

ESTIMATION OF GENETIC PARAMETERS ARISING IN NONLINEAR MODELS

V. Ducrocq

Station de Génétique Quantitative et Appliquée
CRJ-INRA 78350 Jouy-en-Josas. France

Summary

Over the past ten years, considerable attention has been paid to the genetic analysis of binary or ordered polychotomous data. But potential applications of nonlinear models in animal breeding settings go far beyond. Examples are reviewed here. Initially, methods and algorithms for estimating dispersion parameters in nonlinear models were designed as mere extensions of procedures developed for linear models. Recent work, often resting on Bayesian methodology, has helped to justify and understand the interest and limitations of these methods. Beyond the estimation of dispersion parameters, many problems important to animal breeders and related to the interpretation and the use of these parameters remain to be solved.

Introduction

Statistics applied to animal breeding problems has for a long time concentrated almost exclusively on traits that can be described by linear models, i.e., by models for which the vector of observations \mathbf{y} is written as:

$$\mathbf{y} = \mathbf{W}\Theta + \mathbf{e} \quad (1)$$

where Θ is the vector of parameters of interest. Traditionally $\Theta = (\mathbf{b}', \mathbf{u}')$ includes fixed effects \mathbf{b} (e.g., environmental effects for which no distributional assumptions are made) and random effects \mathbf{u} (e.g., additive genetic values). \mathbf{W} is an incidence matrix relating each observation y_i in \mathbf{y} to parameters in Θ . Note that \mathbf{W} may include covariables or even nonlinear functions of covariables (e.g., squares or products) but $\mathbf{W}\Theta$ remains linear in Θ . \mathbf{e} is a vector of residuals and \mathbf{u} and \mathbf{e} are assumed to have a joint distribution which is a known function of a vector of dispersion parameters τ (variance components). There is nowadays a broad agreement on the fact that the statistical methods of choice are Best Linear Unbiased Prediction (BLUP; Henderson, 1973) for jointly estimating the fixed effects \mathbf{b} and predicting the random effects \mathbf{u} , and Restricted Maximum Likelihood (REML, Patterson and Thompson, 1971) for estimating the variance components in τ . The considerable amount of work dedicated to the study of their properties and their practical implementation has led to a worldwide understanding, acceptance and use of BLUP and REML.

However, we are in a world in which economically important traits are not always adequately described by linear models. In many situations, animal breeders did get around the difficulty of nonlinear statistical analyses by replacing observations with obvious nonlinear characteristics (growth curve points, daily egg or milk production) by elaborate observations (weight gains between two fixed ages, yearly cumulated productions). Practical efficiency, in terms of expected or realized genetic gains, has been considered satisfactory but the permanent desire to improve this efficiency, often motivated by commercial considerations, and the existence of some traits for which no reasonable function of the original observations can be adequately described by equation (1) have led to the development of new statistical techniques for a wider class of models summarizing the data as:

$$g(\mathbf{y}) = f(\Theta | \tau) + \mathbf{e} \quad (2)$$

where \mathbf{e} is not necessarily normally distributed. Often, either $g(\mathbf{y}) = \mathbf{y}$ or $f(\Theta | \tau) = \mathbf{W}\Theta$, but this is not always the case (Goto et al., 1989). If $g(\mathbf{y}) = \mathbf{y}$ and $f(\Theta | \tau)$ is a function of $\mathbf{W}\Theta$ only, (2) describes a generalized linear model with link function f^{-1} (McCullagh and Nelder, 1983).

We illustrate here a wide variety of nonlinear models with potential animal breeding applications and consider the problem of the estimation of the dispersion parameters in τ .

Nonlinear models

Discrete data

By far the most studied nonlinear models in the 80's deal with the analysis of discrete dichotomous or polychotomous data. There are at least two reasons for that : Many traits of economical importance in animal production have a discrete expression (e.g., calving scores, survival to a given age, perinatal mortality, disease prevalence,...). Also, although robust, mixed linear models directly applied to original or transformed scores (Schaeffer and Wilton, 1976; Berger and Freeman, 1978) do not adequately handle the heterogeneity of incidence rate and of variance among fixed effect subclasses (Gianola, 1982). Other approaches have been proposed to circumvent this difficulty (Van Vleck and Edlin, 1984; Beitler and Landis, 1985) but are either tedious to implement or induce undesirable constraints (Foulley, 1987, Foulley et al. in Gianola and Hammond (thereafter noted G&H), 1990).

The threshold-liability model for the analysis of binary traits or ordered polychotomies does not have these drawbacks. Originally suggested by Wright (1934), it stipulates the existence of a conceptual continuous underlying variable (called liability), whose value y_i for record i is described using a mixed linear model like in (1):

$$y_i = \mathbf{w}_i' \Theta = \mathbf{x}_i' \mathbf{b} + \mathbf{z}_i' \mathbf{u} + e_i \quad (3)$$

and $\mathbf{w}_i' = (\mathbf{x}_i', \mathbf{z}_i')$ is the i th row of $\mathbf{W} = (\mathbf{X} \ \mathbf{Z})$.

