A MAXIMUM LIKELIHOOD ESTIMATION PROCEDURE FOR BINARY DATA FROM CROSS-OVER STUDIES

G.B. Matthews
School of Statistics and Actuarial Science, University of KwaZulu-Natal, Durban, 4001, South Africa
and
N.A.S. Crowther
Department of Statistics, University of Pretoria, Pretoria, 2000, South Africa

Key words: Binary cross-over design; constraints; estimation; loglinear models; log-nonlinear models.

Summary: A maximum likelihood estimation procedure is presented for the expected frequencies when modelling binary data from two-period cross-over studies. This procedure provides a simple approach to parameter estimation in models for binary data from cross-over experiments, and in particular for a log-nonlinear model proposed by Becker and Balagtas (1993).

1. Introduction

A cross-over trial is a clinical trial in which the patients receive different treatments in each of a number of periods. The simplest design is the 2×2 cross-over design, where each patient/subject receives two different treatments A and B. Subjects either receive A in the first period followed by B in the second period, or B in the first period followed by A in the second.

The two periods of administration are separated by a "washout" period to enable the subject's condition to return to a level uninfluenced by the previous treatment. This is to reduce the chance of the effect of the first treatment

AMS: 62F10

being carried over into the second period. In a binary cross-over experiment the response observed on each subject at the end of each period can have only one of two possible outcomes, usually labelled 0 or 1, hence a binary response variable. An example of such an experiment is the administering of two different pain killers, say A and B and the outcome experienced by the subject is either "relief" from pain or "no relief" from pain. The pain killers can be administered in the order A first followed by a washout period and then B or in the order B first followed by the washout period and then A. Kenward and Jones (1987), discuss the 2×2 cross-over design and use a loglinear and logit model to describe the data. Becker and Balagtas (1993) propose a model which they call a log-nonlinear model, in which the marginal logits and the log-odds ratios are modelled as a linear model. For the models where no closed-form ML estimates exist, ML estimation of the parameters is not straightforward. A nonlinear maximization routine is required. Becker and Balagtas (1993) show how PROC NONLIN in SAS may be used to find the ML estimates.

The purpose of this paper is to show how the ML estimation procedure of Matthews and Crowther (1995) provides a simple method for estimation in the log-nonlinear and loglinear model for binary cross-over data. In Section 2 the ML estimation procedure of Matthews and Crowther (1995) is presented as it provides the framework for finding the ML estimates for the parameters. This procedure provides the ML estimate for the mean vector of the multivariate exponential family when modelling in terms of constraints on the elements of the mean vector. Details of the proofs and other applications may be found in the latter reference. The models proposed by Becker and Balagtas (1993) are discussed in Section 3 and we show how these models can be written in terms of their implied constraints in order to provide a function $\mathbf{g}(\mathbf{F}) = \mathbf{0}$ as

required by Proposition 1. Section 3 also shows how the multivariate delta method can be used to find the variance covariance matrix of the vector of parameter estimates. An example is given to illustrate the application of the estimation procedure and confirm the results obtained by Becker and Balagtas (1993). In Section 4 we show how the ML estimation procedure can be used for loglinear models and Section 5 is the conclusion.

2. A maximum likelihood estimation procedure

We first define the multivariate exponential family and then describe the ML estimation procedure.

Let $\mathbf{y}$ be a $p \times 1$ random vector and $\boldsymbol{\theta}$ be a $p \times 1$ vector of parameters. The probability function for the exponential family is given by

$$\begin{aligned} p(\mathbf{y}, \boldsymbol{\theta}) &= a(\boldsymbol{\theta}) b(\mathbf{y}) \exp(\mathbf{y}'\boldsymbol{\theta}) \ , \quad \mathbf{y} \in \mathbb{R}^p \ , \quad \boldsymbol{\theta} \in \mathcal{N} \\ &= b(\mathbf{y}) \exp\{\mathbf{y}'\boldsymbol{\theta} - \kappa(\boldsymbol{\theta})\} \ , \end{aligned} \tag{2.1}$$

where $\kappa(\boldsymbol{\theta}) = -\ln a(\boldsymbol{\theta})$ and $\mathcal{N}$ is the natural parameter space for the canonical parameter $\boldsymbol{\theta}$. The mean vector and covariance matrix of $\mathbf{y}$ are given by

$$E(\mathbf{y}) = \boldsymbol{\mu} = \frac{\partial \kappa(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{ and } \mathrm{Cov}(\mathbf{y}) = \frac{\partial^2 \kappa(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \left\{ \frac{\partial^2 \kappa(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\} = \mathbf{V}_\mu \ .$$

The following proposition proved in Matthews and Crowther (1995), provides an ML estimation procedure which gives the estimate of $\boldsymbol{\mu}$ subject to the constraints $\mathbf{g}(\boldsymbol{\mu}) = \mathbf{0}$.

