THE ECONOMIC AND SOCIAL RESEARCH INSTITUTE

Memorandum Series No. 176

OBTAINING EXPECTED MAXIMUM LOG LIKELIHOOD ESTIMATORS

DENIS CONNIFFE

February 1988

OBTAINING EXPECTED MAXIMUM LOG LIKELIHOOD ESTIMATORS

DENIS CONNIFFE

The Economic and Social Research Institute, Dublin 4, Ireland

Abstract. This paper describes a widely applicable method of obtaining expected maximum log likelihood estimators (EMLE). Approximate numerical solutions are required in the most general cases, but exact algebraic solutions are possible in special cases. It is shown that in an important special case the EMLE solution coincides with ML as applied to conditional likelihood. The method of restricted or residual maximum likelihood (REML) is compared with EMLE by means of an example.

I Introduction

The likelihood $L(\theta\ x)$, where θ and x are vectors of parameter variable and data respectively, is not directly a function of the true parameter values θ_o. However, the expected log likelihood is a function of θ and θ_o and its maximum occurs at $\theta=0_o$. So determining the true value is the same as determining the maximum of the expected log likelihood. Given regularity assumptions, maximisation implies:

$$E\left(\frac{\partial \log L}{\partial \theta_i}\right) = 0, \text{ at } \theta_j=\theta_{jo} \text{ for all i and j.} \qquad (1)$$

If there is only one component of θ, this equation says that the expectation of the derivative of the log likelihood is zero at the true value. Hence, equating the derivative of the sample likelihood to zero and solving for θ is a plausible way of estimating θ_o. Of course, it also maximises the sample likelihood, so the estimate is the usual ML one. With multiple parameters, the derivative for θ_i will usually be a function of other parameter variables besides

θ_i. If these are replaced by estimates, perhaps obtained by equating other derivatives to zero, then, in general,

$$E\left[\frac{\partial \log L}{\partial \theta_i}(\theta_i \tilde{\theta}_j \ldots \tilde{\theta}_k x)\right] \neq 0$$

because the estimates differ from the true values. Then the values that maximise the sample likelihood are not the same as estimates obtained by equating functions of likelihood derivatives to the expectations they would have at the true values of the parameters. This paper develops a generally applicable method for obtaining the latter (EMLE) estimates.

Of course, it may be possible to manipulate from the likelihood equations

$$\frac{\partial \log L}{\partial \theta_i} = \ell_i(\theta x) \tag{2}$$

to the equations

$$W_i\left(\frac{\partial \log L}{\partial \theta}, \theta\right) = k_i(\theta_i x) \tag{3}$$

so that θ_i may be estimated by equating k_i to the expectation of W_i at θ_o. The expectation could involve parameters other than θ_i, but then estimates could be inserted. If W_i happened to be a linear function

$$\Sigma A_i(\theta) \frac{\partial \log L}{\partial \theta_i}$$

then the expectations of W_i would be zero and in fact EMLE and ML would coincide. Examples of EMLE solutions via the progression from (2) to (3) were given in Conniffe (1987) for fairly simple cases. However, it is not certain that unique algebraically explicit expressions for (3) can always be obtained from (2), nor that exact expectations can be derived for the W_i. Approximations and numerical approaches are possible, but then it seems simpler to commence the approximations at the stage of the initial equations (2).

2 Approximating Derivative Expectations

Suppose θ has two components ϕ and ψ and that an initial estimate $\tilde{\phi}$ is available. The derivative of the log likelihood w.r.t. ψ, evaluated at $\tilde{\phi}$, is approximately:

$$\frac{\partial \log L}{\partial\psi}(\phi\psi x) + (\tilde{\phi}-\phi)\frac{\partial^2 \log L}{\partial\phi\partial\psi} + \tfrac{1}{2}(\tilde{\phi}-\phi)^2 \frac{\partial^3 \log L}{\partial\phi^2\partial\psi} \qquad (4)$$

