

An Iterative Procedure for Analyzing a Combination of Censored and Uncensored Traits in Mixed Models: Estimation of Location Parameters

Kenji TOGASHI^{a)} and J.E.O. REGE^{b)}

^{a)} *Livestock and Grassland Division,
Japan International Research Center for Agricultural Sciences
(Tsukuba, Ibaraki, 305 Japan)*

^{b)} *International Livestock Research Institute,
(Addis Ababa , P.O. Box 5689, Ethiopia)*

Received May 8, 1997

Abstract

A mixed model simultaneous procedure is described to estimate location parameters for the right-censored and uncensored traits in the practical animal breeding data set for herd-life and milk production. Location parameters of fixed effects represented by herd -year effects and random effects represented by sire additive genetic effects were given as the mode of posterior distribution resulting from a combination of a multivariate normal density and a standard normal cumulative distribution function. An iterative method for solving mixed model equations is presented for this kind of data analysis without explicitly setting up the equations.

Additional key words: sire model, censored data, herd-life, milk-yield

Introduction

Low genetic potential is often quoted as a major constraint to improving milk and meat production in Africa. Consequently, most livestock improvement programmes have, over the years, resorted to importation of exotic breeds for crossbreeding or for direct replacement of

indigenous genotypes. This trend continues as human population and urbanisation increased and demand for animal products mounts. Consequently, indigenous breeds of livestock continue to be at risk. Fortunately, there are some pockets of pure indigenous breeds left, particularly in the harsher environments. Moreover, the general failure of livestock development projects

^{a)} Present address: Hokkaido National Agricultural Experimental Station (Sapporo, 062 Japan)

based on imported 'high producing' breeds is increasingly shifting attention to the adapted breeds. For example, helminth resistance, tick resistance, trypanotolerance, ability to survive long period without feed or water, and resistance/tolerance to many other biotic and abiotic stresses have been reported in some African livestock breeds^{10, 12, 14)}. It is generally realized that adaptation is a crucial component of overall productivity in stressful environments. Thus, increased utilization is the most rational conservation strategy. There has, however, been limited research into how genetic adaptation can be harnessed in practical breeding programmes. There is a need to implement genetic evaluation methods which incorporate productivity and adaptation traits. In stressful environments and especially in smallholder systems, selection pressure on females for such production traits as milk yield is nearly zero. Thus, how long a cow remains in the herd is closely linked to her overall adaptation. Under such circumstances herd-life can be considered a measure of adaptation. At any point in time, the period during which a particular animal has stayed in the herd can be regarded as its herd-life. However, any such animal still has the possibility of staying in the herd beyond the period of measurement. Thus, herd-life is right-censored. Togashi and Rege²⁴⁾ reported that heritability for herd-life and genetic correlation between herd-life and milk yield using a similar data set were lower in treating right-censored points as continuous measurements than in treating them as right-censored data. Censoring, if ignored in analyses, can cause loss of information and has a potential of introducing bias.

Several methods have been proposed for analyses of survival data in cattle (e.g. van Vleck²⁵⁾, Famula³⁾, Smith and Quaas²⁰⁾, Delorenzo and Everett²⁾, Smith and Allaire²¹⁾). However, none of these account for censoring. Carriquiry et al.¹⁾ presented a mixed model procedure for censored data in a univariate case. With regard to obtaining solutions for location parameters, Schaeffer and Kennedy¹⁷⁾ and Misztal¹³⁾ proposed indirect

procedure in continuous traits without explicitly setting up the mixed model equations.

The present paper presents a procedure to estimate, simultaneously, location parameters for censored and uncensored traits in a mixed bivariate model accounting for a censored trait without explicitly setting up the mixed model equations.

Materials and Methods

1) Statistical model

The model (model 1) for one continuous uncensored trait (y_1) and one right-censored trait (y_2) is represented as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = X \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + Z \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

where y_1 and y_2 are $q \times 1$ observation vectors, b_1 and b_2 are fixed effect vectors for a continuous trait and a right-censored trait, respectively, u_1 and u_2 are $s \times 1$ vectors of random additive genetic effects of animals for two respective traits, such as $u_i \sim N(0, A\sigma_{ui}^2)$, A is a known $s \times s$ matrix of additive genetic relationships, e_i is a $q \times 1$ vector of unobserved random residuals for i th trait so that $e_i \sim N(0, Iq\sigma_{ei}^2)$, X and Z are the known incidence matrices. That is, censored data are treated as data from normal distribution whose errors are normally and independently distributed with constant variance.

The model (model 2) for one continuous trait (y^*1) and one uncensored trait (y^*2) is

$$\begin{pmatrix} y^*1 \\ y^*2 \end{pmatrix} = X^* \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + Z^* \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} e^*1 \\ e^*2 \end{pmatrix}$$

where y^*1 and y^*2 are $p \times 1$ vectors, e^*i is a $p \times 1$ vector of residuals for i th trait so that $e^*i \sim N(0, Ip\sigma_{ei}^2)$, X^* and Z^* are known incidence matrices. The levels of fixed effects for b_1 and b_2 in model 1 and 2 are defined equally and the size of these

vectors is defined as the maximum number in model 1 or 2. This definition also applies to levels

or subscripts for random effects of u1 and u2, and maximum number of level is defined as s.