Proposition 1.

Consider a random vector $\mathbf{y}$, with probability function belonging to the exponential family. Let $\mathbf{g}(\boldsymbol{\mu})$ be a continuous vector valued function of $\boldsymbol{\mu}$, for which the first order partial derivatives exist. Let $\mathbf{G}_\mu = \dfrac{\partial \mathbf{g}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$ be the derivative of $\mathbf{g}(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ and $\mathbf{G}_y = \left. \dfrac{\partial \mathbf{g}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \right|_{\boldsymbol{\mu}=\mathbf{y}}$.

The ML estimate of $\boldsymbol{\mu}$ subject to the constraints $\mathbf{g}(\boldsymbol{\mu}) = \mathbf{0}$, is given by

$$\widehat{\boldsymbol{\mu}}_c = \mathbf{y} - (\mathbf{G}_\mu \mathbf{V})'(\mathbf{G}_y \mathbf{V} \mathbf{G}'_\mu)^{-1}\mathbf{g}(\mathbf{y}) + o(||\mathbf{y} - \boldsymbol{\mu}||) \; . \tag{2.2}$$

This result implies that the ML estimate must, in general, be obtained iteratively.

It is also useful to have the asymptotic covariance matrix of the ML estimate, $\widehat{\boldsymbol{\mu}}_c$. This result is found by making use of the multivariate delta method and is also proved in the above-mentioned reference.

Proposition 2.

The asymptotic covariance matrix of $\widehat{\boldsymbol{\mu}}_c$ is given by

$$\sum\nolimits_c = \mathbf{V}_\mu - (\mathbf{G}_\mu \mathbf{V}_\mu)'(\mathbf{G}_\mu \mathbf{V}_\mu \mathbf{G}'_\mu)^{-1}\mathbf{G}_\mu \mathbf{V}_\mu \; . \tag{2.3}$$

The estimated asymptotic covariance matrix of $\widehat{\boldsymbol{\mu}}_c$ is

$$\widehat{\sum\nolimits_c} = \mathbf{V}_{\widehat{\mu}_c} - (\mathbf{G}_{\widehat{\mu}_c} \mathbf{V}_{\widehat{\mu}_c})'(\mathbf{G}_{\widehat{\mu}_c} \mathbf{V}_{\widehat{\mu}_c} \mathbf{G}'_{\widehat{\mu}_c})^{-1}\mathbf{G}_{\widehat{\mu}_c} \mathbf{V}_{\widehat{\mu}_c} \; ,$$

where $\widehat{\boldsymbol{\mu}}_c$ is obtained on convergence of the iterative procedure in (2.2) and $\mathbf{V}_{\widehat{\mu}_c}$ is $\mathbf{V}_\mu$ evaluated at $\widehat{\boldsymbol{\mu}}_c$.

Note: It is implicitly assumed that the restrictions $\mathbf{g}(\boldsymbol{\mu})$ are linearly independent. If this is not the case, then the inverse may be replaced by any generalized inverse.

We now show how Proposition 1 may be implemented to find the ML estimate for models suitable for binary cross-over data.

3. Estimation for the log-nonlinear model

Adopting the notation of Becker and Balagtas (1993), define the group of subjects receiving the treatment order AB as group 1, the group of subjects receiving the treatment BA as group 2, and let n_{ijk} denote the number of subjects in group k having response i to the first treatment and response j

to the second treatment. The data may then be summarized in a $2 \times 2 \times 2$ contingency table given in Table 1.

Table 1. $2 \times 2 \times 2$ table from binary cross-over experiment

Treatment 1 response	AB: Treatment 2 response +	AB: Treatment 2 response −	Treatment 1 response	BA: Treatment 2 response +	BA: Treatment 2 response −
+	n_{111}	n_{121}	+	n_{112}	n_{122}
−	n_{211}	n_{221}	−	n_{212}	n_{222}

The 2×2 tables shown in Table 1 are independent multinomial samples of sizes n_{++k}, where a subscript + is used to denote summation over the replaced subscript. Let π_{ijk} denote the true multinomial probabilities, where $\sum_{ij} \pi_{ijk} = 1,\ k = 1, 2$. The quantities of interest in a cross-over experiment are the marginal probabilities or equivalently, logits of marginal response probabilities. A saturated model for the marginal logits and the log-odds ratios for the two 2×2 tables in Table 1, is

$$\begin{aligned}
&\ln(\pi_{1+1}/\pi_{2+1}) = \alpha + \tau_1 + \rho_1\ , \\
&\ln(\pi_{+11}/\pi_{+21}) = \alpha + \tau_2 + \rho_2 + \gamma_1\ , \\
&\ln(\pi_{1+2}/\pi_{2+2}) = \alpha + \tau_2 + \rho_1\ , \\
&\ln(\pi_{+12}/\pi_{+22}) = \alpha + \tau_1 + \rho_2 + \gamma_2\ ,
\end{aligned}$$

$$\ln[\pi_{11k}\pi_{22k}/(\pi_{12k}\pi_{21k})] = \psi_k, \quad k = 1, 2, \tag{3.1}$$

where $\pi_{+jk} = \sum_i \pi_{ijk}, \pi_{i+k} = \sum_j \pi_{ijk}$, α denotes the intercept, τ_m and ρ_m denote the mth treatment effect and mth period effect, respectively, γ_m describes a treatment-by-period interaction, sometimes called the carryover effect for the mth treatment. The ψ_k are log-odds ratios measuring the association between response to treatment A and response to treatment B in the kth group.

