Original article

Inference for threshold models with variance components from the generalized linear mixed model perspective

B Engel [1], W Buist [2], A Visscher [2]

[1] *DLO Agricultural Mathematics Group (GLW-DLO),*
PO Box 100, 6700 AC Wageningen;
[2] *DLO Institute for Animal Science and Health (ID-DLO),*
PO Box 501, 3700 AM Zeist, The Netherlands

Summary – The analysis of threshold models with fixed and random effects and associated variance components is discussed from the perspective of generalized linear mixed models (GLMMs). Parameters are estimated by an interative procedure, referred to as iterated re-weighted REML (IRREML). This procedure is an extension of the iterative re-weighted least squares algorithm for generalized linear models. An advantage of this approach is that it immediately suggests how to extend ordinary mixed-model methodology to GLMMs. This is illustrated for lambing difficulty data. IRREML can be implemented with standard software available for ordinary normal data mixed models. The connection with other estimation procedures, eg, the maximum *a posteriori* (MAP) approach, is discussed. A comparison by simulation with a related approach shows a distinct pattern of the bias of MAP and IRREML for heritability. When the number of fixed effects is reduced, while the total number of observations is kept about the same, bias decreases from a large positive to a large negative value, seemingly independently of the sizes of the fixed effects.

binomial data / threshold model / variance components / generalized linear model / restricted maximum likelihood

Résumé – *Inférence sur les composantes de variance des modèles à seuil dans une perspective de modèle linéaire mixte généralisé. L'analyse des modèles à seuils avec effets fixes et aléatoires et des composantes de variance correspondantes est ici placée dans la perspective des modèles linéaires mixtes généralisés (GLMMs). Les paramètres sont estimés par une procédure itérative, appelée maximum de vraisemblance restreinte repondéré obtenu par itération (IRREML). Cette procédure est une extension de l'algorithme itératif des moindres carrés repondérés pour les modèles linéaires généralisés. Elle a l'avantage de suggérer immédiatement une manière d'étendre la méthodologie habituelle du modèle mixte aux GLMMs. Une application à des données de difficultés d'agnelage est présentée. IRREML peut être mis en œuvre avec les logiciels standard disponibles pour les modèles linéaires mixtes normaux habituels. Le lien avec d'autres procédures d'estimation, par exemple l'approche du maximum a posteriori (MAP), est discuté. Une comparaison*

par simulation avec une méthode voisine montre un biais caractéristique du MAP et de l'IRREML pour l'héritabilité. Quand le nombre des effets fixés est diminué, à nombre to- tal d'observations constant, le biais passe d'une valeur fortement positive à une valeur fortement négative, apparemment indépendantes de l'importance des effets fixés.

distribution binomiale / modèle à seuil / composante de variance / modèle linéaire généralisé / maximum de vraisemblance restreinte

INTRODUCTION

In his paper on sire evaluation Thompson (1979) already pointed out the potential interest for binomial data in modifying the generalized linear model (GLM) esti- mating equations to allow for random effects. He conjectured that if modification is feasible, generalization towards other distributions such as the Poisson or gamma distribution should be easy. The iterated re-weighted restricted maximum likeli- hood (IRREML) procedure (Schall, 1991; Engel and Keen, 1994) for generalized linear mixed models (GLMM) proves to be exactly such a modification. IRREML is motivated by the fact that in GLMMs the adjusted dependent variate in the iterated re-weighted least squares (IRLS) algorithm (McCullagh and Nelder, 1989, § 2.5) approximately follows an ordinary mixed-model structure with weights for the residual errors and, in the absence of under- or overdispersion, residual error variance fixed at a constant value (typically 1). IRREML is quite flexible and not only covers a variety of underlying distributions for the threshold model but also easily extends to other types of data such as count data, for example, litter size. This entails simple changes in the algorithm with respect to link and variance func- tion employed. When the residual error variance for the adjusted dependent variate is not fixed, it represents an additional under- or overdispersion parameter which is a useful feature, for example, under- or overdispersed Poisson counts. Calculations in this paper are performed with REML (Patterson and Thompson, 1971) facilities for ordinary mixed models in Genstat 5 (1993). Software for animal models such as DFREML (Meyer, 1989), after some modification, can be used for IRREML as well.

Methods for inference in ordinary normal data mixed models, *eg*, the Wald test (Cox and Hinkley, 1974, p 323) for fixed effects, are also potentially useful for GLMMs, as will be illustrated for the lambing difficulty data. Simulation results for the Wald test in a GLMM for (overdispersed) binomial data were presented in Engel and Buist (1995).

For threshold models with normal underlying distributions and known compo- nents of variance, Gianola and Foulley (1983) observe that their Bayesian maximum *a posteriori* (MAP) approach produces estimating equations for fixed and random effects such as those anticipated by Thompson. Under normality assumptions, for fixed components of variance, IRREML will be shown to be equivalent to MAP. IRREML therefore offers an alternative, non-Bayesian, derivation of MAP. The MAP approach was also presented in Harville and Mee (1984), including estima- tion of variance components. Their updates of the components of variance are akin to those of the estimation maximization (EM) algorithm (Searle *et al*, 1992, § 8.3) for REML. The algorithm presented in Engel and Keen (1994), which is used in this

paper, is related to Fisher scoring. Both algorithms solve the same final estimating equations, but the latter is considerably faster than the former.

Gilmour *et al* (1985) presented an iterative procedure for threshold models with normal underlying distributions, which also uses an adjusted dependent variate and residual weights. This approach, which will be referred to as GAR, is different from MAP and IRREML. In the terminology of Zeger *et al* (1988) MAP and IRREML are closely related to the subject-specific nature of the GLMM, while GAR is of a population-averaged nature, as will be explained in more detail in this paper.

A number of authors, *eg*, Preisler (1988), Im and Gianola (1988), and Jansen (1992), have discussed maximum likelihood estimation for threshold models. Apart from the fact that straightforward maximum likelihood estimation does not correct for loss of degrees of freedom due to estimation of fixed effects, as REML does in the conventional mixed model, it is also handicapped by the need for high-dimensional numerical integration. Maximum likelihood estimation for models with several components of variance, especially with crossed random effects, is practically impossible. IRREML is more akin to quasi-likelihood estimation (McCullagh and Nelder, 1989, chap 9; McCullagh, 1991): conditional upon the random effects only; the relationship between the first 2 moments is employed while no full distributional assumptions are needed beyond existence of the first 4 moments.

Since practical differences between various methods proposed pertain mainly to their subject-specific or population-averaged nature, we will give some attention to a comparison between GAR and IRREML. Simulation studies were reported in Gilmour *et al* (1985), Breslow and Clayton (1993), Hoeschele and Gianola (1989), and Engel and Buist (1995). Conclusions from the Hoeschele and Gianola study differ from conclusions from the other studies with respect to bias of MAP/IRREML and GAR. Since the Hoeschele and Gianola study was rather modest in size, it was decided to repeat it here in more detail, *ie* under a variety of parameter configurations and for larger numbers of simulations.

