*Biometrika* (1991), **78**, 4, *pp.* 719–27
*Printed in Great Britain*

# Estimation in generalized linear models with random effects

BY ROBERT SCHALL

*Division of Biometry, Department of Pharmacology, University of the Orange Free State, Bloemfontein 9300, Republic of South Africa*

SUMMARY

A conceptually very simple but general algorithm for the estimation of the fixed effects, random effects, and components of dispersion in generalized linear models with random effects is proposed. Conditions are described under which the algorithm yields approximate maximum likelihood or quasi-maximum likelihood estimates of the fixed effects and dispersion components, and approximate empirical Bayes estimates of the random effects. The algorithm is applied to two data sets to illustrate the estimation of components of dispersion and the modelling of overdispersion.

*Some key words*: Component of dispersion; Empirical Bayes; Quasi-likelihood; Restricted maximum likelihood; Variance component.



## 1. INTRODUCTION

In this paper estimation of the fixed effects, the random effects, and the components of dispersion (McCullagh & Nelder, 1989, p. 432) in generalized linear models with random effects is considered.

A generalized linear model with random effects is defined as follows: let $y = [y_1, \ldots, y_n]'$ be a vector of $n$ observations, which can be written as

$$y = \mu + e, \tag{1·1}$$

where $e$ is a vector of random errors with zero expectation and covariance matrix $V$. Let further $g(.)$ be a monotonic function, the link (McCullagh & Nelder, 1989, p. 27), such that $g(\mu)$ can be written as the linear model

$$g(\mu) = \eta = X\beta + U_1 b_1 + \ldots + U_c b_c. \tag{1·2}$$

Here $X_{n\times p}$ is a known design matrix, $\beta$ is a vector of fixed effects, the $U_i$ are known $n \times q_i$ matrices, and the $b_i$ are $q_i \times 1$ vectors of random effects ($i = 1, \ldots, c$). If conditionally on $\mu$ the components of $y$ are independently distributed, and if their distribution is a member of the exponential family, then (1·1) and (1·2) define a generalized linear model with random effects.

To achieve an economical notation, let

$$q_1 + \ldots + q_c = q, \quad U = [U_1 \vdots \ldots \vdots U_c], \quad b = [b_1', \ldots, b_c']'.$$

The random vectors $b_1, \ldots, b_c$ are assumed to be uncorrelated with zero expectation. The random effects are also assumed to be uncorrelated with $e$. Further, $\mathrm{cov}\,(b_i) = \sigma_i^2 I_{q_i}$ ($i = 1, \ldots, c$) so that

$$\mathrm{cov}\,(b) = D = \mathrm{diag}\,(\sigma_1^2 I_1, \ldots, \sigma_c^2 I_c), \tag{1·3}$$

where $I_1, \ldots, I_c$ are identity matrices of orders $q_1 \times q_1, \ldots, q_c \times q_c$. The only difference from the usual definition of a generalized linear model is the introduction of random effects into the linear predictor (1·2). The special case when $y$ follows a normal distribution and $g(.)$ is the identity link leads to the well-known linear random effects model

$$y = X\beta + U_1 b_1 + \ldots + U_c b_c + e. \tag{1·4}$$

In model (1·4) we have $E(y) = X\beta$, $\mathrm{cov}\,(b) = D$, $\mathrm{cov}\,(e) = V = \sigma^2 I_n$ and consequently $\mathrm{cov}\,(y) = V + UDU'$.

The analysis of a generalized linear model with random effects can proceed along the following lines: a distribution, or a parameterized class of distributions, for the random effects $b_1, \ldots, b_c$ can be specified, and the parameters $\beta$, and the parameters specifying the distribution of the random effects can be estimated by maximum likelihood based on the marginal distribution of the observations $y$. This approach is typically used in the linear random effects model (1·4), see e.g. Harville (1977), where both the distribution of the random effects, and the conditional distribution of $y$ are assumed to be normal. Anderson & Aitkin (1985) and Im & Gianola (1988) use maximum likelihood estimation in logistic and probit models where the random effects are assumed to follow a normal distribution, and the conditional distribution of $y$ is binomial.