The model in (3.1) is overparameterized unless suitable restrictions are used to uniquely identify the τ_m, ρ_m, and γ_m. Consider the restrictions $\tau_1 = 0$, $\rho_1 = 0$, and $\gamma_1 = 0$. The model in (3.1) then becomes

$$\begin{aligned}
&\ln(\pi_{1+1}/\pi_{2+1}) = \alpha \ , \\
&\ln(\pi_{+11}/\pi_{+21}) = \alpha + \tau_2 + \rho_2 \ , \\
&\ln(\pi_{1+2}/\pi_{2+2}) = \alpha + \tau_2 \ , \\
&\ln(\pi_{+12}/\pi_{+22}) = \alpha + \rho_2 + \gamma_2 \ ,
\end{aligned}$$

$$\ln[(\pi_{11k}\pi_{22k})/(\pi_{12k}\pi_{21k})] = \psi_k, \quad k = 1, 2 \ . \tag{3.2}$$

Note that this model is specified in terms of the logits of the marginal probabilities and the ψ_1 and ψ_2 which are the logs of the odds ratios for the two groups. There is a closed form expression for the maximum likelihood estimators for the parameters for the model in (3.2) and the estimates are as follows

$$\widehat{\alpha} = \ln\left(\frac{n_{1+1}}{n_{2+1}}\right), \ \widehat{\tau}_2 = \ln\left(\frac{n_{1+2}n_{2+1}}{n_{1+1}n_{2+2}}\right), \ \widehat{\rho}_2 = \ln\left(\frac{n_{+11}n_{2+2}}{n_{+21}n_{1+2}}\right),$$
$$\widehat{\gamma}_2 = \widehat{\tau}_2 - \ln\left(\frac{n_{+11}n_{+22}}{n_{+12}n_{+21}}\right), \ \widehat{\psi}_1 = \ln\left(\frac{n_{111}n_{221}}{n_{121}n_{211}}\right),$$

$$\widehat{\psi}_2 = \ln\left(\frac{n_{112}n_{222}}{n_{122}n_{212}}\right). \tag{3.3}$$

Thus, under the restrictions, $\hat{\alpha}$ corresponds to the odds of success on treatment A in group 1, τ_2 and ρ_2 are estimated from log-odds ratios in 2×2 tables formed from marginal totals in Table 1 and γ_2 is estimated from a corresponding difference of log-odds ratios. The ψ_k are estimated by the log-odds ratios.

We now reformulate the model given in (3.2) and show how Proposition 1 may be used to find the ML estimates for this model and other reduced models, for which there is not a closed form expression for the ML estimates and which require an iterative procedure. Let $\mathbf{F} =$

$(F_{111}, F_{121}, F_{211}, F_{221}, F_{112}, F_{122}, F_{212}, F_{222})'$, be the expected frequencies for the cross-over experiment in Table 1 and $(a, b, c, d, e, f, g, h)'$, be the vector of marginal frequencies as indicated in Table 2.

Table 2. $2 \times 2 \times 2$ table of expected frequencies from binary cross-over experiment

Treatment 1 response	AB Treatment 2 response			Treatment 1 response	BA Treatment 2 response		
	+	−			+	−	
+	F_{111}	F_{121}	a	+	F_{112}	F_{122}	e
−	F_{211}	F_{221}	b	−	F_{212}	F_{222}	f
	c	d			g	h	

In terms of the expected frequencies and marginals, the model in (3.2) becomes

(i) $\ln(a/b) = \alpha$ (i) $\ln(a) - \ln(b) = \alpha$
(ii) $\ln(c/d) = \alpha + \tau_2 + \rho_2$ or (ii) $\ln(c) - \ln(d) = \alpha + \tau_2 + \rho_2$
(iii) $\ln(e/f) = \alpha + \tau_2$ (iii) $\ln(e) - \ln(f) = \alpha + \tau_2$
(iv) $\ln(g/h) = \alpha + \rho_2 + \gamma_2$ (iv) $\ln(g) - \ln(h) = \alpha + \rho_2 + \gamma_2$.

and (v) $\ln(F_{11k}F_{22k}/F_{12k}F_{21k}) = \psi_k$, $k = 1, 2$.