GLMMs and threshold models

The GLMM model

Suppose that random effects are collected in a random vector $\mathbf{u}$, with zero means and dispersion matrix $\mathbf{G}$, *eg*, for a sire model $\mathbf{G} = \mathbf{A}\sigma_s^2$, where $\mathbf{A}$ is the additive relationship matrix and σ_s^2 the sire component of variance. Conditional upon $\mathbf{u}$, *eg*, for given sires, observations y are assumed independent, with variances proportional to known functions V of the means μ:

$$\mathrm{E}(y|u) = \mu \quad \text{and} \quad \mathrm{Var}(y|u) = \phi V(\mu) \qquad [1]$$

For binary data, $y = 1$ may denote a difficult birth and $y = 0$ a normal birth. The mean μ is the probability of a difficult birth for offspring of a particular sire. The conditional variance is $\mathrm{Var}(y|u) = V(\mu) = \mu(1 - \mu)$ and ϕ equals 1. For proportions $y = x/n$, an appropriate choice may be:

$$\mathrm{Var}(y|u) = \phi V(\mu) = \phi\mu(1 - \mu)/n \qquad [2]$$

Parameter ϕ may be included to allow for under- or overdispersion relative to binomial variation (McCullagh and Nelder, 1989, § 4.5). Observe that [2] is inappropriate when n is predominantly small or large: for $n = 1$ no overdispersion is possible and ϕ should equal 1 and for $n \to \infty$, [2] vanishes to 0 while extra-binomial variation should remain. More complicated variances (Williams, 1982) may be obtained by replacing ϕ in [1] by $\{1 + (n-1)\sigma_0^2\}$ or by $\{1 + (n-1)\sigma_0^2\mu(1-\mu)\}$. In both expressions σ_0^2 is a variance corresponding to a source of overdispersion (for a discussion of underdispersion see Engel and Te Brake, 1993). Limits for $n \to \infty$ of the variances are $\sigma_0^2\mu(1-\mu)$ and $\sigma_0^2\mu^2(1-\mu)^2$, respectively. Both can be accomodated in a GLMM for continuous proportions, *eg*, motility of spermatozoa, and are covered by IRREML.

The mean μ is related to a linear predictor η by means of a known link function g: $\eta = g(\mu)$. The linear predictor is a combination of fixed and random effects: $\eta = \mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u}$, where $\mathbf{x}$ and $\mathbf{z}$ are design vectors for fixed and random effects collected in vectors $\boldsymbol{\beta}$ and $\mathbf{u}$, respectively. For difficulty of birth, for instance, η may include main effects for parity of the dam and a covariable for birthweight as fixed effects and the genetic contribution of the sire as a random effect. Popular link functions for binary or binomial data are the logit and probit link functions: $\text{logit}(\mu) = \log(\mu/(1-\mu)) = \eta$ and $\text{probit}(\mu) = \Phi^{-1}(\mu) = \eta$, where Φ^{-1} is the inverse of the cumulative density function (cdf) of the standard normal distribution.

The threshold model

Suppose that r is the 'liability', an underlying random variable such that $y = 1$ when r exceeds a threshold value θ and $y = 0$ otherwise. Without loss of generality it may be assumed that $\theta = 0$. Let η be the mean of r, conditional upon u. Furthermore, let the cdf of the residual $\varepsilon = (r - \eta)$, say F, be independent of u. Then

$$\mu = P(y = 1|u) = P(r > 0|u) = 1 - F(-\eta) \quad \text{and so} \quad \eta = -F^{-1}(1 - \mu)$$

where F^{-1} is the inverse of F. It follows that the threshold model is a GLMM with link function $g(\mu) = -F^{-1}(1 - \mu)$, which simplifies to $g(\mu) = F^{-1}(\mu)$ when ε is symmetrically distributed. Residual ε may represent variation due to Mendelian sampling and environment. Probabilities μ do not change when r is multiplied by an arbitrary positive constant and the variance of ε can be fixed at any convenient constant value, say σ^2. When F is the cdf L of the standard logistic distribution, *ie* $F(\varepsilon) = L(\varepsilon) = 1/(1 + \exp(-\varepsilon))$, g is the logit link and $\sigma^2 = \pi^2/3$. When F is the cdf Φ of the standard normal distribution, g will be the probit link and $\sigma^2 = 1$. Although the logistic distribution has relatively longer tails than the normal distribution, to a close approximation (Jonhson and Kotz, 1970, p 6):

$$\Phi(t) = L(ct) \tag{3}$$

where $c = (15/16)\pi/\sqrt{3}$. Results of analyses with a probit or logit link are usually virtually equivalent, apart from the scaling factor c for the effects and c^2 for the components of variance. Heritability may be defined on the liability scale, *eg*, for a sire model: $h^2 = 4\sigma_s^2/(\sigma_s^2 + \sigma^2)$. As a function of σ_s^2/σ^2 heritability does not depend on the choice of σ^2. Hence, estimates $\widehat{h}^2$ for the probit and logit link are often about the same.

Conditional and marginal effects

In a GLMM, effects are introduced in the link-transformed conditional means, *ie* in the linear predictor $\eta = g(\mu)$. Consequently, effects refer to subjects or individuals. The GLMM and the threshold model are both subject-specific models, using the terminology of Zeger *et al* (1988). This is in contrast with a population-averaged model where effects are introduced in the link-transformed marginal means $g(E(\mu))$ and refer to the population as a whole. In animal breeding, where sources of variation have a direct physical interpretation and are of primary interest, a subject-specific model, which explicitly introduces these sources of variation through random effects, seems the natural choice. For fixed effects however, presentation in terms of averages over the population is often more appropriate. In the threshold model, there is no information in the data about the phenotypic variance of the liability, allowing σ^2 to be fixed at an arbitrary value. Intuitively one would expect the expressions for marginal effects to involve some form of scaling by the underlying phenotypic standard deviation. For normally distributed random effects and probit link this is indeed so. From $r \sim N(\mathbf{x}'\boldsymbol{\beta}, \mathbf{z}'\mathbf{Gz} + 1)$ the marginal probability, say p, follows directly:

$$p = P(y = 1) = P(r > 0) = \Phi(\mathbf{x}'\boldsymbol{\beta}(\mathbf{z}'\mathbf{Gz} + 1)^{-0.5})$$

Hence, the probit link also holds for marginal probabilities, but the effects are shrunken by a factor $\lambda_p = (\mathbf{z}'\mathbf{Gz} + 1)^{-0.5}$. For a sire model $\lambda_p = (\sigma_s^2 + 1)^{-0.5}$. That the same link applies for both conditional means μ and marginal means p is rather exceptional. For the logit link, the exact integral expression for p cannot be reduced to any simple form (Aitchison and Shen, 1980). However, from [3] it follows that the logit link holds approximately for p, with shrinkage factor $\lambda_L = ((\mathbf{z}'\mathbf{Gz}/c^2)+1)^{-0.5}$. Without full distributional assumptions, for relatively small components of variance, marginal moments may also be obtained by a Taylor series expansion (see Engel and Keen, 1994).

