On sire evaluation with uncertain paternity

Jl Foulley, R Thompson, D Gianola

To cite this version:

Jl Foulley, R Thompson, D Gianola. On sire evaluation with uncertain paternity. Genetics Selection Evolution, BioMed Central, 1990, 22 (3), pp.373-376. <hal-00893852>

HAL Id: hal-00893852

https://hal.archives-ouvertes.fr/hal-00893852

Submitted on 1 Jan 1990

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Short note

On sire evaluation with uncertain paternity

JL Foulley*, 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 \( \theta \) of a normal distribution in situations where there is uncertainty with respect to the assignment of data to some effects (eg sire) of \( \theta \). 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 \( \theta \). 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:

\[
\dot{L}(\theta) = \sum_{j} W_j v_j - \Sigma^{-1} (\Theta - \alpha)
\]  

(1)

where:

\[
W_{j(n,p)} = (w_{1j}, w_{2j}, \ldots, w_{ij}, \ldots, w_{n_j})'
\]

(2)

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

\[
v_{j(n,1)} = \{v_{ij} = q_{ij}(y_i - \mu_{ij})\}
\]

(3)

* Correspondence and reprints
Differentiating (1) again, one obtains the expression for the negative Hessian of $L(\theta)$, i.e:

$$-\ddot{L}(\theta) = \sigma_e^{-2}(\Sigma^{-1}\sigma_e^2 + \sum_j \sum_k W_j' R_{jk} W_k)$$ (4)

where:

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

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

or, explicitly:

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

where $q_{ij}$ is the same as in FGP and $\delta_{jk}$ 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:

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

The Newton-Raphson algorithm can be written as:

$$\left[ \sum_j \sum_k W_j' R_{jk} W_k \right] + \Sigma^{-1} \sigma_e^2 \Delta \theta^{(t+1]} = \sum_j W_j' v_j^{[t]} - \Sigma^{-1} \sigma_e^2 (\theta^{[t]} - \alpha)$$ (7)

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

$$\begin{bmatrix}
X'K^{[t]}X & X'L^{[t]} \\
L^{[t]}X & M + A^{-1} \lambda
\end{bmatrix}
\begin{bmatrix}
\Delta \beta^{(t+1]} \\
\Delta u^{(t+1]}
\end{bmatrix}
= \begin{bmatrix}
X'y & X'X & X'Q^{[t]} & \beta^{[t]} \\
Q^{[t]}y & Q^{[t]}X & Q^{[t]}X & u^{[t]}
\end{bmatrix}
- \begin{bmatrix}
0 \\
A^{-1} \lambda u^{[t]}
\end{bmatrix}$$ (8)

where:

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

$$L_{(n \times m)} = \{l_{ij} = \sum_k r_{i,jk}; \ i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m\}$$ (9b)
On sire evaluation with uncertain paternity

\[
M_{(m \times m)} = \{m_{jk} = \sum_i r_{i,jk}; j; k = 1, 2, \ldots, m \} \quad (9c)
\]

and, as before:

\[
Q_{(n \times m)} = \{q_{ij} \}; \quad D_{c(m \times m)} = \text{Diag}\{d_{cj} = \sum_i q_{ij} \} \quad (9d)
\]

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:

\[
E(y_i - \mu_{ij})(y_i - \mu_{ik}) = \sigma_e^2 + \left[ \sum_{h=1}^{m} p_{ih}(z_{ih} - z_{ij})'A(z_{ih} - z_{ik}) \right] \sigma_u^2 \quad (10a)
\]
or, more explicitly:

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

where \(a_{jk}\) is the \(jk\) element of the numerator relationship matrix \(A\); if \(j = k\), formulae (6a and b) apply with \(z_{ij} = z_{ik}\) 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:

\[
\bar{L}(\theta) = -\Gamma^{-1} - \sum_k \sum_l W_k' R_{kl} W_l \quad (11)
\]

with, in [24b]:

\[
R_{kl(m \times m)} = \text{Diag}\{r_{m,kl}\}; \quad m = 1, 2, \ldots, M
\]

and, in [26c]:

\[
r_{m,kl} = -\frac{\partial s_{ml}}{\partial \mu_{mk}} \quad (12a)
\]
or:

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

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

\[
\left[ (\sum_k \sum_l W_k' R_{kl}^{[t]} W_l) + \Gamma^{-1} \right] \Delta \theta^{[t+1]} = \sum_k W_k' S_k^{[t]} - \Gamma^{-1}(\theta^{[t]} - \theta_0) \quad (13)
\]
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)