Original article

Mixture model for inferring susceptibility to mastitis in dairy cattle: a procedure for likelihood-based inference

Daniel GIANOLA[a,b*], Jørgen ØDEGÅRD[b], Bjørg HERINGSTAD[b],
Gunnar KLEMETSDAL[b], Daniel SORENSEN[c], Per MADSEN[c], Just JENSEN[c],
Johann DETILLEUX[d]

[a] Department of Animal Sciences, University of Wisconsin-Madison,
Madison, WI 53706, USA

[b] Department of Animal Science, Agricultural University of Norway, P.O. Box 5025,
1432 Ås, Norway

[c] Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences,
P.O. Box 50, 8830 Tjele, Denmark

[d] Faculté de Médecine Vétérinaire, Université de Liège, 4000 Liège, Belgium

Abstract – A Gaussian mixture model with a finite number of components and correlated random effects is described. The ultimate objective is to model somatic cell count information in dairy cattle and to develop criteria for genetic selection against mastitis, an important udder disease. Parameter estimation is by maximum likelihood or by an extension of restricted maximum likelihood. A Monte Carlo expectation-maximization algorithm is used for this purpose. The expectation step is carried out using Gibbs sampling, whereas the maximization step is deterministic. Ranking rules based on the conditional probability of membership in a putative group of uninfected animals, given the somatic cell information, are discussed. Several extensions of the model are suggested.

mixture models / maximum likelihood / EM algorithm / mastitis / dairy cattle

1. INTRODUCTION

Mastitis is an inflammation of the mammary gland associated with bacterial infection. Its prevalence can be as large as 50%, *e.g.*, [16, 30]. Its adverse economic effects are through a reduction in milk yield, an increase in veterinary costs and premature culling of cows [39]. Milk must be discarded due

---

* Corresponding author: gianola@calshp.cals.wisc.edu

to contamination with antibiotics, and there is a deterioration of milk quality. Further, the disease reduces an animal's well being.

Genetic variation in susceptibility to the disease exists. Studies in Scandinavia report heritability estimates between 0.06 and 0.12. The most reliable estimate is the 0.07 of Heringstad *et al.* [17], who fitted a threshold model to more than 1.6 million first-lactation records in Norway. These authors reported genetic trends equivalent to an annual reduction of 0.23% in prevalence of clinical mastitis for cows born after 1990. Hence, increasing genetic resistance to the disease *via* selective breeding is feasible, albeit slow.

Routine recording of mastitis is not conducted in most nations, *e.g.*, France and the United States. Instead, milk somatic cell score (SCS) has been used in genetic evaluation as a proxy measure. Heritability estimates of SCS average around 0.11 [29]. Pöso and Mäntysaari [32] found that the genetic correlation between SCS and clinical mastitis ranged from 0.37 to 0.68. Hence, selection for a lower SCS is expected to reduce prevalence of mastitis. On this basis, breeders have been encouraged to choose sires and cows having low estimated breeding values for SCS.

Booth *et al.* [4] reported that 7 out of 8 countries had reduced bulk somatic cell count by about 23% between 1985 and 1993; however, this was not accompanied by a reduction in mastitis incidence. Schukken *et al.* [38] stated that a low SCS might reflect a weak immune system, and suggested that the dynamics of SCS in the course of infection might be more relevant for selection. Detilleux and Leroy [8] noted that selection for low SCS might be harmful, since neutrophils intervene against infection. Also, a high SCS may protect the mammary gland. Thus, it is not obvious how to use SCS information optimally in genetic evaluation.

Some of the challenges may be met using finite mixture models, as suggested by Detilleux and Leroy [8]. In a mixture model, observations (*e.g.*, SCS, or milk yield and SCS) are used to assign membership into groups; for example, putatively "diseased" *versus* "non-diseased" cows. Detilleux and Leroy [8] used maximum likelihood; however, their implementation is not flexible enough.

Our objective is to give a precise account of the mixture model of Detilleux and Leroy [8]. Likelihood-based procedures are described and ranking rules for genetic evaluation are presented. The paper is organized as follows. The second section gives an overview of finite mixture models. The third section describes a mixture model with additive genetic effects for SCS. A derivation of the *EM* algorithm, taking into account presence of random effects, is given in the fourth section. The fifth section presents restricted maximum likelihood (REML) for mixture models. The final section suggests possible extensions.

2. FINITE MIXTURE MODELS: OVERVIEW

Suppose that a random variable y is drawn from one of K mutually exclusive and exhaustive distributions ("groups"), without knowing which of these underlies the draw. For instance, an observed SCS may be from a healthy or from an infected cow; in mastitis, the case may be clinical or subclinical. Here $K = 3$ and the groups are: "uninfected", "clinical" and "sub-clinical". The density of y can be written [27, 45] as:

$$p(y|\theta) = \sum_{i=1}^{K} P_i p_i(y|\theta_i),$$

where K is the number of components of the mixture; P_i is the probability that the draw is made from the ith component ($\sum_{i=1}^{K} P_i = 1$); $p_i(y|\theta_i)$ is the density under component i; θ_i is a parameter vector, and $\theta = \left[\theta'_1, \theta'_2, ..., \theta'_K, P_1, P_2, ..., P_K \right]'$ includes all distinct parameters, subject to $\sum_{i=1}^{K} P_i = 1$. If $K = 2$ and the distributions are normal with component-specific mean and variances, then θ has 5 elements: P, the 2 means and the 2 variances. In general, the y may be either scalar or vector valued, or may be discrete as in [5, 28].

Methods for inferring parameters are maximum likelihood and Bayesian analysis. An account of likelihood-based inference applied to mixtures is in [27], save for models with random effects. Some random effects models for clustered data are in [28, 40]. An important issue is that of parameter identification. In likelihood inference this can be resolved by introducing restrictions in parameter values, although creating computational difficulties. In Bayesian settings, proper priors solve the identification problem. A Bayesian analysis with Markov chain Monte Carlo procedures is straightforward, but priors must be proper. However, many geneticists are refractory to using Bayesian models with informative priors, so having alternative methods of analysis available is desirable. Hereafter, a normal mixture model with correlated random effects is presented from a likelihood-based perspective.

3. A MIXTURE MODEL FOR SOMATIC CELL SCORE

3.1. Motivation

Detilleux and Leroy [8] argued that it may not be sensible viewing SCS as drawn from a single distribution. An illustration is in [36], where different trajectories of SCS are reported for mastitis-infected and healthy cows. A randomly drawn SCS at any stage of lactation can pertain to either a healthy or

to an infected cow. Within infected cows, different types of infection, including sub-clinical cases, may produce different SCS distributions.

Genetic evaluation programs in dairy cattle for SCS ignore this heterogeneity. For instance, Boichard and Rupp [3] analyzed weighted averages of SCS measured at different stages of lactation with linear mixed models. The expectation is that, on average, daughters of sires with a lower predicted transmitting ability for somatic cell count will have a higher genetic resistance to mastitis. This oversimplifies how the immune system reacts against pathogens [7].

Detilleux and Leroy [8] pointed out advantages of a mixture model over a specification such as in [3]. The mixture model can account for effects of infection status on SCS and produces an estimate of prevalence of infection, plus a probability of status ("infected" *versus* "uninfected") for individual cows, given the data and values of the parameters. Detilleux and Leroy [8] proposed a 2−component mixture model, which will be referred to as DL hereafter. Although additional components may be required for finer statistical modelling of SCS, our focus will be on a 2−component specification, as a reasonable point of departure.

3.2. Hierarchical DL

The basic form of DL follows. Let $\mathbf{y}$ and $\mathbf{a}$ be random vectors of observations and of additive genetic effects for SCS, respectively. In the absence of infection, their joint density is

$$p_0\left(\mathbf{y}, \mathbf{a}|\beta_0, \mathbf{A}, \sigma_a^2, \sigma_e^2\right) = p_0\left(\mathbf{y}|\beta_0, \mathbf{a}, \sigma_e^2\right) p\left(\mathbf{a}|\mathbf{A}, \sigma_a^2\right). \tag{1}$$

The subscript 0 denotes "no infection", β_0 is a set of fixed effects, $\mathbf{A}$ is the known additive genetic relationship matrix between members of a pedigree, and σ_a^2 and σ_e^2 are additive genetic and environmental components of variance, respectively. Since $\mathbf{A}$ is known, dependencies on this matrix will be suppressed in the notation. Given $\mathbf{a}$, the observations will be supposed to be conditionally independent and homoscedastic, *i.e.*, their conditional variance-covariance matrix will be $\mathbf{I}\sigma_e^2$. A single SCS measurement per individual will be assumed, for simplicity. Under infection, the joint density is

$$p_1\left(\mathbf{y}, \mathbf{a}|\beta_1, \sigma_a^2, \sigma_e^2\right) = p_1\left(\mathbf{y}|\beta_1, \mathbf{a}, \sigma_e^2\right) p\left(\mathbf{a}|\sigma_a^2\right), \tag{2}$$

where subscript 1 indicates "infection". Again, the observations are assumed to be conditionally independent, and β_1 is a location vector, distinct (at least some elements) from β_0. DL assumed that the residual variance and the distribution of genetic effects were the same in healthy and infected cows. This can be relaxed, as described later.

