On sire evaluation with uncertain paternity

Jl Foulley, R Thompson, D Gianola

Genet Sel Evol (1990) 22, 373-376

Short note

On sire evaluation with uncertain paternity

JL Foulley [1*], R Thompson [2], D Gianola [3]

[1] *INRA, Station de Génétique Quantitative, 78350 Jouy-en-Josas, France;*
[2] *Institute of Animal Physiology and Genetics Research, Edinburgh EH9 3JN, UK;*
[3] *Department of Animal Science, Urbana, IL 61801, USA*

(Received 28 November 1989; accepted 21 May 1990)

Foulley *et al* (1987) – from now on referred to as FGP – have described a first order algorithm (functional iteration) for computing the maximum *a posteriori* (MAP) estimator of the location parameter θ of a normal distribution in situations where there is uncertainty with respect to the assignment of data to some effects (*eg* sire) of θ. This algorithm has a simple form, related to the mixed model equations, and is easy to program and to apply. Second order algorithms can also be used for computing MAP estimates of θ. These algorithms are needed for getting estimates of the asymptotic accuracy of these modal estimators, or for variance component estimation.

The objective of this note is to correct formulae needed for such algorithms given by FGP, and to describe an alternative computing procedure based on the method of scoring.

Let $L(\theta)$ be the logposterior density; the first derivatives can be written as:

$$\dot{L}(\theta) = \sum_{j}^{m} \mathbf{W}_j \mathbf{v}_j - \Sigma^{-1}(\Theta - \alpha) \tag{1}$$

where:

$$\mathbf{W}_{j(n,p)} = (\mathbf{w}_{1j}, \mathbf{w}_{2j}, \ldots, \mathbf{w}_{ij}, \ldots, \mathbf{w}_{nj})' \tag{2}$$

$\mathbf{w}_{ij}$ being an incidence column vector $(p,1)$ pertaining to the ith observation $(i = 1, 2, \ldots, n)$, given j is the true sire, and

$$\mathbf{v}_{j(n,1)} = \{v_{ij} = q_{ij}(y_i - \mu_{ij})\} \tag{3}$$

* Correspondence and reprints

Differentiating (1) again, one obtains the expression for the negative Hessian of $L(\theta)$, *ie*:

$$-\ddot{L}(\theta) = \sigma_e^{-2}(\Sigma^{-1}\sigma_e^2 + \sum_j \sum_k \mathbf{W}_j'\mathbf{R}_{jk}\mathbf{W}_k) \tag{4}$$

where:

R_{jk} is an $(n \times n)$ diagonal matrix pertaining to sires j and k with ith element

$$r_{i,jk} = -\frac{\partial v_{ik}}{\partial \mu_{ij}}; \ i = 1, 2, \ldots, n \text{ observations}$$

or, explicity:

$$r_{i,jk} = \delta_{jk}q_{ij} - [q_{ij}(\delta_{jk} - q_{ik})(y_i - \mu_{ij})(y_i - \mu_{ik})]/\sigma_e^2 \tag{5}$$

where q_{ij} is the same as in FGP and δ_{jk} is the Kronecker delta, equal to 1 if $j = k$ or 0 otherwise.

Note that these formulae are slightly different from those given by FGP (Appendix B, p 99). Actually, formula (5) reduces to their expression [B4] when $j = k$ and to:

$$r_{i,jk} = q_{ij}q_{ik}(y_i - \mu_{ij})(y_i - \mu_{ik})/\sigma_e^2 \text{ if } j \neq k \tag{6}$$

The Newton-Raphson algorithm can be written as:

$$\left[\left(\sum_j \sum_k \mathbf{W}_j'\mathbf{R}_{jk}^{[t]}\mathbf{W}_k\right) + \Sigma^{-1}\sigma_e^2\right]\Delta\theta^{[t+1]} = \sum_j \mathbf{W}_j'\mathbf{v}_j^{[t]} - \Sigma^{-1}\sigma_e^2(\theta^{[t]} - \alpha) \tag{7}$$