At the true values $\phi_o\psi_o$, the expectation of the first term is zero. The expectations of the second and third terms would depend on the composition of $\tilde{\phi}$, considered as a function of the sample data, but can be approximated by the expectations of

$$(\hat{\phi} - \phi)\frac{\partial^2 \log L}{\partial\phi\partial\psi} \text{ and } \tfrac{1}{2}(\hat{\phi}-\phi)^2 \frac{\partial^3 \log L}{\partial\phi^2\partial\psi} \qquad (5)$$

where

$$\hat{\phi}-\phi = V_\phi\frac{\partial \log L}{\partial\phi} + C\frac{\partial \log L}{\partial\psi} \qquad (6)$$

and
$$\begin{bmatrix} V_\phi & C \\ C & V_\psi \end{bmatrix}^{-1} = -E\begin{bmatrix} \frac{\partial^2 \log L}{\partial\phi^2} & \frac{\partial^2 \log L}{\partial\phi\partial\psi} \\ \frac{\partial^2 \log L}{\partial\phi\partial\psi} & \frac{\partial^2 \log L}{\partial\psi^2} \end{bmatrix} \qquad (7)$$

Assuming that the data consist of n independent values drawn from the same distribution $f(\phi\psi x)$ the equation (6) is asymptotically valid for the EMLE (or the MLE since they coincide asymptatically) estimator. So the expectation of the first term of (5) is:

$$V_\phi E\left(\frac{\partial \log L}{\partial\phi}\,\frac{\partial^2 \log L}{\partial\phi\partial\psi}\right) + C\,E\left(\frac{\partial \log L}{\partial\psi}\,\frac{\partial^2 \log L}{\partial\phi\partial\psi}\right)$$

$$\text{or}\quad n(V_\phi M_{10010} + C\,M_{01010}) \qquad (8)$$

where

$$M_{ijk\ell m} = E\left[\left(\frac{\partial \log f}{\partial\phi}\right)^i\left(\frac{\partial \log f}{\partial\psi}\right)^j\left(\frac{\partial^2 \log f}{\partial\phi^2}\right)^k\left(\frac{\partial^2 \log f}{\partial\phi\partial\psi}\right)^\ell\left(\frac{\partial^2 \log f}{\partial\psi^2}\right)^m\right]$$

The expectation of the second term of (5) is

$$\tfrac{1}{2}V_\phi \; E\left(\frac{\partial^3 \log L}{\partial\phi^2\partial\psi}\right) + \tfrac{1}{2}\mathrm{Cov}\left[(\hat{\phi}-\phi)^2, \frac{\partial^3 \log L}{\partial\phi^2\partial\psi}\right] \tag{9}$$

Using

$$E\left(\frac{\partial^3 \log L}{\partial\phi^2\partial\psi}\right) = -2E\left(\frac{\partial \log L}{\partial\phi}\frac{\partial^2 \log L}{\partial\phi\partial\psi}\right) - E\left(\frac{\partial \log L}{\partial\psi}\frac{\partial^2 \log L}{\partial\phi^2}\right) - E\left[\left(\frac{\partial \log L}{\partial\phi}\right)^2 \frac{\partial \log L}{\partial\psi}\right]$$

the expectation of the first term of (9) is

$$-\tfrac{1}{2}n\; V_\phi(2M_{10010} - M_{01100} - M_{21000}) \tag{10}$$

The expectation of the second term of (9) is

$$\tfrac{1}{2}n\;\mathrm{Cov}\left[V_\phi^2\left(\frac{\partial \log f}{\partial\phi}\right)^2 + 2V_\phi C\,\frac{\partial \log f}{\partial\phi}\frac{\partial \log f}{\partial\psi} + C^2\left(\frac{\partial \log f}{\partial\psi}\right)^2, \frac{\partial^3 \log f}{\partial\phi^2\partial\psi}\right] \tag{11}$$