The mixture model is developed hierarchically now. Let P be the probability that a SCS is from an uninfected cow. Unconditionally to group membership, but given the breeding value of the cow, the density of observation i is

$$p\left(y_i|\beta, a_i, \sigma_e^2, P\right) = P p_0\left(y_i|\beta_0, a_i, \sigma_e^2\right) + (1-P)\, p_1\left(y_i|\beta_1, a_i, \sigma_e^2\right); \quad i = 1, 2, ..., n,$$

where y_i and a_i are the SCS and additive genetic value, respectively, of the cow on which the record is taken, and $\beta = \left[\beta_0', \beta_1'\right]'$. The probability that the draw is made from distribution 0 is supposed constant from individual to individual.

Assuming that records are conditionally independent, the density of all n observations, given the breeding values, is

$$p\left(\mathbf{y}|\beta, \mathbf{a}, \sigma_e^2, P\right) = \prod_{i=1}^{n}\left[P p_0\left(y_i|\beta_0, a_i, \sigma_e^2\right) + (1-P)\, p_1\left(y_i|\beta_1, a_i, \sigma_e^2\right)\right]. \quad (3)$$

The joint density of $\mathbf{y}$ and $\mathbf{a}$ is then

$$p\left(\mathbf{y}, \mathbf{a}|\beta, \sigma_a^2, \sigma_e^2, P\right) =$$
$$\left\{\prod_{i=1}^{n}\left[P p_0\left(y_i|\beta_0, a_i, \sigma_e^2\right) + (1-P)\, p_1\left(y_i|\beta_1, a_i, \sigma_e^2\right)\right]\right\} p\left(\mathbf{a}|\sigma_a^2\right), \quad (4)$$

and the marginal density of the data is

$$p\left(\mathbf{y}|\beta, \sigma_a^2, \sigma_e^2, P\right) = \int p\left(\mathbf{y}, \mathbf{a}|\beta, \sigma_a^2, \sigma_e^2, P\right) p\left(\mathbf{a}|\sigma_a^2\right) \mathbf{da}. \quad (5)$$

When viewed as a function of the parameters $\theta = \left[\beta_0', \beta_1', \sigma_a^2, \sigma_e^2, P\right]'$, (5) is Fisher's likelihood. This can be written as the product of n integrals only when individuals are genetically unrelated; here, σ_a^2 would not be identifiable. On the other hand, if a_i represents some cluster effect (*e.g.*, a sire's transmitting ability), the between-cluster variance can be identified.

DL assume normality throughout and take $y_i|\beta_0, \mathbf{a}, \sigma_e^2 \sim N_0\left(\mathbf{x}_{0i}'\beta_0 + a_i, \sigma_e^2\right)$ and $y_i|\beta_1, \mathbf{a}, \sigma_e^2 \sim N_1\left(\mathbf{x}_{1i}'\beta_1 + a_i, \sigma_e^2\right)$. Here, $\mathbf{x}_{0i}'$ and $\mathbf{x}_{1i}'$ are known incidence vectors relating fixed effects to observations. The assumption about the genetic effects is $\mathbf{a}|\mathbf{A}, \sigma_a^2 \sim N\left(0, \mathbf{A}\sigma_a^2\right)$. Let now $z_i \sim Bernoulli\,(P)$, be an independent (*a priori*) random variable taking the value $z_i = 0$ with probability P if the datum is drawn from process N_0, or the value $z_i = 1$ with probability $1 - P$ if

from N_1. Assuming all parameters are known, one has

$$\Pr\left(z_i = 0|y_i,\beta_0,\beta_1,a_i,\sigma_e^2,P\right) = \frac{Pp_0\left(y_i|\beta_0,a_i,\sigma_e^2\right)}{Pp_0\left(y_i|\beta_0,a_i,\sigma_e^2\right) + (1-P)\,p_1\left(y_i|\beta_1,a_i,\sigma_e^2\right)}.$$
$$(6)$$

Thus, $\Pr\left(z_i = 1|y_i,\beta_0,\beta_1,a_i,\sigma_e^2,P\right) = 1-(6)$ is the probability that the cow belongs to the "infected" group, given the observed SCS, her breeding value and the parameters.

A linear model for an observation (given z_i) can be written as

$$y_i = (1 - z_i)\,\mathbf{x}'_{0i}\beta_0 + z_i\mathbf{x}'_{1i}\beta_1 + a_i + e_i.$$

A vectorial representation is

$$\mathbf{y} = [\mathbf{I} - Diag\,(z_i)]\,\mathbf{X}_0\beta_0 + [Diag\,(z_i)]\,\mathbf{X}_1\beta_1 + \mathbf{a} + \mathbf{e}$$
$$= \mathbf{X}_0\beta_0 + [Diag\,(z_i)]\,(\mathbf{X}_1\beta_1 - \mathbf{X}_0\beta_0) + \mathbf{a} + \mathbf{e}$$

where $Diag\,(z_i)$ is a diagonal matrix with typical element z_i; $\mathbf{X}_0$ is an $n \times p_0$ matrix with typical row $\mathbf{x}'_{0i}$; $\mathbf{X}_1$ is an $n \times p_1$ matrix with typical row $\mathbf{x}'_{1i}$; $\mathbf{a} = \{a_i\}$ and $\mathbf{e} = \{e_i\}$. Specific forms of β_0 and β_1 (and of the corresponding incidence matrices) are context-dependent, but care must be exercised to ensure parameter identifiability and to avoid what is known as "label switching" [27]. For example, DL take $\mathbf{X}_0\beta_0 = \mathbf{1}\mu_0$ and $\mathbf{X}_1\beta_1 = \mathbf{1}\mu_1$.

4. MAXIMUM LIKELIHOOD ESTIMATION: EM ALGORITHM

4.1. Joint distribution of missing and observed data

We extremize (5) with respect to θ *via* the expectation-maximization algorithm, or *EM* [6, 25]. Here, an *EM* version with stochastic steps is developed.

The *EM* algorithm augments (4) with n binary indicator variables z_i ($i = 1, 2, ..., n$), taken as independently and identically distributed as *Bernoulli*, with probability P. If $z_i = 0$, the SCS datum is generated from the "uninfected" component; if $z_i = 1$, the draw is from the other component. Let $\mathbf{z} = [z_1, z_2, ..., z_n]'$ denote the realized values of all z variables. The "complete" data is the vector $[\mathbf{a}', \mathbf{y}', \mathbf{z}']'$, with $[\mathbf{a}', \mathbf{z}']'$ constituting the "missing" part and $\mathbf{y}$ representing the "observed" fraction. The joint density of $\mathbf{a}$, $\mathbf{y}$ and $\mathbf{z}$ can be written as

$$p\left(\mathbf{a}, \mathbf{y}, \mathbf{z}|\beta_0,\beta_1,\sigma_a^2,\sigma_e^2,P\right) = p\left(\mathbf{z}|P\right)p\left(\mathbf{a}|\sigma_a^2\right)p\left(\mathbf{y}|\mathbf{z},\mathbf{a},\beta_0,\beta_1,\sigma_e^2\right). \qquad (7)$$

Given $\mathbf{z}$, the component of the mixture generating the data is known automatically for each observation. Now

$$p(\mathbf{z}|P) = \prod_{i=1}^{n} P^{1-z_i}(1-P)^{z_i},$$

$$p\left(y_i|\beta_0,\beta_1,a_i,\sigma_e^2,Z_i=0\right) = p_0\left(y_i|\beta_0,a_i,\sigma_e^2\right),$$

$$p\left(y_i|\beta_0,\beta_1,a_i,\sigma_e^2,Z_i=1\right) = p_1\left(y_i|\beta_1,a_i,\sigma_e^2\right),$$

for $i = 1, 2, ..., n$. Then, (7) becomes

$$p\left(\mathbf{a},\mathbf{y},\mathbf{z}|\beta_0,\beta_1,\sigma_a^2,\sigma_e^2,P\right) = p\left(\mathbf{y},\mathbf{z}|\beta_0,\beta_1,\mathbf{a},\sigma_a^2,\sigma_e^2,P\right)p\left(\mathbf{a}|\sigma_a^2\right)$$

$$= \left\{\prod_{i=1}^{n}\left[Pp_0\left(y_i|\beta_0,\mathbf{a},\sigma_e^2\right)\right]^{1-z_i}\left[(1-P)\,p_1\left(y_i|\beta_1,\mathbf{a},\sigma_e^2\right)\right]^{z_i}\right\}p\left(\mathbf{a}|\sigma_a^2\right). \quad (8)$$

4.2. Fully conditional distributions of missing variables

The form of (8) leads to conditional distributions needed for implementing the Monte Carlo *EM* algorithm.

