Department of Statistical Science

Southern Methodist University

Technical report SMU-TR-330

Jan 24th 2005

# Restricted maximum likelihood estimators as Bayes estimators in a mixed linear model with two variance components.

Ian R. Harris

*Department of Statistical Science, Southern Methodist University*

*Dallas, TX 75275*

iharris@smu.edu

Brent D. Burch

*Department of Mathematics and Statistics, Northern Arizona University*

*Flagstaff, AZ 86011*

Brent.Burch@nau.edu

## Summary

It is well known (Searle, Casella, and McCulloch, 1992, pages 321-325) that restricted maximum likelihood estimators of variance components in a two component mixed linear model are equivalent to Bayes estimators using a flat or improper prior for the fixed effects. This is due to a certain algebraic identity involving location families. A similar algebraic identity involving scale families is introduced, and used to show that the restricted maximum likelihood estimator of the intraclass correlation coefficient is also a Bayes estimator using an improper prior for one of the variance components.

*Some key words*: Bayesian inference, improper priors, intraclass correlation coefficient, marginal likelihood.

## 1. Introduction

This note demonstrates a connection between restricted maximum likelihood estimation of the intraclass correlation, $\rho$, in a mixed linear model and "objective" Bayesian inference using improper priors for the nuisance parameters. A novel identity for scale families is introduced and applied in demonstrating the equivalence of the seemingly different methods of deriving estimators. A one-dimensional posterior distribution for $\rho$, or equivalently a marginal likelihood function for $\rho$, is derived. There are practical implications to the result which are briefly described at the end of the paper.

## 2. Restricted likelihood and Bayesian inference

For brevity and simplicity, we will illustrate the connection using the one-way random effects model. Extension to the general mixed linear model is conceptually simple and involves just a little more algebraic complexity in dealing with the fixed effects. Consider the one-way random effects model given by

$$Y_{ij} = \mu + A_i + e_{ij}, \tag{1}$$

where $i = 1, ..., a$, $j = 1, ..., b_i$, and $\sum_{i=1}^{a} b_i = n$. $Y_{ij}$ is the $j^{th}$ observation associated with the $i^{th}$ class or group of factor $A$. The $a$ groups of $A$ in the model are assumed to be randomly selected from some large population of groups. Furthermore, a random sample of size $b_i$ has been obtained from the $i^{th}$ group. The random error is $e_{ij}$. It is assumed that the $A_i$ are a random sample from a $N(0, \sigma_1^2)$ distribution, the $e_{ij}$ are a random sample from a $N(0, \sigma_2^2)$ distribution, and that $A_i$ and $e_{ij}$ are mutually independent. In addition, $\sigma_1^2 \geq 0$ and $\sigma_2^2 > 0$. The overall mean of $Y_{ij}$, $\mu$ is a fixed but unknown quantity.

Often the main objects of the inference are the variance components, with $\mu$ representing a nuisance parameter. Under this scenario, a preferred mode of inference is to use restricted likelihood, see for example Harville (1974). There are several ways to obtain the restricted likelihood function. One way is to transform the data to a set of minimal sufficient statistics $(\bar{Y}_{..}, Q_1, ..., Q_d)$, where $\bar{Y}_{..}$ is the overall mean and the $Q_i$ are quadratic forms that are a set of minimal sufficient statistics for the model void of the fixed effect. The $Q_i$ are independent $\sigma_2^2(1 + \Delta_i \rho/(1-\rho))\chi^2_{r_i}$ random variables where $\Delta_i$ and $r_i$ are related to the design of the experiment. See Burch and Harris

(2001) for more details on the derivation and form of the $Q_i$ and the values of $\Delta_i$ and $r_i$. The $Q_i$ are independent of $\bar{Y}_{..}$, so integrating with respect to $\bar{Y}_{..}$ produces a marginal likelihood which is a function of $\sigma_1, \sigma_2$ alone. See Searle et al. (1992) for details. The marginal likelihood, or joint density of the $Q_i$ is then just a product of scaled $\chi^2$ densities.

As shown by Searle et al. (1992), one can also obtain the restricted likelihood function from the full likelihood by integrating out the unwanted fixed effect parameters using flat or improper priors. That this is equivalent to the marginal likelihood derivation is due to the algebraic identity,

$$\int f(m-\mu)dm = \int f(m-\mu)d\mu, \tag{2}$$

with $m = \bar{Y}_{..}$. So the restricted likelihood can be obtained from the full likelihood function by integration with respect to the overall sample mean or the unknown population mean.

Often the parameter of interest is the intraclass correlation, defined as $\rho = \sigma_1^2/(\sigma_1^2+\sigma_2^2)$. The purpose of this note is to demonstrate that there is a further connection between Bayesian inference and restricted likelihood in finding the restricted maximum likelihood estimator for $\rho$. To see the connection we must first derive a marginal likelihood function for $\rho$ whose maximum is the restricted maximum likelihood estimator for $\rho$.