The usual assumptions about polygenic inheritance are considered to hold for this liability variable. If \mathbf{u} is the vector of additive genetic effects, $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$ with $\mathbf{G} = \mathbf{A}\sigma_u^2$ and \mathbf{A} is the relationship matrix. The e_i 's are independent normally distributed residuals with mean 0 and variance 1, hence defining, without loss of generality, a unit of measurement.

The observed discrete variable Y_i depends on the position of y_i with respect to a set of thresholds $-\infty = t_0 < t_1 < \dots < t_{I+1} = +\infty$, where $I+1$ is the total number of categories, through the condition :

$$Y_i = j \quad \text{if} \quad t_j \leq y_i < t_{j+1} \quad (4)$$

Statistically, these thresholds are not identifiable but their differences are, so one of them can be arbitrarily set to 0. In a way, model (3) mimics "traditional" linear models. Nonlinearity stems from the fact that $(y_i - t_j)$ is not observed. Also:

$$\Pr(Y_i = j | \Theta) = \Phi(t_{j+1} - \mathbf{w}_i' \Theta) - \Phi(t_j - \mathbf{w}_i' \Theta) \quad (5)$$

where Φ is the normal cumulative distribution function. In the binary case with $t_1 = 0$:

$$f(\Theta | \tau) = E(Y_i | \Theta) = \Pr(Y_i = 1 | \Theta) = \Phi(\mathbf{w}_i' \Theta) \quad (6)$$

Therefore the threshold model can be envisioned as a generalized linear model with probit link function Φ^{-1} (Thompson, 1979 ; Gianola and Foulley, 1983). To reduce computations, the normal integral has been approximated in some cases by a logistic function (Gianola and Foulley, 1983; Foulley et al., 1983; De Lorenzo and Everett, 1986) hence defining a logit link function for the generalized linear model.

Generalization of the threshold model

There are cases for which there is a strong interest in the simultaneous study of two or more discrete traits. The threshold concept is readily generalized to multiple dichotomous responses (Foulley and Gianola, 1984; Hoeschele et al., 1986; Foulley et al., 1987b). Then, one underlying liability variable is defined for each trait and the joint distributions of the random effects u and e on the underlying scales are multivariate normal. A threshold for each trait is used to convert the continuous underlying variables into discrete categories. The probability to observe a given combination of categories for an observation on all these traits requires the computation of a multivariate normal integral. As for linear models, situations for which some observations are missing on some traits can be dealt with (Foulley and Gianola, 1986).

Another important generalization of the threshold model is the joint consideration of discrete and continuous traits. A typical example is the improvement of cattle evaluations on dystocia traits using birth weight data. Then, random effects of the continuous trait(s) and of the underlying variable(s) of the discrete trait(s) are again assumed to follow a joint multivariate normal distribution (Foulley et al., 1983; Gilmour and Raadsma, 1986; Zhao, 1987; Simianer and Schaeffer, 1989).

Other discrete traits

It has been suggested (Foulley et al., 1987c; Foulley et al., 1988) that some discrete traits such as litter size or ovulation rate in some species would be better described using a Poisson model rather than a threshold model as described above. For the i th observation Y_i on such a trait, we write:

$$\Pr(Y_i = n | \Theta) = \frac{\lambda_i^n \exp(-\lambda_i)}{n!} \quad \text{with } n = 0, 1, \dots \text{ and } \lambda_i = \exp(\mathbf{w}_i' \Theta) \quad (7)$$

The Poisson model (7) is another example of a generalized linear model with a log link function. Such a model can be included in the analysis of the components of prolificacy. Ovulation rate or litter size is then described using (7) and viability is a correlated trait described using a threshold model (Foulley et al., 1987c).

Box-Cox transformation

Some continuous traits exhibit markedly skewed distributions. Examples include egg production traits in poultry, age at sexual maturity and intervals between parturitions in mammals. If nonnormality is not taken into account, biased estimation and reduced efficiency of selection can result as a consequence of a nonlinear relationship between records of related animals and of heterogeneous error variances (Ibe and Hill, 1987). In some special cases, a log transformation of the data has been proposed to achieve normality of the data and destroy the relationship between mean and variance (Hill et al., 1983) but the analysis usually proceeds as if the transformed data were the observed ones, without any further reference to the transformation. Gianola et al. (in G&H, 1990) proposed a general treatment of the Box-Cox transformation (Box and Cox, 1964) replacing y_i by:

$$g(y_i) = y_i^{[\lambda]} = (y_i - 1) / \lambda \quad \text{for } \lambda \neq 0 \quad \text{and} \quad y_i^{[\lambda]} = \log y_i \quad \text{for } \lambda = 0 \quad (8)$$

The transformation parameter λ is estimated jointly with the other parameters of interest in a model such that :

$$y_i^{[\lambda]} = \mathbf{w}_i' \Theta + e_i \quad (9)$$

In (9), the data vector itself is transformed but situations where only $E(y_i)$ is transformed can be considered, defining a generalized linear model with power link

function (Thompson *in* G&H, 1990). Simultaneous power-normal transformations of the data and of their expected value have also been suggested (Goto et al., 1989).

Survival data.