Expectations and variances are:

$$E \begin{pmatrix} u1 \\ u2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, E \begin{pmatrix} e1 \\ e2 \\ e*1 \\ e*2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{var} \begin{pmatrix} u1 \\ u2 \end{pmatrix} = G = \begin{pmatrix} A\sigma_{u1}^2 & A\sigma_{u12}^2 \\ A\sigma_{u21}^2 & A\sigma_{u2}^2 \end{pmatrix}$$

$$G = G_0 \Phi A, G_0 = \begin{pmatrix} \sigma_{u1}^2 & \sigma_{u12}^2 \\ \sigma_{u21}^2 & \sigma_{u2}^2 \end{pmatrix}$$

$$\text{var} \begin{pmatrix} e1 \\ e2 \\ e*1 \\ e*2 \end{pmatrix} = \begin{pmatrix} Iq\sigma_{e1}^2 & Iq\sigma_{e12} & 0 & 0 \\ Iq\sigma_{e21} & Iq\sigma_{e2}^2 & 0 & 0 \\ 0 & 0 & Ip\sigma_{e1}^2 & Ip\sigma_{e12} \\ 0 & 0 & Ip\sigma_{e21} & Ip\sigma_{e2}^2 \end{pmatrix} = \begin{pmatrix} R_0 \Phi Iq & 0 \\ 0 & R_0 \Phi Ip \end{pmatrix}$$

$$R_0 = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{e12} \\ \sigma_{e21} & \sigma_{e2}^2 \end{pmatrix}, \Phi \text{ is direct product operator of Searle }^{18)}$$

Likelihood function:

With location and dispersion parameters known, the probability density function of y*1 and

y*2 can be written, by assuming that residuals follow the multivariate normal distribution, f(y*1, y*2/parameters) as:

$$= \infty -1/2 \left(\begin{pmatrix} y*1 \\ y*2 \end{pmatrix} - \begin{pmatrix} X* & 0 \\ 0 & X* \end{pmatrix} \begin{pmatrix} b1 \\ b2 \end{pmatrix} - \begin{pmatrix} Z* & 0 \\ 0 & Z* \end{pmatrix} \begin{pmatrix} u1 \\ u2 \end{pmatrix} \right)' \\ \times R_0^{-1} \Phi Ip \times \left(\begin{pmatrix} y*1 \\ y*2 \end{pmatrix} - \begin{pmatrix} X* & 0 \\ 0 & X* \end{pmatrix} \begin{pmatrix} b1 \\ b2 \end{pmatrix} - \begin{pmatrix} Z* & 0 \\ 0 & Z* \end{pmatrix} \begin{pmatrix} u1 \\ u2 \end{pmatrix} \right)$$

The probability density function of y1 and y2 is f(y1,y2/parameters)=f(y1/parameters)xf(y2/y1,parameters) can be similarly presented, noting that:

$$f(y2/y1,parameters) = \prod_i Q(x_i), i=1,..,q$$

where Q(.)=is the standard normal cumulative distribution function, and $x_i = (t_i - t_{mi}) / \sigma_e^*$,

t_i is the right censored point for ith observation, $t_{mi} = b_0(y1_i - x_i' b1 - z_i' u1) + x_i' b2 + z_i' u2$, $b_0 = \sigma_{e12} / \sigma_{e1}^2$, $\sigma_e^* = \sigma_{e2}^2 (1 - r_{e12}^2)$, $y1_i$ is the ith element of vector of y1, x_i' and z_i' are ith row vector for X and Z, respectively, r_{e12} is the residual correlation coefficient.

Random additive genetic effect vector (U'=(u1,u2)') is assumed to follow a priori a multivariate normal distribution, i.e., U~N(0,G) which is independent of the distribution of fixed effects. Assuming flat priors for fixed effects, the prior distribution is taken to be uniform over the whole space. If dispersion parameters are known, the joint posterior density function for location parameters, that is,

f(b1, b2, u1, u2/y1, y2, y*1, y*2, dispersion parameters) can be written as:

$$\propto f(y*1, y*2/parameters) \times f(y1/parameters) \times f(y2/y1, parameters)$$

$$\times \exp \left[-1/2 \begin{vmatrix} u1 \\ u2 \end{vmatrix}' G^{-1} \begin{vmatrix} u1 \\ u2 \end{vmatrix} \right]$$

Therefore, log-posterior likelihood function (L), given data and dispersion parameters, are expressed as:

$$\begin{aligned} &\propto -1/2 \left[\begin{vmatrix} y^{*1} \\ y^{*2} \end{vmatrix} - \begin{vmatrix} X^* & 0 \\ 0 & X^* \end{vmatrix} \begin{vmatrix} b1 \\ b2 \end{vmatrix} - \begin{vmatrix} Z^* & 0 \\ 0 & Z^* \end{vmatrix} \begin{vmatrix} u1 \\ u2 \end{vmatrix} \right]' \times R^{-1} \\ &\times \left[\begin{vmatrix} y^{*1} \\ y^{*2} \end{vmatrix} - \begin{vmatrix} X^* & 0 \\ 0 & X^* \end{vmatrix} \begin{vmatrix} b1 \\ b2 \end{vmatrix} - \begin{vmatrix} Z^* & 0 \\ 0 & Z^* \end{vmatrix} \begin{vmatrix} u1 \\ u2 \end{vmatrix} \right] \\ &-1/2(y1-Xb1-Zu1)'R^{-1}(y1-Xb1-Zu1)-1/2 \begin{vmatrix} u1 \\ u2 \end{vmatrix}' G^{-1} \begin{vmatrix} u1 \\ u2 \end{vmatrix} + \sum_i \log(Q(x_i)) \dots (1) \end{aligned}$$

where $R=R_0\Phi Iq$.