- The density of the distribution $[\mathbf{z}|\beta_0,\beta_1,\mathbf{a},\sigma_a^2,\sigma_e^2,P,\mathbf{y}] \equiv [\mathbf{z}|\beta_0,\beta_1,\mathbf{a},\sigma_e^2,P,\mathbf{y}]$ is

$$p\left(\mathbf{z}|\beta_0,\beta_1,\mathbf{a},\sigma_e^2,P,\mathbf{y}\right)$$

$$\propto \prod_{i=1}^{n}\left\{\left[Pp_0\left(y_i|\beta_0,\mathbf{a},\sigma_e^2\right)\right]^{1-z_i}\left[(1-P)\,p_1\left(y_i|\beta_1,\mathbf{a},\sigma_e^2\right)\right]^{z_i}\right\}.$$

This is the distribution of n independent Bernoulli random variables with probability parameters (6).

- The density of distribution $\left[\mathbf{a},\mathbf{z}|\beta_0,\beta_1,\sigma_a^2,\sigma_e^2,P,\mathbf{y}\right]$ can be written as

$$p\left(\mathbf{a},\mathbf{z}|\beta_0,\beta_1,\sigma_a^2,\sigma_e^2,P,\mathbf{y}\right)$$

$$\propto \prod_{i=1}^{n}\left\{\left[Pp_0\left(y_i|\beta_0,\mathbf{a},\sigma_e^2\right)\right]^{1-z_i}\left[(1-P)\,p_1\left(y_i|\beta_1,\mathbf{a},\sigma_e^2\right)\right]^{z_i}\right\}p\left(\mathbf{a}|\sigma_a^2\right)$$

$$\propto \prod_{i=1}^{n}\exp\left[-\frac{(1-z_i)\left[a_i-\left(y_i-\mathbf{x}_{0i}'\beta_0\right)\right]^2 + z_i\left[a_i-\left(y_i-\mathbf{x}_{1i}'\beta_1\right)\right]^2}{2\sigma_e^2}\right]$$

$$\times\left[\prod_{i=1}^{n}P^{1-z_i}(1-P)^{z_i}\right]p\left(\mathbf{a}|\sigma_a^2\right). \quad (9)$$

As shown in the Appendix, the density of the distribution $[\mathbf{a}|\mathbf{z}, \beta_0, \beta_1,$ $\sigma_a^2, \sigma_e^2, P, \mathbf{y}]$ is

$$
p\left(\mathbf{a}|\mathbf{z}, \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right) \propto \exp\left[-\frac{\left(\mathbf{a} - \widehat{\lambda}\right)'\left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)\left(\mathbf{a} - \widehat{\lambda}\right)}{2\sigma_e^2}\right]
$$

with $\widehat{\lambda}$ as in (41). This indicates that the distribution is the Gaussian process

$$
\mathbf{a}|\mathbf{z}, \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y} \sim \mathbf{N}\left[\widehat{\lambda}, \left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)^{-1}\sigma_e^2\right]. \tag{10}
$$

Recall that, given $\mathbf{z}$, the probability P does not intervene in any distribution.

- Integrating (42) in the Appendix with respect to $\mathbf{a}$ gives the conditional distribution $\left[\mathbf{z}|\beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right]$ with probability mass function

$$
\Pr\left(\mathbf{z}|\beta_0, \beta_1, \sigma_a^2, \sigma_e^2, \mathbf{y}\right) = \frac{g\left(\mathbf{z}\right)}{\sum_{z_1}\sum_{z_2}\cdots\sum_{z_n} g\left(\mathbf{z}\right)}, \tag{11}
$$

where

$$
g\left(\mathbf{z}\right) = \exp\left[-\frac{\lambda'\mathbf{M}\left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)^{-1}\mathbf{A}^{-1}\lambda}{2\sigma_a^2}\right]\prod_{i=1}^{n} P^{1-z_i}\left(1 - P\right)^{z_i}.
$$

Computing the denominator of (11) is tedious because of the many sums involved.

4.3. Complete data log-likelihood

The logarithm of (8) is called the "complete data" log-likelihood

$$
L_{complete} = \log\left[p\left(\mathbf{a}|\sigma_a^2\right)\right] + \sum_{i=1}^{n}\left\{(1 - z_i)\left[\log P + \log p_0\left(y_i|\beta_0, \mathbf{a}, \sigma_e^2\right)\right]\right.
$$

$$
\left. + z_i\left[\log\left(1 - P\right) + \log p_1\left(y_i|\beta_1, \mathbf{a}, \sigma_e^2\right)\right]\right\}. \tag{12}
$$

4.4. The *E* and *M* steps

The $E-$step [6, 25, 27] consists of finding the expectation of $L_{complete}$ taken over the conditional distribution of the missing data, given $\mathbf{y}$ and some current (in the sense of iteration) values of the parameters, say $\theta^{[k]} = \left[\beta'^{[k]}, \sigma_a^{2[k]}, \sigma_e^{2[k]}, P^{[k]} \right]'$, where k denotes round of iteration. This expectation is known as the Q function

$$Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right) = E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}}\left[L_{complete}\right]. \tag{13}$$

The $M-$step finds parameter values maximizing $E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}}\left[L_{complete}\right]$. These are called the "complete data" maximum likelihood estimates. Taking partial derivatives of $Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right)$ with respect to all elements of θ gives

$$\frac{\partial E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}} Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right)}{\partial P} = E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}}\left[\frac{\sum_{i=1}^{n}(1 - z_i)}{P} - \frac{\sum_{i=1}^{n} z_i}{1 - P} \right], \tag{14}$$

$$\frac{\partial E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}} Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right)}{\partial \beta_0} = \sum_{i=1}^{n} E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}}\left[\frac{(1 - z_i)\mathbf{x}_{0i}\left(y_i - \mathbf{x}'_{0i}\beta_0 - a_i\right)}{\sigma_e^2} \right], \tag{15}$$

$$\frac{\partial E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}} Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right)}{\partial \beta_1} = \sum_{i=1}^{n} E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}}\left[\frac{z_i \mathbf{x}_{1i}\left(y_i - \mathbf{x}'_{1i}\beta_1 - a_i\right)}{\sigma_e^2} \right],$$

$$\frac{\partial E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}} Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right)}{\partial \sigma_e^2} = -\frac{n}{2\sigma_e^2}$$

$$+ E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}} \sum_{i=1}^{n} \left\{ \frac{(1 - z_i)\left(y_i - \mathbf{x}'_{0i}\beta_0 - a_i\right)^2}{2\sigma_e^4} + \frac{z_i\left(y_i - \mathbf{x}'_{1i}\beta_1 - a_i\right)^2}{2\sigma_e^4} \right\}, \tag{16}$$

and

$$\frac{\partial E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}} Q\left(\beta, \sigma_a^2, \sigma_e^2, P; \theta^{[k]}\right)}{\partial \sigma_a^2} = -\frac{q}{2\sigma_a^2} + E_{\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}}\left[\frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}{2\sigma_a^4} \right]. \tag{17}$$

Setting derivatives to 0 and solving for the parameters leads to the "complete data" maximum likelihood estimates; these provide new values for the next

$E-$step. The updates are obtained iterating with

$$P^{[k+1]} = 1 - \frac{\sum\limits_{i=1}^{n} E^{[k]}(z_i)}{n}, \tag{18}$$

$$\beta_0^{[k+1]} = E^{[k]} \left\{ \mathbf{X}_0' \left[\mathbf{I} - Diag(z_i) \right] \mathbf{X}_0 \right\}^{-1} \mathbf{X}_0' \left[\mathbf{I} - Diag(z_i) \right] (\mathbf{y} - \mathbf{a}_D), \tag{19}$$

$$\beta_1^{[k+1]} = E^{[k]} \left\{ \mathbf{X}_1' Diag(z_i) \mathbf{X}_1 \right\}^{-1} \mathbf{X}_1' Diag(z_i) (\mathbf{y} - \mathbf{a}_D), \tag{20}$$

$$\sigma_e^{2[k+1]} = \sum_{i=1}^{n} \frac{E^{[k]} \left[(1 - z_i)\left(y_i - \mathbf{x}_{0i}'\beta_0 - a_i\right)^2 + z_i\left(y_i - \mathbf{x}_{1i}'\beta_1 - a_i\right)^2 \right]}{n} \tag{21}$$

and

$$\sigma_a^{2[k+1]} = \frac{E^{[k]}\left(\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}\right)}{q}, \tag{22}$$

where $E^{[k]}$ (expression) $= E_{\mathbf{a},\mathbf{z}|\theta^{[k]},\mathbf{y}}$ (expression).