However, in generalized linear models other than linear random effects models there are two problems with such an approach: first, it is often difficult to justify a particular distribution or class of distributions for the random effects. Secondly, maximum likelihood estimation based on the marginal distribution of the observations involves the 'integrating out' of the random effects which, except in certain special cases, is numerically not feasible (Stiratelli, Laird & Ware, 1984, p. 964). For example, when random effects are not nested but crossed, numerical integration is usually not practical.

In the following an algorithm for estimation in generalized linear models with random effects is proposed. The algorithm is conceptually simple and easy to program, and generally applicable to generalized linear models regardless of the structure of the linear predictor. Thus random effects may be nested or crossed, or even occur in regression or analysis of covariance models. Further, it is not necessary to specify the distribution of the random effects beyond the weak assumptions on their expectation and variance made above.



## 2. THE ALGORITHM

Fellner (1986, 1987), based on work by Henderson (1963) and Harville (1977), proposed the following algorithm for maximum likelihood estimation in the normal variance components model (1·4).

*Algorithm* 1:

1. Given estimates $\hat{\sigma}^2$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_c^2$ for $\sigma^2$ and $\sigma_1^2, \ldots, \sigma_c^2$, compute estimates $\hat{\beta}$ and $\hat{b}_1, \ldots, \hat{b}_c$ for $\beta$ and $b_1, \ldots, b_c$ as least-squares solutions to the set of overdetermined linear equations

$$C\begin{bmatrix}\hat{\beta}\\ \hat{b}\end{bmatrix} = \begin{bmatrix} V^{-\frac{1}{2}}X & V^{-\frac{1}{2}}U \\ 0 & D^{-\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{b}\end{bmatrix} = \begin{bmatrix} V^{-\frac{1}{2}}y \\ 0\end{bmatrix},$$

where $V = \sigma^2 I$ and $D$ from (1·3) are evaluated at the current estimates of the variance components.

2. Let $T^*$ be the inverse of the matrix formed by the last $q$ rows and columns of $C'C$, partitioned conformably with $D$ as
$$\begin{bmatrix} T^*_{11} & \cdots & T^*_{1c} \\ \vdots & & \vdots \\ T^*_{c1} & \cdots & T^*_{cc} \end{bmatrix}.$$
Given estimates $\hat{\beta}$ and $\hat{b}_1, \ldots, \hat{b}_c$ for $\beta$ and $b_1, \ldots, b_c$, compute estimates $\hat{\sigma}^2$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_c^2$ for $\sigma^2$ and $\sigma_1^2, \ldots, \sigma_c^2$ as
$$\hat{\sigma}^2 = (y - X\hat{\beta} - U\hat{b})'(y - X\hat{\beta} - U\hat{b}) \Big/ \left\{ n - \sum_{i=1}^{c} (q_i - v_i^*) \right\}, \quad \hat{\sigma}_i^2 = \frac{\hat{b}_i'\hat{b}_i}{q_i - v_i^*},$$
where $v_i^* = \operatorname{tr}(T^*_{ii})/\sigma_i^2$ is evaluated at the current estimate of $\sigma_i^2$.

This algorithm yields maximum likelihood estimates of the parameters. To obtain restricted maximum likelihood estimates (Patterson & Thompson, 1971) replace Step 2 above by the following.

2′. Let $T$ be the matrix formed by the last $q$ rows and columns of the inverse of $C'C$, partitioned conformably with $D$ as
$$\begin{bmatrix} T_{11} & \cdots & T_{1c} \\ \vdots & & \vdots \\ T_{c1} & \cdots & T_{cc} \end{bmatrix}.$$
Given estimates $\hat{\beta}$ and $\hat{b}_1, \ldots, \hat{b}_c$ for $\beta$ and $b_1, \ldots, b_c$, compute estimates $\hat{\sigma}^2$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_c^2$ for $\sigma^2$ and $\sigma_1^2, \ldots, \sigma_c^2$ as
$$\hat{\sigma}^2 = (y - X\hat{\beta} - U\hat{b})'(y - X\hat{\beta} - U\hat{b}) \Big/ \left\{ n - p - \sum_{i=1}^{c} (q_i - v_i) \right\}, \quad \hat{\sigma}_i^2 = \frac{\hat{b}_i'\hat{b}_i}{q_i - v_i},$$
where $v_i = \operatorname{tr}(T_{ii})/\sigma_i^2$ is evaluated at the current estimate of $\sigma_i^2$.