Taking the joint posterior model as point estimator of location parameters requires maximizing the log-posterior likelihood function

(1) with respect to these location parameters.

Maximization of the log-likelihood function can be done using the Newton-Raphson algorithm. The equations are ;

$\begin{vmatrix} e^{11}X^*X^*+(1/\sigma^2e1)X'X+b_0^2X'SX & e^{12}X^*X^*-b_0X'SX & e^{11}X^*Z^*+(1/\sigma^2e1)X'Z+b_0^2X'SZ & e^{12}X^*Z^*-b_0X'SX \\ & e^{22}X^*X^*+X'SX & e^{12}X^*Z^*-b_0X'SZ & e^{22}X^*Z^*+X'SZ \\ & & b_0^2Z'SZ+e^{11}Z^*Z^* & -b_0Z'SZ+e^{12}Z^*Z^* \\ & & +(1/\sigma^2e1)Z'Z+g^{11}A^{-1} & +g^{12}A^{-1} \\ & & & Z'SZ+e^{22}Z^*Z^* \\ & & & +g^{22}A^{-1} \end{vmatrix}$	$\begin{vmatrix} b1 \\ b2 \\ u1 \\ u2 \end{vmatrix}$
<p style="text-align: center;">symmetry</p>	
<p>= RHS</p>	

Equations in RHS

$$\begin{aligned} b1: & e^{11}X^*y^{*1}+e^{12}X^*y^{*2}+(1/\sigma^2e1)X'y1 \\ & +X'S(b_0^2Xb1^{(t-1)}+b_0^2Zu1^{(t-1)}-b_0S^{-1(t-1)}h^{(t-1)} \\ & -b_0Xb2^{(t-1)}-b_0Zu2^{(t-1)}) \\ b2: & e^{22}X^*y^{*2}+e^{12}X^*y^{*1}-X'S^{(t-1)}(b_0Xb1^{(t-1)} \\ & +b_0Zu1^{(t-1)}-Xb2^{(t-1)}-Zu2^{(t-1)}-S^{-1(t-1)}h^{(t-1)}) \\ u1: & e^{11}Z^*y^{*2}+e^{12}Z^*y^{*1}+(1/\sigma^2e1)Z'y1 \\ & +Z'S^{(t-1)}(b_0^2Xb1^{(t-1)}+b_0^2Zu1^{(t-1)} \\ & -b_0Xb2^{(t-1)}-b_0S^{-1(t-1)}h^{(t-1)}-b_0Zu2^{(t-1)}) \\ u2: & e^{22}Z^*y^{*2}+e^{12}Z^*y^{*1}-Z'S^{(t-1)}(b_0Xb1^{(t-1)} \\ & +b_0Zu1^{(t-1)}-Xb2^{(t-1)}-Zu2^{(t-1)}-S^{-1(t-1)}h^{(t-1)}) \end{aligned}$$

element denoted as s_i given by

$s_i=h_i(h_i-x_i/\sigma e^*)$, and h is a vector of order q the i th value of which is $:h_i=\{Z(x_i)/Q(x_i)\}/\sigma e^*$, where $Z(.)$ is the standard normal probability density function.

The e^{ij} and g^{ij} in the above sets of equations for maximization of log-likelihood are elements of the inverse R_0 and G_0 , respectively, that is:

$$\begin{vmatrix} e^{11} & e^{12} \\ e^{21} & e^{22} \end{vmatrix} \text{ is inverse of } R_0, \text{ and } \begin{vmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{vmatrix} \text{ is inverse of } G_0.$$

where S is a $q \times q$ diagonal matrix with i th diagonal

2) Iterative procedure

As has been shown above, the diagonal elements of matrix S are defined as $s_i = h_i(h_i - x_i / \sigma_e^2)$. And the elements of a vector h are defined as $h_i = (Z(x_i) / Q(x_i)) / \sigma_e^2$. These values can be quite small. One way to avoid this is to standardize the data to $N(0,1)$ values. The other possibility is to eliminate the denominator from these equations by multiplying both left hand side and right hand side (RHS) by σ_e^2 . When either is done, s_i and h_i do not become so small. Moreover, when this is done, s_i and h_i can be represented more simply as $s_i = h_i(h_i - x_i)$ and $h_i = Z(x_i) / Q(x_i)$.

The implementation of this procedure requires that data be prepared so that the levels of each factor (fixed and random) are numbered consecutively. Additionally, the data must be sorted by levels of fixed effect. Within the same level of the fixed effect, data from model 1 are followed by data from model 2. Initial solutions for fixed and random effects are set to zero.

