ARTICLE IN PRESS LAA:12540

Linear Algebra and its Applications ••• (••••) •••–•••

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Restricted likelihood representation for multivariate heterogeneous random effects models

Andrew L. Rukhin

Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-0001, USA

ARTICLE INFO

Article history:
Received 25 September 2013
Accepted 8 January 2014
Available online xxxx
Submitted by M. Drton

MSC:
primary 62E15
secondary 62J05, 62P99, 62C15

Keywords:
Confidence regions
Inadmissibility
Maximum likelihood
Meta-analysis
Polynomial roots
Stein effect
Variance components

ABSTRACT

In the random effects model of meta-analysis for heterogeneous multidimensional data a canonical representation of the restricted likelihood function is obtained. This representation is related to a linear data transform which is based on the algebraic characteristics of error covariance matrices which are supposed to commute. The relationship between the heterogeneity covariance matrix estimators and the mean effect estimators is explored. It is noted that the sample mean exhibits the Stein-type phenomenon being an inadmissible estimator of the effect size under the quadratic loss when the number of studies exceeds three.

Published by Elsevier Inc.

1. Introduction: meta-analysis model

One of the important applications of random effects models is meta-analysis where one has to combine information in multivariate measurements made in several studies

E-mail address: andrew.rukhin@nist.gov.

0024-3795/$ – see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.laa.2014.01.008

which commonly exhibit not only non-negligible between-study variability, but also have different within-study precision.

Consider a model where several independent sources provide the estimates of q-dimensional parameter θ (representing the treatment effect or the common mean). Let the i-th study vector estimate of θ be X_i, $i = 1, \ldots, n$. In the random effects model of meta-analysis

$$X_i = \theta + \ell_i + \epsilon_i, \tag{1}$$

where the independent vectors ℓ_i represent random *between-study* effects with zero mean and some unknown $q \times q$ covariance matrix Ξ (which may have rank smaller than q). If the errors ϵ_i are assumed to be independent and normally distributed, $\epsilon_i \sim N(0, S_i)$, then $X_i \sim N(\theta, \Xi + S_i)$.

This model appears under scenario where each study measures its linear functions of θ, i.e., when the i-th study data vector consists of n_i measurements,

$$Y_i = B_i[\theta + \ell_i] + \epsilon_i. \tag{2}$$

Here B_i is the known i-th laboratory design matrix having the rank q and the size $n_i \times q$. The meaning of θ, ℓ_i and ϵ_i remains the same as in (1), and statistics $X_i = (B_i^T B_i)^{-1} B_i^T Y_i$ (the classical least squares estimators) satisfy this model. Unlike the general mixed effects model, the condition in (2) is that the random between-study effect with probability one belongs to the space spanned by columns of B_i. See [9] for further motivation of (2) and for some examples.

In many applications, e.g. [4,5,8], estimates of the full covariance matrices S_i are not available but estimators V_i of the variances are given. In view of the lack of appropriate data, simplifying assumptions are to be made. For example, one may impose the condition that $S_i = V_i^{1/2} R V_i^{1/2}$ for some given correlation matrix R and a diagonal matrix V_i. Then the results obtained for several correlation matrices R can be compared (see Section 5). The assumption made in this work is that all given matrices S_i as well as unknown Ξ commute.

In the setting with known covariance matrices S_i's, the parameters to be estimated are the matrix Ξ and θ itself. If Ξ is known, the best linear unbiased estimator of θ is

$$\tilde{X} = \left[\sum_i (S_i + \Xi)^{-1}\right]^{-1} \sum_i (S_i + \Xi)^{-1} X_i, \tag{3}$$

so the traditional methods seek to estimate Ξ using a plug-in estimator of θ afterwards.

We discuss some of these traditional estimators in Section 3 where a wider class of θ-estimators is suggested. This class is motivated by the form of Bayes procedures and by the representation of the restricted likelihood function derived in Section 2. This canonical representation makes use of the polynomials determined by the matrices S_i, $i = 1, \ldots, n$.

The main statistical goal is not only to estimate the parameter θ but to give a confidence region for this parameter or for a function thereof. These objectives are also touched upon in Section 3. It is noticed in Section 4 that when the number of studies exceeds three the sample mean exhibits the Stein-type effect being an inadmissible estimator under the quadratic loss. A practical example in Section 5 and the discussion Section 6 conclude the paper.