This algorithm is now adapted for estimation in generalized linear models with random effects. To motivate this adaptation, the link function $g(.)$ applied to the data $y$ (McCullagh & Nelder, 1989, p. 31) is linearized, giving to the first order
$$g(y) = g(\mu) + (y - \mu)g'(\mu) = z. \tag{2·1}$$
McCullagh & Nelder (1989, p. 31) call $z$ the adjusted dependent variable. Now, from (1·2) and (2·1) a linear random effects model for $z$ is
$$z = X\beta + Ub + eg'(\mu). \tag{2·2}$$
Here $E(z) = X\beta$, $\operatorname{cov}(b) = D$ and
$$\operatorname{cov}\{eg'(\mu)\} = V(\partial\eta/\partial\mu)^2 = W^{-1} \tag{2·3}$$
so that $\operatorname{cov}(z) = W^{-1} + UDU'$. Thus model (2·2) is a linear random effects model with the same first and second order structure as model (1·4). The natural way of adapting Algorithm 1 to model (2·2) is to replace $y$ by $z$, and $V = \operatorname{cov}(e)$ by $W^{-1} = \operatorname{cov}\{eg'(\mu)\}$. This yields the following algorithm for the estimation of fixed effects, random effects and components of dispersion in a generalized linear model with random effects.

*Algorithm* 2:

1. Given estimates $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_c^2$ for $\sigma_1^2, \ldots, \sigma_c^2$, compute least-squares estimates $\hat{\beta}$ and $\hat{b}_1, \ldots, \hat{b}_c$ for $\beta$ and $b_1, \ldots, b_c$ as solutions to the overdetermined set of

linear equations

$$C\begin{bmatrix}\hat{\beta}\\ \hat{b}\end{bmatrix}=\begin{bmatrix}W^{\frac{1}{2}}X & W^{\frac{1}{2}}U\\ 0 & D^{-\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{b}\end{bmatrix}=\begin{bmatrix}W^{\frac{1}{2}}z\\ 0\end{bmatrix}, \tag{2·4}$$

where $D$, $z$ and $W$ are respectively given by (1·3), (2·2) and (2·3), and all unknown quantities are evaluated at their current estimates.

2. Let $T^*$ be the inverse of the matrix formed by the last $q$ rows and columns of $C'C$, partitioned conformably with $D$ as

$$\begin{bmatrix}T^*_{11} & \cdots & T^*_{1c}\\ \vdots & & \vdots\\ T^*_{c1} & \cdots & T^*_{cc}\end{bmatrix}.$$

Given estimates $\hat{b}_1,\ldots,\hat{b}_c$ for $b_1,\ldots,b_c$, compute estimates $\hat{\sigma}_1^2,\ldots,\hat{\sigma}_c^2$ for $\sigma_1^2,\ldots,\sigma_c^2$ as

$$\hat{\sigma}_i^2=\frac{\hat{b}_i'\hat{b}_i}{q_i-v_i^*},$$

where $v_i^*=\operatorname{tr}(T^*_{ii})/\sigma_i^2$ is evaluated at the current estimate of $\sigma_i^2$.

This algorithm is analogous to the algorithm yielding maximum likelihood estimates in the normal random effects model. To obtain an algorithm analogous to the algorithm yielding restricted maximum likelihood estimates replace Step 2 above by the following.

2′. Let $T$ be the matrix formed by the last $q$ rows and columns of the inverse of $C'C$, partitioned conformably with $D$ as

$$\begin{bmatrix}T_{11} & \cdots & T_{1c}\\ \vdots & & \vdots\\ T_{c1} & \cdots & T_{cc}\end{bmatrix}.$$

Given estimates $\hat{b}_1,\ldots,\hat{b}_c$ for $b_1,\ldots,b_c$, compute estimates $\hat{\sigma}_1^2,\ldots,\hat{\sigma}_c^2$ for $\sigma_1^2,\ldots,\sigma_c^2$ as

$$\hat{\sigma}_i^2=\frac{\hat{b}_i'\hat{b}_i}{q_i-v_i},$$

where $v_i=\operatorname{tr}(T_{ii})/\sigma_i^2$ is evaluated at the current estimate of $\sigma_i^2$.