Illustration of Method

1) Sire model

Pedigree information is not always available. Sire model can be adopted by replacing the random additive genetic effect in animal model by sire genetic effect. In this case G_0 in the coefficient matrix is replaced by G_s , the sire genetic variance-covariance matrix and new residual (co) variances corresponding to the sire model.

2) Numerical example

A subset of data on Kenyan Friesian analysed by Rege¹⁶⁾ was used to illustrate this procedure. The traits analysed were first lactation milk yield (trait 1) which was corrected by adjustment factors of Rege and Mosi¹⁵⁾ and herd-life (trait 2). Herd-life was calculated as the number of days a cow stayed in a herd since first calving. In practice herd-life data can have definite length or can be right-censored if recording is terminated before cows left the herd. Therefore, herd-life data are a mixture of censored (model 1) and uncensored (model 2) data. Only data from one herd were used to illustrate the procedure.

The numerical example applied a simple sire model with only one fixed effect. The data structure is given in Table 1. The linear model to describe both traits is:

$$Y_{ijk} = b_i + s_j + e_{ijk}$$

where b_i = fixed effect of calving year ($i=1,2$) and s_j = random effect of j -th sire ($j=1,2,\dots,9$). Heritability for milk yield, heritability for herd-life, genetic correlation and the residual correlation were assumed to be 0.36, 0.21, 0.24 and 0.21, respectively.

All the relationships in the pedigree were ignored; i.e. $A=I$. Sire and fixed effect solutions were obtained by the iterative method. Stopping criterion for convergence in original scale was defined as the sum of squares of the difference of the present and previous estimates divided by the

Table 1. Data structure of numerical example

	First lactation milk yield	Herd-life
Total (N)	60	60
Sires	9	9
records/sire	6.7	6.7
No. of uncensored data	49	49
No. of censored data	11	11
mean	3697.9 kg	2800.8 day
σ_x	781.2	390.1
Minimum value	2441	2256
Maximum value	5561	4030

sum of squares of present estimates. Convergence was considered to have occurred when a value of 1.0×10^{-6} was reached.

3) Iterative procedure in numerical example

Firstly, the data are normal-standardized and first solution for random and fixed effects set to zero. Mean and standard deviations for milk yield and herd-life were assumed to be 3637 and 644.5, and 2856.2 and 761.6, respectively, these values having been obtained from the entire data set of Rege¹⁶⁾.

3)-1 : Model 1 iteration

1. As an example of one right-censored datum in the first level of fixed effect, x_1 , h_1 and s_1 are calculated:

Milk yield (y_{1j}) : $(3023\text{kg} - 3637)/644.5$

Right - censored point for herd-life (t_j) : $(3178 \text{ days} - 2856.2)/761.6$

G_s as normal-standardization~

$$\begin{vmatrix} 0.0900 & 0.0165 \\ 0.0165 & 0.0525 \end{vmatrix}$$

R_0 as normal-standardization~

$$\begin{vmatrix} 0.9100 & 0.1950 \\ 0.1950 & 0.9475 \end{vmatrix}$$

$$r_{e12} = 0.1950 / (0.91 \times 0.9475)^{0.5} = 0.21$$

$$\sigma_e^2 = 0.9475(1 - r_{e12}^2), t_{\mu 1} = b_0(y_{11} - x_1'0 - z_1'0) + x_1'0 + z_1'0 = b_0 y_{11}$$

$$b_0 = 0.1950 / 0.91$$

$$x_1 = (t_1 - t_{\mu 1}) / \sigma_e^* = 0.698280$$

$$Q(x_1) = \int_{x_1} 1 / \sqrt{2\pi} \cdot \exp(-x_1^2/2) / dx = 0.255002$$

$$h_1 = 1 / \sqrt{2\pi} \cdot \exp(-x_1^2/2) / Q(x_1) = 1.25991$$

$$s_1 = h_1 (h_1 - x_1) = 0.757645.$$

The other right - censored data are processed in the same way.

2. The elements of a 2×1 work vector (B) for the observation in the first level of fixed effect are calculated and accumulated as:

$$\begin{aligned} B &= \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} \begin{vmatrix} s_1 \cdot b_{1(i)} \\ \sigma_e^* \cdot h_1 + s_1 \cdot b_{2(i)} \end{vmatrix} \\ &+ \sigma_e^2 e \begin{vmatrix} (1/\sigma_e^2) \cdot (y_{11} - u_{1(9)}) \\ 0 \end{vmatrix} + B \\ &= \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} \begin{vmatrix} s_1 \cdot 0 \\ \sigma_e^* \cdot h_1 + s_1 \cdot 0 \end{vmatrix} \\ &+ \sigma_e^2 e \begin{vmatrix} (1/0.91) \cdot (y_{11} - 0) \\ 0 \end{vmatrix} + B \end{aligned}$$

where $b_{1(i)}$ and $b_{2(i)}$ are the first level of fixed effects for b_1 and b_2 ; $u_{1(9)}$ is the 9-th level of sire effect for u_1 , and sire number of the first observation is 9. The multiplier 0 in the value $s_1 \cdot 0$ of the above equation represents the starting values of the fixed effects. For the other observations in model 1 in the first level of fixed effect, calculations proceed in the same way.