From (7) it is clear that V_ϕ and C are $O(\frac{1}{n})$ so that (8) and (10) are $O(1)$, but (11) is $O(\frac{1}{n})$. So (11) will be treated as negligible in comparison to the other terms. Combining terms gives the approximate expectation

$$n[C\,M_{01010} - \tfrac{1}{2}V_\phi(M_{01100} + M_{21000})] \tag{12}$$

By equating

$$\frac{\partial \log L}{\partial\psi}(\tilde{\phi}\;\psi\;x) \tag{13}$$

to (12), which is approximately the expectation it would have at $\psi=\psi_o$, an equation is obtained for ψ. If (12) is a function of ϕ_o as well as of ψ_o, the estimate $\tilde{\phi}$ must be substituted for it.

Given an initial estimate $\tilde{\psi}$, an equation for ϕ can be obtained by equating

$$\frac{\partial \log L}{\partial\phi}(\phi\;\tilde{\psi}\;x) \tag{14}$$

to the result corresponding to (12), which is

$$n[C\,M_{10010} - \tfrac{1}{2}V_\psi(M_{10001} + M_{12000})] \tag{15}$$

Alternatively, the result for ψ of equating (13) to (12) could be taken as $\tilde{\psi}$ and used in (14) and (15) to obtain a new $\tilde{\phi}$ which could be inserted again in (13) and (12), so defining an iterative process.

The procedure to generalise the approach to any number of parameters is obvious in principle and not mathematically difficult in practice. But the expressions, especially the 'M' terms, quickly become very clumsy; indicating the need for a more concise, though perhaps less transparent, notation.

It may be useful to restate what has been done in the steps from (4) to (12) and (13). The knowledge that the expected values of derivatives of the log likelihood are zero at the true values of parameters does not (except in very large samples) justify replacing one parameter in the derivative by an estimate and solving for the other parameter by equating to zero. But since the derivative at the estimated value can be related to the derivative at the true values by the approximation (4), the non-zero expectation to which the former derivative ought to be equated can be deduced, at least approximately. The procedure adopted here resembles that used in deriving corrections for the biases of ML estimators [see, for example, Cox and Hinkley (1974), Chapter 9].

3 An Important Special Case and Conditional Likelihood

If one of the derivative equations, say for ϕ, takes the form

$$\frac{\partial \log L}{\partial \phi} = AG(\phi\ x) \tag{16}$$

where A is not a function of x, it follows that

$$E[G(\phi\ x)] = 0$$

and ϕ may be estimated by just equating $G(\phi\ x)$ to zero and solving for ϕ. But it is then also true that

$$E\left(\frac{\partial^2 \log L}{\partial\phi\,\partial\psi}\right) = \frac{\partial A}{\partial\psi} E[G(\phi\ x)] = 0$$

and

$$E\left(\frac{\partial^3 \log L}{\partial\phi^2\,\partial\psi}\right) = \frac{\partial A}{\partial\psi} E\left(\frac{\partial G}{\partial\phi}\right)$$

So C in (6) is zero and final expressions simplify considerably. In particular, if (16) takes the form

$$\frac{\partial \log L}{\partial\phi} = A(\psi)(\tilde{\phi} - \phi) \tag{17}$$

associated with suitably parametrised distributions belonging to the exponential family, (4) becomes

$$\frac{\partial \log L}{\partial\psi}(\phi\psi x) + (\tilde{\phi} - \phi)^2 \frac{dA}{d\psi} - \tfrac{1}{2}(\tilde{\phi} - \phi)^2 \frac{dA}{d\psi}$$

with expectation

$$\tfrac{1}{2} V_\phi \frac{dA}{d\psi} = \tfrac{1}{2} \frac{1}{A} \frac{dA}{d\psi} \tag{18}$$