Sometimes $\operatorname{cov}(e)=\sigma^2V$, that is, the conditional variance function of $y$ is a function of an unknown nuisance parameter, as for example in the normal random effects model (1·4). In that case the weight matrix is $\sigma^{-2}W$. The extra component of dispersion can be estimated by

$$\hat{\sigma}^2=(z-X\hat{\beta}-U\hat{b})'W(z-X\hat{\beta}-U\hat{b})\Big/\left\{n-\sum_{i=1}^{c}(q_i-v_i^*)\right\}$$

in Step 2 of Algorithm 2, or by

$$\hat{\sigma}^2=(z-X\hat{\beta}-U\hat{b})'W(z-X\hat{\beta}-U\hat{b})\Big/\left\{n-p-\sum_{i=1}^{c}(q_i-v_i)\right\}$$

in Step 2′ of Algorithm 2. In Step 1, then, $W$ must be replaced by $\hat{\sigma}^{-2}W$.

Even in cases where the variance function of the standard generalized linear model does not contain a nuisance parameter, as in logistic or log linear models, it is advisable to compute $\hat{\sigma}^2$ using the above formulae at convergence of the algorithm. This would provide a check for over- or underdispersion, which would respectively be indicated by values of $\hat{\sigma}^2$ significantly larger or smaller than 1.

## 3. Discussion

In the special case of the normal linear random effects model (1·4), Algorithm 2 reduces to Algorithm 1, and yields maximum likelihood estimates. Otherwise, Algorithm 2 does not in general yield maximum likelihood estimates, but it is shown in the Appendix that Algorithm 2 yields the same estimates as the algorithm proposed by Stiratelli et al. (1984). Specifically, given a diffuse prior for the fixed effects $\beta$, and a normal prior $N(0, D)$ for the random effects $b$, Step 1 of Algorithm 2 constitutes an iteration of Fisher's scoring method to maximize the posterior likelihood with respect to $\beta$ and $b$. Step 2 constitutes an EM iteration to compute the dispersion components, where the posterior distribution of $b$ is approximated by a multivariate normal distribution with the same mode and curvature as the true posterior (Stiratelli et al., 1984, pp. 964–5). This property of Step 2 implies that the estimates of the dispersion components are nonnegative.

Because Step 1 of Algorithm 2 involves only the variance function $V$ of the conditional distribution of $y$ given $b$, and the variance function $D$ of the prior distribution of $b$, the interpretation of Step 1 remains essentially the same if only a conditional quasi-likelihood for $y$ given $b$ with some variance function $V$, and a prior quasi-likelihood for $b$ with some variance function $D$, not necessarily of form (1·3), are specified. In that case, $\hat{\beta}$ and $\hat{b}$ maximize the posterior quasi-likelihood. In summary, therefore, when both the conditional distribution of $y$ given $b$, and the prior distribution for $b$ are in the exponential family, with $b$ not necessarily normal, then $\hat{\beta}$ can be interpreted as an approximate maximum likelihood estimate for $\beta$, the estimates of the dispersion components as approximate maximum likelihood or restricted maximum likelihood estimates, and $\hat{b}$ as an approximate empirical Bayes estimate for $b$, in the manner of Stiratelli et al. (1984). When working with quasi-likelihoods, $\hat{\beta}$ and the estimates of the dispersion components can be interpreted as approximate quasi-likelihood estimates.

It may also be interesting to note that application of the 'restricted maximum likelihood' version of Algorithm 2 for the special case of a single dispersion component, binomial data and a probit link yields the 'joint-maximization method' of Gilmour, Anderson & Rae (1985). However, Algorithm 2 is perfectly general with respect to the conditional distribution of $y$ providing it is in the exponential family, the link function, and the number of dispersion components.

The proposed algorithm is appealing as it allows a very flexible and general extension of the GLIM package to random effects models, with a single additional command and relatively little additional programming. To apply and run Algorithm 2 with a data set, as usually in GLIM the link function $g(.)$, the linear predictor, and the conditional distribution of $y$, or at least the variance function $V$, must be specified. The only additional command to invoke random effects estimation would be to specify the random effects, namely by specifying the decomposition $[X \,\vdots\, U_1 \,\vdots\, \ldots \,\vdots\, U_c]$ of the linear predictor into fixed and random effects, and, possibly, to specify the variance function $D$. On a programming level, Algorithm 2 uses the same quantities as the usual iterative reweighted least squares algorithm in GLIM, namely the adjusted dependent variable $z$, and the weight matrix $W$. The only modification to the usual iterative reweighted least squares algorithm would be to solve an extended set of normal equations in Step 2, and the computations necessary to perform Steps 2 or 2′.