Binary observations y_i and y_j corresponding to, for instance, the same sire will be correlated. For the probit link the covariance follows from:

$$\begin{aligned}
\mathrm{cov}(y_i, y_j) &= E_u(\mathrm{cov}(y_i, y_j|u)) + \mathrm{cov}_u(E(y_i|u), E(y_j|u)) = \mathrm{cov}_u(\mu_i, \mu_j) \\
&= E_u[P(r_i > 0, r_j > 0|u)] - p_i p_j = P(r_i > 0, r_j > 0) - p_i p_j \\
&= \Phi_2(\lambda_p \mathbf{x}_i'\boldsymbol{\beta}, \lambda_p \mathbf{x}_j'\boldsymbol{\beta}; \rho_{ij}) - p_i p_j \qquad\qquad [4]
\end{aligned}$$

Here $\Phi_2(a, b; \rho)$ is the cdf of the bivariate normal distribution with zero means, unit variances and correlation coefficient ρ, ρ_{ij} is the correlation on the underlying scale, *eg*, in a simple sire model $\rho_{ij} = \sigma_s^2/(\sigma_s^2 + 1)$. For the logit link, using [3], λ_p should be replaced by λ_L/c, while the value of the correlation, expressed in terms of the components of variance in the logit model, is about the same. The double integral in Φ_2 may effectively be reduced to a single integral (Sowden and Ashford, 1969), which can be evaluated by Gauss quadrature (Abramowitz and Stegun, 1965, p 924). Alternatively, for small ρ^2, a Taylor expansion (Pearson, 1901; Abramowitz

and Stegun, 1965, 26.3.29, p 940) may be used:

$$\Phi_2(a,b;\rho) = \Phi(a)\Phi(b) + \tau(a)\tau(b)\rho + \sum_{t=1}^{\infty} \tau^{(t)}(a)\tau^{(t)}(b)\rho^{t+1}/(t+1)! \qquad [5]$$

where $\tau^{(t)}$ is the tth derivative of the probability density function (pdf) τ of the standard normal distribution. For a sire model, under normality assumptions, the first-order approximation appears to be satisfactory, except for extreme incidence rates p (Gilmour *et al*, 1985). By grouping of n binary observations pertaining to the same fixed and random effect, moments for binomial proportions y immediately follow from [4], *eg*:

$$\mathrm{Var}(y) = p(1-p)[1+(n-1)\{\Phi_2(\lambda_p \mathbf{x}'\boldsymbol{\beta}, \lambda_p \mathbf{x}'\boldsymbol{\beta}; \rho) - p^2\}/p(1-p)]/n \qquad [6]$$

where ρ is the intra-class correlation on the liability scale. Expression [6] can be simplified by using [5]. Results for the logit link follow from [3].

Estimation of parameters

The algorithm for IRREML

The algorithm will be described briefly. For details see Engel and Keen (1994) and Engel and Buist (1993a). Suppose that $\boldsymbol{\beta}_0$ and $\mathbf{u}_0$ are starting values obtained from an ordinary GLM fit with, for example, random effects treated as if they were fixed or with random effects ignored, *ie* $\mathbf{u}_0 = 0$. After the initial GLM has been fitted by IRLS, the adjusted dependent variate ζ and iterative weights w (McCullagh and Nelder, 1989, § 2.5) are saved:

$$\zeta = (y - \mu_0)g'(\mu_0) + g(\mu_0) \quad \text{and} \quad w = \{g'(\mu_0)^2 V(\mu_0)\}^{-1} \qquad [7]$$

where g' is the derivative of the link function with respect to μ, *eg*, for the probit link: $w = n\tau(\eta_0)^2/\{\mu_0(1-\mu_0)\}$, ζ approximately follows an ordinary mixed-model structure with weights w for the residual errors and residual variance ϕ. Now a minimum norm quadratic unbiased estimation (MINQUE) (Rao, 1973, § 4j) is applied to ζ, employing the Fisher scoring algorithm for REML (1 step of this algorithm corresponds to MINQUE). From the mixed-model equations (MMEs) (Henderson, 1963; Searle *et al*, 1992, § 7.6) new values $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ for the fixed and random effects are solved:

$$\begin{bmatrix} \mathbf{X'WX} & \mathbf{X'WZ} \\ \mathbf{Z'WX} & \mathbf{Z'WZ} + \phi\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'W}\zeta \\ \mathbf{Z'W}\zeta \end{bmatrix}$$

Here, $\mathbf{X}$ and $\mathbf{Z}$ are the design matrices for the fixed and random effects respectively, $\mathbf{W}$ is a diagonal matrix with weights w along the diagonal and ζ denotes the vector of values of the adjusted dependent variate. $\boldsymbol{\beta}_0$ and $\mathbf{u}_0$ are replaced by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$, ζ and w are updated and a new MINQUE step is performed. This is repeated until convergence. Note that MINQUE does not require full

distributional assumptions beyond the existence of the first 4 moments and may be presented as a weighted least-squares method (Searle *et al*, 1992, Ch 12).

Some properties of IRREML

When the MMEs are expressed in terms of the original observations y, it is readily shown that at convergence the following equations are solved:

$$\mathbf{X}'\widehat{\mathbf{W}}(g'(\boldsymbol{\mu})^*(\mathbf{y} - \boldsymbol{\mu})) = 0,$$
$$\mathbf{Z}'\widehat{\mathbf{W}}(g'(\boldsymbol{\mu})^*(\mathbf{y} - \boldsymbol{\mu})) = \phi\,\mathbf{G}^{-1}\mathbf{u} \qquad [8]$$

where * denotes a direct elementwise (Hadamard) product. These equations are similar to the GLM equations for fixed $\mathbf{u}$ (with appropriate side conditions) except for the term $\phi\,\mathbf{G}^{-1}\mathbf{u}$ on the right-hand side. Equations [8] may also be obtained by setting first derivatives with respect to elements of $\boldsymbol{\beta}$ and $\mathbf{u}$ of $D + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u}$ equal to zero, where D is the (quasi) deviance (see McCullagh and Nelder, 1989, § 2.3 and § 9.2.2.) conditional upon $\mathbf{u}$. The assumption of randomness for $\mathbf{u}$ imposes a 'penalty' on values which are 'too far' from 0. When the pdf of observations $\mathbf{y}$ conditional upon normally distributed random effects $\mathbf{u}$ is in the GLM exponential family, *eg*, a binomial or Poisson distribution, maximization of $D + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u}$ is easily shown to be equivalent to maximization of the joint pdf of $\mathbf{y}$ and $\mathbf{u}$.

Suppose that we have a sire model with q sires and sire variance component σ_s^2. The IRREML estimating equations for σ_s^2 and ϕ (see, for example, Engel, 1990) are:

$$\text{trace}(\mathbf{Z}_k'\mathbf{P}\mathbf{Z}_k\mathbf{A}_k) = \boldsymbol{\zeta}'\mathbf{P}\mathbf{Z}_k\mathbf{A}_k\mathbf{Z}_k'\mathbf{P}\boldsymbol{\zeta},\ k = 0,\ 1 \qquad [9]$$

Here $\mathbf{Z}_0 = \mathbf{I}$, $\mathbf{Z}_1$ is the design matrix for the sires, $\mathbf{A}_0 = \mathbf{W}^{-1}$, $\mathbf{A}_1 = \mathbf{A}$, $\mathbf{P} = \Omega^{-1} - \Omega^1\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}$ and $\Omega = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \phi\mathbf{W}^{-1}$. The difference with ordinary REML equations is that $\boldsymbol{\zeta}$ depends on the parameter values as well. The MINQUE/Fisher scoring update of IRREML can be recovered from [9], by using $\mathbf{P} = \mathbf{P}\Omega\mathbf{P} = \sigma_s^2\mathbf{P}\mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'\mathbf{P} + \phi\mathbf{P}\mathbf{W}^{-1}\mathbf{P}$ on the left-hand side:

$$\sum_k \text{trace}(\mathbf{Z}_{k'}'\mathbf{P}\mathbf{Z}_k\mathbf{A}_k\mathbf{Z}_k'\mathbf{P}\mathbf{Z}_{k'}\mathbf{A}_{k'})\sigma_k^2 = \boldsymbol{\zeta}'\mathbf{P}\mathbf{Z}_{k'}\mathbf{A}_{k'}\mathbf{Z}_{k'}'\mathbf{P}\boldsymbol{\zeta},\ k' = 0,\ 1 \qquad [10]$$

where $\sigma_0^2 = \phi$ and $\sigma_1^2 = \sigma_s^2$. When ϕ is fixed at value 1, the equation for $k' = 0$ is dropped from [10]. Alternative updates related to the EM algorithm may also be obtained from [9] (see, for example, Engel, 1990), and will be of interest when other estimation procedures are discussed:

$$\widehat{\sigma}_s^2 = \{\widehat{\mathbf{u}}'\mathbf{A}^{-1}\widehat{\mathbf{u}} + \text{trace}(\mathbf{A}^{-1}\mathbf{T})\}/q \qquad [11]$$

Here $\mathbf{T}/\phi$ is the part of the inverse of the MME coefficient matrix corresponding to $\mathbf{u}$.

