson log-linear model instead of the correct multinomial log-linear model, our inferences about the parameter α_2^D would have remained the same. Also, if we had not included the α_2^T-term in our model the inferences, including point estimates, would have been different for the two models, since assumption $\mathcal{A}1$ would not be satisfied.

For these data we computed parameter estimators and their estimated variances under both sampling schemes. The results for the independence model are shown in Table 2.

TABLE 1
Cell counts and underlying distributions

T	*D* 1	*D* 2		*T*	*D* 1	*D* 2	
1	30	20	50	1	π_{11}	π_{12}	1.0
2	60	15	75	2	π_{21}	π_{22}	1.0
			125				

TABLE 2
Independence log-linear model estimates

Parameter	*Estimate*	*Multinomial standard error*	*Poisson standard error*
π_{11}	0.720	0.040	0.110
π_{21}	0.720	0.040	0.092
α	3.584	0.056	0.152
α_2^T	0.405	0.000	0.183
α_2^D	−0.944	0.199	0.199

TABLE 3
Covariance estimates for the independence model

	Multinomial covariance estimates			*Poisson covariance estimates*		
	α	α_2^T	α_2^D	α	α_2^T	α_2^D
α	0.0031			0.0231		
α_2^T	0.0000	0.0000		−0.0200	0.0333	
α_2^D	−0.0111	0.0000	0.0397	−0.0111	0.0000	0.0397

Table 3 contains the covariance estimates for the independence model freedom parameter estimators.

From Table 2, it is evident that the inference about the parameter α_2^D is the same for both sampling schemes; this we know to be the case from result 4. Table 3 shows that the last row and last column of the variance–covariance estimates are identical for the two sampling assumptions. This also must be the case by result 4.

7.2. *Example 2*

For the same data, if the saturated model $\{\boldsymbol{\xi}: \boldsymbol{\xi} = \boldsymbol{\beta}, \text{samp}(\boldsymbol{\xi}) = \mathbf{0}\}$ were fitted, we would have that $\mathbf{X} = \mathbf{I}_4$ and $\mathbf{U} = \mathbf{0}$. For this parameterization, all four columns of $\mathbf{X}$ are needed to span a set containing $R(\oplus_{k=1}^2 \mathbf{1}_2)$. Thus,

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Gamma}(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}_{ij} = 0$, if $(i, j) \notin \{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$, i.e. all elements in $\boldsymbol{\Lambda}$ are generally non-zero. We conclude that, if this saturated Poisson log-linear model is fitted, the inferences about all four parameters are different. In contrast, by the first remark following result 3, inferences about any linear contrast in $\boldsymbol{\beta}$ are the same for both sampling schemes.

A better, and more common, parameterization might be

$$\xi_{ij} = \alpha + \alpha_k^T + \alpha_j^D + \alpha_{kj}^{TD},$$

where $\alpha_1^T = \alpha_1^D = \alpha_{1j}^{TD} = \alpha_{k1}^{TD} = 0$. Corresponding to the identifiable freedom parameters in $\boldsymbol{\beta}' = (\alpha, \alpha_2^T, \alpha_2^D, \alpha_{22}^{TD})$ is the design matrix

$$\mathbf{X}=\begin{pmatrix}1&0&0&0\\1&0&1&0\\1&1&0&0\\1&1&1&1\end{pmatrix}.$$

In this case, the first two columns of $\mathbf{X}$ span $R(\oplus_{k=1}^{2}\mathbf{1}_2)$ so that the adjustment matrix $\boldsymbol{\Lambda}$ has the form

$$\boldsymbol{\Lambda}=\begin{pmatrix}x&y&0&0\\y&z&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}.$$

and the inferences about both α_2^D and α_{22}^{TD} are the same under both sampling assumptions.

8. DISCUSSION

We have outlined an alternative approach to showing the equivalence between product multinomial and Poisson log-linear models. Using constraint equations to formulate the model, we have derived the asymptotic distributions of several estimators of interest. This alternative approach, which is based on the seminal papers by Aitchison and Silvey (1958, 1960), has several advantages over the original approach (Birch, 1963; Palmgren, 1981).

A comparison of the large sample behaviour of Poisson and product multinomial estimators was carried out. With this approach, adjustment matrices are computed as a by-product. The adjustment matrices enable us to modify the Poisson variance estimates so that they match product multinomial variance estimates. Therefore, valid inferences about multinomial cell probabilities can be conducted by using the Poisson log-linear model along with the adjustments. Also, the asymptotic covariances have intuitive forms in that we can readily see the effect of parsimonious modelling.

For pedagogical reasons, this approach has some advantages over the original approach. As illustrated by example, we can immediately see the effect of different freedom parameterizations; it is simple to determine when inferences about different freedom parameters will be invariant with respect to the sampling scheme.

An important feature of this method, which does not require reparameterization of the likelihood in terms of freedom parameters, is that we can extend these equivalence results to non-standard models of the form $\mathbf{C}\log(\mathbf{A}\boldsymbol{\mu})=\mathbf{X}\boldsymbol{\beta}$ where $\mathbf{C}$ and $\mathbf{A}$ are known conformable matrices. See Lang and Agresti (1994) for a discussion of the applicability of these generalized log-linear models.