Note that with the form (17), there is no need to distinguish $\tilde{\phi}$ and $\hat{\phi}$ since (6) holds for $\tilde{\phi}$. For a vector of parameters ϕ, so that A in (17) is a matrix, the generalisation of (18) is easily seen to be

$$\tfrac{1}{2} T_r\left(A^{-1} \frac{dA}{d\psi}\right) \tag{19}$$

The EMLE approach will not always provide an estimator of ψ, given $\tilde{\phi}$. The form (17) corresponds to

$$\log L = -\tfrac{1}{2} A(\psi)(\tilde{\phi} - \phi)^2 + B(\psi\ x)$$

But if B is not a function of x

$$\frac{\partial \log L}{\partial\psi} = -\tfrac{1}{2} \frac{dA}{d\psi} (\tilde{\phi} - \phi)^2 + \frac{dB}{d\psi}$$

So that, evaluated at $\tilde{\phi}$, the derivative is just a constant which, of course, equals (18). For x to provide information about a parameter it is obviously necessary that the distribution of x depend on the parameter, but this is not sufficient. The derivative (13), or the k_i in (3), must be a function of x. The case may seem trivial but it is related to an important result. Suppose

$$L(\phi\psi x) = k(\psi x|\tilde{\phi})g(\phi\psi\tilde{\phi}) \tag{20}$$

so that

$$\frac{\partial \log L}{\partial \phi} = \frac{\partial \log g}{\partial \phi}$$

$$\frac{\partial \log L}{\partial \psi} = \frac{\partial \log k}{\partial \psi} + \frac{\partial \log g}{\partial \psi}$$

and further suppose

$$\log g = -\tfrac{1}{2}A(\psi)(\tilde{\phi} - \phi)^2 + B(\psi)$$

The EMLE method equates

$$\frac{\partial \log L}{\partial \psi}(\tilde{\phi}\ \psi\ x) \text{ to } \tfrac{1}{2}\frac{1}{A}\frac{dA}{d\psi} \tag{21}$$

from (18). But

$$\frac{\partial \log g}{\partial \psi} = -\tfrac{1}{2}\frac{dA}{d\psi}(\tilde{\phi} - \phi)^2 + \frac{dB}{d\psi} \tag{22}$$

so at $\tilde{\phi}$ this derivative equals the second term on the right hand side of (22). But since the expectation of (21) must be zero at the true values as g is a marginal likelihood

$$\frac{dB}{d\psi} = \tfrac{1}{2}\frac{1}{A}\frac{dA}{d\psi}$$

Since the distribution conditional on $\tilde{\phi}$ is assumed in (20) not to be a function of ϕ the derivative of log k is unaffected by replacing ϕ by $\tilde{\phi}$. So (21) reduces to equating

$$\frac{\partial \log k}{\partial \psi}$$

to zero. That is, EMLE on the joint likelihood (20) gives the same result as ML on the conditional distribution of x given $\tilde{\phi}$. Remembering the small sample optimality sometimes attained by EMLE [Conniffe (1987)], this may explain why inference based on a conditional likelihood is sometimes claimed to have advantages over inference based on a joint likelihood.

4 An Example of EMLE and of REML

Consider the problem of estimating the parameters μ, σ_1^2 and σ_2^2 given independent samples of sizes n_1 and n_2 from two normal distributions with the same mean. The log likelihood, apart from a constant, is

$$-\frac{n_1}{2}\log\sigma_1^2 - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu)^2 - \frac{n_2}{2}\log\sigma_2^2 + \frac{1}{2\sigma_2^2}\sum_{i=1}^{n_2}(y_{2i}-\mu)^2 \qquad (23)$$

$$\frac{\partial \log L}{\partial\mu} = \frac{n_1}{\sigma_1^2}(\bar{y}_1-\mu)+\frac{n_2}{\sigma_2^2}(\bar{y}_2-\mu) \qquad \frac{\partial^2\log L}{\partial\mu^2} = -\left(\frac{n_1}{\sigma_1^2}+\frac{n_2}{\sigma_2^2}\right)$$