To find the marginal likelihood function, transform the vector of $Q$'s to $Q, Y_2, ..., Y_d$, where $Q = Q_1$ and $Y_i = Q_i/Q_1, i = 2, ..., d$. By simple Jacobian transformation, the joint density function of $Q, Y_2, ..., Y_d$ is

$$f_{Q,\underline{Y}}(q,\underline{y}) = \prod_{i=1}^{d} 2^{-r_i/2}\Gamma(r_i/2)^{-1} \prod_{i=2}^{d} y_i^{r_i/2-1} \prod_{i=1}^{d} \left(1+\frac{\rho}{1-\rho}\Delta_i\right)^{-r_i/2} \sigma_2^{-\sum_{i=1}^d r_i}$$
$$\times q^{\sum_{i=1}^d r_i/2-1} \exp\left\{-\frac{q}{2\sigma_2^2}\left(\frac{1-\rho}{1+\rho(\Delta_1-1)} + \sum_{i=2}^{d} \frac{y_i(1-\rho)}{1+\rho(\Delta_i-1)}\right)\right\} \quad (3)$$

Maximization of $f_{Q,\underline{Y}}$ with respect to $\rho$ and $\sigma_2^2$ will give the restricted maximum likelihood estimators of these parameters. Now factorize this joint density function into $f_{\underline{Y}} f_{Q|\underline{Y}}$, where the marginal density function of $\underline{Y}$ is given by

$$f_{\underline{Y}} = \frac{\Gamma(\sum_{i=1}^d r_i/2)}{\prod_{i=1}^d \Gamma(r_i/2)} \prod_{i=2}^{d} y_i^{r_i/2-1} \frac{\prod_{i=1}^d (1+\rho(\Delta_i-1))^{-r_i/2}}{(1+\rho(\Delta_1-1))^{-\sum_{i=1}^d r_i/2}}$$
$$\times \left(1+(1+\rho(\Delta_1-1))\sum_{i=2}^{d} \frac{y_i}{1+\rho(\Delta_i-1)}\right)^{-\sum_{i=1}^d r_i/2}, \quad (4)$$

and the conditional density of $Q$ given $\underline{Y}$ is given by

$$f_{Q|\underline{Y}} \propto \left(\frac{h(\rho)}{\sigma_2^2}\right)^{\sum_{i=1}^d r_i/2} \exp\left(-\frac{qh(\rho)}{2\sigma_2^2}\right).$$

Here $q$ is the observed value of $Q_1$ and $h(\rho)$ is a somewhat complicated function of $\rho$, not involving $q$ or $\sigma_2^2$.

Note that $f_{\underline{Y}}$ is a function of $\rho$ alone as $\underline{Y}$ are a set of pivots for $\rho$. If $f_{Q|\underline{Y}}$ was a function of $\sigma_2^2$ alone then finding the restricted maximum likelihood estimator of $\rho$ would be easy, we would just maximize $f_{\underline{Y}}$ with respect to $\rho$. This of course is equivalent to saying $\underline{Y}$ is sufficient for $\rho$. However, $f_{Q|\underline{Y}}$ is also a function of $\rho$, so this term cannot just be dropped. Curiously, this term does not play a role in the maximization, as one can see using a profile approach to the maximization. First maximize with respect to $\sigma_2^2$, with $\rho$ fixed to obtain $\hat{\sigma_2^2}(\rho) = Bh(\rho)/A$ for some constants $A, B$. Then substituting this back into $f_{Q|\underline{Y}}$ we obtain an expression devoid of $\rho$. The end result is that one can find the restricted maximum likelihood estimator of $\rho$ by directly maximizing $f_{\underline{Y}}$ with respect to $\rho$.

The above is not a Bayesian argument. To make the connection we argue as follows. The marginal density $f_{\underline{Y}}$ is obtained by integrating out $Q$ from the joint density $f_{Q,\underline{Y}}$. Now $f_{Q,\underline{Y}}$ is of the scale density form $\tau^{-1}f(q/\tau)$ with respect to $q$ and $\tau = \sigma_2^2$. There exists a simple, but apparently unknown algebraic identity for such scale densities which is

$$\int_0^\infty \tau^{-1} f(q/\tau)\tau^{-1} d\tau = q^{-1} \int_0^\infty \tau^{-1} f(q/\tau) dq. \tag{5}$$

This follows by simple substitution of $t = q/\tau$ in both integrals. Although (5) is a straightforward result, the Authors are not aware of it being reported in the literature. Arguably, (5) is as important as (2) in that like (2) it relates integration with respect to the parameter to integration with respect to the random variable.

Now in terms of deriving a likelihood function for $\rho$ alone, the $q^{-1}$ term on the right hand side of (5) plays no role, so from a likelihood approach, integrating with respect to $q$ is equivalent to integrating with respect to $\sigma_2^2$ with an improper prior $\sigma_2^{-2}$. But integrating with respect to $q$ is precisely what is needed to obtain $f_{\underline{Y}}$. Hence the restricted maximum likelihood estimator of $\rho$ is also a Bayesian posterior mode using an improper prior for the nuisance parameter $\sigma_2^2$.

## 3. Implications and uses of the result

There are two interesting theoretical features of the connection here. The first is the demonstration of a novel identity for scale families, and an illustration of it's use. It remains to be seen whether there are any more applications of the identity. The second interesting result is a further connection between restricted maximum likelihood estimation, marginal likelihoods and "objective" Bayesian inference.

In addition to the theoretical results there are some practical uses for the marginal likelihood for $\rho$. This is a one-dimensional function on a bounded parameter space, so simple and reliable bisection search algorithms can be used to find the restricted maximum likelihood estimator for $\rho$. A second practical use is in finding Bayesian means via numerical integration. Again, this is a one-dimensional function, so the integration is relatively simple and should be more reliable and faster than algorithms used for two-dimensional posterior distributions. This could be a particular advantage in large data sets.

The posterior mean for $\rho$ will generally need to be found using numerical methods, except for very specific cases. Although these posterior means may make good estimators from a frequentist viewpoint, the study of their properties is well beyond the scope of this paper.

## References

Burch, B. D. and Harris, I. R. (2001), Closed-form approximations to the REML estimator of a variance ratio (or heritability) in a mixed linear model, *Biometrics* **57**, 1148–1156.

HARVILLE, D. A. (1974), Bayesian inference for variance components using only error contrasts, *Biometrika* **61**, 383–385.

SEARLE, S. R., CASELLA, G., AND MCCULLOCH, C. E. (1992), *Variance Components*, New York: Wiley.