2. Sufficient statistics and restricted likelihood function

The model (1) leads to the (negative) log-likelihood function having the form

$$\frac{1}{2}\left[\sum_1^n (X_i-\theta)^T(\Xi+S_i)^{-1}(X_i-\theta)+\sum_1^n \log|\Xi+S_i|\right].$$

Then with $\tilde{X}$ defined by (3) the (negative) restricted log-likelihood function has the form

$$\mathcal{L}=\frac{1}{2}\left[\sum_1^n (X_i-\tilde{X})^T(\Xi+S_i)^{-1}(X_i-\tilde{X}) + \sum_1^n \log|\Xi+S_i| + \log\left|\sum_1^n (\Xi+S_i)^{-1}\right|\right],$$

[12, Sec. 6.6]. The classical maximum likelihood estimators as well as the restricted likelihood estimators of θ can be found as roots of polynomial equations and thus are amenable to the methodology of algebraic statistics [1]. However the degrees of these polynomials grow fast, the corresponding Groebner basis gets very complicated, which makes efficient determination of the likelihood estimators via solving polynomial equations quite difficult even for moderate n [3,11].

We pursue a different approach to estimate θ which is also based on polynomial algebra. Our main condition is that all matrices $S_1,\dots,S_n$ and Ξ *commute*. Of course this assumption is quite restrictive but it holds automatically when $q=1$ or when all matrices are diagonal. After an orthogonal transform, one can assume that $S_i=\mathrm{diag}(s_{i1},\dots,s_{iq})$, $\Xi=\mathrm{diag}(\xi_1,\dots,\xi_q)$. We write $X_i=(x_{i1},\dots,x_{iq})^T$, $i=1,\dots,n$, for the vector which corresponds to the same basis as the diagonal matrices S_i and Ξ, i.e., for the orthogonal transform of the original X_i.

For k, $1\leqslant k\leqslant q$, denote by ν_{ik}, $\nu_{ik}\geqslant 1$, the multiplicity of s_{ik} in the series $s_{1k},\dots,s_{nk}$, so that $\sum_i \nu_{ik}=n$. Put p_k, $1\leqslant p_k\leqslant n$, to be the number of distinct s_{ik}, and let $\hat{x}_{ik}=\sum_{m:\, s_{mk}=s_{ik}} x_{mk}/\nu_{ik}$ represent the average of ν_{ik} $x_{\cdot k}$'s corresponding to the particular s_{ik}, $i=1,\dots,p_k$. Denote by v_{ik} their sample variance, $v_{ik}=\sum_{m:\, s_{mk}=s_{ik}}(x_{mk}-\hat{x}_{ik})^2/(\nu_{ik}-1)$, when $\nu_{ik}>1$.

In this notation the k-th coordinate of $\tilde{X}$ is

$$\tilde{x}_k = \frac{\sum_i \frac{\nu_{ik}\hat{x}_{ik}}{\xi_k+s_{ik}}}{\sum_i \frac{\nu_{ik}}{\xi_k+s_{ik}}},$$

with the index i varying from 1 to p_k. Then $Var(\tilde{x}_k) = [\sum_i \nu_{ik}(\xi_k + s_{ik})^{-1}]^{-1}$.

The following representation of $\mathcal{L}$ obtains

$$\mathcal{L} = \frac{1}{2}\Bigg[\sum_k\sum_i \frac{\nu_{ik}(\hat{x}_{ik} - \tilde{x}_k)^2}{\xi_k + s_{ik}} + \sum_k\sum_i \nu_{ik}\log(\xi_k + s_{ik}) + \sum_k \log\Bigg(\sum_i \frac{\nu_{ik}}{\xi_k + s_{ik}}\Bigg) + \sum_k\sum_i \frac{(\nu_{ik}-1)v_{ik}}{\xi_k + s_{ik}}\Bigg].$$

To keep the number of subscripts to the minimum we write e for a vector with unit coordinates whose dimension is clear from context with the same convention for the identity matrix I. The p_k-dimensional vector $\hat{X}_k$, $k = 1,\dots,q$, with coordinates $\hat{x}_{ik}$, $i = 1,\dots,p_k$, has the diagonal covariance matrix C_k whose non-zero elements are $(\xi_k + s_{ik})/\nu_{ik}$, $i = 1,\dots,p_k$, and $\tilde{x}_k = (e^T C_k^{-1}\hat{X}_k)/(e^T C_k^{-1}e)$ (with p_k-dimensional e). Clearly $e^T\hat{X}_k = \sum_i x_{ik}$ is the k-th coordinate of the q-dimensional vector $\sum_i X_i$.