Let

$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} = \mathbf{I}_2 \otimes \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} ,$$

then $\mathbf{CF} = (a, b, c, d, e, f, g, h)'$.

Now let

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} = \mathbf{I}_4 \otimes (1, -1) .$$

Then the equations (i) to (iv) may be expressed as

$$\mathbf{B}\ln(\mathbf{CF}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \tau_2 \\ \rho_2 \\ \gamma_2 \end{pmatrix} = \mathbf{X}_1\boldsymbol{\beta}_1. \tag{3.4}$$

The two expressions in (v) may be written as

$$\mathbf{A}\ln(\mathbf{F}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \mathbf{X}_2\boldsymbol{\beta}_2\ , \tag{3.5}$$

where

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

In order to use Proposition 1 we need to write the model given in equations (3.4) and (3.5) in terms of the implied constraints. This is done by letting $\mathbf{P}_1 = \mathbf{I}_4 - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$ and $\mathbf{P}_2 = \mathbf{I}_2 - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$. Now multiplying equations (3.4) and (3.5) by $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively we get

$$\mathbf{g}_1(\mathbf{F}) = \mathbf{P}_1\mathbf{B}\ln(\mathbf{CF}) = \mathbf{P}_1\mathbf{X}_1\boldsymbol{\beta}_1 = \mathbf{0}, \text{ since } \mathbf{P}_1\mathbf{X}_1 = \mathbf{0}\ , \tag{3.6}$$

and

$$\mathbf{g}_2(\mathbf{F}) = \mathbf{P}_2\mathbf{A}\ln(\mathbf{F}) = \mathbf{P}_2\mathbf{X}_2\boldsymbol{\beta}_2 = \mathbf{0}, \text{ since } \mathbf{P}_2\mathbf{X}_2 = \mathbf{0}\ . \tag{3.7}$$

Note that in this case $\mathbf{X}_2 = \mathbf{I}_2$, so that $\mathbf{P}_2 = \mathbf{0}$. This will mean that the function $\mathbf{g}_2(\mathbf{F})$ will not be required in the iterative procedure for this case, but we will keep this function in the further discussion, since other models with restrictions on the log-odds ratios will not have $\mathbf{X}_2 = \mathbf{I}_2$.

We now have the model written in terms of the implied constraints $\mathbf{g}_1(\mathbf{F}) = \mathbf{0}$ and $\mathbf{g}_2(\mathbf{F}) = \mathbf{0}$, which are used in Proposition 1. We also require

$$\mathbf{G}_{1F} = \frac{\partial \mathbf{g}_1(\mathbf{F})}{\partial \mathbf{F}} = \mathbf{P}_1\mathbf{B}\mathbf{D}_{CF}^{-1}\mathbf{C}\ , \tag{3.8}$$

and

$$\mathbf{G}_{2F} = \frac{\partial \mathbf{g}_2(\mathbf{F})}{\partial \mathbf{F}} = \mathbf{P}_2\mathbf{A}\mathbf{D}_F^{-1}\ , \tag{3.9}$$

where $\mathbf{D}_{CF}$ and $\mathbf{D}_F$ are diagonal matrices with the elements of $\mathbf{CF}$ and $\mathbf{F}$, respectively on the principal diagonal.

Now let

$$\mathbf{g}(\mathbf{F}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{F}) \\ \mathbf{g}_2(\mathbf{F}) \end{pmatrix} . \tag{3.10}$$

Then

$$\mathbf{G}_F = \frac{\partial \mathbf{g}\,(\mathbf{F})}{\partial \mathbf{F}} = \begin{pmatrix} \mathbf{G}_{1F} \\ \mathbf{G}_{2F} \end{pmatrix} . \tag{3.11}$$

If we use a Poisson sampling procedure, then $\mathbf{V} = \mathbf{D}_F$ in the ML estimation procedure of equation (2.2), whereas if a multinomial sampling procedure is used then $\mathbf{V} = \mathbf{D}_F - \frac{1}{n}\mathbf{F}\mathbf{F}'$. In terms of the frequencies and Poisson sampling, the ML estimation procedure of Proposition 1 is given by

$$\widehat{\mathbf{F}}_c = \mathbf{y} - (\mathbf{G}_F\mathbf{D}_F)'(\mathbf{G}_y\mathbf{D}_F\mathbf{G}_F')^{-1}\mathbf{g}(\mathbf{y}) + o(||\mathbf{y} - \mathbf{F}||) . \tag{3.12}$$

The vector $\mathbf{y}$ is the vector of observed frequencies at the start of the iterative procedure, but changes during the iteration procedure. The ML estimates of the expected frequencies for the model are obtained by applying the expression above, using a double iterative procedure. The double iterative procedure is explained in Matthews and Crowther (1995). The rate of convergence of the procedure is very fast.