4.5. Monte Carlo implementation of the *E*-step

In mixture models without random effects (*i.e.*, with fixed $\mathbf{a}$), calculation of $E_{\mathbf{z}|\theta^{[k]},\mathbf{y}}$ (expression) is direct, as the distribution $\left[\mathbf{z}|\mathbf{a}, \theta^{[k]}, \mathbf{y}\right]$ is Bernoulli with probability (6). In the fixed model $\mathbf{a}$ is included in the parameter vector, assuming it is identifiable; *e.g.*, when the $a's$ represent fixed cluster effects. Here, the iterates are linear functions of the missing $\mathbf{z}$, and the computation involves replacing z_i by its conditional expectation, which is (6) evaluated at $\theta = \theta^{[k]}$. This was employed by DL, but it is not correct when $\mathbf{a}$ is random.

In a mixture model with random effects, the joint distribution $[\mathbf{a}, \mathbf{z}|\theta, \mathbf{y}]$, with density (42) is not recognizable, so analytical evaluation of $E_{\mathbf{a},\mathbf{z}|\theta^{[k]},\mathbf{y}}$ (expression) is not possible. We develop a Monte Carlo $E-$step using the Gibbs sampler. Observe that distributions $[\mathbf{z}|\mathbf{a}, \theta, \mathbf{y}]$ and $[\mathbf{a}|\mathbf{z}, \theta, \mathbf{y}]$ are recognizable. Given $\mathbf{a}$, each of the elements of $\mathbf{z}$ is independent Bernoulli, with probability (6). Likewise, $[\mathbf{a}|\mathbf{z}, \theta, \mathbf{y}]$ is multivariate normal, as in (10). The Gibbs sampler [35, 41] draws from $[\mathbf{a}, \mathbf{z}|\theta, \mathbf{y}]$ by successive iterative sampling from $[\mathbf{z}|\mathbf{a}, \theta, \mathbf{y}]$ and $[\mathbf{a}|\mathbf{z}, \theta, \mathbf{y}]$. For example, draw each of the z_i from their Bernoulli distributions, then sample additive effects conditional on these realizations, update the Bernoulli distributions, and so on. The process requires discarding early samples ("burn-in"), and collecting m additional samples, with or without thinning. The additive effects can be sampled simultaneously or in a piece-wise manner [41].

Let the samples from $\left[\mathbf{a}, \mathbf{z}|\theta^{[k]}, \mathbf{y}\right]$ be $\mathbf{a}^{(j,k)}, \mathbf{z}^{(j,k)}$, $j = 1, 2, ..., m$, recalling that k is iterate number. Then, form Monte Carlo estimates of the complete data

maximum likelihood estimates in (18)–(22) as

$$\widehat{P}^{[k+1]} = 1 - \frac{\sum\limits_{i=1}^{n} \frac{1}{m} \sum\limits_{j=1}^{m} z_i^{(j,k)}}{n}, \tag{23}$$

$$\widehat{\beta}_0^{[k+1]} = \frac{1}{m} \sum\limits_{j=1}^{m} \left\{ \mathbf{X}_0' \left[\mathbf{I} - Diag\left(z_i^{(j,k)} \right) \right] \mathbf{X}_0 \right\}^{-1} \mathbf{X}_0' \left[\mathbf{I} - Diag\left(z_i^{(j,k)} \right) \right] \left(\mathbf{y} - \mathbf{a}_D^{(j,k)} \right) \tag{24}$$

$$\widehat{\beta}_1^{[k+1]} = \frac{1}{m} \sum\limits_{j=1}^{m} \left\{ \mathbf{X}_1' Diag\left(z_i^{(j,k)} \right) \mathbf{X}_1 \right\}^{-1} \mathbf{X}_1' Diag\left(z_i^{(j,k)} \right) \left(\mathbf{y} - \mathbf{a}_D^{(j,k)} \right), \tag{25}$$

$$\sigma_e^{2[k+1]} =$$

$$\sum\limits_{i=1}^{n} \frac{\sum\limits_{j=1}^{m} \left[\left(1 - z_i^{(j,k)}\right)\left(y_i - \mathbf{x}_{0i}'\widehat{\beta}_0^{[k]} - a_i^{(j,k)}\right)^2 + z_i^{(j,k)}\left(y_i - \mathbf{x}_{1i}'\widehat{\beta}_1^{[k]} - a_i^{(j,k)}\right)^2 \right]}{mn} \tag{26}$$

and

$$\sigma_a^{2[k+1]} = \frac{\sum\limits_{j=1}^{m} \mathbf{a}^{(j,k)\prime} \mathbf{A}^{-1} \mathbf{a}^{(j,k)}}{mq}. \tag{27}$$

The E and M steps are repeated until parameter values do not change appreciably.

When $m \rightarrow \infty$, the limiting form of the algorithm is the standard EM. The deterministic EM algorithm increases the likelihood function monotonically [6, 25] until convergence to some stationary point, although this may not be a global maximum. Monte Carlo implementations of EM are discussed in [12, 46]. For finite m, however, it is not true that the likelihood increases monotonically, although this may reduce the chance that the algorithm gets stuck in some local stationary point of little inferential value. Tanner [44] suggests keeping m low at early stages of the algorithm and then increase sample size in the neighborhood of some maximum. Due to Monte Carlo error, convergence may be declared when fluctuations of the iterates appear to be random about some value. At that point it may be worthwhile to increase m [44].

4.6. Inference about additive genetic effects

Genetic evaluation for SCS would be based on $\widehat{a}_i$, the ith element of $\widehat{\mathbf{a}}$, this being the mean vector of the distribution $\left[\mathbf{a} | \beta_0 = \widehat{\beta}_0, \beta_1 = \widehat{\beta}_1, \sigma_a^2 = \widehat{\sigma}_a^2, \sigma_e^2 = \widehat{\sigma}_e^2, \mathbf{y} \right]$. While $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma}_e^2, \widehat{\sigma}_a^2$ and $\widehat{P}$ follow from the maximum likelihood

procedure, $\widehat{\mathbf{a}}$ must be calculated using Monte Carlo methods. From standard theory

$$E\left(\mathbf{a}|\beta_0 = \widehat{\beta}_0, \beta_1 = \widehat{\beta}_1, \sigma_a^2 = \widehat{\sigma}_a^2, \sigma_e^2 = \widehat{\sigma}_e^2, \mathbf{y}\right) =$$

$$E_{\mathbf{z}|\beta_0=\widehat{\beta}_0,\beta_1=\widehat{\beta}_1,\sigma_a^2=\widehat{\sigma}_a^2,\sigma_e^2=\widehat{\sigma}_e^2,\mathbf{y}} E\left(\mathbf{a}|\beta_0 = \widehat{\beta}_0, \beta_1 = \widehat{\beta}_1, \mathbf{z}, \sigma_a^2 = \widehat{\sigma}_a^2, \sigma_e^2 = \widehat{\sigma}_e^2, \mathbf{y}\right). \quad (28)$$

Using (10), one then has that

$$\widehat{E}\left(\mathbf{a}|\beta_0 = \widehat{\beta}_0, \beta_1 = \widehat{\beta}_1, \sigma_a^2 = \widehat{\sigma}_a^2, \sigma_e^2 = \widehat{\sigma}_e^2, \mathbf{y}\right) =$$

$$\left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)^{-1} \left[\begin{array}{c} [\mathbf{I}_n - Diag\,(\bar{z}_i)]\,(\mathbf{y} - \mathbf{X}_0\beta_0) + Diag\,(\bar{z}_i)\,(\mathbf{y} - \mathbf{X}_1\beta_1) \\ \mathbf{0} \end{array} \right] \quad (29)$$

where $\bar{z}_i$ is the mean of m draws from $\left[z_i|\beta_0 = \widehat{\beta}_0, \beta_1 = \widehat{\beta}_1, \sigma_a^2 = \widehat{\sigma}_a^2, \sigma_e^2 = \widehat{\sigma}_e^2, \mathbf{y}\right]$, obtained from running a Gibbs sampler applied to the process $\left[\mathbf{a}, \mathbf{z}|\theta = \widehat{\theta}, \mathbf{y}\right]$. One can also estimate (28) directly from the m draws for $\mathbf{a}$ obtained from such sampler. In theory, however, (29) is expected to have a smaller Monte Carlo error.

Another issue is how the SCS information is translated into chances of a cow belonging to the "uninfected" group. A simple option is to estimate (6) as

$$\Pr\left(Z_i = 0|y_i, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{a}_i, \widehat{\sigma}_e^2, \widehat{P}\right) = \frac{\widehat{P}p_0\left(y_i|\widehat{\beta}_0, \widehat{a}_i, \widehat{\sigma}_e^2\right)}{\widehat{P}p_0\left(y_i|\widehat{\beta}_0, \widehat{a}_i, \widehat{\sigma}_e^2\right) + \left(1 - \widehat{P}\right)p_1\left(y_i|\widehat{\beta}_1, \widehat{a}_i, \widehat{\sigma}_e^2\right)}. \quad (30)$$

For cow i (30) induces the log odds-ratio

$$k_{01}\left(y_i, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{a}_i, \widehat{\sigma}_e^2, \widehat{P}\right) = \log\frac{\widehat{P}}{\left(1 - \widehat{P}\right)} + \log\frac{p_0\left(y_i|\widehat{\beta}_0, \widehat{a}_i, \widehat{\sigma}_e^2\right)}{p_1\left(y_i|\widehat{\beta}_1, \widehat{a}_i, \widehat{\sigma}_e^2\right)}.$$

The first term is constant across cows, so rankings based on (30) are driven by the ratios between the densities of the SCS record of the cow in question under the "healthy" and "diseased" components.