Typically, survival data are measures of the time separating an origin point (birth, first parturition) from the occurrence of a specific event, called failure (death or culling). A main problem related to life data analyses is censoring, i.e., the frequent existence of only partial records of length of life. Animals still alive at the end of the study period are censored. To avoid censoring, animal breeders have considered survival at a given age, excluding censored records and analyzing the resulting 0-1 data with linear or nonlinear (De Lorenzo and Everett, 1986 ; Hoeschele et al., 1986) models. The correct statistical treatment of partially censored data sets involves the modelling of a hazard function $h(t)$ which specifies the instantaneous failure rate at time t . The hazard function is related to the survivor function $S(t) = 1 - F(t)$ where $F(t)$ is the cumulative distribution function of failure time: $h(t) = - \partial \log S(t) / \partial t$. The most common model for the hazard function is Cox's proportional hazard model (Cox, 1972) :

$$h(t) = h_0(t) \exp (\mathbf{w}_1' \Theta) \quad (10)$$

If \mathbf{w}_1 does not vary with time, the ratio of hazards for two individuals is a constant independent of time. $h_0(t)$ is called the baseline hazard function and can either be completely arbitrary (Smith and Quaas, 1984; Smith and Allaire, 1986; Ducrocq et al., 1988a) or have a parametric form, e.g., a Weibull distribution $h_0(t) = \lambda \rho (\lambda t)^{\rho-1}$ as in Ducrocq et al. (1988b). For uncensored records, (10) leads to a linear model for the log of failure time ($g(y_i) = \log y_i$) with a residual proportional to an extreme value distribution in the case of a Weibull baseline hazard.

Note that the existence of censoring is not limited to survival data, but can affect any kind of traits. One example is age at puberty, studied by Carriquiry et al. (1987). Some individuals had reached puberty before the beginning of the study. To avoid obvious biases in the estimation, this partial information has to be incorporated in the analysis. Nonlinearity results from the fact that the probability of a censored record is a function of parameters in Θ , which are to be estimated.

Other nonlinear models

There is no doubt that nonlinear models will play a central role in the future in the analysis of traits combining major gene and polygenic inheritance. Directions for studies of the resulting mixtures of distributions have been proposed by, e.g., Hoeschele (1988) and Foulley and Elsen (1988). Similar mixtures of distributions also arise in sire evaluations with uncertain paternity (Foulley et al., 1987a).

Another important application of nonlinear models is the analysis of traits on which only the ranking of animals is available. Different approaches have been proposed. Tavernier (1989, 1990) assumed the existence of an underlying normal variable y_i for each observed rank in horse competitions. y_i was analyzed using a mixed linear model like in (1) and Θ was estimated from the rank information. Smith (in G&H, 1990) also proposed to use rank information to estimate the parameters Θ in a survival data analysis using a model like (2) with $g(y)$ being a monotone increasing, otherwise arbitrary function and $f(\Theta | \tau) = \mathbf{w}_1' \Theta$. Other topics for which linear models have been sufficiently flexible for genetic evaluation purposes but where potential improvements may be found with nonlinear approaches include the analysis of growth curves (Kachman, 1986; Laird in G&H, 1990), lactation curves (e.g, Masselin et al., 1987; Grossman and Koops, 1988) and egg production.

Estimation of dispersion parameters

General considerations

Most suggestions for the estimation of dispersion parameters in nonlinear models are based on approaches which have been successful for linear models. This may be due to the frequent strong resemblance between the mixed model equations arising in linear model analyses and the system of equations to be solved at each iteration for the estimation of location parameters when dispersion parameters are known. Location parameters in nonlinear models are usually estimated as the mode of a function which can be either the likelihood function or a joint or marginal posterior density in a Bayesian context. But the maximum of a nonlinear function is commonly computed iteratively using a second order algorithm defined at iteration k as :

$$\begin{cases} \text{Solve } \mathbf{H}^{[k]} \Delta \Theta = \mathbf{v}^{[k]} \text{ for } \Delta \Theta \\ \Theta^{[k+1]} = \Theta^{[k]} + \Delta \Theta \end{cases} \quad (11)$$

where $\mathbf{v}^{[k]}$ is the vector of first derivatives with respect to Θ of the function to maximize, evaluated at $\Theta = \Theta^{[k]}$. $\mathbf{H}^{[k]}$ is either the negative of the corresponding matrix of second derivatives (defining a Newton-Raphson-type algorithm) or its expectation (Fisher's scoring algorithm). Define $\mathbf{v}^{*[k]} = \mathbf{v}^{[k]} + \mathbf{H}^{[k]} \Theta^{[k]}$. Then, (11) is equivalent to:

$$\text{Solve } \mathbf{H}^{[k]} \Theta^{[k+1]} = \mathbf{v}^{*[k]} \text{ for } \Theta^{[k+1]} \quad (12)$$

Indeed, if $\text{var}(\mathbf{u}) = \mathbf{G}$ and $\mathbf{y}^{*[k]}$ represents a "working variable" which is a function of the data \mathbf{y} and the estimates of \mathbf{b} and \mathbf{u} at iteration k , (12) can often be written as :

$$\begin{bmatrix} \mathbf{X}' \mathbf{Q}^{[k]} \mathbf{X} & \mathbf{X}' \mathbf{Q}^{[k]} \mathbf{Z} \\ \mathbf{Z}' \mathbf{Q}^{[k]} \mathbf{X} & \mathbf{Z}' \mathbf{Q}^{[k]} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b}^{[k+1]} \\ \mathbf{u}^{[k+1]} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{Q}^{[k]} \mathbf{y}^{*[k]} \\ \mathbf{Z}' \mathbf{Q}^{[k]} \mathbf{y}^{*[k]} \end{bmatrix} \quad (13)$$

where $\mathbf{Q}^{[k]}$ is a diagonal matrix, also function of the estimates \mathbf{b} and \mathbf{u} at iteration k .