The ML estimates for the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are

$$\widehat{\boldsymbol{\beta}}_{1c} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{B}\ln(\mathbf{C}\widehat{\mathbf{F}}_c) \text{ and } \widehat{\boldsymbol{\beta}}_{2c} = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{A}\ln(\widehat{\mathbf{F}}_c) . \tag{3.13}$$

The covariance matrix for $\widehat{\boldsymbol{\beta}}_{1c}$ and $\widehat{\boldsymbol{\beta}}_{2c}$ can be found by applying the multivariate delta method. First consider the covariance matrix for $\widehat{\boldsymbol{\beta}}_{2c}$. It follows that

$$\mathcal{V}(\widehat{\boldsymbol{\beta}}_{2c}) = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\,\mathbf{A}\,\mathcal{V}[\ln(\widehat{\mathbf{F}}_c)]\mathbf{A}'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}$$

and using the multivariate delta method

$$\begin{aligned}\mathcal{V}[\ln(\widehat{\mathbf{F}}_c)] &= \left[\frac{\partial}{\partial \mathbf{F}}\ln(\mathbf{F})\right]\mathcal{V}(\widehat{\mathbf{F}}_c)\left[\frac{\partial}{\partial \mathbf{F}}\ln(\mathbf{F})\right]' \\ &= \mathbf{D}_F^{-1}\mathcal{V}(\widehat{\mathbf{F}}_c)\mathbf{D}_F^{-1}\ .\end{aligned}$$

For Poisson sampling $\mathbf{V}_F = \mathbf{D}_F$ and using the asymptotic covariance matrix $\mathbf{\Sigma}_c$ in Proposition 2, we get

$$\begin{aligned}\mathcal{V}(\widehat{\mathbf{F}}_c) = \textstyle\sum_c &= \mathbf{V}_F - (\mathbf{G}_F\mathbf{V}_F)'(\mathbf{G}_F\mathbf{V}_F\mathbf{G}_F')^{-1}\mathbf{G}_F\mathbf{V}_F \\ &= \mathbf{D}_F - (\mathbf{G}_F\mathbf{D}_F)'(\mathbf{G}_F\mathbf{D}_F\mathbf{G}_F')^{-1}\mathbf{G}_F\mathbf{D}_F\ .\end{aligned}$$

Here $\mathbf{G}_F = \begin{pmatrix}\mathbf{G}_{1F}\\ \mathbf{G}_{2F}\end{pmatrix}$ with $\mathbf{G}_{1F}$ and $\mathbf{G}_{2F}$ given in equations (3.8) and (3.9). We now find the covariance matrix for $\widehat{\boldsymbol{\beta}}_{1c}$. It follows that

$$\widehat{\boldsymbol{\beta}}_{1c} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{B}\ln(\mathbf{C}\widehat{\mathbf{F}}_c) = \mathbf{H}\ln(\mathbf{C}\widehat{\mathbf{F}}_c),\ \text{where}\ \mathbf{H} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\ \mathbf{B}\ .$$

Thus

$$\mathcal{V}(\widehat{\boldsymbol{\beta}}_{1c}) = \mathbf{H}\mathcal{V}[\ln(\mathbf{C}\widehat{\mathbf{F}}_c)]\mathbf{H}' = \mathbf{H}\mathbf{D}_{\mathbf{CF}}^{-1}\mathbf{C}\mathcal{V}(\widehat{\mathbf{F}}_c)\mathbf{C}'\mathbf{D}_{\mathbf{CF}}^{-1}\mathbf{H}'$$

using the multivariate delta method with

$$\frac{\partial}{\partial \mathbf{F}}\ln(\mathbf{CF}) = \mathbf{D}_{\mathbf{CF}}^{-1}\mathbf{C}\ \text{and}\ \mathbf{D}_{\mathbf{CF}} = \text{diag}(\mathbf{CF})\ .$$

As in the case above $\mathcal{V}(\widehat{\mathbf{F}}_c) = \mathbf{D}_F - (\mathbf{G}_F\mathbf{D}_F)'(\mathbf{G}_F\mathbf{D}_F\mathbf{G}_F')^{-1}\mathbf{G}_F\mathbf{D}_F$.

We consider the data of Zimmerman and Rahlfs (1978) to illustrate the application of the estimation procedure.

Example:

Table 3 gives the frequencies of the four possible outcomes within the treatment sequences AB and BA. The symbol "+" indicates a favourable response, while "-" indicates an unfavourable response.