Statistically (30) does not take into account the error of the maximum likelihood estimates of all parameters. If the likelihood function is sharp and unimodal (large samples), this would be of minor concern. However, asymptotic arguments in finite mixtures are more subtle than in linear or generalized linear models [27] and multimodality is to be expected in small samples, as illustrated in [1].

5. RESTRICTED MAXIMUM LIKELIHOOD

5.1. General

Maximum likelihood ignores uncertainty associated with simultaneous estimation of nuisance parameters [41]. For example, if variance components are inferred, no account is taken of degrees of freedom lost for estimating fixed effects [31]. This is clear when casted from a Bayesian perspective [14]. Maximum likelihood estimates can be viewed as components of the mode of a joint posterior distribution obtained after assigning flat (improper) priors to all parameters. On the other hand, restricted maximum likelihood estimates (of variance components) correspond to the mode of the posterior distribution of the variance components after integration of fixed effects, these viewed as nuisances. We extend this idea to our mixture model, *i.e.*, account for uncertainty about fixed effects.

5.2. Distributions

The sampling model is as in (3). Following the *EM−REML* algorithm in [6], we treat β_0 and β_1 as missing data and assign an improper uniform prior to each of these two vectors. The missing data includes now $\beta_0, \beta_1, \mathbf{a}$ and $\mathbf{z}$, and the joint density of the missing data and of the observations is proportional to (8). Pertinent distributions follow.

- The density of $\left[\beta_0, \beta_1, \mathbf{a}, \mathbf{z} | \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right]$ is proportional to (9), but it is convenient to rewrite it as

$$
p\left(\beta_0, \beta_1, \mathbf{a}, \mathbf{z} | \sigma_a^2, \sigma_e^2, \mathbf{y}\right)
$$
$$
\propto \exp\left[-\frac{(\mathbf{y} - \mathbf{X}_0\beta_0 - \mathbf{a})' \left[\mathbf{I} - Diag\left(z_i\right)\right] (\mathbf{y} - \mathbf{X}_0\beta_0 - \mathbf{a})}{2\sigma_e^2}\right]
$$
$$
\times \exp\left[-\frac{(\mathbf{y} - \mathbf{X}_1\beta_1 - \mathbf{a})' Diag\left(z_i\right) (\mathbf{y} - \mathbf{X}_1\beta_1 - \mathbf{a})}{2\sigma_e^2}\right] p\left(\mathbf{a} | \sigma_a^2\right)
$$
$$
\times \prod_{i=1}^{n} P^{1-z_i} (1 - P)^{z_i}. \tag{31}
$$

- The distribution $\left[\mathbf{z} | \beta_0, \beta_1, \mathbf{a}, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right]$ is as before, *i.e.*, group membership variables are independent Bernoulli with parameters (6).
- The conditional posterior distribution $\left[\mathbf{a} | \mathbf{z}, \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right]$ is also as before, *i.e.*, normal, with parameters given in (10).

- From (31) it can be deduced that the distribution $\left[\beta_0|\mathbf{a}, \mathbf{z}, \beta_1, \sigma_a^2, \right.$ $\left. \sigma_e^2, P, \mathbf{y} \right]$ has density

$$p\left(\beta_0|\mathbf{a}, \mathbf{z}, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right)$$

$$\propto \exp\left[-\frac{(\mathbf{y} - \mathbf{a} - \mathbf{X}_0\beta_0)'\left[\mathbf{I} - Diag\left(z_i\right)\right](\mathbf{y} - \mathbf{a} - \mathbf{X}_0\beta_0)}{2\sigma_e^2}\right]. \quad (32)$$

Given $\mathbf{z}$, the only observations contributing information about β_0 are those for which $z_i = 0$. Hence, (32) can be recognized [41] as the density of the normal distribution

$$\beta_0|\mathbf{a}, \mathbf{z}, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y} \sim \mathbf{N}\left[\left(\mathbf{X}_0'\mathbf{T}_z\mathbf{X}_0\right)^{-1}\mathbf{X}_0'\mathbf{T}_z\left(\mathbf{y} - \mathbf{a}\right), \left(\mathbf{X}_0'\mathbf{T}_z\mathbf{X}_0\right)^{-1}\sigma_e^2\right], \quad (33)$$

where $\mathbf{T}_z = \mathbf{I} - Diag\left(z_i\right)$.
- Likewise

$$p\left(\beta_1|\mathbf{a}, \mathbf{z}, \beta_0, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right)$$

$$\propto \exp\left[-\frac{(\mathbf{y} - \mathbf{a} - \mathbf{X}_1\beta_1)'\,Diag\left(z_i\right)(\mathbf{y} - \mathbf{a} - \mathbf{X}_1\beta_1)}{2\sigma_e^2}\right], \quad (34)$$

is the density of the normal distribution

$$\beta_1|\mathbf{a}, \mathbf{z}, \beta_0, \sigma_a^2, \sigma_e^2, P, \mathbf{y} \sim \mathbf{N}\left[\left(\mathbf{X}_1'\mathbf{D}_z\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{D}_z\left(\mathbf{y} - \mathbf{a}\right), \left(\mathbf{X}_1'\mathbf{D}_z\mathbf{X}_1\right)^{-1}\sigma_e^2\right], \quad (35)$$

where $\mathbf{D}_z = Diag\left(z_i\right)$.

5.3. The E and M steps

The complete data log-likelihood, apart from a constant, is as in (12). Since the fixed effects must now be integrated out, the pertinent complete data maximum likelihood estimates are those of P, σ_e^2 and σ_a^2. The integration over the missing data $\beta_0, \beta_1, \mathbf{a}$ and $\mathbf{z}$ needed for the E−step is done by sampling from $\left[\mathbf{a}, \mathbf{z}, \beta_0, \beta_1|P^{[k]}, \sigma_e^{2[k]}, \sigma_a^{2[k]}, \mathbf{y}\right]$. The Gibbs sampler draws successively from (6), (10), (33) and (35), leading to samples $\mathbf{a}^{(j,k)}, \mathbf{z}^{(j,k)}, \beta_0^{(j,k)}, \beta_1^{(j,k)},$

$j = 1, 2, ..., m$. The Monte Carlo estimates of the complete data maximum like-lihood estimates are

$$\widehat{P}^{[k+1]} = 1 - \frac{\sum\limits_{i=1}^{n} \frac{1}{m} \sum\limits_{j=1}^{m} z_i^{(j,k)}}{n},$$

$$\sigma_e^{2[k+1]} =$$

$$\sum\limits_{i=1}^{n} \frac{\frac{1}{m} \sum\limits_{j=1}^{m} \left[\left(1 - z_i^{(j,k)}\right)\left(y_i - \mathbf{x}'_{0i}\beta_0^{(j,k)} - a_i^{(j,k)}\right)^2 + z_i^{(j,k)}\left(y_i - \mathbf{x}'_{1i}\beta_1^{(j,k)} - a_i^{(j,k)}\right)^2 \right]}{n},$$

and

$$\sigma_a^{2[k+1]} = \frac{\sum\limits_{j=1}^{m} \mathbf{a}^{(j,k)\prime}\mathbf{A}^{-1}\mathbf{a}^{(j,k)}}{mq}.$$

Convergence issues are as for full maximum likelihood.

5.4. Inference about additive genetic effects

From now on, let the REML estimates be $\widehat{P}$, $\widehat{\sigma}_e^2$, $\widehat{\sigma}_a^2$. Genetic evaluation for SCS may be based on Monte Carlo estimates of the mean vector of $\left[\mathbf{a}|\widehat{P}, \widehat{\sigma}_e^2, \widehat{\sigma}_a^2, \mathbf{y}\right]$. These can be obtained from running a second Gibbs sampler applied to $\left[\mathbf{a}, \mathbf{z}, \beta_0, \beta_1|\widehat{P}, \widehat{\sigma}_e^2, \widehat{\sigma}_a^2\right]$. This takes into account uncertainty about β_0 and β_1 (in the usual REML sense), but not that associated with the variance component estimates and with P.