$$\frac{\partial \log L}{\partial\sigma_1^2} = -\frac{n_1}{2\sigma_1^2}+\frac{1}{2\sigma_1^4}\Sigma(y_{1i}-\mu)^2 \qquad \frac{\partial^2\log L}{\partial(\sigma_1^2)^2} = \frac{n_1}{2\sigma_1^4}-\frac{1}{\sigma_1^6}\Sigma(y_{1i}-\mu)^2$$

$$\frac{\partial \log L}{\partial\sigma_2^2} = \frac{-n_2}{2\sigma_2^2}+\frac{1}{2\sigma_2^4}\Sigma(y_{2i}-\mu)^2 \qquad \frac{\partial^2\log L}{\partial(\sigma_2^2)^2} = \frac{n_2}{2\sigma_2^4}-\frac{1}{\sigma_2^6}\Sigma(y_{2i}-\mu)^2$$

Also

$$\frac{\partial^2\log L}{\partial\sigma_1^2\partial\sigma_2^2} = 0$$

and

$$E\left(\frac{\partial^2\log L}{\partial\mu\partial\sigma_1^2}\right) = E\left(\frac{\partial^2\log L}{\partial\mu\partial\sigma_2^2}\right) = 0$$

So the equations corresponding to (6) are

$$\hat{\mu} - \mu = \frac{\sigma_1^2 \sigma_2^2}{n_1\sigma_1^2+n_2\sigma_2^2} \frac{\partial \log L}{\partial \mu} \tag{24}$$

$$\hat{\sigma}_1^2 - \sigma_1^2 = \frac{2\sigma_1^4}{n_1} \frac{\partial \log L}{\partial \sigma_1^2} \tag{25}$$

$$\hat{\sigma}_2^2 - \sigma_2^2 = \frac{2\sigma_2^4}{n_2} \frac{\partial \log L}{\partial \sigma_2^2} \tag{26}$$

The EMLE equation for μ is obtained by equating

$$\frac{\partial \log L}{\partial \mu} (\mu\ \tilde{\sigma}_1^2\ \tilde{\sigma}_2^2) = \frac{n_1}{\tilde{\sigma}_1^2} (\bar{y}_1-\mu) + \frac{n_2}{\tilde{\sigma}_2^2} (\bar{y}_2-\mu)$$

to the expectation corresponding to (12) which is

$$E\left[(\hat{\sigma}_1^2 - \sigma_1^2) \frac{\partial^2 \log L}{\partial\mu\partial\sigma_1^2} + (\hat{\sigma}_2^2 - \sigma_2^2)\frac{\partial^2 \log L}{\partial\mu\partial_2^2}\right] + \frac{1}{2} E(\hat{\sigma}_1^2 - \sigma_1^2)^2\, E\left[\frac{\partial^3 \log L}{\partial\mu\partial(\sigma_1^2)^2}\right]$$
$$+ E\left[(\hat{\sigma}_1^2 - \sigma_1^2)(\hat{\sigma}_2^2 - \sigma_2^2)\right] E\left[\frac{\partial^3 \log L}{\partial\mu\partial\sigma_1^2\partial\sigma_2^2}\right] + \frac{1}{2} E(\hat{\sigma}_2^2 - \sigma_2^2)^2\, E\left[\frac{\partial^3 \log L}{\partial\mu\partial(\sigma_2^2)^2}\right] \tag{27}$$

Using (25), (26) and the derivatives of the log likelihood, (27) turns out to be zero. So for initial estimates $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ the resulting initial estimate of μ is

$$\tilde{\mu} = \frac{1}{n_1\tilde{\sigma}_2^2+n_2\tilde{\sigma}_1^2} (n_1\tilde{\sigma}_2^2\, \bar{y}_1 + n_2\tilde{\sigma}_1^2\, \bar{y}_2)$$

The 'natural' initial estimators to take for $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ would seem to be

$$\frac{1}{n_1-1} \Sigma(y_{1i} - \bar{y}_1)^2 \quad \text{and} \quad \frac{1}{n_2-1} \Sigma(y_{2i} - \bar{y}_2)^2$$

Strictly, of course, only the ratio of $\tilde{\sigma}_1^2$ to $\tilde{\sigma}_2^2$ is needed in (28).