## 4. EXAMPLES OF APPLICATION

### 4·1. *General*

The algorithm is applied to two sets of data. The first example is used to illustrate the estimation of components of dispersion, and to compare the results to an alternative analysis of McCullagh & Nelder (1989). The second example illustrates the modelling of overdispersion through the fitting of random effects.

### 4·2. *Salamander mating data*

As a first example of application the salamander mating data described and analysed by McCullagh & Nelder (1989, Ch. 14.5) are used. The experiment involved two populations of salamander, Rough Butt and Whiteside. Ten Rough Butt males and ten Whiteside males were sequestered as pairs with ten Rough Butt females and ten Whiteside females on six occasions according to the design given in Table 14.3 of McCullagh & Nelder (1989). For each pair it was then recorded whether mating actually occurred. In total therefore, 120 binary observations are available. The logit of the mating probability is assumed to be the sum of a fixed cross effect, plus an individual random effect due to the female involved, plus an individual random effect due to the male involved. For the cross of the $i$th Rough Butt female with the $j$th Rough Butt male, for example, the logit of the mating probability is

$$\log \frac{\pi_{RRij}}{1-\pi_{RRij}} = \beta_{RR} + a_i + b_j,$$

where $\beta_{RR}$ is the average logit of the mating probabilities between a female Rough Butt and a male Rough Butt, and $a_i$ and $b_j$ are the random effects due to the female and male individuals in the pair.

The experiment was performed three times, and Table 1 summarizes the results. For the data from the first experiment, Algorithm 2 converged very slowly, i.e. more than 100 iterations were needed to achieve convergence at 4 decimal digits, possibly because the estimate for the male dispersion component is close to zero. In a second application,

Table 1. *Salamander mating data: summary of parameter estimates for three experiments*

| Experiment | Method | Parameter estimate $\hat{\pi}_{RR}$ | $\hat{\pi}_{RW}$ | $\hat{\pi}_{WR}$ | $\hat{\pi}_{WW}$ | Dispersion component $\hat{\sigma}_1^2$ | $\hat{\sigma}_2^2$ |
|---|---|---|---|---|---|---|---|
| 1 | M&N | 0·7333 | 0·6667 | 0·2333 | 0·7000 | 1·3704 | 0·6963 |
| | REML | 0·7619 | 0·6865 | 0·1959 | 0·7340 | 1·4099 | 0·0896 |
| | REML* | 0·7609 | 0·6870 | 0·1956 | 0·7350 | 1·4679 | 0 |
| | ML | 0·7563 | 0·6836 | 0·2028 | 0·7280 | 1·1511 | 0·0066 |
| 2 | M&N | 0·6000 | 0·4667 | 0·2333 | 0·6667 | 0·9787 | 0·5997 |
| | REML | 0·6101 | 0·4629 | 0·1830 | 0·6955 | 1·2584 | 0·6161 |
| | ML | 0·6079 | 0·4638 | 0·1943 | 0·6897 | 0·9497 | 0·4404 |
| 3 | M&N | 0·6667 | 0·5333 | 0·1667 | 0·6333 | 0·3954 | 1·3440 |
| | REML | 0·6985 | 0·5402 | 0·1339 | 0·6584 | 0·2618 | 1·4988 |
| | ML | 0·6917 | 0·5386 | 0·1401 | 0·6541 | 0·1517 | 1·1864 |

M&N, estimates from McCullagh & Nelder (1989, Table 14.10).
* Results for fitting only the female random effects.

therefore, the male random effects and dispersion component were dropped from the model. In that case, and in all other applications of Algorithm 2 convergence was fairly rapid.

The estimates of the four mating probabilities $\pi_{RR}$, $\pi_{RW}$, $\pi_{WR}$ and $\pi_{WW}$ differ slightly from those given by McCullagh & Nelder (1989, Table 14.10). However, McCullagh & Nelder (1989) estimate the population-averaged mating probabilities, while the present estimates are based on a subject-specific model (Zeger, Liang & Albert, 1988). The estimates of the dispersion components differ more markedly, but this is not surprising as the dispersion components estimated here represent variances of effects on the logistic scale, while McCullagh & Nelder (1989) consider variances of probabilities. However, qualitatively the results are the same: in the first two experiments the female dispersion component is larger than the male, while this relationship is reversed in the last experiment.