This is exactly the form of the mixed model equations introduced by Henderson (1973). Such form has been found for the iterative estimation of location parameters for most nonlinear problems mentioned above (binary data, Poisson data, normal data with censored records, mixtures of normal distributions). Similarly, analogues of the multivariate mixed model equations (Henderson and Quaas, 1976) have been reported in multitrait analyses (multiple binary responses; joint analysis of discrete and continuous traits). In some cases, (13) is only a part of the entire iterative algorithm: the parameter λ for a power-normal transformed trait in (8) is estimated at the same time as \mathbf{b} and \mathbf{u} in (13) via a specific algorithm (Gianola *in* G&H, 1990). For polychotomies, (13) is extended in a way such that the thresholds appear as "fixed effects" (Gianola and Foulley, 1983; Harville and Mee, 1984). The resemblance with linear mixed models can even be more acute: in an analysis of calving scores, Zhao (1987) used an Expectation-Maximization (EM)-type argument (Dempster et al., 1977) and a Taylor series approximation of the working variate which strongly reduce computations - the coefficient matrix in (13) remains unchanged across iterations. Note however that in survival analysis, iterative algorithms (11) or (12) cannot be easily written as (13) and that alternative maximizing strategies have been used (Cyclic maximization, Smith and Allaire, 1986; Quasi-Newton algorithm, Ducrocq, 1987).

The similarity between mixed model equations and iterates of the nonlinear maximization problem like (13) suggests that methods for estimating variance comp-

ponents in linear models could be extended to estimate τ . For example, Smith and Allaire advocated the use of an extension of Henderson's method 3 (1953) for estimating sire variance for survival data. With a threshold model, Manfredi et al. (1990) estimated 4 variance components (for sire, maternal grand-sire, cow within grand-sire and herd-year effects on dystocia) and one covariance component using an extension of Van Raden and Jung's (1988) pseudo-expectation approach, which can be validated in the case of nonlinear models using asymptotic arguments. This approach was motivated by the desire to get reasonable estimates of variance components with a large data set for a complex model which is thought to better describe the data. More sophisticated methods would have required tedious computations on smaller data sets and with less satisfactory models. However, it is well known that simple methods of moments applied to variance components estimation have few optimal properties in the case of linear models. Even their most desired one - unbiasedness - is unlikely to be obtained with nonlinear models. There is a definite need for better methods.

Since REML does have many desirable properties with linear models, its use was advocated for nonlinear models by several authors (e.g., Foulley et al., 1983; Harville and Mee, 1984). As it will be seen, the same types of algorithms as those derived for REML estimation with linear models have been suggested using arguments other than pure analogy. But this review would not be complete without considering first Gilmour's approach, whose originality does not lie so much in the variance components estimation part but rather on the way location parameters are estimated.

Quasi-likelihood procedures

As already mentioned, the threshold model can be viewed as a special case of generalized linear models with probit link function and therefore, it is natural to look at specific techniques developed for this kind of models. From model (3) with binary data ($\tau = (\sigma_u^2, 1)$), Gilmour et al. (1985) derived the marginal expectation μ_i of the observation y_i after integrating out the random effects \mathbf{u} : $\mu_i = \Phi(\mathbf{x}_i' \mathbf{b} / \sqrt{1 + \sigma_u^2})$ and they proposed an approximation \mathbf{V} of the true marginal variance-covariance matrix $\text{var}(\mathbf{y})$ of the y_i 's which can be written as:

$$\text{var}(\mathbf{y}) \approx \mathbf{V} = \mathbf{R} + \mathbf{D} \mathbf{Z} \mathbf{G} \mathbf{Z}' \mathbf{D} \quad (14)$$

where \mathbf{D} is a diagonal matrix function of $\mathbf{x}_i' \mathbf{b}$ and $\mathbf{R} = \mathbf{V} - \mathbf{D} \mathbf{Z} \mathbf{G} \mathbf{Z}' \mathbf{D}$.