From a group allocation perspective, $\Pr\left(Z_i = 0|y_i, a_i, \sigma_e^2, \sigma_a^2, P\right)$ is arguably more relevant than $\Pr\left(Z_i = 0|y_i, \beta_0, \beta_1, a_i, \sigma_e^2, P\right)$, since cows and bulls must be selected across a range of environmental circumstances. Here, one can use the (approximate) Bayesian argument

$$\Pr\left(Z_i = 0|\mathbf{y}, a_i, \sigma_e^2, \sigma_a^2, P\right)$$

$$\approx \int \Pr\left(Z_i = 0|\mathbf{y}, \beta_0, \beta_1, a_i, \sigma_e^2, P\right) p\left(\beta_0, \beta_1|\sigma_e^2, \sigma_a^2, P, \mathbf{y}\right) d\beta_0 d\beta_1$$

$$= \int \Pr\left(Z_i = 0|y_i, \beta_0, \beta_1, a_i, \sigma_e^2, P\right) p\left(\beta_0, \beta_1|\sigma_e^2, \sigma_a^2, P, \mathbf{y}\right) d\beta_0 d\beta_1 \qquad (36)$$

and calculate the estimated probability

$$\widehat{\Pr}\left(Z_i = 0|\mathbf{y}, a_i, \sigma_e^2, \sigma_a^2, P\right) = \frac{1}{m}\sum\limits_{j=1}^{m} \Pr\left(Z_i = 0|y_i, \beta_0^{(j)}, \beta_1^{(j)}, a_i, \widehat{\sigma}_e^2, \widehat{P}\right)$$

where the m samples of the fixed effects are draws from $\left[\mathbf{a}, \mathbf{z}, \beta_0, \beta_1 | \widehat{P}, \widehat{\sigma}_e^2, \right.$
$\left. \widehat{\sigma}_a^2, \mathbf{y}\right]$. The "exact" integration involves sampling from $\left[\mathbf{a}_{-i}, \mathbf{z}, \beta_0, \beta_1 | a_i, \widehat{P}, \widehat{\sigma}_e^2, \right.$
$\left. \widehat{\sigma}_a^2, \mathbf{y}\right]$, which requires running an "auxiliary sampler"; $\mathbf{a}_{-i}$ is $\mathbf{a}$ without a_i. This
is of doubtful additional value, in view of the computational burden.

6. EXTENSIONS OF MODEL

A first extension is to allow for heterogeneous residual variance in the two
components. The parameter vector would now be $\theta = \left[\beta_0', \beta_1', \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P\right]$,
and the complete data log-likelihood would take the form

$$L_{complete} = \log\left[p\left(\mathbf{a}|\sigma_a^2\right)\right] + \sum_{i=1}^{n}\left\{(1 - z_i)\left[\log P + \log p_0\left(y_i|\beta_0, \mathbf{a}, \sigma_{e0}^2\right)\right]\right.$$
$$\left. +z_i\left[\log(1 - P) + \log p_1\left(y_i|\beta_1, \mathbf{a}, \sigma_{e1}^2\right)\right]\right\}.$$

The Monte Carlo complete data full maximum likelihood estimates (23)–(27)
have the same form as before, except for the residual components of variance,
which are now

$$\sigma_{e0}^{2[k+1]} = \frac{1}{m}\sum_{j=1}^{m}\left\{\sum_{i=1}^{n_0^{(j,k)}}\frac{\left[\left(1 - z_i^{(j,k)}\right)\left(y_i - \mathbf{x}_{0i}'\beta_0^{[k]} - a_i^{(j,k)}\right)^2\right]}{n_0^{(j,k)}}\right\}$$

$$\sigma_{e1}^{2[k+1]} = \frac{1}{m}\sum_{j=1}^{m}\left\{\sum_{i=1}^{n_1^{(j,k)}}\frac{\left[z_i^{(j,k)}\left(y_i - \mathbf{x}_{1i}'\beta_1^{[k]} - a_i^{(j,k)}\right)^2\right]}{n_1^{(j,k)}}\right\},$$

where $n_0^{(j,k)}$ and $n_1^{(j,k)} = n - n_0^{(j,k)}$ are the number of indicator variables $z_i^{(j,k)}$
falling into each of the two groups in sample j at iteration k. The imputa-
tions of $\mathbf{z}$ and $\mathbf{a}$ in the E–step need to be modified as well. The probability of
membership (6) must be amended to reflect residual heteroscedasticity; this is
straightforward. The density of the distribution $\left[\mathbf{a}, \mathbf{z}|\beta_0, \beta_1, \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P, \mathbf{y}\right]$,

counterpart of (9), can be expressed as

$$p\left(\mathbf{a}, \mathbf{z}|\beta_0, \beta_1, \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P, \mathbf{y}\right)$$

$$\propto \prod_{i=1}^{n} \exp\left[-\frac{(1-z_i)\left[a_i - \left(y_i - \mathbf{x}_{0i}'\beta_0\right)\right]^2 + z_i\left[a_i - \left(y_i - \mathbf{x}_{1i}'\beta_1\right)\right]^2}{2\sigma_{ez_i}^2}\right]$$

$$\times \left[\prod_{i=1}^{n} P^{1-z_i}(1-P)^{z_i}\right] p\left(\mathbf{a}|\sigma_a^2\right). \qquad (37)$$

Using similar algebra as before, this can be written as

$$p\left(\mathbf{a}, \mathbf{z}|\beta_0, \beta_1, \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P, \mathbf{y}\right)$$

$$\propto \exp\left[-\sum_{i=1}^{n} \frac{(a_i - \lambda_i)^2}{2\sigma_{ez_i}^2}\right]\left[\prod_{i=1}^{n} P^{1-z_i}(1-P)^{z_i}\right] p\left(\mathbf{a}|\sigma_a^2\right).$$

Now

$$\sum_{i=1}^{n} \frac{(a_i - \lambda_i)^2}{\sigma_{ez_i}^2} + \frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}{\sigma_a^2} = (\mathbf{a} - \lambda)' \mathbf{M}_z (\mathbf{a} - \lambda) + \frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}{\sigma_a^2},$$

where

$$\mathbf{M}_z = \left[\begin{array}{cc} Diag\left(\frac{1}{\sigma_{ez_i}^2}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right].$$

Then, as in (40) and (42), as in the Appendix, one is led directly to

$$\mathbf{a}|\beta_0, \beta_1, \mathbf{z}, \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P, \mathbf{y} \sim \mathbf{N}\left[\widehat{\lambda}_z, \left(\mathbf{M}_z + \frac{\mathbf{A}^{-1}}{\sigma_a^2}\right)^{-1}\right],$$

where

$$\widehat{\lambda}_z = \left(\mathbf{M}_z + \frac{\mathbf{A}^{-1}}{\sigma_a^2}\right)^{-1} \mathbf{M}_z \lambda.$$

With $\left[\mathbf{a}|\beta_0, \beta_1, \mathbf{z}, \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P, \mathbf{y}\right]$ recognized as above, and with

$$\mathbf{z}|\beta_0, \beta_1, \mathbf{a}, \sigma_a^2, \sigma_{e0}^2, \sigma_{e1}^2, P, \mathbf{y} \equiv \mathbf{z}|\beta_0, \beta_1, \mathbf{a}, \sigma_{e0}^2, \sigma_{e1}^2, P$$

identified, this completes the ingredients for carrying out the $E-$step *via* Gibbs sampling.

A second extension is as follows. It is known from microarray studies (*e.g.*, [47]) that there is differential expression of genes in diseased and healthy individuals. For example, some genes are expressed in tissues sampled from cancerous tumors whereas other genes are expressed in negative biopsies. Mixture models for diseases with a genetic basis should be flexible enough to allow for differential expression. A cow cannot be mastitic and healthy at the same time, but there may be genes that are expressed when infection takes place, but to a different extent in the absence of disease. This can be modelled by introducing a genetic correlation, much in the same way that one can think of a genetic covariance between ovulation rate and scrotal circumference in sheep [24], or between direct and maternal effects in cattle [48]. In the context of mastitis, and in a 2−component mixture, a sufficient condition for statistical identification of the genetic correlation is that some healthy animals have relatives that contract the disease. If a K−component mixture model is fitted, one would need that at least some "families" (in a loosely defined sense) are represented in all components of the mixture. In dairy cattle, where large half-sib families are common, this requirement can be met, unless K is large. Also, the assumption that genetic variances are the same for each of the two components of the mixture can be relaxed.

In mixture models, it is not always clear whether asymptotic properties of the maximum likelihood estimates hold well when one departs from standard settings (*e.g.*, [15, 22]). Hosmer [20] reported poor mean squared error behavior of parameter estimates in a 2−component mixture model, especially when the distributions overlapped and when P was close to 1. DL found that estimates of P were not reliable when the means of the mixture components were close. This implies that posterior probabilities based on maximum likelihood estimates should be viewed with caution. DL reported that posterior probabilities of infection were lower in "truly" non-infected that in infected cows, with the differences increasing as mixing proportion increased. Posterior probabilities were close to zero both for non-infected and infected cows with small SCS values, and close to 1 for infected cows with large SCS. From plots in [8], it seems that if one used the rule that the posterior probability should be larger than $\frac{1}{2}$ to assign a cow to the "possibly infected" class, only a few cows would be false positives. It is difficult to assess the proportion of false negatives (infected cows classified as non-infected) that would result from such an assignment. DL assumed homogeneous variances in the two populations, and the maximum likelihood estimates agreed with the simulated values, particularly when the two distributions were well separated and P was far away from $\frac{1}{2}$. The assumption of homogeneity of variances was to ensure a global maximum of the likelihood, but perhaps at the expense of realism, as one would expect the variances to be larger in the "infected" component, merely from scaling

considerations. A Bayesian analysis, exploring the entire posterior distribution of the parameters, could be useful here. If the model calls for heterogeneous dispersion parameters, likelihood-based inference would run the risk of placing the (conceptual) asymptotic distribution at a local stationary point, whereas the Bayesian analysis would integrate over the entire parameter space.