The EMLE equation for σ_1^2 is obtained by equating

$$\frac{\partial \log L}{\partial \sigma_1^2}(\tilde{\mu}\ \sigma_1^2\ \tilde{\sigma}_2^2) = \frac{n_1}{2\sigma_1^2} + \frac{1}{2\sigma_1^4}\Sigma(y_{1i} - \tilde{\mu})^2$$

to

$$E\left[(\hat{\mu} - \mu)\frac{\partial^2 \log L}{\partial\mu\partial\sigma_1^2}\right] + \frac{1}{2}E(\hat{\mu} - \mu)^2\ E\left[\frac{\partial^3 \log L}{\partial\mu^2\partial(\sigma_1^2)^2}\right]$$

which, using (24) and the derivatives of the log likelihood, turns out to be

$$-\tfrac{1}{2}\frac{n_1\sigma_2^2}{\sigma_1^2(n_1\sigma_2^2 + n_2\sigma_1^2)} \tag{29}$$

This contains σ_2^2 which has to be replaced by $\tilde{\sigma}_2^2$ showing that σ_1^2 is estimated by equating

$$-n_1\sigma_1^2 + \Sigma(y_{1i} - \tilde{\mu})^2 \quad \text{to} \quad -\sigma_1^2\left(1 - \frac{1}{\dfrac{n_1}{n_2}\dfrac{\tilde{\sigma}_2^2}{\sigma_1^2} + 1}\right) \tag{30}$$

This equation is intuitively plausible. The second sample can only provide information about the variance of the first distribution via improving the estimate of the mean. If n_2 is zero, or if the variability in the second sample is huge, the right hand side of (30) becomes $-\sigma_1^2$ and $\tilde{\mu}$ in (28) becomes $\bar{y}_1$, so that the new estimate of σ_1^2 is still $\tilde{\sigma}_1^2$. But if n_2 is extremely large relative to n_1, or $\tilde{\sigma}_2^2$ negligible, then the right hand side of (30) becomes zero and the new estimate of σ_1^2 is

$$\frac{1}{n_1}\Sigma(y_{1i} - \mu)^2$$

Described a little loosely, it could be said the second sample has determined μ exactly, so that there is no loss of degrees of freedom in estimating σ_1^2 in the first sample.

Formulae similar to (29) and (30) hold for deriving a new estimate of σ_2^2 and then the estimate of μ can be updated via (28). The whole procedure is computationally somewhat tedious, but even ML treatment of this example would have required interative computations.

The REML - restricted or residual maximum likelihood - procedure is difficult to compare with EMLE in a general way, because REML seems to be only defined for certain types of likelihood functions occurring in the interblock analysis of experimental designs (Patterson and Thompson, 1971) and in certain ARMA processes (Cooper and Thompson, 1977). If the log likelihood of y is

$$\log L = \text{constant} - \tfrac{1}{2}\log|H| - \tfrac{1}{2}(y - x\alpha)' H^{-1}(y - x\alpha)$$

where x is a matrix of known constants, α is a vector of parameters to be estimated and the elements of H are functions of another vector σ^2 of parameters, then REML estimates α by maximising L'', the likelihood or joint distribution of

$$(x'H^{-1}x)^{-1}x'H^{-1}y \qquad (31)$$

and then estimates σ^2 by maximising a 'residual likelihood' L' defined as

$$\log L' = \log L - \log L''$$

Now it is reasonably obvious that if the derivatives of L'' w.r.t. α are analagous to (17) so that the resulting $\tilde{\alpha}$ are functions only of y, and if L' is not a function of α, then the original likelihood must have been of the form (20) with α for ϕ and σ^2 for ψ. So REML would then just be EMLE, or ML based on the conditional distribution in the case of σ^2 and on the marginal distribution of $\tilde{\alpha}$ in the case of α. However, L'', considered as a function of σ^2, will usually be such that the derivatives equated to zero give equations for α that are functions of σ^2. REML surmounts this difficulty by replacing these by initial estimates. Similarly $\sigma^2_.$ estimated from L' will usually be a function of μ. Yet applying REML to the example of this section suggests that the procedure can still give EMLE estimates, at least in some situations.