With quasi-likelihood (QL) for independent data, it is suggested (McCullagh and Nelder, 1989, § 4.5 and chap 9) that one can estimate ϕ from Pearson's (generalized) chi-square statistic. From [9] it may be shown that $\widehat{\phi} = X_p^2/d$, where $X_p^2 = \sum_j (y_j - \widehat{\mu}_j)^2/V(\widehat{\mu}_j)$ is Pearson's chi-square in terms of conditional means and

variances and $d = N - \text{rank}(\mathbf{X}) - \{q - \text{trace}(\mathbf{A}^{-1}\mathbf{T}/\hat{\sigma}_s^2)\}$ is an associated 'number of degrees of freedom'.

Application to birth difficulties in sheep

The data are part of a study into the scope for a Texel sheep breeding program in the Netherlands employing artificial insemination. Lambing difficulty will be analyzed as a binary variable: 0 for a normal birth and 1 for a difficult birth. There are 43 herd-year-season (HYS) effects. Herds are nested within regions and regions are nested within years. There are 2 years, 3 regions per year, about 4 herds per region, and 3 seasons. The 33 sires are nested within regions. The 433 dams are nested within herds with about 20 dams per herd. Observations are available from 674 offspring of the sires and dams.

Variability on the liability scale may depend on litter size. Therefore, observations corresponding to a litter size of 1 and litter sizes of 2 or more are analyzed separately. Corresponding data sets are referred to as the S-set (single; 191 observations) and M-set (multiple; 483 observations). The M-set is reproduced in Engel and Buist (1993) and is available from the authors. Some summary statistics are shown in table I.

Table I. Summary statistics for the lambing difficulty data and 'raw' fractions for difficult births and corresponding numbers of progeny for the S- and M-set.

	S-set		*M-set*	
	Mean	*Number of progeny*	*Mean*	*Number of progeny*
Year 1	0.76	59	0.50	204
Year 2	0.84	132	0.77	279
Season 1	0.78	93	0.64	286
Season 2	0.83	84	0.71	180
Season 3	0.93	14	0.41	17
Overall	0.82	191	0.66	483

Table II shows some results for components of variance, for models fitted to the S- and M-sets. Dam effects are absorbed. To stabilize convergence, the occurrence of extreme weights was prevented by limiting fitted values on the probit scale to the range $[-3.5, 3.5]$. In addition to fixed HYS effects, factors for age and parity of the dam (P), sex of the lamb (S), and for the M-set a covariate for litter size $(L = $ litter size $- 2)$ and included. Levels for factor P consist of the following 6 combinations of age and parity: $(1;1)$, $(2;1)$, $(2;2)$, $(3; \leqslant 2)$, $(4; \geqslant 3)$ and $(\geqslant 5; \geqslant 4)$. In models 3, 4 and 5 a factor D for pelvic dimension of the dam ('wide', 'normal' or 'narrow'), and in models 4 and 5 a covariate $W = $ birthweight $-$ average birthweight of the lamb is also included, with separate averages of 4.27 and 3.63 for the S- and M-sets respectively.

Fixed effects may be screened by applying the Wald test to the values of ζ saved from the last iteration step. Some results for the M-set are shown in

Table II. Estimated components of variance σ_S^2 and σ_D^2 for sires and dams, $h^2 = 4\sigma_S^2$ $(1 + \sigma_S^2 + \sigma_D^2)$, $\rho = \sigma_D^2/(1 + \sigma_S^2 + \sigma_D^2)$.

	σ_S^2	σ_D^2	h^2	ρ
S-set, model				
1. HYS + S + P	0.25 (0.30)		0.81 (0.79)	
2. HY + S + P	0.13 (0.21)		0.46 (0.66)	
3. HY + S + P + D	0.09 (0.24)		0.34 (0.81)	
4. HY + S + P + D + W	0.10 (0.29)		0.38 (0.95)	
5. HY + S + P + D + W + D × W	0.01 (0.26)		0.04 (1.00)	
M-set, model				
1. HYS + S + P + L + S × L	0.00 (−)	0.84 (0.25)	0.00 (−)	0.46 (0.07)
2. HY + S + P + L + S × L	0.00 (−)	0.69 (0.20)	0.00 (−)	0.41 (0.07)
3. HY + S + P + L + S × L + D	0.00 (−)	0.69 (0.21)	0.00 (−)	0.41 (0.07)
4. HY + S + P + L + S × L + D + W	0.01 (0.09)	0.71 (0.22)	0.02 (0.21)	0.42 (0.07)
5. HY + S + P + L + S × L +D + W + D × W	0.00 (0.09)	0.74 (0.22)	0.00 (0.21)	0.43 (0.07)

Standard errors are in parentheses. S, P and D denote main effects, W and L are covariables, S × L and D × W are interactions. When an estimate is negative ($-$), the component is assumed to be negligible and set to 0.

table III. In all cases, test statistics are calculated for the values of the variance components obtained for the corresponding full model, *ie* model 1–5. Variability due to estimation of the variance components is ignored. For each line in the table the corresponding test statistic accounts for effects above that line, but ignores effects below the line. Referring to a chi-square distribution, in model 1 seasonal effects seem to be unimportant and are excluded from the subsequently fitted models.

In model 3 for the M-set, the following contrasts for pelvic opening (D) are found: 0.520 (0.231), 2.019 (0.547) and 1.499 (0.531), for 'normal' *versus* 'wide', 'narrow' *versus* 'wide' and 'narrow' *versus* 'normal', respectively. Pairwise comparison, with a normal approximation, shows that any 2 levels are significantly $(P < 0.05)$ different. The effects refer to the probits of the conditional probabilities. For the probits of the marginal probabilities, effects have to be multiplied by $(1 + \sigma_S^2 + \sigma_D^2)^{-0.5} = 0.769$ (from table II). The difference between 'narrow' and 'wide', for example, becomes 1.55 (0.43). In model 5 for the M-set, separate coefficients for birthweight are fitted for the 3 levels of pelvic opening. The estimated coefficient for birthweight for a dam with a narrow pelvic opening is 0.72 (0.61); this is about 0.47 (0.63) higher than the estimates for the other 2 levels, which are about the same. Although a larger coefficient is to be expected for a narrow pelvic opening, the difference found is far from significant. Fitting a common coefficient, *ie* dropping the interaction $D \times W$ between pelvic opening and birthweight in model 4, gives an estimated coefficient for birthweight of 0.28 (0.15), which becomes 0.21 (0.12) after shrinkage. By comparison, the coefficient for the S-set, after shrinkage, is 0.92 (0.30). The reduced effect of birthweight for the M-set agrees with the negligible values found for the component of variance for sires.

Table III. Some test results for the fixed effects for the M-set.