Our approach does not allow for hierarchical modelling of the prior probability of membership to the components, assumed constant for every observation. In the context of longitudinal information, evidence indicates that the incidence of mastitis is much higher around calving than at other stages of lactation (*e.g.*, [18]). In such a setting, it would be sensible to allow for a time-varying probability of membership.

Mastitis is complex, and more than two components may be needed for adequate modelling. Different types of bacterial agents may lead to pathogen-specific distributions of SCS. Hence, a mixture model with an unknown number of components may be in order, although the difficulty of implementing such a specification must be recognized. For example, a Bayesian approach with an unknown number of components may require a reversible jump Markov chain Monte Carlo algorithm [11,34], but it is difficult to tune this type of procedure [13]. A possibly more feasible alternative is to fit a sequence of models with $2, 3, ...K$ components, and carry out the analysis for each of the settings. Then, assuming that models are equi-probable *a priori*, choose the one with the largest Bayes factor support [21] or infer breeding values or future observations using Bayesian model averaging [19]. Bayesian model averaging accounts for uncertainty about the models, and is expected to lead to better prediction of future observations in some well-defined sense [26,33].

Although SCS is claimed to be nearly normally distributed [2], the problem of outlier protection in multivariate data is a difficult one [23]. A possibility is to use thick-tailed distributions, instead of normal processes, in mixture model analysis. Robust distributions were introduced for cross-sectional data in quantitative genetics by [10, 42, 43]; extensions are in [37, 41]. McLachlan and Peel [27] have described mixtures using the t−distribution, one of the many possible thick-tailed processes, and warn that this distribution, like the normal, cannot accommodate asymmetrically distributed errors. However, an asymmetric t-distribution [9], or skewed-Gaussian distributions can be used in mixture models without much difficulty. Bayesian Markov chain Monte Carlo procedures allow all these extensions.

A mixture model may be even more powerful when other traits are brought into the picture. For example, the analysis might involve SCS, protein yield and some udder type trait simultaneously. A mixture model for vectors would need to be fitted here.

7. CONCLUSION

In conclusion, finite mixture models provide an elegant manner of dealing with SCS data in the context of genetic improvement of resistance to mastitis. Irrespective of whether Bayesian or likelihood-based methods are employed, the main challenges may reside in developing meaningful mixture models. These will need to be validated using information from cows where the disease outcome is known. In short, research is needed to establish their usefulness, to identify models having an adequate predictive ability, and to develop feasible computational procedures. For example, a non-Bayesian analysis may be carried out more efficiently with algorithms other than EM.

REFERENCES

[1] Aitkin M., Wilson G.T., Mixture models, outliers and the EM algorithm, Technometrics 22 (1980) 325–331.

[2] Ali A.K.A., Shook G.E., An optimum transformation for somatic cell concentration in milk, J. Dairy Sci. 63 (1980) 487–490.

[3] Boichard D., Rupp R., Genetic analysis and genetic evaluation for somatic cell score in French dairy cattle, Interbull Bull. 15 (1997) 54–60, International Bull Evaluation Service, Uppsala, Sweden.

[4] Booth J.M., Progress in the control of mastitis, in: Proceedings of the 3rd International Mastitis Seminar, 1995, Tel Aviv, Israel. S4.3–S4.11, International Dairy Federation, Brussels, Belgium.

[5] Dellaportas P., Bayesian classification of neolithic tools, Appl. Stat. 47 (1998) 279–297.

[6] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data *via* the EM algorithm (with discussion), J. Royal Stat. Soc. B 39 (1977) 1–38.

[7] Detilleux J., Mathematical modeling of inflammatory reaction during bovine mastitis, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, 19-23 August 2002, Vol. 31, Inra, pp. 711–714.

[8] Detilleux J., Leroy P.L., Application of a mixed normal mixture model to the estimation of mastitis-related parameters, J. Dairy Sci. 83 (2000) 2341–2349.

[9] Fernández C., Steel M.F.J., On Bayesian modelling of fat tails and skewness, J. Am. Stat. Assoc. 93 (1998) 359–371.

[10] Gianola D., Strandén I., Foulley J.L., Modelos lineales mixtos con distribuciones-t: potencial en genetica cuantitativa, in: Actas, Quinta Conferencia Española de Biometria, Sociedad Española de Biometria, pp. 3–4 , Valencia, Spain.

[11] Green P., Reversible jump MCMC computation and Bayesian model determination, Biometrika 82 (1995) 711–732.

[12] Guo S.W., Thompson E.A., Monte Carlo estimation of mixed models for large complex pedigrees, Biometrics 50 (1994) 417–432.

[13] Han C., Carlin B.P., Markov chain Monte Carlo methods for computing Bayes factors: a comparative review, J. Am. Stat. Assoc. 96 (2001) 1122–1132.

[14] Harville D.A., Bayesian inference of variance components using only error contrasts, Biometrika 61 (1974) 383–385.

[15] Hathaway R.J., A constrained formulation of maximum likelihood estimation for normal mixture distributions, Ann. Math. Stat. 13 (1985) 795–800.

[16] Heringstad B., Klemetsdal G., Ruane J., Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries, Livest. Prod. Sci. 64 (2000) 95–106.

[17] Heringstad B., Rekaya R., Gianola D., Klemetsdal G., Weigel K.A., Genetic change for clinical mastitis in Norwegian cattle: a threshold model analysis, J. Dairy Sci. 86 (2003) 369–375.

[18] Heringstad B., Chang Y.M., Gianola D., Klemetsdal G., Genetic analysis of longitudinal trajectory of clinical mastitis in first-lactation Norwegian cattle, J. Dairy Sci. 86 (2003) 2676–2683.

[19] Hoeting J.A., Madigan D., Raftery A.E., Volinsky C.T., Bayesian model averaging: a tutorial, Stat. Sci. 14 (1999) 382–417.

[20] Hosmer D.W., On MLE of the parameters of a mixture of two normal distributions when the sample size is small, Comput. Stat. 1 (1973) 217–227.

[21] Kass R.E., Raftery A.E., Bayes factors, J. Am. Stat. Assoc. 90 (1995) 773–795.

[22] Kiefer J., Wolfowitz J., Consistency of the maximum likelihood estimates in the presence of infinitely many incidental parameters, Ann. Math. Stat. 27 (1956) 887–906.

[23] Kosinski A., A procedure for the detection of multivariate outliers, Comput. Stat. Data Anal. 29 (1999) 145–161.

[24] Land R.B., The expression of female sex-limited characters in the male, Nature 241 (1973) 208–209.

[25] Little R.J.A., Rubin D.B., Statistical Analysis with Missing Data, 1st edn., John Wiley and Sons, New York, 1987.

[26] Madigan D., Raftery A.E., Model selection and accounting for model uncertainty in graphical models using Occam's window, J. Am. Stat. Assoc. 89 (1994) 1535–1546.

[27] McLachlan G., Peel D., Finite Mixture Models, John Wiley and Sons, New York, 2000.

[28] Militino A.F., Ugarte M.D., Fean C.B., The use of mixture models for identifying high risks in disease mapping, Stat. Med. 20 (2001) 2035–2049.

[29] Mrode R.A., Swanson G.J.T., Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle, Anim. Breed. Abstr. 64 (1996) 847–857.

[30] Myllys V., Asplund K., Brofeldt E., Hirvela-Koski V., Honkanen-Buzalski T., Junttila J., Kulkas L., Myllykangas O., Niskanen M., Saloniemi H., Sandholm M., Saranpaa T., Bovine mastitis in Finland in 1998 and 1995: Changes in prevalence and antibacterial resistance, Acta Vet. Scand. 39 (1998) 119–126.

[31] Patterson H.D., Thompson R., Recovery of interblock information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[32] Pöso J., Mantysaari E.A., Relationships between clinical mastitis, somatic cell score and production for the first three lactations of Finnish Ayrshire, J. Dairy Sci. 79 (1996) 1284–1291.

[33] Raftery A.E., Madigan D., Hoeting J.A., Model selection and accounting for model uncertainty in linear regression models, J. Am. Stat. Assoc. 92 (1997) 179–191.

[34] Richardson S., Green P., On Bayesian analysis of mixtures with an unknown number of components (with discussion), J. Royal Stat. Soc. B 59 (1997) 731–792.

[35] Robert C.P., Casella G., Monte Carlo Statistical Methods, Springer-Verlag, New York, 1999.

[36] Rodriguez-Zas S.L., Gianola D., Shook G.E., Evaluation of models for somatic cell score lactation patterns in Holsteins, Livest. Prod. Sci. 67 (2000) 19–30.