Gilmour and coauthors suggested to estimate fixed effects \mathbf{b} using a quasi-likelihood procedure (McCullagh, 1983). Then, the expression of the true likelihood function is not required and is replaced by a function of the mean and the approximate variance. Instead of maximizing a complex nonlinear function via, e.g., a Newton-Raphson algorithm, estimates of \mathbf{b} are obtained as iterative weighted least squares:

$$\mathbf{X}' \mathbf{W}^{[k]} \mathbf{X} \mathbf{b}^{[k+1]} = \mathbf{X}' \mathbf{W}^{[k]} \mathbf{y} \quad (15)$$

where $\mathbf{W}^{[k]} = \mathbf{D}^{[k]} \mathbf{V}^{-1[k]} \mathbf{D}^{[k]}$ and $\mathbf{y}^{*[k]}$ is a working variable at iteration k which depends on $\mathbf{b}^{[k]}$ and σ_u^2 only. By analogy with standard BLUP results, Gilmour et al. (1985) showed that the solutions for \mathbf{b} in (15) are the same as those obtained from a system identical to (13) with $\mathbf{Q}^{[k]} = \mathbf{D}^{[k]} \mathbf{R}^{-1[k]} \mathbf{D}^{[k]}$. As a consequence, they proposed to use the solutions for \mathbf{u} in (13) as predictors of the random effects and to estimate the variance component σ_u^2 iteratively as :

$$\sigma_u^2 [k+1] = \{ \mathbf{u}^{*[k]} \mathbf{A}^{-1} \mathbf{u}^{*[k]} + \text{tr}(\mathbf{A}^{-1} \mathbf{C}_{\mathbf{u}\mathbf{u}}^{[k]}) \} / q \quad (16)$$

where q is the size of \mathbf{u} and $\mathbf{C}_{\mathbf{u}\mathbf{u}}^{[k]}$ is the $q \times q$ diagonal submatrix corresponding to random effects in the inverse of the coefficient matrix in (13). Expression (16) coincides with the expression arising in EM-type algorithms used to find REML solutions.

This reasoning was extended to the joint analysis of binary and normal data (Gilmour and Raadsma, 1986) and polychotomous data (Gilmour et al., 1987). Gilmour's approach has been discussed by various authors (Foulley, 1987; Hoeschele and Gianola, 1989; Foulley et al., Thompson, Knuiiman and Laird in G&H, 1990). Good estimates of fixed effects \mathbf{b} are thought to be obtained as a result of the approximate integration of the random effects. But Im (quoted by Hoeschele and Gianola, 1989), Foulley (1987) and Foulley et al. (in G&H, 1990) gave an exact expression of $\text{var}(\mathbf{y})$ and showed that the approximation (14) is obtained as a first order Taylor expansion of terms involving bivariate normal integrals. Exact calculations would not lead to an expression like (14) and would invalidate the use of system (13). Knuiiman and Laird (1989) indicated that estimating \mathbf{u} from (13) has no formal justification in Gilmour's approach. Similarly, a REML-type of estimator for the variance seems to be chosen more for convenience than on theoretical grounds.

The estimation of dispersion parameters in generalized linear models is indeed an area of intensive research. To account for extra random components in the linear predictor, Nelder and Pregibon (1987) have defined the concept of *extended* quasi-likelihood (see also Nelder, 1988). Then (over)dispersion of the data is modelled in a way similar to the mean. Unfortunately, Williams (1988) reported cases where maximum extended quasi likelihood estimates of dispersion parameters are inconsistent.

Bayesian estimation

The Bayes methodology offers a global and complete framework for the estimation of both location and dispersion parameters (Gianola and Fernando, 1986) which has proven to be quite successful for most types of nonlinear traits. Dispersion parameters in τ are considered as hyperparameters. A prior distribution is defined for the random effects \mathbf{u} (often, $\mathbf{u} | \tau \sim N(\mathbf{0}, \mathbf{A} \sigma_u^2)$). The prior distribution is combined with the likelihood of the data, e.g., with a product of binomials for binary data, with a function accounting for censoring in survival data, with a weighted average of likelihood functions for mixtures of distributions, etc. Assuming "flat" priors for hyperparameters and fixed effects, a joint posterior distribution $\pi(\Theta, \tau | \mathbf{y})$ is obtained, with $\Theta = (\mathbf{b}', \mathbf{u}')$.

For genetic evaluation purposes, Goffinet and Elsen (1984) showed that the selection criterion leading to maximum genetic progress is $E(\mathbf{u} | \mathbf{y})$. This requires to integrate \mathbf{b} and τ out of $\pi(\Theta, \tau | \mathbf{y})$. This is usually not feasible because the posterior distribution has no tractable form. Instead, an empirical Bayes procedure is preferred: first, an estimate $\hat{\tau}$ of τ is obtained as the mode of the marginal posterior distribution $\pi(\tau | \mathbf{y})$. Then, the posterior mean of Θ given $\tau = \hat{\tau}$ is approximated as the posterior mode of $\pi(\Theta | \mathbf{y}, \tau = \hat{\tau})$. Indeed, asymptotically (Berger, 1985):

$$\Theta | \mathbf{y}, \tau = \hat{\tau} \sim N(\hat{\Theta}, \mathbf{C}) \quad (17)$$

where $\hat{\Theta}$ and \mathbf{C} are respectively the solutions for Θ and the inverse of the coefficient matrix in (13) at convergence, when $\tau = \hat{\tau}$.