A

	Wald statistics Model 1	Degrees of freedom
S = sex of lamb	0.4	1
P = parity / age of dam	2.8	5
L = litter size	2.2	1
S × L	0.1	1
HYS		
Years	18.0*	1
Seasons	0.9	2
Years × Regions	20.0*	4
Years × Seasons	1.0	2
Years × Regions × Herds	29.1*	16
Years × Regions × Seasons	1.3	5
Years × Regions × Seasons × Herds	4.4	12

B

	Wald statistics				Degrees of freedom
	Model 2	Model 3	Model 4	Model 5	
Years	18.7*	18.0*	16.2*	16.7*	1
Years × Regions	21.8*	22.6*	21.8*	22.0*	4
Years × Regions × Herds	46.5*	39.7*	37.6*	37.5*	16
S = sex of lamb	3.0	2.6	2.7	2.7	1
P = parity/age of dam	7.2	5.6	5.5	5.4	5
D (pelvic opening)	–	15.6*	14.6*	13.0*	2
L = litter size	2.5	2.3	2.1	2.1	1
W (birth weight lamb)	–	–	3.3	3.2	1
S × L	0.1	0.0	0.0	0.0	1
D × W	–	–	–	0.6	2

* $P < 0.05$, referring to a chi-square distribution, ignoring effects in the following rows.

Relation to other methods

We will mainly concentrate on differences between GAR and IRREML. GAR is based on QL for the marginal moments with a probit link and normally distributed random effects (see also Foulley *et al*, 1990). QL-estimating equations for dependent data are (McCullagh and Nelder, 1989, § 9.3): $\mathbf{D}'\mathrm{Var}(\mathbf{y})^{-1} (\mathbf{y} - \mathbf{p}) = \mathbf{0}$, where the matrix of derivatives $\mathbf{D} = (d_{ij})$, $d_{ij} = (\partial p_i/\partial \beta_{*j})$, follows from $\mathbf{p} = \Phi(\mathbf{X}\boldsymbol{\beta}_*)$, and $\boldsymbol{\beta}_* = \lambda_p\boldsymbol{\beta}$ denotes the vector of marginal fixed effects. It follows from [6] that:

$$\mathrm{Var}(\mathbf{y}) = \mathbf{R} + \mathbf{B}(\mathbf{ZCZ}')\mathbf{B}' + 0(\rho^2) \qquad [12]$$

where $\mathbf{R} = \mathrm{diag}(\{p(1-p) - \rho\tau^2(\lambda_p \mathbf{X}\boldsymbol{\beta})\}/n)$, $\mathbf{B} = \mathrm{diag}(\tau(\lambda_p \mathbf{X}\boldsymbol{\beta}))$, $\mathbf{C} = \lambda_p^2 \mathbf{G}$ and ρ represents the appropriate intraclass correlation. $\mathrm{Var}(\mathbf{y})$ will be replaced by [12], ignoring terms of order ρ^2 and higher. The QL equations may be solved by iterated generalized least squares on an adjusted dependent variate, say ζ_{GAR}:

$$\zeta_{\mathrm{GAR}} = \mathbf{X}\widehat{\boldsymbol{\beta}}_* + \widehat{\mathbf{B}}^{-1}(\mathbf{y} - \widehat{\mathbf{p}}) = \mathbf{X}\widehat{\boldsymbol{\beta}}_* + \mathrm{diag}(\tau(\mathbf{X}\widehat{\boldsymbol{\beta}}_*))^{-1} (\mathbf{y} - \widehat{\mathbf{p}})$$

where $\widehat{\mathbf{B}}$ is $\mathbf{B}$ evaluated at $\widehat{\boldsymbol{\beta}}_*$. By comparison, the adjusted dependent variate from [7] is:

$$\zeta = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\mathbf{u}} + \mathrm{diag}(\tau(\mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\mathbf{u}}))^{-1} (\mathbf{y} - \widehat{\boldsymbol{\mu}})$$

The latter variate relates to a first-order approximation of the conditional means μ, while the former relates to a first-order approximation of the marginal means p. Approximately $\mathrm{Var}(\zeta_{\mathrm{GAR}}) = \widehat{\mathbf{B}}^{-1}\mathrm{Var}(\mathbf{y})\widehat{\mathbf{B}}^{-1} = \widehat{\mathbf{B}}^{-1}\mathbf{R}\widehat{\mathbf{B}}^{-1} + \mathbf{Z}\mathbf{C}\mathbf{Z}' = \mathbf{W}_{\mathrm{GAR}}^{-1} + \mathbf{Z}\mathbf{C}\mathbf{Z}'$, and Gilmour *et al* (1985) solve MMEs in terms of ζ_{GAR} and $\mathbf{W}_{\mathrm{GAR}}$. For example, in a sire model with q sires, predictions from these MMEs for $\lambda_p \mathbf{u}$, say $\widehat{\mathbf{u}}_{\mathrm{GAR}}$, are used to update the intraclass correlation ρ. Analogous to [11] with $\phi = 1$:

$$\widehat{\rho}_{\mathrm{GAR}} = \{\widehat{\mathbf{u}}_{\mathrm{GAR}}' \mathbf{A}^{-1}\widehat{\mathbf{u}}_{\mathrm{GAR}} + \mathrm{trace}(\mathbf{A}^{-1}\mathbf{T}_{\mathrm{GAR}})\}/q \qquad [3]$$

Both GAR and IRREML are based on an approximate mixed-model structure for an adjusted dependent variate. Note however that in GAR, in contrast to IRREML, predictions of the random effects are only used to update the components of variance and do not enter $\mathbf{W}_{\mathrm{GAR}}$ and ζ_{GAR} directly. At the end of this section we shall see that this may be an advantage for GAR over IRREML in situations where there is little information in the data about individual random effects, *eg*, in a sire scheme with small family sizes or extreme incidence. The marginal quasi-likelihood method (MQL) presented in Breslow and Clayton (1993) can be regarded as a generalization of GAR. However, in its more general setting, MQL does not have the benefit of some of the exact results for binomial data and underlying normal distributions used to derive GAR.

In MAP, assuming a vague prior for $\boldsymbol{\beta}$ and a normal prior for $\mathbf{u}$, the posterior mean for $(\boldsymbol{\beta}', \mathbf{u}')'$ is approximated by the posterior mode. Hence, $(\widehat{\boldsymbol{\beta}}', \widehat{\mathbf{u}}')'$ maximize the joint pdf of $\mathbf{y}$ and $\mathbf{u}$. Under normality this is equivalent to maximization of a penalized deviance. Hence, the estimating equations for $\boldsymbol{\beta}$ and $\mathbf{u}$ for MAP and IRREML are the same and given by [8]. As shown in Foulley *et al* (1987) an estimator for, for instance, a sire component σ_{S}^2, may be solved from: $\mathrm{E}_0[\partial/\partial\sigma_{\mathrm{S}}^2 \log(f(\mathbf{u}|\sigma_{\mathrm{S}}^2))] = 0$. Here, $f(.|.)$ denotes a (conditional) pdf for the variables indicated, and expectation E_0 is with respect to $f(\mathbf{u}|\mathbf{y}, \sigma_{\mathrm{S}}^2)$. The latter pdf may be approximated by a normal density with mean $\widehat{\mathbf{u}}$ and dispersion $\mathbf{T}$. In an iterative scheme this leads to the EM-type update [11]. Consequently, MAP and IRREML also solve the same final estimating equations with respect to the components of variance, although the algorithms used are different.