3. The elements of a 2×2 work matrix (D: for the observation in the first level of fixed effect) are

$$\begin{aligned} D &= \sigma_e^2 e \begin{vmatrix} 1/\sigma_e^2 e & 0 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} s_1 + D \\ &= \sigma_e^2 e \begin{vmatrix} 1/0.91 & 0 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} s_1 + D. \end{aligned}$$

Same process is applied for the other observations in the first level of fixed effect.

4. A 2×1 work vector (E) for the first level of fixed effect, after accumulating the elements of D and B in model 1 and 2, is obtained as:

$$E = D^{-1}B = \begin{vmatrix} -0.952677 \\ 1.378386 \end{vmatrix}.$$

5. A 2×1 work vector (C_9) is created and C_j is calculated and accumulated for the other observations in the first level of fixed effect. Sire number (j) of the first observation is 9.

$$\begin{aligned}
C_9 &= \sigma^{2*}e \begin{vmatrix} (1/\sigma^2 e1) \cdot (y1_1 - b1_{(1)}) \\ 0 \end{vmatrix} \\
&+ \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} \begin{vmatrix} s_1 \cdot b1_{(1)} + s_1 \cdot u1_{(9)} - s_1 \cdot b1_{(1)} \\ \sigma_e^* h_1 + s_1 \cdot b2_{(1)} + s_1 \cdot u2_{(9)} - s_1 \cdot b2_{(1)} \end{vmatrix} + C_9 \\
&= \sigma^{2*}e \begin{vmatrix} (1/0.91) \cdot \{(3023-3637)/644.5+0.952677\} \\ 0 \end{vmatrix} \\
&+ \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} \begin{vmatrix} s_1 \cdot 0 + s_1 \cdot 0 - s_1 \cdot (-.952677) \\ \sigma_e^* h_1 + s_1 \cdot 0 + s_1 \cdot 0 - s_1 \cdot (1.3783864) \end{vmatrix} + C_9
\end{aligned}$$

where $b1_{(1)}$ and $b2_{(1)}$ are updated solutions for $b1_{(1)}$ and $b2_{(1)}$.

6. Calculation within the next level of fixed effects, until the final level of fixed effects, i.e. steps 2 to 5 are processed similarly.

7. A 2×2 work matrix (F_9) for 9th sire effects is calculated as:

$$\begin{aligned}
F_9 &= \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} \begin{vmatrix} s_1 + \sigma^{2*}e & 1/\sigma^2 e1 & 0 \\ 0 & 0 & 0 \end{vmatrix} \\
&= \begin{vmatrix} b_0^2 & -b_0 \\ -b_0 & 1 \end{vmatrix} \begin{vmatrix} s_1 + \sigma^{2*}e & 1/0.91 & 0 \\ 0 & 0 & 0 \end{vmatrix}
\end{aligned}$$

The same procedure is repeated for all the observations and the elements are accumulated for the corresponding sire in the F matrix.

8. Calculate k from the pedigree information of the sire.

$$\begin{aligned}
k &= (n1/3) + \\
&\begin{cases} 1, & \text{if the male parent of the sire is not known} \\ 4/3, & \text{if the male parent of the sire is known} \end{cases}
\end{aligned}$$

where n1 is the number of male progeny of the sire assuming no inbreeding and only male-side information is available.

In this example, there is no inbreeding and no relationship among sires, i.e. $n1 = 0$. Therefore k is unity for all sires. After that add $k\sigma^{2*}eG_s^{-1}$ to the

present F_9 , i.e.

$$\begin{aligned}
F_9 &= k\sigma^{2*}eG_s^{-1} + F_9 \\
&= \begin{vmatrix} 0.09 & 0.0165 \\ 0.0165 & 0.0525 \end{vmatrix}^{-1} \sigma^{2*}e + F_9.
\end{aligned}$$

9. Add some contribution to C_9 .

Type of contribution to C_j is as follows;

Parent

$$\text{male parent known: } (2/3)G_s^{-1} \begin{vmatrix} \text{pr1} \\ \text{pr2} \end{vmatrix} \sigma^{2*}e.$$

else

$$0$$

Progeny

$$\text{male progeny known: } (2/3)G_s^{-1} \begin{vmatrix} \text{pg1} \\ \text{pg2} \end{vmatrix} \sigma^{2*}e.$$

else

$$0$$

where pri = effect of male parent on i-th trait, pgi = effect of male progeny on i-th trait.

As there is no inbreeding and no relationship, the contribution to C_9 is zero.

10. Sire solution vector (R_9) is calculated after accumulating the elements of F_9 and C_9 in model 1 and model 2:

$$R_9 = F_9^{-1}C_9.$$

The remaining sires are processed in the same way.

The next round of iteration is continued until convergence.

3)-2 : Model 2 iteration

Matrices B, D, C and F are calculated differently from the process outlined for model 1. There are, however, many similarities between models 1 and 2.