Ducrocq et al. (1988b) used a log-gamma distribution as a prior distribution for \mathbf{u} , which was later found to closely approximate a normal distribution. This made possible the actual algebraic integration of \mathbf{u} , thus leading to substantial simplifications in

the analysis. Estimates of \mathbf{b} and τ (including the Weibull parameter ρ) were then obtained as the joint mode of $\pi(\mathbf{b}, \tau | \mathbf{y})$ using a Newton-Raphson algorithm (Ducrocq, 1987) but it was not possible to obtain the true marginal posterior distribution of τ .

Foulley et al. (1987b, 1989) found an interesting result which avoids the actual integration of \mathbf{b} and \mathbf{u} out of $\pi(\Theta, \tau | \mathbf{y})$. They showed that if a flat prior is chosen for τ , the mode $\hat{\tau}$ of the marginal posterior distribution $\pi(\tau | \mathbf{y})$ satisfies the equation:

$$E \left\{ \frac{\partial}{\partial \tau} \log [\pi(\mathbf{y}, \Theta | \tau)] \right\} = 0 \tag{18}$$

where the expectation is taken over the distribution of $\Theta | \mathbf{y}, \tau$. (18) can be solved iteratively taking at iteration m the expectation in (18) with respect to the distribution of $\Theta | \mathbf{y}, \tau = \tau^{[m]}$. Then, a new estimate $\tau^{[m+1]}$ is computed as:

$$\tau^{[m+1]} = \text{Arg Max}_{\tau} \{ E_{\Theta | \mathbf{y}, \tau = \tau^{[m]}} [\log \pi(\mathbf{y}, \Theta | \tau)] \} \tag{19}$$

If we consider the situation where $\tau = (\sigma_u^2, \sigma_e^2)$ (remember that sometimes, σ_e^2 is known beforehand) and using the asymptotic normal distribution of Θ indicated in (17):

$$\Theta | \mathbf{y}, \tau = \tau^{[m]} \sim \mathbf{N}(\hat{\Theta}^{[m]}, \mathbf{C}^{[m]}) \tag{20}$$

expression (19) leads to the following iterative algorithms:

$$\sigma_u^2 [^{m+1}] = \{ \hat{\mathbf{u}}^{[m]} \mathbf{A}^{-1} \hat{\mathbf{u}}^{[m]} + \text{tr}(\mathbf{A}^{-1} \mathbf{C}_{\mathbf{u}\mathbf{u}}^{[m]}) \sigma_e^2 [^m] \} / q \tag{21}$$

$$\text{and } \sigma_e^2 [^{m+1}] = \{ \hat{\mathbf{e}}^{[m]} \hat{\mathbf{e}}^{[m]} + [p + q - \alpha \text{tr}(\mathbf{A}^{-1} \mathbf{C}_{\mathbf{u}\mathbf{u}}^{[m]})] \sigma_e^2 [^m] \} / n \tag{22}$$

Of course, for binary or Poisson data, only (21) is used. In (22), p and n are the size of \mathbf{b} and \mathbf{y} , $\alpha = (\sigma_e^2 / \sigma_u^2)^{[m]}$ and $\hat{\mathbf{e}}^{[m]} = \mathbf{y} - \mathbf{f}(\hat{\Theta}^{[m]} | \tau^{[m]})$. Indeed, there is no need to wait for actual convergence of (13) before applying (21) and (22): in practice, new values of the dispersion parameters can be computed after a few iterations of (13) (Stiratelli et al., 1984; Knuiman and Laird *in* G&H, 1989).

Algorithm (21) or its extension to the multitrait case has been reported in many instances (Harville and Mee, 1984; Foulley et al., 1987a,b,c; Foulley and Elsen, 1988; Hoeschele et al., 1987; Carriquiry et al., 1987). For binary data, (21) is identical to expression (16) from Gilmour, except that the computations of $\mathbf{u}^{[m]}$ are different.

Alternative strategies such as Newton-Raphson (Foulley et al., 1987c), Fisher-scoring (Foulley et al., 1987b), or empirical (Simianer and Schaeffer, 1989) algorithms or other expressions equivalent to (22) (e.g., Foulley et al., 1987a) have also been proposed for the estimation of residual correlations in multitrait settings.

It can be shown that formula (19) is the one arising from a generalized EM algorithm, considering \mathbf{y} as the incomplete data and \mathbf{y}, \mathbf{b} and \mathbf{u} as the complete data (Foulley et al., 1989). EM-type algorithms are known to be slow to converge. If this is the case, second-order algorithms may be obtained extending the reasoning which leads to (18) (Hoeschele et al., 1987; Foulley et al., 1989). Note also that informative priors have been considered by some authors (e.g. Gianola et al. *in* G&H, 1990), leading to expressions not very different from (21) and (22).

Conclusions

The Bayesian methodology may offer a unified approach for the estimation of dispersion parameters. But current algorithms are based on asymptotic approximations. In particular, it is essential to keep in mind that the validity of (21-22) depends

on assumption (20) (Knuiman and Laird *in* G&H, 1990). With few observations for each element in u and / or extreme nonlinearity of the distributions, this approximation may not be appropriate. Biased estimates have been reported in such cases (Gilmour et al., 1985; Hoeschele et al., 1987; Simianer and Schaeffer, 1989; Thompson *in* G&H, 1990). New ways to get more exact marginal distributions are needed.