One may note the relatively big differences between the approximate maximum likelihood and approximate restricted maximum likelihood estimates of the dispersion components. This is probably due to the relatively large bias of maximum likelihood estimates for this rather sparse data.

The approximate variance-covariance matrix of $\hat{\beta}$ is given by the first $p$ rows and columns of $(C'C)^{-1}$, that is

$$\operatorname{cov}(\hat{\beta}) = \{X'(W^{-1} + UDU')^{-1}X\}^{-1}.$$

Its estimate for the first experiment is

$$\begin{bmatrix} 0{\cdot}3553 & 0{\cdot}1453 & 0{\cdot}0087 & 0 \\ 0{\cdot}1453 & 0{\cdot}3282 & 0 & 0{\cdot}0090 \\ 0{\cdot}0087 & 0 & 0{\cdot}3774 & 0{\cdot}1303 \\ 0 & 0{\cdot}0090 & 0{\cdot}1303 & 0{\cdot}3446 \end{bmatrix}.$$

The approximate restricted maximum likelihood estimate of the contrast $\beta_{RW} - \beta_{WR}$ is 2·1958, with estimated standard error 0·8399. Thus, as given by McCullagh & Nelder (1989, p. 450), 'the evidence for a non-zero mixed contrast is strong'.

### 4·3. *Cell irradiation data*

The second data set comes from an experiment to measure the mortality of cancer cells under radiation, and were provided by Dr G. Blekkenhorst, Department of Radiology, University of Cape Town. Four hundred cells were placed on a dish, and three dishes were irradiated at a time, or occasion. After the cells were irradiated, the surviving cells were counted. Since cells would also die naturally, dishes with cells were put in the radiation chamber without being irradiated, to establish the natural mortality. For the purposes of this example, only these zero-dose data are analysed. Twenty seven dishes on nine occasions, or three per occasion were available, resulting in 27 binomial observations. The data are given in Table 2.

A concern in this data was the presence of extra-binomial variation. Treating the 27 counts as independent binomial observations leads to a Pearson chi-squared statistic of 470·34 with 26 degrees of freedom, or an estimate of the residual variance of $\hat{\sigma}^2 = 18{\cdot}09$, which is considerably larger than 1. One way to explain this overdispersion is through random differences in the mortality between occasions, and possibly even between dishes. Initially modelling random effects due to occasions, the model for the logits of the

Table 2. *Cell irradiation data*

| Occasion | Dish | No. cells surviving out of 400 placed | Occasion | Dish | No. cells surviving out of 400 placed |
|---|---|---|---|---|---|
| 1 | 1 | 178 | 6 | 16 | 115 |
| 1 | 2 | 193 | 6 | 17 | 130 |
| 1 | 3 | 217 | 6 | 18 | 133 |
| 2 | 4 | 109 | 7 | 19 | 200 |
| 2 | 5 | 112 | 7 | 20 | 189 |
| 2 | 6 | 115 | 7 | 21 | 173 |
| 3 | 7 | 66 | 8 | 22 | 88 |
| 3 | 8 | 75 | 8 | 23 | 76 |
| 3 | 9 | 80 | 8 | 24 | 90 |
| 4 | 10 | 118 | 9 | 25 | 121 |
| 4 | 11 | 125 | 9 | 26 | 124 |
| 4 | 12 | 137 | 9 | 27 | 136 |
| 5 | 13 | 123 | | | |
| 5 | 14 | 146 | | | |
| 5 | 15 | 170 | | | |

27 survival probabilities is

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \mu + b_{1i} \quad (i=1,\ldots,9;\, j=1,\ldots,3),$$

where $\mu$ is the intercept and the vector $b_1$ represents the 9 random effects due to occasions.

From this model, the estimate of the dispersion component is $\hat{\sigma}_1^2 = 0{\cdot}225$, and an estimate of the residual variance is $\hat{\sigma}^2 = 1{\cdot}810$. Because $\hat{\sigma}^2$ is still slightly larger than 1, random effects due to dishes were also fitted, to yield the model

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \mu + b_{1i} + b_{2ij} \quad (i=1,\ldots,9;\, j=1,\ldots,3).$$

Here $\mu$ and $b_1$ are as before, and the vector $b_2$ represents the 27 random effects due to each dish. The estimates of the dispersion components are now $\hat{\sigma}_1^2 = 0{\cdot}222$ and $\hat{\sigma}_2^2 = 0{\cdot}010$, with a residual variance of $\hat{\sigma}^2 = 0{\cdot}937$, which is very close to 1.