As an example of one observation in model 2, in the first level of fixed effect, a work vector (B) is calculated as:

$$B = \sigma^2 * e \begin{vmatrix} e^{11} & e^{12} \\ e^{21} & e^{22} \end{vmatrix} \begin{vmatrix} y^*1 - u1_{(7)} \\ y^*2 - u2_{(7)} \end{vmatrix} + B$$

$$= \sigma^2 * e \begin{vmatrix} 0.91 & 0.195 \\ 0.195 & 0.9475 \end{vmatrix}^{-1} \begin{vmatrix} (3956.5 - 3637)/644.5 - 0 \\ (2427 - 2856.2)/761.6 - 0 \end{vmatrix} + B$$

where $u1_{(7)}$ and $u2_{(7)}$ are 7th sire effects for $u1$ and $u2$, and sire number of the observation is 7.

Matrix D is calculated for the data in the first level of fixed effect:

$$D = \sigma^2 * e \begin{vmatrix} e^{11} & e^{12} \\ e^{21} & e^{22} \end{vmatrix} = \sigma^2 * e \begin{vmatrix} 0.91 & 0.195 \\ 0.195 & 0.9475 \end{vmatrix}^{-1} + D.$$

Sire number(j) of the first observation in model 2 in the first level of fixed effect is seven, therefore a work vector (C_7) is created and C_j is calculated and accumulated for the other observations in model 2 in the first level of fixed effect:

$$C_7 = \sigma^2 * e \begin{vmatrix} e^{11} & e^{12} \\ e^{21} & e^{22} \end{vmatrix} \begin{vmatrix} y^*1 - b1_{(1)} \\ y^*2 - b2_{(1)} \end{vmatrix} + C_7$$

$$= \sigma^2 * e \begin{vmatrix} 0.91 & 0.195 \\ 0.195 & 0.9475 \end{vmatrix}^{-1} \begin{vmatrix} (3956.5 - 3637)/644.5 + .952677 \\ (2427 - 2856.2)/761.6 - 1.378386 \end{vmatrix} + C_7,$$

and F_7 is obtained as:

$$F_7 = \sigma^2 * e \begin{vmatrix} e^{11} & e^{12} \\ e^{21} & e^{22} \end{vmatrix} = \sigma^2 * e \begin{vmatrix} 0.91 & 0.195 \\ 0.195 & 0.9475 \end{vmatrix}^{-1} + F_7.$$

The process is repeated for all data points and elements of F matrix are accumulated.

Solutions by round of iteration are listed in

Table 2 which shows that, for this example, convergence was achieved at the 67th iteration.

Table 2. Solutions by round of iteration^{a)}

Parameters ^{b)}	Round			
	37	47	57	67
b1 ₍₁₎	3648.1	3645.7	3644.2	3643.1
b1 ₍₂₎	3625.2	3627.7	3629.4	3630.5
b2 ₍₁₎	2853.1	2853.8	2854.2	2854.5
b2 ₍₁₎	2868.9	2866.3	2864.5	2863.3
s1 ₍₁₎	48.5	48.7	48.9	49.1
s1 ₍₂₎	62.5	62.8	63.0	63.2
s1 ₍₃₎	30.2	30.2	30.2	30.2
s1 ₍₄₎	-100.6	-100.9	-101.1	-101.2
s1 ₍₅₎	64.3	64.6	64.9	65.0
s1 ₍₆₎	-77.8	-78.0	-78.1	-78.3
s1 ₍₇₎	40.3	40.5	40.6	40.6
s1 ₍₈₎	24.7	24.7	24.7	24.6
s1 ₍₉₎	-39.6	-39.7	-39.9	-39.9
s2 ₍₁₎	15.0	15.0	15.0	15.0
s2 ₍₂₎	-5.5	-5.5	-5.6	-5.6
s2 ₍₃₎	6.9	7.0	7.0	7.0
s2 ₍₄₎	6.9	7.1	7.2	7.3
s2 ₍₅₎	-15.0	-15.0	-15.0	-15.1
s2 ₍₆₎	18.4	18.5	18.6	18.6
s2 ₍₇₎	-0.1	-0.1	-0.1	-0.1
s2 ₍₈₎	8.7	8.7	8.7	8.7
s2 ₍₉₎	9.2	9.2	9.3	9.3

a) unit: kg for the first trait (milk yield); days for the second trait (herd-life)

b) $b_{j(i)}$ indicates fixed effects solution for i th level of j th trait and $s_{j(i)}$ indicates sire solution for i th level of j th trait

Discussion

Censored data of herd-life (y_2) can be written as $y_2 = -b_0 Xb_1 + Xb_2 - b_0 Zu_1 + Zu_2 + S^{-1}h$. With this modification, the coefficient matrix and LHS are directly analogous to Henderson's⁸⁾ mixed model equations. This correspondence has previously been recognized (Foulley and Gianola⁴⁾; Carriquiry et al.¹⁾). The diagonal elements of the matrix S can be considered as the relative reduction of variance in a normal distribution truncated at the point of censoring and the variance of h_i is restored by multiplying s^{-1}_i by the variance before the truncation. That is, the matrix S can be a matrix of weights that takes into account loss of information due to censoring. In this paper, it is assumed that each animal has both traits (milk yield measured

on a continuous scale and herd-life either censored or uncensored). Not every individual will have a complete set of records on both traits. However, the procedure can be easily extended using the generalization applied by Togashi et al.²³⁾

The complexity of the Bayesian approach in the context of the given model and data structure often necessitates the use of certain assumptions (Simianer and Schaeffer¹⁹⁾). Basing the inference about location parameters on the conditional distribution of the dispersion parameters has proven to be very useful (Gianola and Fernando⁶⁾, Gianola et al.⁷⁾, Hoeschele et al.⁹⁾) but this requires that the marginal posterior distribution, $f(\text{dispersion parameters} \mid Y)$, be sharp so that "most of the probability mass is concentrated over a small region about the mode", implying that the

approximation will hold in "large" data sets (Hoeschele et al.⁹⁾). Thus, in the present study, the location parameters were made conditional on the value of the (co) variance component which was assumed to be the mode of $f(\text{dispersion parameters} \mid Y)$.