In this review, the computational burden involved in analyses of nonlinear traits has not been considered, even though it is often deemed insuperable. This point should probably not be overemphasized, considering the rapid progress in software and hardware over the past years. Given the algorithms proposed such as (13), (21) and (22), this is typically an area where a lot can be gained from the research conducted in the context of linear models (e.g., Smith *in* G&H, 1990; Meyer, 1990).

Then, the difficulty lies on how to interpret these estimates of dispersion parameters. The heritability concept is still useful for the threshold model with a normally distributed underlying variable, for censored normal variables or mixtures of normal variables. Note that for binary traits, heritability estimates can also be derived using parent-offspring regression (Thompson et al., 1985; Im and Gianola, 1989; Thompson *in* G&H, 1990) but such estimates cannot be used to calculate correlation between relatives. Also, the heritability on the observed scale depends on the estimates of fixed effects. In the case of censored survival data, the residual variance cannot even be computed, at least without strong assumptions on the future value of fixed effects after censoring (Ducrocq, 1987). Is it sensible to report heritability in such situations ?

Other specific problems have only been partially tackled for discrete data and are unsolved in other nonlinear cases. These include, among others: a) the way results should be reported to practitioners - backtransformed evaluations are more appealing but may lose some properties essential for selection decisions (Gianola et al. *in* G&H, 1990); b) the sensitivity of the estimates to selection in the population (Gianola and Fernando, 1986); c) the computation of probability statements about future observations (Harville and Mee, 1984; Foulley et al., 1988; Foulley and Im, 1989); d) the computation of expected genetic gain from selection - which is often asymmetric and tedious to compute (Foulley, 1987) - and e) the design of optimum selection schemes.

References

- BEITLER, P. and LANDIS, J.R. 1985. *Biometrics* 41: 991-1000.
- BERGER, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New-York.
- BERGER, P.J. and FREEMAN, A.E. 1978. *J. Dairy Sci.* 61:1156-1150.
- BOX, G.E.P. and COX, D.R. 1964. *J. Roy. Stat. Soc. B*, 26 : 211-252.
- CARRIQUIRY, A.L., GIANOLA, D. and FERNANDO, R.L. 1987. *Biometrics* 43:929-940.
- COX, D.R. 1972. *J. Roy. Stat. Soc. B*, 34 : 269-276.
- DE LORENZO, M.A. and EVERETT, R.W., 1986. *J. Dairy Sci.*, 69 : 501-509.
- DEMPSTER, A.P., LAIRD, N.M. and RUBIN, R.B., 1977. *J. Roy. Stat. Soc. B*, 39 : 1-20.
- DUCROCQ, V. 1987. An analysis of length of productive life in dairy cattle. Ph.D. Thesis. Cornell University, Ithaca, New-York.
- DUCROCQ, V., QUAAS, R.L., POLLAK, E.J. and CASELLA, G. 1988a. *J. Dairy Sci.*, 71 : 3061-3070.
- DUCROCQ, V., QUAAS, R.L., POLLAK, E.J. and CASELLA, G. 1988b. *J. Dairy Sci.*, 71 : 3071-3079.
- FOULLEY, J.L., GIANOLA, D. and THOMPSON, R. 1983. *Génét. Sé. Evol.*, 15:407-424.
- FOULLEY, J.L. and GIANOLA, D., 1984. *Génét. Sé. Evol.*, 16 : 285-306.
- FOULLEY, J.L. and GIANOLA, D., 1986. *J. Dairy Sci.*, 69 : 2681 - 2695.
- FOULLEY, J.L. 1987. Méthodes d'évaluation des reproducteurs pour des caractères discrets à déterminisme polygénique en sélection animale. Thèse d'Etat. Université Paris-Sud.
- FOULLEY, J.L., GIANOLA, D. and PLANCHENAULT, D. 1987a. *Génét. Sé. Evol.*, 19: 83-102.
- FOULLEY, J.L., IM, S., GIANOLA, D. and HOESCHELE, I. 1987b. *Génét. Sé. Evol.*, 19: 197-204.
- FOULLEY, J.L., GIANOLA, D. and IM, S. 1987c. *Theor. Appl. Genet.*, 73 : 870 - 877.
- FOULLEY, J.L. and ELSEN, J.M., 1988. *Génét. Sé. Evol.*, 20 : 227-238.