Then $\mathcal{L}$ can be written in the following form:

$$\mathcal{L} = \frac{1}{2}\Bigg[\sum_k(\hat{X}_k - \tilde{x}_k e)^T C_k^{-1}(\hat{X}_k - \tilde{x}_k e) + \sum_k \log\Bigg(\sum_i \frac{\nu_{ik}}{\xi_k + s_{ik}}\Bigg) + \sum_k\sum_i \nu_{ik}\log(\xi_k + s_{ik}) + \sum_k\sum_i \frac{(\nu_{ik}-1)v_{ik}}{\xi_k + s_{ik}}\Bigg].$$

Theorem 2.1. *In the model (1) assume that the matrices $S_1,\dots,S_n$ and Ξ are diagonal. Under the notation above let $k = 1,\dots,q$, $\hat{X}_k \sim N_{p_k}(\theta, C_k)$, and*

$$Y_k = \big(A_k^T N_k^{-1} A_k\big)^{-1/2} A_k^T \hat{X}_k$$

with $p_k \times (p_k - 1)$ matrix A_k determined by its elements in (9), $N_k = \operatorname{diag}(\nu_{1k},\dots,\nu_{p_k k})$, and $T_k = \operatorname{diag}(t_{1k},\dots,t_{p_k-1k})$. Here $-t_{ik}$ denote the roots of the polynomial Q_k defined in (5). Then $Y_k \sim N_{p_{k-1}}(0, T_k + \xi_k I)$, and the restricted log-likelihood function $\mathcal{L}$ for model (1) admits the representation

$$\mathcal{L} = \frac{1}{2}\Bigg\{\sum_k\big[Y_k^T(T_k + \xi_k I)^{-1}Y_k + \log|T_k + \xi_k I|\big] + \sum_k\sum_i(\nu_{ik}-1)\Bigg[\frac{v_{ik}}{\xi_k + s_{ik}} + \log(\xi_k + s_{ik})\Bigg]\Bigg\}. \tag{4}$$

The $(p_k - 1) \times (p_k - 1)$ matrices $A_k^T N_k^{-1} A_k$ and $A_k^T C_k A_k$ are diagonal, $A_k^T N_k^{-1} A_k = \operatorname{diag}(b_1^{(k)},\dots,b_{p_k-1}^{(k)})$, $A_k^T C_k A_k = \operatorname{diag}(b_1^{(k)}(\xi_k + t_{1k}),\dots,b_{p_k-1}^{(k)}(\xi_k + t_{p_k-1k}))$ with $b_j^{(k)}$ given in (10).

Proof. For fixed k each polynomial $P_k(v) = \prod_i (v + s_{ik})^{\nu_{ik}}$ has the degree $n = \sum_i \nu_{ik}$. Let $M_k(v) = \prod_i (v + s_{ik})$ be the minimal annihilating polynomial (of degree p_k). Define

$$Q_k(v) = M_k(v)\frac{P_k'(v)}{P_k(v)} = \sum_i \nu_{ik} \prod_{\ell:\ell\neq i} (v + s_{\ell k}). \tag{5}$$

Thus Q_k is a degree $p_k - 1$ polynomial which has only real (negative) roots, say, $-t_{1k}, \dots, -t_{p_k-1k}$ (coinciding with the roots of $P_k'(v)$ different from $-s_{1k}, \dots, -s_{p_k k}$). Thus $t_{1k}, \dots, t_{p_k-1k}$ interlace $s_{1k}, \dots, s_{p_k k}$, $P_k'(-t_{jk}) = 0$, and $Q_k(v) = n\prod_{j=1}^{p_k-1}(v+t_{jk})$. If $\nu_{1k} = \dots = \nu_{p_k k} = 1$, $M_k(v) = P_k(v)$, and $Q_k(v) = P_k'(v)$.

It follows from (5) that for all positive v,

$$\sum_i \log(v + s_{ik}) + \log\left(\sum_i \frac{\nu_{ik}}{v + s_{ik}}\right) = \sum_j \log(v + t_{jk}) + \log n, \tag{6}$$

where the index j varies from 1 to $p_k - 1$.