Thus, by fitting random effects due to occasions, and possibly due to dishes, the extra-binomial variation can be explained.

## Acknowledgements

I am grateful to three referees whose comments have improved an earlier version of the paper, and to my colleagues Stephan Maritz and Carl Lombard for valuable discussions. This article was written while the author was a member of the Institute for Biostatistics of the South African Medical Research Council.

## Appendix

*Proof that Algorithm* 2 *maximizes the posterior likelihood under a normal prior*

Let $f(y \mid \beta, b)$ be the conditional density of $y$, which by assumption is in the exponential family. Let further $p(b \mid D)$ be the $q$-variate normal density with zero mean and dispersion matrix $D$ given in (1·3). Under a diffuse prior for $\beta$, the posterior density for $\beta$ and $b$ is proportional to $f(y|\beta, b)p(b|D)$ (Stiratelli et al., 1984, p. 964). The posterior log likelihood is then, up to a constant, the sum of $\log f(y|\beta, b)$ and $\log p(b|D)$. We want to show that Step 1 of Algorithm 2 constitutes an iteration of Fisher's scoring method to maximize the posterior log likelihood with respect to $\beta$ and $b$.

From general results for exponential families (McCullagh & Nelder, 1989, pp. 41-2), the score vector $S$ and expected information matrix $I$ are

$$\begin{aligned} S &= \frac{\partial \log f(y|\beta, b)}{\partial(\beta, b)} + \frac{\partial \log p(b|D)}{\partial(\beta, b)} \\ &= \begin{bmatrix} X' \\ U' \end{bmatrix} W(z - X\beta - Ub) + \begin{bmatrix} 0 \\ -D^{-1}b \end{bmatrix} \\ &= \begin{bmatrix} X' \\ U' \end{bmatrix} Wz - C'C \begin{bmatrix} \beta \\ b \end{bmatrix}, \end{aligned} \tag{A·1}$$

$$I = E(SS') = \begin{bmatrix} X'WX & X'WU \\ U'WX & U'WU + D^{-1} \end{bmatrix} = C'C, \tag{A·2}$$

where all quantities have been defined in §§ 1 and 2. Given current estimates $\beta^{(m-1)}$ and $b^{(m-1)}$, Fisher's scoring method yields new estimates $\beta^{(m)}$ and $b^{(m)}$ as the solutions to the system of equations

$$I \begin{bmatrix} \beta^{(m)} \\ b^{(m)} \end{bmatrix} = I \begin{bmatrix} \beta^{(m-1)} \\ b^{(m-1)} \end{bmatrix} + S.$$

Using (A·1) and (A·2) this leads to

$$C'C \begin{bmatrix} \beta^{(m)} \\ b^{(m)} \end{bmatrix} = \begin{bmatrix} X' \\ U' \end{bmatrix} Wz,$$

which are the normal equations equivalent to (2·4).

## References


ANDERSON, D. A. & AITKIN, M. (1985). Variance component models with binary response. *J. R. Statist. Soc.* B **47**, 203-10.

FELLNER, W. H. (1986). Robust estimation of variance components. *Technometrics* **28**, 51-60.

FELLNER, W. H. (1987). Sparse matrices, and the estimation of variance components by likelihood methods. *Comm. Statist.* B **16**, 439-63.

GILMOUR, A. R., ANDERSON, R. D. & RAE, A. L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika* **72**, 593-9.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *J. Am. Statist. Assoc.* **72**, 320-40.

HENDERSON, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, National Research Council Publication No. 892, pp. 141-63. National Academy of Sciences.

IM, S. & GIANOLA, D. (1988). Mixed models for binomial data with an application to lamb mortality. *Appl. Statist.* **37**, 196-204.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of interblock information when the block sizes are unequal. *Biometrika* **58**, 545-54.

STIRATELLI, R., LAIRD, N. & WARE, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40**, 961-71.

ZEGER, S. L., LIANG, K. Y. & ALBERT, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049-60.


[*Received July* 1990. *Revised April* 1991]