Table 3. $2 \times 2 \times 2$ data set of Zimmerman and Rahlfs

Treatment 1 response	AB Treatment 2 response		Treatment 1 response	BA Treatment 2 response	
	+	−		+	−
+	6	33	+	15	6
−	4	7	−	11	18

Fitting the model described in equations (3.4) and (3.5), and using the expressions for the closed form solutions given in (3.3), gives the following parameter estimates

$\widehat{\alpha}$=1.266, $\widehat{\tau}_2$=–1.588, $\widehat{\rho}_2$=–1.064, $\widehat{\gamma}_2$=–0.122, $\widehat{\psi}_1$=–1.145, $\widehat{\psi}_2$=1.409.

These estimates can also be obtained by applying the ML estimation procedure in (3.12) and require only one iteration.

Testing $H_0 : \gamma_2 = 0$, i.e. there is no treatment-by-period carry over effect and using a Wald test

$$z = \frac{\widehat{\gamma}_2}{s.e.(\widehat{\gamma}_2)} = \frac{-0.122}{\sqrt{0.405}} = -0.19,$$

we see that γ_2 is not significant in the model. This indicates that there is no treatment-by-period carry over effect. We now fit the model given in equation (3.4) with $\gamma_2 = 0$, namely

$$\mathbf{B}\ln(\mathbf{CF}) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \tau_2 \\ \rho_2 \end{pmatrix} \tag{3.14}$$

along with equation (3.5).

The estimation procedure given in equation (3.12) is used to find the ML estimates for the expected frequencies for the model. On convergence of the iterative procedure, the vector of estimated expected frequencies is

$$\widehat{\mathbf{F}}_c = (5.78,\ 32.99,\ 3.98,\ 7.25,\ 15.39,\ 6.01,\ 11.01,\ 17.59)'$$

and the ML estimates for the parameters in the model are

$$\widehat{\alpha} = 1.239,\ \widehat{\tau}_2 = -1.529,\ \widehat{\rho}_2 = -1.126,\ \widehat{\psi}_1 = -1.144 \text{ and } \widehat{\psi}_2 = 1.409\ .$$

The estimated covariance matrix for $\widehat{\boldsymbol{\beta}}_{1c} = (\widehat{\alpha}, \widehat{\tau}_2, \widehat{\rho}_2)'$ is

$$\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}_{1c}) = \begin{bmatrix} 0.095 & -0.072 & -0.072 \\ -0.072 & 0.101 & 0.047 \\ -0.072 & 0.047 & 0.101 \end{bmatrix},$$

while the estimated covariance matrix for $\widehat{\boldsymbol{\beta}}_{2c} = (\widehat{\psi}_1, \widehat{\psi}_2)'$ is

$$\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}_{2c}) = \begin{bmatrix} 0.592 & 0.000 \\ 0.000 & 0.379 \end{bmatrix}.$$

For this model the likelihood-ratio statistic, $G^2 = 0.04$ with $df = 1$ and p-value $= 0.85$, which agrees with the values reported in Becker and Balagtas (1993). Wald tests for $H_0 : \rho_2 = 0$ and $H_0 : \tau_2 = 0$ give $z = -3.55$ and $z = -4.82$, respectively, indicating significant treatment and period effects. From these estimates it can be seen that treatment A is the better of the two treatments, and that each group did worse on their second treatment than the other group had done on that same treatment.

The model in (3.14) can also be reparameterized and written as

$$\mathbf{B}\ln(\mathbf{CF}) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

and will produce the same vector of estimated expected frequencies, $\widehat{\mathbf{F}}_c$ given above.

To fit the model $\gamma_2 = 0$ and $\psi_1 = \psi_2 = \psi$, the matrix $\mathbf{X}_2$ used earlier, becomes $\mathbf{X}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{\beta}_2 = \psi$. Once again, using the estimation procedure in (3.12), we find the ML estimates on convergence of the iterative

procedure. The ML estimates for the expected frequencies are

$$\widehat{\mathbf{F}}_c = (8.17,\ 30.53,\ 1.53,\ 9.77,\ 12.84,\ 8.48,\ 13.48,\ 15.21)'$$

and the ML estimates for the parameters are

$$\widehat{\alpha} = 1.231,\ \widehat{\tau}_2 = -1.529,\ \widehat{\rho}_2 = -1.126,\ \widehat{\psi} = 0.536.$$

The likelihood-ratio statistic, $G^2 = 6.57$ with $df = 2$ and the p-value $= 0.04$ indicating a poor model fit. This suggests that $\psi_1 = \psi_2 = \psi$ is not tenable and indicates that there is not homogeneous within-group association as reflected by the log-odds ratios for the two groups.

A more realistic model is $\gamma_2 = 0$ and $\psi_1 = -\psi_2$. Here the matrix $\mathbf{X}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\boldsymbol{\beta}_2 = \psi_1$. The ML estimates for the expected frequencies are

$$\widehat{\mathbf{F}}_c = (5.51, 33.27, 4.26, 6.97, 15.12, 6.28, 11.29, 17.31)' .$$

$\widehat{\psi}_1 = -1.306$ and the likelihood-ratio statistic, $G^2 = 0.11$ with $df = 2$, and p-value $= 0.95$, which indicates a good fit.