The joint posterior mode is taken as point estimators of location parameters because the integration of fixed effects-nuisance parameters out of the joint posterior density function for location parameters is technically complex. Lindley and Smith¹¹⁾ suggested that the mode of the joint posterior density of location parameters was an approximation to the mean of posterior function for location parameters. However, posterior mode may not have general invariant properties. An alternative approximation to the mean of location parameters would be to apply Gibbs sampling method (Gelfand and Smith⁵⁾) which would eliminate the necessity for integration as a means of making inferences on location parameters. However, Sorenson et al.²²⁾ have alluded to the need for more research to facilitate full exploitation of the benefits of the Gibbs sampler.

The data set of the numerical example is very small. Furthermore, the number of progeny per sire was also small. This could have caused bias in estimating location parameters. However, the location parameters converged satisfactorily. In a practical application of an iterative procedure, there would be some room to clarify the relationship between the aspect of convergence and data structure. Application of relaxation factors or common intercept approach may become necessary. However, the numerical example presented is simply to demonstrate the procedure. Censored data are treated as right-censored in this paper. The extension of this procedure to the left-censored case, for example, age at the onset of puberty, should be straightforward. In that case, $\log Q(x_i)$ in the likelihood function would be simply replaced by $\log (1-Q(x_i))$. In any censored data, the effect on inferences of treating the data as censored would depend upon the extent to which observations are censored. However, most of the

survival data in dairy cattle would be right-censored. Therefore, genetic evaluations taking account of the covariance of both traits by using the proposed procedure would lead to more correct inferences. The simplicity of the proposed iterative procedure would make it possible to evaluate a large number of animals.

Examples of endurance measures are longevity, productive life span, survival from birth to breeding, prenatal and postnatal survival. These traits can be measured on continuous or discrete scales. The "all or none" scoring on the basis of whether an animal has survived to some prescribed age has the effect of reducing information. Loss of information in treating such traits as censored with different censoring points would be lower than treating them as discrete data.

One might argue that selection for herd-life is likely to increase generation interval and decrease annual genetic progress accrued in such traits as milk yield. However, genetic evaluations which include both herd-life (a measure of adaptability) and productivity (e.g. milk yield) would make significant contributions in situations where adaptation is considered crucial. An understanding of the mechanism responsible for each component that constitutes the adaptability such as trypanotolerance, heat-tolerance, food-utilization and water conservation was not examined individually. However, total benefit of each component is key criterion for survival of African cattle in harsh environment. Many African countries place their indigenous livestock populations at risk through programmes of exotic breed importation and/or crossbreeding. Rarely has adequate attention been given to evaluating and setting realistic and optimum breeding objectives prior to embarking on breed improvement programmes. Therefore, breeding programmes should be based on the fact that indigenous cattle are genetically very adaptable to African conditions as well as on improving productivity. From this point of view, simultaneous estimation procedure for adaptability and productivity is crucial for African cattle. Any

measure can not be processed properly without simultaneous estimation procedure for productivity and adaptability developed in this study. It is not too much to say that African indigenous cattle have a high degree of heat tolerance, resistance to many of the diseases represented by trypanosomiasis and the ability to survive long period of feed and water shortage through natural selection over hundreds of generations. It is, however, quite obvious that the potential for milk and/or meat production is poorly developed in African indigenous breeds of cattle. Therefore, breeding programme to improve this potential to a satisfactory level without sacrificing adaptability should be initiated and continued as exemplified by the international collaboration efforts.

Acknowledgements

Togashi acknowledges Dr. Fitzhugh, Director General of ILRI and Drs. Hayakawa and Kawashima, present and former Directors of Japan International Research for Agricultural Sciences, respectively. The authors also acknowledge the helpful support of Ms Tirsit Mamo and Ms Roman Tirfie.