Letting $\mathbf{W}_j = (\mathbf{X}, \mathbf{z}_j)$ where $\mathbf{X}$ and $\mathbf{z}_j$ are (n, p) and (n, m) incidence matrices (given j being the true sire) pertaining to the β and $\mathbf{u}$ elements of $\theta = (\beta', \mathbf{u}')'$, this system can be expressed more explicitly as:

$$\begin{bmatrix} \mathbf{X}'\mathbf{K}^{[t]}\mathbf{X} & \mathbf{X}'\mathbf{L}^{[t]} \\ \mathbf{L}'^{[t]}\mathbf{X} & \mathbf{M} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \Delta\beta^{[t+1]} \\ \Delta\mathbf{u}^{[t+1]} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}'y \\ \mathbf{Q}'^{[t]}\mathbf{y} \end{bmatrix} - \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Q}^{[t]} \\ \mathbf{Q}'^{[t]}\mathbf{X} & \mathbf{D}_c^{[t]} \end{bmatrix} \begin{bmatrix} \beta^{[t]} \\ \mathbf{u}^{[t]} \end{bmatrix} - \begin{bmatrix} 0 \\ \mathbf{A}^{-1}\lambda\mathbf{u}^{[t]} \end{bmatrix} \tag{8}$$

where:

$$\mathbf{K}_{(n\times n)} = \text{Diag}\{k_i = \sum_{j,k} r_{i,jk}; \ i = 1, 2, \ldots, n \tag{9a}$$

$$\mathbf{L}_{(n\times m)} = \{l_{ij} = \sum_k r_{i,jk}\}; \ i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m \tag{9b}$$

$$\mathbf{M}_{(m\times m)} = \{m_{jk} = \sum_i r_{i,jk}\};\ j;k = 1,2,\ldots,m \tag{9c}$$

and, as before:

$$\mathbf{Q}_{(n\times m)} = \{q_{ij}\};\ \mathbf{D}_{c(m\times m)} = \mathrm{Diag}\{d_{c_j} = \sum_i q_{ij} \tag{9d}$$

Another possibility would be to develop a scoring procedure by taking in (4) the unconditional expectation of $r_{i,jk}$ based on the following expression:

$$E(y_i - \mu_{ij})(y_i - \mu_{ik}) = \sigma_e^2 + \Big[\sum_{h=1}^{m} p_{ih}(\mathbf{z}_{ih} - \mathbf{z}_{ij})'\mathbf{A}(\mathbf{z}_{ih} - \mathbf{z}_{ik})\Big]\sigma_u^2 \tag{10a}$$

or, more explicitly:

$$E(y_i - \mu_{ij})(y_i - \mu_{ik}) = \sigma_e^2 + \Big[\sum_{h=1}^{m} p_{ih}(a_{hh} + a_{jk} - a_{hj} - a_{hk})\Big]\sigma_u^2 \tag{10b}$$

where a_{jk} is the jk element of the numerator relationship matrix $\mathbf{A}$; if $j = k$, formulae (6a and b) apply with $\mathbf{z}_{ij} = \mathbf{z}_{ik}$ and $a_{hj} = a_{hk}$ respectively.

Finally, it must be kept in mind that "regular" mixed model equations can also be used as an alternative to (4) as shown recently by Im (1989).

The same corrections apply to Foulley and Elsen (1988) on p 233. Their expression [23b] should read:

$$\ddot{L}(\theta) = -\Gamma^{-1} - \sum_k \sum_l \mathbf{W}_k'\mathbf{R}_{kl}\mathbf{W}_l \tag{11}$$

with, in [24b]:

$$\mathbf{R}_{kl_{(M\times M)}} = \mathrm{Diag}\{r_{m,kl}\};\ m = 1,2,\ldots,M$$

and, in [26c]:

$$r_{m,kl} = -\frac{\partial s_{ml}}{\partial \mu_{mk}} \tag{12a}$$

or:

$$r_{m,kl} = -\delta_{kl} q_{ml}\frac{\partial v_{ml}}{\partial \mu_{ml}} - q_{ml}(\delta_{kl} - q_{mk})v_{ml}v_{mk} \tag{12b}$$

The Newton-Raphson algorithm consists of iterating from round t to $t+1$ with:

$$\Big[\Big(\sum_k \sum_l \mathbf{W}_k'\mathbf{R}_{kl}^{[t]}\mathbf{W}_l\Big) + \Gamma^{-1}\Big]\Delta\theta^{[t+1]} = \sum_k \mathbf{W}_k'\mathbf{S}_k^{[t]} - \Gamma^{-1}(\theta^{[t]} - \theta_0) \tag{13}$$

The expression in (13) replaces that in [25]. Formula [26b] for v_{ml} is unaltered and reduces to $v_{ml} = (y_m - \mu_{ml})/\sigma_e^2$ in the normal case.

REFERENCES

Foulley JL, Gianola D, Planchenault (1987) Sire evaluation with uncertain paternity. *Genet Sel Evol* 19, 83-102

Foulley JL, Elsen JM (1988) Posterior probability at a major locus based on progeny-test results for discrete characters. *Genet Sel Evol* 20, 227-238

Im S (1989) A note on sire evaluation with uncertain paternity. *Appl Stat* (in press)