A further interesting application of the ML estimation procedure is for the model given in equation (3.14), namely $\mathbf{B}\ln(\mathbf{CF}) = \mathbf{X}_1\boldsymbol{\beta}_1$ with $\psi_1 = 0$ and $\psi_2 = 0$. In this case we use the constraints $\mathbf{g}_1(\mathbf{F})$ in equation (3.6) and the matrix $\mathbf{G}_{1F}$, given in (3.8) as well as the natural constraints imposed on $\mathbf{F}$ by $\psi_1 = 0$ and $\psi_2 = 0$. This is done as follows:

$$\psi_1 = \ln\left(\frac{F_{111}\ \ F_{221}}{F_{121}\ \ F_{211}}\right) = 0 \ \text{ implies that } \ F_{111}\ \ F_{221} - F_{121}\ \ F_{211} = 0.$$

Arranging $\mathbf{F} = (F_{111}, F_{121}, F_{211}, F_{221}, F_{112}, F_{122}, F_{212}, F_{222})'$ and letting

$$g_{21} = F_{111}\ \ F_{221} - F_{121}\ \ F_{211},$$

it follows that

$$\frac{\partial g_{21}}{\partial \mathbf{F}} = (F_{221}, -F_{211}, -F_{121}, F_{111}, 0, 0, 0, 0) .$$

Similarly $\psi_2 = \ln\left(\dfrac{F_{112}\ F_{222}}{F_{122}\ F_{212}}\right) = 0$, implies that $F_{112}\,F_{222} - F_{122}\,F_{212} = 0$.

Letting $g_{22} = F_{112}F_{222} - F_{122}F_{212}$, we find

$$\frac{\partial g_{22}}{\partial \mathbf{F}} = (0, 0, 0, 0, F_{222}, -F_{212}, -F_{122}, F_{112})\ .$$

The constraints $\mathbf{g}_2(\mathbf{F})$ in equation (3.10) are now replaced by

$$\mathbf{g}_2(\mathbf{F}) = \begin{pmatrix} g_{21} \\ g_{22} \end{pmatrix}$$

and $\mathbf{G}_{2F}$ in (3.9) is replaced by

$$\mathbf{G}_{2F} = \begin{bmatrix} F_{221} & -F_{211} & -F_{121} & F_{111} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & F_{222} & -F_{212} & -F_{122} & F_{112} \end{bmatrix}\ .$$

Using these expressions in (3.12), the iterative procedure gives

$$\widehat{\mathbf{F}}_c = (7.51,\ 31.19,\ 2.19,\ 9.11,\ 11.21,\ 10.10,\ 15.10,\ 13.60)$$

as the ML estimates of the expected frequencies. The parameter estimates are $\widehat{\alpha} = 1.231, \widehat{\tau}_2 = -1.529,\ \widehat{\rho}_2 = -1.127$ and $G^2 = 7.78$ with $df = 3$ and p-value=0.05. This model does not fit well, but shows how the natural constraints imposed by $\psi_1 = 0$ and $\psi_2 = 0$ can be used in the ML estimation procedure.

A program for estimating the parameters for the model given in (3.5) and (3.14), is given in the Appendix. The program utilizes SAS/IML, which is a powerful matrix language and is a procedure of the SAS system. The other models fitted are also obtained in a similar manner by simply using the appropriate functions $\mathbf{g}(\mathbf{F})$ and the corresponding matrix of partial derivatives $\mathbf{G}_F$ defined earlier.

4. Estimation in the loglinear model

Kenward and Jones (1987) propose loglinear models for the analysis of binary cross-over data. The ML estimation procedure of Matthews and Crowther

(1995) can also be easily applied to the analysis of loglinear models and is discussed in the foregoing reference. The methodology involves writing the loglinear model in terms of the implied constraints expressed as a function $\mathbf{g}(\mathbf{F}) = \mathbf{0}$ in Proposition 1. The loglinear model can be written as $\ln(\mathbf{F}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ and $\boldsymbol{\beta}$ depend on the model that is specified for the data. By letting $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, we find a function $\mathbf{g}(\mathbf{F}) = \mathbf{P}\ln(\mathbf{F}) = \mathbf{0}$, for which $\mathbf{G}_F = \dfrac{\partial \mathbf{g}(\mathbf{F})}{\partial \mathbf{F}} = \mathbf{P}\mathbf{D}_F^{-1}$. Assuming Poisson sampling, for which $\mathbf{V}_F = \mathbf{D}_F$, the ML estimates for the expected frequencies for the model are found by applying (3.12), which gives

$$\widehat{\mathbf{F}}_c = \mathbf{y} - \mathbf{P}'(\mathbf{P}\mathbf{D}_y^{-1}\mathbf{P})^{-1}\,\mathbf{P}\ln(\mathbf{y}) + o(||\mathbf{y} - \mathbf{F}||)\,.$$

At the start of the iterative procedure $\mathbf{y}$ is the vector of observed frequencies. However, in the iterative procedure iteration takes place over $\mathbf{y}$ until convergence is attained. Estimation in the loglinear model, the cumulative logit model, adjacent categories logit model, and the Poisson regression model is discussed in some detail in Matthews and Crowther (1995).