References

- 1) Carriquiry, A.L., Gianola, D. and Fernando, R. L.(1987). Mixed -model analysis of a censored normal distribution with reference to animal breeding. *Biometrics* **43**:929–939.
- 2) DeLorenzo, M. A. and Everett, R. W.(1986). Prediction of sire effects for probability of survival to fixed ages with a logistic linear model. *J.Dairy Sci.* **69**:501–509.
- 3) Famula, T. R.(1981). Exponential stayability model with censoring and covariates. *J.Dairy Sci.* **64**:538–545.
- 4) Foulley, J. L. and Gianola, D.(1984). Estimates of genetic merit from bivariate “all-or-none” responses. *Genet Sel Evol.* **16**:285–306.
- 5) Gelfand, A. E. and Smith, A. F. M.(1990). Sampling-based approaches to calculating marginal densities. *J.Am.Stat Assoc.* **85**:398–409.
- 6) Gianola, D. and Fernando, R. L.(1986). Bayesian methods in animal breeding theory. *J Anim Sci.* **63**:217–244.
- 7) Gianola, D., Foulley, J. L. and Fernando, R. L. (1986). Prediction of breeding values when variances are not known. *Genet Sel Evol.* **18**:485–498.
- 8) Henderson, C. R.(1973). Sire evaluation and genetic trends. *In:Proc. Anim. Breed. Genet. Symp. in Honor of Dr. J. L. Lush.* ASAS and ADSA, Champaign, Illinois: 10–41.
- 9) Hoeschele, I., Gianola, D. and Foulley, J. L. (1987). Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet.* **104**:334–349.
- 10) ILCA (1994). Improving livestock production in Africa, “Evolution of ILCA’s programme”. ILCA. Addis Ababa, Ethiopia. 233.
- 11) Lindley, D. V. and Smith, A. F. M.(1972). Bayes estimates for the linear model. *J R Statist Soc B.* **34**:1–41.
- 12) Marry, M., Clifford, D. J., Gettinby, G., Snow, W. F. and McIntyne, W. I. M.(1981). A study of the susceptibility to African trypanosomiasis of N’Dama and Zebu cattle in an area of *Glossina morsitans submorsitans* challenge. *Veterinary Record.* **109**:503–510.
- 13) Misztal, I. and Gianola, D.(1987). Indirect solution of mixed model equations. *J Dairy Sci.* **70**:716–723.
- 14) Otchere, E.(1983). The productivity of white Fulani cattle in pastoralist herds on the Kaduna plains of Nigeria. ILCA programme, Kaduna, Nigeria, ILCA, Addis Ababa, Ethiopia.
- 15) Rege, J. E. O. and Mosi, R. O.(1989). An analysis of the Kenyan Friesian breed from 1968-1984:Genetic and environmental trends and related parameters of milk production. *Bulletin Anim Health Prod Africa* **37**:267–278.
- 16) Rege, J. E. O.(1991). Genetic analysis of reproductive and productive performance of Friesian cattle in Kenya. Genetic and phenotypic trends. *J Anim Breed Genet.*

- 108:424-433.
- 17) Schaeffer, L. R. and Kennedy, B. W. (1986). Computing strategies for solving mixed model equations. *J Dairy Sci.* **69**:575-579.
 - 18) Searle, S R.(1982). Matrix algebra useful for statistics. John Wiley & Sons, New york, NY.438.
 - 19) Semianer, H. and Schaeffer, L. R.(1989). Estimation of covariance components between one continuous and one binary trait. *Genet Sel Evol.* **21**:303-315.
 - 20) Smith, S. P. and Quass, R. L. (1984). Productive lifespan of bull progeny groups: Failure time analysis. *J Dairy Sci.* **67**:2999-3007.
 - 21) Smith, S. P. and Allaire, F. R.(1986). Analysis of failure time measured in dairy cows: Theoretical considerations in animal breeding. *J Dairy Sci.* **69**:217-227.
 - 22) Sorensen, D., Anderson, S., Jensen, J., Wang, C. S. and Gianola, D.(1994). Inferences about genetic parameters using the Gibbs sampler. Proc. 5th World Cong Genet Appl Livestock Prod. **18**:321-328.
 - 23) Togashi, K., Smith, S. P. and Hammond, K. (1992). Application of an approximate EM type algorithm and some devices for estimation of (co)variances into the data set with combinations of measurements on a continuous and a binary trait. *Res Bull Hokkaido Nat Agric Exp Stn.* **157**:1-23.
 - 24) Togashi, K. and Rege, J. E. O.(1996). Maximum likelihood and an approximate expectation and maximization procedures to estimate dispersion parameters in a data set of combination of censored and uncensored traits. *JIRCAS J.* [submitted].
 - 25) Van Vleck, D.(1980). Stayability evaluation as a categorical trait and by considering other traits. *J Dairy Sci.* **63**:1172-1180.

打ち切りおよび非打ち切り型データ混合モデル式での パラメータ効果の反復推定法

富樫研治^{a)}, J.E.O. Rege^{b)}

a) 国際農林水産業研究センター畜産草地部
(〒305 茨城県つくば市大わし1-2)

b) ILRI (International Livestock Research Institute),
(Addis Ababa, POBox 5689, Ethiopia)

摘 要

打ち切り型と非打ち切り型データを含む混合モデル式において、牛群効果のような母数効果と父牛効果のような変数効果を係数行列式を作らずに反復法で推定する方法を開発した。本手法は、係数行列式が大きくなっても、その解を直接に逆行列を解かないで反復して求めるもの

で、効果の数が大きい場合に特に、有効となる。数値例として、打ち切り型データとして牛群滞在期間、非打ち切り型データとして乳量を取りあげてアフリカ牛に適用した。

キーワード：父牛モデル，打ち切り型データ，反復法，牛群滞在期間，乳量

^{a)} 現在：北海道農業試験場畜産部（〒062 札幌市羊が丘1）