Finally, we have not discussed model assessment statistics such as goodness-of-fit statistics or adjusted residuals in this paper. The well-known equivalences, e.g. χ^2 goodness-of-fit statistics, are the same for either sampling scheme and can be shown by using our method of proof. Using this method, it is also easy to see that, for models satisfying assumption $\mathcal{A}$, the score statistic, the Lagrange multiplier statistic (see Aitchison and Silvey (1958)) and the Pearson χ^2-statistic are identical.

ACKNOWLEDGEMENTS

The author would like to thank two referees for their helpful comments and suggestions.

APPENDIX A

Using the method of Aitchison and Silvey (1958), it can be shown that the solution $\text{vec}(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}})$ is consistent in the following sense: $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} = O_P(n^{-1/2})$, $\hat{\boldsymbol{\lambda}} = O_P(n^{1/2})$. Expanding $n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\hat{\boldsymbol{\xi}}})$ about $\boldsymbol{\xi}$ in a Taylor expansion we obtain

$$\begin{aligned} n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\hat{\boldsymbol{\xi}}}) &= n^{-1/2}\{\mathbf{Y} - \mathrm{e}^{\boldsymbol{\xi}} - \mathbf{D}_{\mathrm{e}^{\boldsymbol{\xi}}}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) + O_P(1)\} \\ &= n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\boldsymbol{\xi}}) - n^{-1}\mathbf{D}_{\mathrm{e}^{\boldsymbol{\xi}}} n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) + O_P(n^{-1/2}) \\ &= n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\boldsymbol{\xi}}) - \mathbf{W}\mathbf{D}_{\boldsymbol{\pi}} n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) + O_P(n^{-1/2}). \end{aligned}$$

Thus, we can approximate the likelihood equations as follows:

$$\begin{aligned} \mathbf{0} &= \begin{pmatrix} n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\hat{\boldsymbol{\xi}}}) + n^{-1/2}\mathbf{U}\hat{\boldsymbol{\lambda}} \\ n^{1/2}\mathbf{U}'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \end{pmatrix} \\ &= \begin{pmatrix} n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\boldsymbol{\xi}}) \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} -\mathbf{W}\mathbf{D}_{\boldsymbol{\pi}} & \mathbf{U} \\ \mathbf{U}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ n^{-1/2}\hat{\boldsymbol{\lambda}} \end{pmatrix} + o_P(1), \end{aligned}$$

i.e.

$$\begin{pmatrix} n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\boldsymbol{\xi}}) \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{W}\mathbf{D}_{\boldsymbol{\pi}} & -\mathbf{U} \\ -\mathbf{U}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ n^{-1/2}\hat{\boldsymbol{\lambda}} \end{pmatrix} + o_P(1).$$

Now, by standard asymptotic arguments for multinomial random variables,

$$n^{-1/2}(\mathbf{Y} - \mathrm{e}^{\boldsymbol{\xi}}) \rightarrow \text{MVN}\{\mathbf{0}, \mathbf{W}(\mathbf{D}_{\boldsymbol{\pi}} - \oplus_{k=1}^{K} \boldsymbol{\pi}_k\boldsymbol{\pi}_k')\} \qquad \text{in distribution.} \tag{A.1}$$

Therefore, by an application of Slutsky's theorem and the delta method,

$$\begin{pmatrix} n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ n^{-1/2}\hat{\boldsymbol{\lambda}} \end{pmatrix} \rightarrow \text{MVN}(\mathbf{0}, \boldsymbol{\Gamma}) \qquad \text{in distribution,}$$

where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{W}\mathbf{D}_{\boldsymbol{\pi}} & -\mathbf{U} \\ -\mathbf{U}' & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{W}(\mathbf{D}_{\boldsymbol{\pi}} - \oplus_{k=1}^{K} \boldsymbol{\pi}_k\boldsymbol{\pi}_k') & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{W}\mathbf{D}_{\boldsymbol{\pi}} & -\mathbf{U} \\ -\mathbf{U}' & \mathbf{0} \end{pmatrix}^{-1}.$$

Following arguments of Aitchison and Silvey (1958), and using the fact that

$$(\oplus_{k=1}^{K} \boldsymbol{\pi}_k')\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{U} = (\oplus_{k=1}^{K} \mathbf{1}_r')\mathbf{U} = \mathbf{0},$$

it is straightforward to show that $\boldsymbol{\Gamma}$ can be rewritten as

$$\begin{pmatrix} \mathbf{W}^{-1}\mathbf{D}_{\boldsymbol{\pi}}^{-1} - \mathbf{W}^{-1}\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{U}(\mathbf{U}'\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{W}^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{W}^{-1} - (\oplus_{k=1}^{K} \mathbf{1}_r\mathbf{1}_r')\mathbf{W}^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{U}'\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{W}^{-1}\mathbf{U})^{-1} \end{pmatrix}.$$

REFERENCES

Agresti, A. (1990) *Categorical Data Analysis*. New York: Wiley.

Aitchison, J. and Silvey, S. D. (1958) Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.*, **29**, 813–828.

——(1960) Maximum-likelihood estimation procedures and associated tests of significance. *J. R. Statist. Soc.* B, **22**, 154–171.

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) *Statistical Modelling in GLIM*. Oxford: Clarendon.

Birch, M. W. (1963) Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc.* B, **25**, 220–233.

Bishop, Y., Fienberg, S. E. and Holland, P. (1975) *Discrete Multivariate Analysis*. Cambridge: Massachusetts Institute of Technology Press.

Lang, J. B. and Agresti, A. (1994) Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *J. Am. Statist. Ass.*, **89**, 625–632.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Palmgren, J. (1981) The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika*, **68**, 563–566.