5. Conclusion

The estimation procedure set out in Proposition 1, provides a simple method for finding the ML estimates of the expected frequencies and the parameters in the log-nonlinear model and loglinear model for data from a binary cross-over design experiment. This is done by expressing the model as a function $\mathbf{g}(\mathbf{F}) = \mathbf{0}$, which satisfies the conditions of the estimation procedure of Proposition 1. This ML estimation procedure can be applied to the generalized linear model and it is particularly useful for cases where the Newton Raphson procedure leads to cumbersome first and second order derivatives.

References

BECKER, M.P. & BALAGTAS, C.C. (1993). Marginal modeling of binary cross-over data. *Biometrics*, **49**, 997–1009.

KENWARD M.J. & JONES, B. (1987). A log-linear model for binary cross-over data. *Applied Statistics*, **36**, 192–204.

MATTHEWS, G.B. & CROWTHER, N.A.S. (1995). A maximum likelihood estimation procedure when modelling in terms of constraints. *S.A Statist. J.*, **29**(1), 29–51.

SAS/STAT *User's Guide, Version 6, Fourth Edition, Volume 2*. Cary NC: SAS Institute Inc., 1990.

ZIMMERMAN, H. & RAHLFS, V. (1978). Testing hypotheses in the two-period change-over with binary data. *Biometrical Journal **2***, 133–141.

APPENDIX

The SAS/IML program for the cross-over design model specified by equations (3.5) and (3.14), is as follows:

```
proc iml worksize=80; options pagesize=500;
*-----> FREQUENCY VECTOR;
*----->; x={ 6,33,4,7,15,6,11,18};
C={1 1 0 0 ,
   0 0 1 1 ,
   1 0 1 0 ,
   0 1 0 1};
C=I(2)@c;
B={1 -1};
B=I(4)@B;
X1={ 1 0 0  ,
   1 1 1 ,
   1 1 0 ,
```

```
   1 0 1 };
X2={1 0,
   0 1};
P1=I(4)-X1*inv(X1'*X1)*X1';
P2=I(2)-X2*inv(X2'*X2)*X2';
P1B=P1*B;
g1x=P1B*log(C*x);
A={1 -1 -1 1 0 0 0 0,
   0  0  0 0 1 -1 -1 1};
P2A=P2*A;
g2x=P2A*log(x);
xi=1/x;
Gx1= P1B*(diag(1/(C*x))*C);
Gx2= P2A*diag(xi);
g=g1x//g2x;
 m=x;   x11=x;
 itr=0;  diff1=1;  i=0;
Sigc=j(8,8,0);
do while (diff1>0.000001);
  i=i+1;
  m=x;  mi=1/m;\vspace{-0.05cm}
  x=x11;
   Gm1= P1B*(diag(1/(C*m))*C);
   Gm2=P2A*diag(mi);
  Gm=Gm1//Gm2;
  j=0;   diff=1;
do while (diff>0.000001);
```

```
  xv=x; j=j+1;
  xi=1/x;
  Gx1= P1B*(diag(1/(C*x))*C);
  Gx2= P2A*diag(xi);
  Gx=Gx1//Gx2;
  g1x=P1B*log(C*x);
  g2x=P2A*log(x);
  g=g1x//g2x;
  x=x-(m'#Gm)'*ginv(Gx*(m#Gm'))*g;
  diff=sqrt((x-xv)'*(x-xv));
  itr=itr+1;
end;
  diff1=sqrt((m-x)'*(m-x));
end;
print x; Sigc=diag(x)-(Gx*diag(x))'*ginv
     (Gx*diag(x)*Gx')*Gx*diag(x);
beta1=inv(X1'*X1)*X1'*B*log(C*x); beta2=inv(X2'*X2)
     *A*log(x);
vbeta2=inv(X2'*X2)*X2'*A*diag(xi)*Sigc*diag(xi
     *A'*X2*inv(X2'*X2);
vbeta1=inv(X1'*X1)*X1'*B*(diag(1/(C*x))*C)
     *Sigc*C'*diag(1/(C*x))*B'*X1*inv(X1'*X1);
print beta1 beta2;
ch2=(x11-m)'*((x11-m)/m);
G2=2*x11'*log(x11/m);
print ch2 G2; print Sigc; print vbeta2; print vbeta1;
```

Manuscript received, 2010.06.29, revised, 2011.05.20, accepted, 2011.06.12.