The penalized quasi-likelihood method (PQL) presented in Breslow and Clayton (1993) is based on Laplace integration. Random effects are assumed to be normally distributed. The log pdf of $\mathbf{y}$ conditional upon $\mathbf{u}$ is replaced by a quasi-likelihood. When parameter ϕ is involved, the extended quasi-likelihood (Nelder and Pregibon,

1987) can be used, see Engel and Buist (1993). In the resulting integral expression, the logarithm of the integrand in terms of $\mathbf{u}$ is approximated by a quadratic function around its optimum in $\hat{\mathbf{u}}$, employing expected second-order derivatives. Now random effects are easily integrated out. Some further approximations, *eg*, approximating conditional deviance residuals by Pearson residuals, result in a normal log likelihood for the adjusted dependent variate ζ from [7]. An REML type adjustment for loss of degrees of freedom due to estimation of fixed effects finally yields an REML log likelihood, which is maximized by Fisher scoring. Hence, although motivated quite differently, PQL is equivalent to IRREML (for details see Engel and Buist, 1993).

A comparison between GAR and MAP/PQL/IRREML by simulation was presented in Gilmour *et al* (1985). They consider a simple one-way model, *eg*, a sire model with q unrelated sires, with n offspring per sire, and with a binary observation per offspring. An overall mean is the only fixed effect. Gilmour *et al* (1985) observe that there is a tendency for MAP/PQL/IRREML to underestimate h^2 for small family size n or extreme incidence p. For extreme incidence p, GAR tends to overestimate h^2. As also noted by Thompson (1990), for this simple set-up, closed-form expressions can be derived from [8] and [11], *viz*, $\widehat{\rho}_{\text{GAR}} = \{MS_{\text{B}} - \overline{y}(1 - \overline{y})\}/\{(n-1)\tau(\Phi^{-1}(\overline{y})^2\}$. Here MS_{B} is the mean sum of squares between sires for the binary observations and $\overline{y}$ is the overall mean. The adjusted dependent variable and weights are: $\zeta_{\text{GAR}} = \Phi^{-1}(\overline{y}) + (y - \overline{y})/\tau(\Phi^{-1}(\overline{y}))$ and $W_{\text{GAR}} = \tau(\Phi^{-1}(\overline{y}))^2/\{\overline{y}(1 - \overline{y}) - \rho\tau(\Phi^{-1}(\overline{y}))^2\}$. As long as the first-order approximation from [5] holds, and q is not too small, $\widehat{\rho}_{\text{GAR}}$ will be nearly unbiased. Actually, for this simple scheme, GAR may be shown to be equivalent to Williams' method of moments for estimating overdispersion (Williams, 1982). Starting from $\mathbf{u}_0 = 0$ and $\eta_0 = \Phi^{-1}(\overline{y})$, the adjusted dependent for IRREML is the same as for GAR, but the weights $w = \tau(\Phi^{-1}(\overline{y}))^2/\{\overline{y}(1 - \overline{y})\}$ are smaller! Consequently, the first estimate $\widehat{\sigma}_S^2 = \{MS_{\text{B}} - \overline{y}(1 - \overline{y})\}/\{n\tau(\Phi^{-1}(\overline{y}))^2\}$ underestimates σ_S^2 approximately by a factor $(n - 1)/n$. Simulation results in Gilmour *et al* (1985) suggest that with further iteration underestimation by this factor persists. For extreme incidence, say close to 1, IRREML weights are approximately $\widehat{\eta}\,\tau(\widehat{\eta})$, as follows from Abramowitz and Stegun (1965, 26.2.12, p 932). Most of the information is in the negative-valued random effects, and over-shrinkage of random effects will yield smaller weights in the next iteration, *ie* 'residual variances' will be too high. Therefore, underestimation by IRREML may be expected to be more serious for extreme incidence, even when the first-order approximation from [5] still holds. Lack of information about individual random effects, because of small family size or extreme incidence, seems a serious problem with MAP/PQL/IRREML. The smaller weights in IRREML are a consequence of 'residual variances' being derived from predicted values of conditional variances. Alternative weights $w_0 = E(g'(\mu)^2 V(\mu))^{-1}$ for IRREML are suggested in Engel and Keen (1994). For the logit link evaluation of alternative weights is straightforward: $w_0 = \{2 + 2\exp(\sigma_1^2/2)\cosh(\mathbf{x}'\boldsymbol{\beta})\}^{-1}$, but for the probit link it is problematic. It is suggested that we use $w_0 = \{g'(\widehat{\mu})^2 E(V(\mu))\}^{-1}$ instead. Note that in the first step for the simple sire scheme this weight equals w_{GAR}. Simulation results in Breslow and Clayton (1993) also show a negative bias for MAP/PQL/IRREML. However, Hoeschele and Gianola (1989) find a positive bias. Their simulation study concerns a sire model which includes 135 fixed HYS

effects. MAP/PQL/IRREML has smaller bias and smaller mean-squared error than GAR. For both methods an upward bias in the order of 20% (based on 45 simulations) of the true $h^2 = 0.25$ is observed. An upward bias is also apparent for some of the models simulated in Hoeschele *et al* (1987). Simulation results for overdispersed binomial data in Engel and Buist (1995) only indicate a serious upward bias for variance components estimated by IRREML when the true component value is small and the non-negativity constraints are active.

Simulation results

Data are generated as described in Hoeschele and Gianola (1989) (referred to as HG). The 135 HYS effects are generated from a $N(0, \sigma^2_{\text{HYS}})$ distribution. For each HYS class, a sire has 0, 1 or 2 offspring with probabilities $1 - p_{\text{HYS}}$, $p_{\text{HYS}}/2$ and $p_{\text{HYS}}/2$ respectively. The HYS effects and design for sires and offspring are generated only once, after which the same HYS values and the same design are used in all subsequent simulations. HYS effects are included in the model as fixed effects. Other fixed effects are sire group effects: 4 groups with effects -0.40, -0.15 (the value -0.10 in HG was assumed to be a typing error), 0.15 and 0.40. The numbers of sires in the groups are 12, 14, 13 and 11. The 50 independent sire effects are generated from an $N(0, \sigma^2_S)$ distribution, where $\sigma^2_S = h^2/(4 - h^2)$. The residuals for the liability are from an $N(0, 1)$ distribution. The overall constant on the probit scale is $-\Phi^{-1}(p_0)\ (1 + \sigma^2_{\text{HYS}} + \sigma^2_S)^{0.5}$, where p_0 determines the overall incidence. Suppose that σ^2_{HYS} corresponds to a proportion f_{HYS} of $1 + \sigma^2_{\text{HYS}} + \sigma^2_S$, then $\sigma^2_{\text{HYS}} = (4 f_{\text{HYS}}/(1 - f_{\text{HYS}}))/(4 - h^2)$. In HG, $h^2 = 0.25$, $f_{HYS} = 0.30$, $p_0 = 0.9$ and $p_{\text{HYS}} = 0.2$. The expected total number of records is 2 025, with about 40 offspring per sire. This is configuration 10, which is presented with some of the other configurations of parameter values studied in table IV.

For each configuration, either 1 or 2 series of 200 simulations are performed. In a series, HYS effects, sire-offspring configuration and data are all generated from a sequence of random numbers from the same seed. The first series corresponds to the same seed for all configurations. Therefore, for the first series, for example for configurations 3 and 8, the design is the same. Seeds of the second series are all different. In table IV bias (%bias) and root mean square error (%$\sqrt{\text{MSE}}$) are both expressed as a percentage of the true value of h^2. In contrast with the results in Gilmour *et al* (1985) and in agreement with Hoeschele and Gianola (1989), with one exception, MAP/PQL/IRREML shows a positive bias. The exception is configuration 13 where HYS effects are not included, neither in the generation of the data nor in the model fitted. The negative value for this configuration indicates that, although the bias seems fairly independent of the size of the fixed effects, it may depend on the (relative) number of fixed effects in the model.