[37] Rosa G.J.M., Gianola D., Padovani C.R., Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation, Biom. J. 54 (2003) 1–18.

[38] Schukken Y.H., Lam T.J.G.M., Barkema H.W., Biological basis for selection on udder health traits, in: Proceedings of the International Workshop on Genetic Improvement of Functional Traits in Cattle, Interbull Bull. 15 (1997) 27–33.

[39] Shook G.E., Selection for disease resistance, J. Dairy Sci. 72 (1989) 2136–2142.

[40] Shoukri M.M., McLachlan G.J., Parametric estimation in a genetic mixture model with application to nuclear family data, Biometrics 50 (1994) 128–139.

[41] Sorensen D., Gianola D., Likelihood, Bayesian and MCMC Methods in Quantitative Genetics, Springer-Verlag, New York, 2002.

[42] Strandén I.J., Robust mixed effects linear models with t-distributions and application to dairy cattle breeding, Ph.D. Thesis, University of Wisconsin-Madison, 1996.

[43] Stranden I., Gianola D., Attenuating effects of preferential treatment with Student-t mixed linear models: a simulation study, Genet. Sel. Evol. 30 (1998) 565–583.

[44] Tanner M.A., Tools for Statistical Inference, 1st edn., Springer-Verlag, New York, 1993.

[45] Titterington D.M., Smith A.F.M., Makov U.E., Statistical Analysis of Finite Mixture Distributions, John Wiley and Sons, Chichester, 1985.

[46] Wei G.C.G., Tanner M.A., A Monte Carlo implementation of the EM algorithm and the poor's man's data augmentation algorithm, J. Am. Stat. Assoc. 85 (1990) 699–704.

[47] West M., Blanchette C., Dressman H., Huang E., Ishida S., Spang R., Zuzan H., Olson J.A. Jr., Marks J.R., Nevins J.R., Predicting the clinical status of human breast cancer by using gene expression profiles, Proc. Natl. Acad. Sci. USA 98 (2001) 11462–11467.

[48] Willham R.L., The covariance between relatives for characters composed of components contributed by related individuals, Biometrics 19 (1963) 18–27.

APPENDIX: CONDITIONAL DENSITY OF GENETIC EFFECTS

The density of $\left[\mathbf{a}, \mathbf{z} | \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right]$ is in (9). The two quadratic forms on a_i can be combined as

$$(1 - z_i)\left[a_i - \left(y_i - \mathbf{x}'_{0i}\beta_0\right)\right]^2 + z_i\left[a_i - \left(y_i - \mathbf{x}'_{1i}\beta_1\right)\right]^2 =$$
$$(a_i - \lambda_i)^2 + (1 - z_i)\, z_i \left(\mathbf{x}'_{0i}\beta_0 - \mathbf{x}'_{1i}\beta_1\right)^2$$

where

$$\lambda_i = (1 - z_i)\left(y_i - \mathbf{x}'_{0i}\beta_0\right) + z_i\left(y_i - \mathbf{x}'_{1i}\beta_1\right).$$

Hence, (9) is expressible as

$$p\left(\mathbf{a}, \mathbf{z}, | \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right) \propto \left[\prod_{i=1}^{n} P^{1-z_i}(1 - P)^{z_i}\right]$$
$$\times \exp\left[-\frac{\sum_{i=1}^{n}(a_i - \lambda_i)^2 + \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}\frac{\sigma_e^2}{\sigma_a^2} + \sum_{i=1}^{n}(1 - z_i)\, z_i\left(\mathbf{x}'_{0i}\beta_0 - \mathbf{x}'_{1i}\beta_1\right)^2}{2\sigma_e^2}\right]. \quad (38)$$

Note that the term $(1 - z_i)\, z_i$ is null for all observations. Thus

$$p\left(\mathbf{a}, \mathbf{z} | \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right)$$
$$\propto \exp\left[-\frac{\sum_{i=1}^{n}(a_i - \lambda_i)^2 + \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}\frac{\sigma_e^2}{\sigma_a^2}}{2\sigma_e^2}\right]\left[\prod_{i=1}^{n} P^{1-z_i}(1 - P)^{z_i}\right]. \quad (39)$$

Now, to combine quadratic forms again, put $\lambda_D = \{\lambda_i\}$, $i = 1, 2, ..., n$, where D denotes the subset of individuals with records, and partition the vector of

breeding values as

$$\mathbf{a} = \left[\begin{array}{c} \mathbf{a}_D \\ \mathbf{a}_{\overline{D}} \end{array} \right]$$

where $\mathbf{a}_{\overline{D}}$ is the vector of additive genetic effects of individuals without observations. Then

$$\sum_{i=1}^{n} (a_i - \lambda_i)^2 + \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \frac{\sigma_e^2}{\sigma_a^2} = (\mathbf{a}_D - \lambda_D)' (\mathbf{a}_D - \lambda_D) + \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \frac{\sigma_e^2}{\sigma_a^2}$$

$$= (\mathbf{a} - \lambda)' \mathbf{M} (\mathbf{a} - \lambda) + \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \frac{\sigma_e^2}{\sigma_a^2}$$

with $\lambda = \left[\begin{array}{cc} \lambda_D' & \mathbf{0}' \end{array} \right]'$ and

$$\mathbf{M} = \left[\begin{array}{cc} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right].$$

The two quadratic forms on $\mathbf{a}$ can be combined into

$$(\mathbf{a} - \lambda)' \mathbf{M} (\mathbf{a} - \lambda) + \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \frac{\sigma_e^2}{\sigma_a^2} = \left(\mathbf{a} - \widehat{\lambda} \right)' \left(\mathbf{M} + \mathbf{A}^{-1} \frac{\sigma_e^2}{\sigma_a^2} \right) \left(\mathbf{a} - \widehat{\lambda} \right)$$

$$+ \lambda' \mathbf{M} \left(\mathbf{M} + \mathbf{A}^{-1} \frac{\sigma_e^2}{\sigma_a^2} \right)^{-1} \mathbf{A}^{-1} \lambda \frac{\sigma_e^2}{\sigma_a^2}, \quad (40)$$

where

$$\widehat{\lambda} = \left(\mathbf{M} + \mathbf{A}^{-1} \frac{\sigma_e^2}{\sigma_a^2} \right)^{-1} \mathbf{M} \lambda$$

$$= \left[\begin{array}{cc} \mathbf{I}_n + \mathbf{A}^{DD} \frac{\sigma_e^2}{\sigma_a^2} & \mathbf{A}^{D\overline{D}} \frac{\sigma_e^2}{\sigma_a^2} \\ \mathbf{A}^{\overline{D}D} \frac{\sigma_e^2}{\sigma_a^2} & \mathbf{A}^{\overline{D}D} \frac{\sigma_e^2}{\sigma_a^2} \end{array} \right]^{-1}$$

$$\times \left[\begin{array}{c} \left[\mathbf{I}_n - Diag\,(z_i) \right] (\mathbf{y} - \mathbf{X}_0 \beta_0) + Diag\,(z_i)\,(\mathbf{y} - \mathbf{X}_1 \beta_1) \\ \mathbf{0} \end{array} \right]. \quad (41)$$

• Further

$$\mathbf{A}^{-1} = \left[\begin{array}{cc} \mathbf{A}_{DD} & \mathbf{A}_{D\overline{D}} \\ \mathbf{A}_{\overline{D}D} & \mathbf{A}_{\overline{D}D} \end{array} \right]^{-1} = \left[\begin{array}{cc} \mathbf{A}^{DD} & \mathbf{A}^{D\overline{D}} \\ \mathbf{A}^{\overline{D}D} & \mathbf{A}^{\overline{D}D} \end{array} \right]$$

Employing the preceding developments in (39), it follows that

$$p\left(\mathbf{a}, \mathbf{z} | \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right) \propto \exp\left[-\frac{\left(\mathbf{a} - \widehat{\lambda}\right)'\left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)\left(\mathbf{a} - \widehat{\lambda}\right)}{2\sigma_e^2}\right]$$

$$\times \exp\left[-\frac{\lambda'\mathbf{M}\left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)\mathbf{A}^{-1}\lambda\frac{\sigma_e^2}{\sigma_a^2}}{2\sigma_e^2}\right]$$

$$\times \left[\prod_{i=1}^{n} P^{1-z_i}(1-P)^{z_i}\right]. \qquad (42)$$

The density of the conditional distribution $\left[\mathbf{a}|\mathbf{z}, \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right]$ is arrived at by fixing $\mathbf{z}$ in (42), yielding

$$p\left(\mathbf{a}|\mathbf{z}, \beta_0, \beta_1, \sigma_a^2, \sigma_e^2, P, \mathbf{y}\right) \propto \exp\left[-\frac{\left(\mathbf{a} - \widehat{\lambda}\right)'\left(\mathbf{M} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)\left(\mathbf{a} - \widehat{\lambda}\right)}{2\sigma_e^2}\right].$$