For the log likelihood (23), x is a vector of units and

$$H = \begin{bmatrix} \sigma_1^2 I_{n_1} & 0 \\ 0 & \sigma_2^2 I_{n_2} \end{bmatrix} \qquad H^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} I_{n_1} & 0 \\ 0 & \frac{1}{\sigma_2^2} I_{n_2} \end{bmatrix}$$

where I_n denotes the nxn identity matrix. The expression (31) is

$$\frac{n_1\sigma_2^2\,\bar{y}_1 + n_2\sigma_1^2\,\bar{y}_2}{n_1\sigma_2^2 + n_2\sigma_1^2}$$

with log likelihood

$$\text{constant} -\frac{1}{2}\log\left(\frac{\sigma_1^2\sigma_2^2}{n_1\sigma_2^2+n_2\sigma_1^2}\right) - \frac{1}{2}\,\frac{n_1\sigma_2^2+n_2\sigma_1^2}{\sigma_1^2\sigma_2^2}\left[\frac{n_1\sigma_2^2\bar{y}_1 + n_2\sigma_1^2\bar{y}_2}{n_1\sigma_2^2 + n_2\sigma_1^2} - \mu\right]^2 = L'' \qquad (32)$$

$\text{Log}L'$ is (23) minus (32). L'' is a function of σ_1^2 and σ_2^2, as well as of μ, and equating the derivatives to zero gives

$$\tilde{\mu} = \frac{1}{n_1\sigma_2^2+n_2\sigma_1^2}\,(n_1\sigma_2^2\,\bar{y}_1 + n_2\sigma_1^2\,\bar{y}_2) \qquad (33)$$

To actually obtain an estimate $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ need to be inserted and then (33) becomes (28). Note that if (33) replaced μ in (32), the second term of that expression and its derivatives with respect to σ_1^2 or σ_2^2 become zero. Thus

L' at $\mu = \tilde{\mu}$ is

$$\text{const} - \frac{n_1}{2}\log\sigma_1^2 - \frac{1}{2\sigma_1^2}\Sigma(y_{1i}-\tilde{\mu})^2 - \frac{n_2}{2}\log\sigma_2^2 - \frac{1}{2\sigma_2^2}\Sigma(y_{2i}-\tilde{\mu})^2 + \frac{1}{2}\log\left(\frac{\sigma_1^2\sigma_2^2}{n_1\sigma_2^2+n_2\sigma_1^2}\right)$$

Differentiating w.r.t. σ_1^2 and simplifying gives

$$-n_1\sigma_1^2 + \Sigma(y_{12}-\tilde{\mu})^2 = -\sigma_1^2\left(1 - \frac{1}{\frac{n_1}{n_2}\frac{\sigma_2^2}{\sigma_1^2}+1}\right) \qquad (34)$$

Now replacing σ_2^2 by $\tilde{\sigma}^2$ gives (30). So at least in this example, EMLE and REML coincide. This suggests there may be general computational methods for implementing EMLE other than the procedure described in Section 2 and used in the first treatment of this example.

References

CONNIFFE, D. (1987) Expected maximum log likelihood estimation, The Statistician, 36, pp. 317-329.

COOPER, D.M. and THOMPSON, R. (1977) A note on the estimation of the parameters of the autoregressive-moving average process, Biometrika, 64, pp. 625-628.

COX, D.R. and HINKLEY, D.V. (1974) Theoretical Statistics, London, Chapman and Hall.

PATTERSON, H.D. and THOMPSON, R. (1971) Recovery of inter-block information when block sizes are unequal, Biometrika, 58, pp. 545-554.