Table V shows that this is indeed so. Starting from the original HG scheme, using new seeds for generating the data, the number of fixed effects is reduced by factors 1/3, 1/2, 2/3 and 3/4. A reduction of 1/3, for instance, is effected by combination of HYS effects (2, 3), (5, 6), (8, 9) and so forth, replacing the original values for 2 levels by their mean value. The first series of 200 simulations refers to the same design used in all subsequent reductions. In the second series, a new design, which is reduced afterwards, is generated for 1, 1/3, ... Bias goes from positive to negative.

B Engel *et al*

Table IV. Simulation results for the HG sire scheme for IRREML and GAR. Bias and mean squared error for estimation of the heritability on the liability scale.

Configuration	h^2	p_0	p_{HYS}	f_{HYS}	%bias IRREML		GAR		%$\sqrt{MSE}$ IRREML	GAR
1	0.25	0.60	0.15	0.30	27.7	(3.6)	22.9	(3.3)	58.5	52.0
2	0.25	0.60	0.20	0.01	7.4	(2.6)	3.8	(2.4)	39.9	34.4
					18.0	(2.6)	13.0	(2.3)	40.4	35.1
3	0.25	0.60	0.20	0.02	7.8	(2.6)	4.1	(2.4)	37.7	34.2
4	0.25	0.60	0.20	0.15	9.1	(2.7)	5.1	(2.5)	39.6	35.6
					15.5	(2.9)	10.8	(2.6)	43.5	38.4
5	0.25	0.60	0.20	0.30	11.3	(2.9)	7.0	(2.7)	42.5	38.2
					16.9	(2.6)	11.8	(2.4)	41.2	36.0
6	0.25	0.60	0.30	0.001	3.6	(2.1)	−0.6	(1.9)	29.4	26.8
					4.5	(1.9)	0.1	(1.8)	27.4	24.7
7	0.25	0.90	0.15	0.30	50.7	(4.9)	52.9	(4.9)	86.1	86.6
8	0.25	0.90	0.20	0.02	21.3	(3.9)	23.5	(4.1)	59.5	62.1
9	0.25	0.90	0.20	0.15	19.2	(3.8)	21.1	(3.9)	57.4	59.3
					24.2	(3.7)	24.3	(3.7)	58.2	58.0
10	0.25	0.90	0.20	0.30	19.2	(4.2)	20.6	(4.2)	62.5	62.6
					35.6	(4.1)	35.6	(4.0)	68.1	66.2
11	0.25	0.90	0.30	0.30	10.8	(3.1)	9.8	(3.1)	45.6	44.5
12	0.40	0.60	0.15	0.30	23.4	(2.6)	14.8	(2.2)	44.0	34.9
13	0.40	0.60	0.20	–	−5.8	(1.7)	−10.9	(1.5)	24.5	23.5
14	0.40	0.60	0.20	0.02	8.7	(2.0)	1.6	(1.8)	30.2	25.3
					13.0	(2.1)	4.4	(1.8)	32.9	26.2
15	0.40	0.60	0.20	0.15	9.1	(2.0)	1.6	(1.8)	30.1	24.9
					11.3	(2.1)	3.1	(1.8)	31.5	25.1
16	0.40	0.60	0.20	0.30	10.1	(2.2)	2.2	(1.9)	32.7	26.9
					15.0	(2.2)	6.2	(1.9)	34.8	27.4
17	0.40	0.60	0.30	0.02	3.4	(1.8)	−3.9	(1.5)	25.0	21.9
					8.7	(1.9)	0.3	(1.6)	28.0	22.5
18	0.40	0.90	0.15	0.30	36.5	(3.5)	36.6	(3.4)	61.5	60.0
19	0.40	0.90	0.20	0.02	16.8	(2.8)	16.9	(2.9)	43.2	43.8
20	0.40	0.90	0.20	0.15	15.5	(2.9)	15.6	(2.9)	43.5	44.2
					21.8	(3.0)	20.7	(3.0)	47.6	47.1
21	0.40	0.90	0.20	0.30	14.5	(3.0)	13.9	(3.0)	45.1	44.0
					18.6	(3.0)	17.7	(2.8)	45.9	43.6
22	0.40	0.90	0.30	0.30	7.7	(2.4)	4.8	(2.3)	35.3	32.7

Standard errors in parentheses; h^2 = heritability; p_0 = incidence rate; p_{HYS} = probability for a sire for progeny in a HYS class; f_{HYS} = proportion of variance explained by the HYS effects.

The negative bias agrees with the calculations in the preceding section for the simple sire model and with the results in Gilmour *et al* (1985), where an overall constant was the only fixed effect in the model. Data for configurations 10 and 16 was also generated with random HYS effects, with a new set of HYS effects for each simulation, and analysed with HYS as a random factor in the model. Estimated bias was −35.8 (2.7)% for configuration 10 and −11.8 (1.7)% for configuration 16. Corresponding root mean square errors were 52.0 and 27.3%, respectively. Bias and root mean square error are reduced when the progeny size is increased, see, for example, configurations 18, 21 and 22. Estimators are also less biased when incidence is less extreme, *eg*, configurations 3 *versus* 8 and 1 *versus* 7. Differences between series within configurations are sometimes greater than would be expected on the basis of the standard errors involved, showing that the configuration of sires and offspring is also of importance.

Table V. Simulation results for IRREML, GAR and IRREML (w_0) for the HG scheme, with smaller numbers of HYS effects.

Reduction in HYS effects	%bias IRREML	%bias GAR	IRREML (w_0)	$\%\sqrt{MSE}$ IRREML	GAR	IRREML (w_0)
Original scheme	21.7 (4.1) 18.7 (3.8)	21.2 (4.0) 19.8 (3.8)	11.6 (3.8)	62.4 58.1	60.3 57.7	54.6
1/3	0.5 (3.6) 3.8 (3.5)	3.5 (3.7) 6.3 (3.6)	−9.0 (3.1)	51.5 49.6	52.0 51.2	45.4
1/2	−5.3 (3.1) −1.9 (3.2)	−3.8 (3.1) 1.1 (3.2)	−13.8 (2.7)	43.8 44.9	44.0 45.8	40.4
2/3	−8.9 (3.2) −8.1 (2.9)	−7.7 (3.4) −5.8 (3.0)	−18.1 (2.9)	46.5 41.6	48.2 42.7	44.2
3/4	−14.9 (6.0) −9.4 (3.1)	−13.2 (6.1) −5.3 (3.3)	−20.9 (3.0)	46.0 45.3	46.8 47.2	47.7
No HYS and sire group effects	−13.1 (3.1)	−2.7 (3.6)	−14.3 (2.8)	45.4	51.1	41.8

Bias and root mean square error for estimated heritability are expressed as a percentage of the true value of h^2. Standard errors in parentheses.

GAR was programmed, similar to MAP/PQL/IRREML, in terms of Fisher scoring for the updates of the sire variance. Estimates are the same as for the original method proposed, which uses EM steps, but speed of convergence is often increased by at least a factor 10. Some results, using the same seeds as for MAP/PQL/IRREML, are presented in tables IV and V. For high incidence ($p_0 = 0.9$) results of MAP/PQL/IRREML and GAR are comparable. For moderate incidence ($p_0 = 0.6$), the bias for GAR is clearly less than for MAP/PQL/IRREML. GAR also has smaller mean square error than MAP/PQL/IRREML at the lower incidence, although the difference is less marked. In all cases, the root

mean square error is considerably larger than the corresponding bias. Plots of the estimated h^2 values of MAP/PQL/IRREML against GAR suggest a strong, fairly linear relationship. Correlations between sire predictions from MAP/PQL/IRREML and GAR are high, *eg*, over 0.99 for the first series of configurations 10 and 16.