- FOULLEY, J.L., GIANOLA, D., IM, S. and MISZTAL, I., 1988. *in* Biométrie et données discrètes. Société Française de Biométrie.
- FOULLEY, J.L. and IM, S. 1989. Génét. Sél. Evol., 21, 359-376.
- FOULLEY, J.L., GIANOLA, D. and IM, S. 1989. 47th session. International Statistical Institute. I :337-338.
- GIANOLA, D. 1982. J. Anim. Sci., 54 : 1079 - 1096.
- GIANOLA, D. and FOULLEY, J.L. 1983. Génét. Sél. Evol., 15: 201-224.
- GIANOLA, D. and FERNANDO, R.L. 1986. J. Anim. Sci., 63 : 217-244.
- GIANOLA, D. and HAMMOND K. (eds) 1990. Advances in Statistical Methods for Genetic Improvement of Livestock. International Symposium. Armidale. Australia. Feb.1987. Springer-Verlag, Heidelberg.
- GILMOUR, A.R., ANDERSON, R.D. and RAE, A.L. 1985. Biometrika, 72 : 593-599.
- GILMOUR, A.R. and RAADSMA, H.1986. Proc. 3rd World Congr. Genet. Appl. Livest. Prod. 12: 460-463.
- GILMOUR, A.R., ANDERSON, R.D. and RAE, A.L. 1987. J. Anim. Breed. Genet., 104 : 149-155.
- GOFFINET, B. and ELSEN, J.M. 1984. Génét. Sél. Evol. 16 : 307 -318.
- GOTO, M., TASAKI, T. and TSUCHIYA, Y. 1989. 47th session. International Statistical Institute. I :391-392
- GROSSMAN, M. and KOOPS, W.J. 1988. J. Dairy Sci. 71 : 1598 - 1608.
- HARVILLE, D.A. and MEE, R.W. 1984. Biometrics, 40 : 393-408.
- HENDERSON, C.R. 1953. Biometrics. 9 : 226 - 252.
- HENDERSON, C.R. 1973. Proc. Anim. Breed. Genet. Symp. in Honor of Dr Lush. ADSA -ASAS, 10-41.
- HENDERSON, C.R. and QUAAS, R.L., 1976. J. Anim. Sci., 43: 1188-1197.
- HILL, W.G., EDWARDS, M.R., AHMED, M.K.A. and THOMPSON, R. 1983. Anim. Prod. 36 : 59-68.
- HOESCHELE, I., FOULLEY, J.L., COLLEAU, J.J. and GIANOLA, D. 1986. Génét. Sél. Evol., 18, 299-320.
- HOESCHELE, I., GIANOLA, D. and FOULLEY, J.L. 1987. J. Anim. Breed. Genet., 104 : 334-349.
- HOESCHELE, I. 1988. Theor. Appl. Genet., 76: 311-319.
- HOESCHELE, I. and GIANOLA, D. 1989. J. Dairy Sci., 72 : 1569-1577.
- IBE, S.N. and HILL, W.G. 1988. J. Anim. Breed. Genet., 105 : 231-240.
- IM, S. and GIANOLA, D. 1988. Theor. Appl. Genet., 73 : 720-722.
- KACHMAN, S.1986. Prediction of genetic merit for growth curve parameters in outbred mice. M.Sc. thesis. U. of Illinois, Urbana-Champaign, Illinois.
- McCULLAGH, P. 1983. Ann. Stat., 11 : 59 - 67.
- McCULLAGH, P. and NELDER, J.A., 1983. Generalized linear models. Chapman & Hall, London.
- MANFREDI, E., DUCROCQ, V. and FOULLEY, J.L. 1990. 41th annual meeting of the EAAP. Toulouse.
- MASSELIN, S., SAUVANT, D., CHAPOUTOT, P. and MILAN, D., 1987. Ann. Zootech., 36 : 171 - 206.
- MEYER, K. 1990. Proc. 4th World Congr. Genet. Appl. Livest. Prod.
- NELDER, J.A. and PREGIBON, D. 1987. Biometrika. 74 : 221-232.
- NELDER, J.A. 1988 14th International Biometric Conference. July 1988, Namur, Belgium. I. 289-299.
- PATTERSON, H.D. and THOMPSON, R., 1971. Biometrika, 58 : 545-554.
- SCHAEFFER, L.R. and WILTON, J.W. 1976. J. Dairy Sci., 59 : 544-551.
- SIMIANER, H. and SCHAEFFER, L.R. 1989. Génét. Sél. Evol., 21 : 303 - 315.
- SMITH, S.P. and QUAAS, R.L., 1984. J. Dairy Sci., 67: 2999-3007.
- SMITH, S.P. and ALLAIRE, F.R., 1986. J. Dairy Sci., 69: 217-227.
- STIRATELLI, R., LAIRD, N. and WARE, J.H., 1984. Biometrics, 40:961-972.
- TAVERNIER, A. 1989. 40th annual meeting of the EAAP. Dublin. 325-326.
- TAVERNIER, A. 1990. Genetic evaluation of horses based on ranks in competitions. Submitted to Génét. Sél. Evol.
- THOMPSON, R. 1979. Biometrics, 35: 339-353.
- THOMPSON, R., McGUIRK B.J. and GILMOUR, A.R. 1985. J. Anim. Breed. Genet., 102 : 342-354.
- VAN RADEN, P. and JUNG, C., 1988. J. Dairy Sci. 71: 187-194.
- VAN VLECK, L.D. and EDLIN, K.M.1984. J. Dairy Sci. 67: 3025-3033.
- WILLIAMS, D.A., 1988. 14th International Biometric Conference. July 1988, Namur, Belgium. I. 301-313.
- WRIGHT, S., 1934. Genetics, 19:506-536 and 537-551.
- ZHAO, Y. 1987. Estimation of parameters in a mixed threshold model : its application to dystocia and birth weight in Simmental cattle. Ph.D. Thesis. Cornell University, Ithaca, New-York.