Table V also includes results obtained for IRREML with the alternative weights $w_0 = \tau(\eta)^2/\{p(1-p) - \rho\tau(\lambda_p\eta)^2\}$ from the previous section. There is a distinct change in the bias due to the change of weights. However, although bias is considerably improved for the top lines in the table, underestimation for the bottom lines has become more serious.

In table VI, approximate standard errors of heritability estimates from MAP/PQL/IRREML are compared with standard errors estimated from the series of 200 simulations (referred to as empirical values). The approximate standard errors, based on Fisher information assuming normality for the adjusted dependent variate, perform quite well, as was also observed in Engel and Buist (1995). These standard errors are standard output in Genstat 5.

Table VI. Empirical standard errors for h^2 from the simulation compared with approximate standard errors.

Configuration	Empirical standard error	Approximate standard error	
		Mean	Percentage points
4	0.096	94.4	[75.8; 115.7]
	0.101	92.1	[71.3; 113.4]
5	0.102	92.7	[72.5; 113.4]
	0.094	103.3	[80.5; 125.0]
6	0.073	100.3	[78.7; 122.2]
	0.068	106.5	[86.5; 128.6]
9	0.135	98.7	[78.7; 119.5]
	0.132	98.3	[78.2; 119.0]
10	0.149	94.4	[75.0; 115.2]
	0.145	99.5	[81.2; 118.7]
15	0.115	101.3	[82.7; 119.8]
	0.118	98.6	[80.4; 118.8]
16	0.124	96.4	[77.8; 114.5]
	0.126	96.7	[78.0; 114.0]
20	0.163	97.2	[77.9; 114.9]
	0.169	96.6	[78.7; 114.0]
21	0.171	95.8	[79.9; 112.4]
	0.168	99.8	[81.3; 115.6]

Mean, 10 and 90% percentage points of the approximate standard errors are presented as percentages of the empirical value.

The approximately linear relationship between IRREML and GAR estimates within a configuration of parameter values, and the similar standard errors, suggest that it should be possible to correct for bias in IRREML, at the least in those cases where GAR performs well. For incidence 0.90 and the full set of 135 HYS effects, IRREML and GAR have large positive bias of similar size. In this case the direction of the bias is in line with results for GLMs, where bias correction often involves

shrinkage towards the origin. Possibly, the results of Cordeiro and McCullagh (1991), involving an extra iteration step for adjusted response variables, may be extended to minimization of the penalized deviance, thus improving estimates $\widehat{\beta}$ and predictions $\widehat{u}$. This would imply modification of the adjusted dependent variate ζ.

Users of MAP/PQL/IRREML should be aware of the problems involved when family sizes are small, or in the context of an animal model, when many animals are only weakly related. With extreme incidence, say over 0.90, and a sizeable number of HYS effects, say in the order of 6% of the number of binary observations, MAP, PQL, IRREML and GAR are liable to seriously overestimate heritability, and actual selection response may be considerably less than expected on the basis of model calculations.

ACKNOWLEDGMENTS

REFERENCES

Aitchison J, Shen SM (1980) Logistic-normal distributions: some properties and uses. *Biometrika* 67, 261-272

Abramowitz M, Stegun I (1965) *Handbook of Mathematical Functions.* Dover, NY, USA

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 88, 9-25

Cordeiro GM, McCullagh P (1991) Bias correction in generalized linear models. *J R Stat Soc B* 53, 629-643

Cox DR, Hinkley DV (1974) *Theoretical Statistics.* Chapman and Hall, London, UK

Engel B (1990) The analysis of unbalanced linear models with variance components. *Statistica Neerlandica* 44, 195-219

Engel B, Te Brake J (1993) Analysis of embryonic development with a model for under- or overdispersion relative to binomial variation. *Biometrics* 49, 269-279

Engel B, Buist W (1993) *Iterated Re-weighted REML Applied to Threshold Models with Components of Variance for Binary Data.* Agricultural Mathematics Group. Research report LWA-93-16, available from first author

Engel B, Buist W (1995) Analysis of a generalized linear mixed model: a case study and simulation results. *Biometr J* (in press)

Engel B, Keen A (1994) A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica* 48, 1-22

Foulley JL, Im S, Gianola D, Hoeschele I (1987) Empirical estimation of parameters for n polygenic binary traits. *Genet Sel Evol* 19, 197-224

Foulley JL, Gianola D, Im S (1990) Genetic evaluation for discrete polygenic traits in animal breeding. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds). Springer Verlag, Berlin, Germany, 361-409

Genstat 5 committee (1993) *Genstat 5 Release 3 Reference Manual.* (RW Payne (Chairman), PW Lane (Secretary)) Clarendon Press, Oxford, UK

Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201-223

Gilmour AR, Anderson RD, Rae AL (1985) The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 593-599

Harville D, Mee RW (1984) A mixed-model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408

Henderson CR (1963) Selection index and expected genetic advance. *NAS-NRC* 982

Hoeschele I, Gianola D (1989) Bayesian *versus* maximum quasi-likelihood methods for sire evaluation with categorical data. *J Dairy Sci* 72, 1569-1577

Hoeschele I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J An Breed Genet* 104, 334-349

Im S, Gianola D (1988) Mixed models for binomial data with an application to lamb mortality *Appl Stat* 2, 196-204

Jansen J (1992) Statistical analysis of threshold data from experiments with nested errors. *Comp Stat Data Anal* 13, 319-330

Johnson NL, Kotz S (1970) *Continuous univariate distributions. -2.* Houghton Mifflin, Boston, USA

McCullagh P (1991) Quasi-likelihood and estimating functions. *In: Statistical Theory and Modelling* (DV Hinkley, N Reid, EJ Snell, eds) Chapman and Hall, London, UK, 265-286

McCullagh P, Nelder JA (1989) *Generalized Linear Models.* Chapman and Hall, London, UK, 2nd edn

Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Sel Evol* 31, 317-340

Nelder J, Pregibon D (1987) An extended quasi-likelihood function. *Biometrika* 74, 221-232

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554

Pearson K (1901) Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measureable. *Phil Trans R Soc (Lond) A* 195, 1-47

Preisler HK (1988) Maximum likelihood estimates for binary data with random effects. *Biometr J* 3, 339-350

Rao CR (1973) *Linear Statistical Inference and its Applications.* John Wiley & Sons, New York, USA, 2nd edn

Schall R (1991) Estimation in generalized linear models with random effects. *Biometrika* 78, 719-728

Searle SR, Casella G, McCulloch CE (1992) *Variance Components.* John Wiley & Sons. New York, USA

Sowden RR, Ashford JR (1969) Computation of the bi-variate normal integral. *Appl Stat* 18, 169-180

Thompson R (1979) Sire evaluation. *Biometrics* 35, 339-353

Thompson R (1990) Generalized linear models and applications to animal breeding. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds) Springer Verlag, Berlin, Germany, 312-328

Williams DA (1982) Extra binomial variation in logistic linear models. *Appl Stat* 31, 144-148

Zeger SL, Liang KY, Albert PS (1988) Models for longitudinal data: a generalized estimating approach. *Biometrics* 44, 1049-1060