The comparison of the coefficients at v^{p_k-2} in (5) gives

$$\sum_j t_{jk} = \sum_i s_{ik} - \frac{\sum_i \nu_{ik} s_{ik}}{n}. \tag{7}$$

We show now that for any i, k and v different from $-t_{jk}$, $j = 1, \dots, p_k - 1$,

$$\frac{\nu_{ik}}{v + s_{ik}}\left[\sum_\ell \frac{\nu_{\ell k}}{v + s_{\ell k}}\right]^{-1} = \frac{\nu_{ik}}{n} - \sum_j \frac{A_{ij}^{(k)}}{v + t_{jk}}, \tag{8}$$

where for $i = 1, \dots, p_k$,

$$A_{ij}^{(k)} = \frac{\nu_{ik} M_k(-t_{jk})}{Q_k'(-t_{jk})(t_{jk} - s_{ik})}. \tag{9}$$

Indeed by the definition of the polynomial Q_k,

$$\begin{aligned}
&\frac{\nu_{ik}}{n} - \frac{\nu_{ik}}{v + s_{ik}}\left[\sum_\ell \frac{\nu_{\ell k}}{v + s_{\ell k}}\right]^{-1} \\
&\quad= \frac{\nu_{ik}}{n} - \frac{\nu_{ik} P_k(v)}{(v + s_{ik})P_k'(v)} \\
&\quad= \frac{\nu_{ik}}{n} - \frac{\nu_{ik} M_k(v)}{(v + s_{ik})Q_k(v)} = \frac{\nu_{ik}[\prod_j (v + t_{jk}) - \prod_{\ell\neq i}(v + s_{\ell k})]}{Q_k(v)},
\end{aligned}$$

with the right-hand side of this identity being the ratio of two polynomials of degrees $p_k - 2$ and $p_k - 1$ respectively. The formulas (8) and (9) follow now from the classical results on partial fraction decomposition for such ratios.

Please cite this article in press as: A.L. Rukhin, Restricted likelihood representation for multivariate heterogeneous random effects models, Linear Algebra Appl. (2014), http://dx.doi.org/10.1016/j.laa.2014.01.008

For any fixed k and j

$$\sum_i \frac{\nu_{ik}}{s_{ik}-t_{jk}} = \frac{P_k'(-t_{jk})}{P_k(-t_{jk})} = 0,$$

so that

$$\sum_i A_{ij}^{(k)} = \frac{M_k(-t_{jk})}{Q_k'(-t_{jk})} \sum_i \frac{\nu_{ik}}{s_{ik}-t_{jk}} = 0.$$

This means that for the $p_k \times (p_k-1)$ matrix A_k is determined by its elements $A_{ij}^{(k)}$ in (9), $A_k^T e = 0$. The formula for the coordinates of $A_k^T e$,

$$\sum_j A_{ij}^{(k)} = \frac{\nu_{ik}}{n}\left(s_{ik} - \frac{\sum_\ell \nu_{\ell k} s_{\ell k}}{n}\right),$$

follows directly from (8).

Since

$$0 = \sum_i \frac{\nu_{ik}}{s_{ik}-t_{jk}} - \sum_i \frac{\nu_{ik}}{s_{ik}-t_{\ell k}} = (t_{jk}-t_{\ell k}) \sum_i \frac{\nu_{ik}}{(s_{ik}-t_{jk})(s_{ik}-t_{\ell k})},$$

when $j \neq \ell$,

$$\sum_i \frac{A_{ij}^{(k)} A_{i\ell}^{(k)}}{\nu_{ik}} = \frac{M_k(-t_{jk})M_k(-t_{\ell k})}{Q_k'(-t_{jk})Q_k'(-t_{\ell k})} \sum_i \frac{\nu_{ik}}{(s_{ik}-t_{jk})(s_{ik}-t_{\ell k})} = 0.$$

For any $j = 1, \dots, p_k - 1$,

$$\frac{d}{dv}\left(\frac{P_k'}{P_k}\right)(-t_{jk}) = \frac{Q_k'(-t_{jk})}{M_k(-t_{jk})} = -\sum_i \frac{\nu_{ik}}{(s_{ik}-t_{jk})^2},$$

and by using (9) one obtains that

$$b_j^{(k)} = \sum_i \frac{(A_{ij}^{(k)})^2}{\nu_{ik}} = \sum_i \frac{\nu_{ik} M_k^2(-t_{jk})}{[Q_k'(-t_{jk})]^2 (t_{jk}-s_{ik})^2} = -\frac{M_k(-t_{jk})}{Q_k'(-t_{jk})}. \tag{10}$$

Thus $A_k^T N_k^{-1} A_k = \mathrm{diag}(b_1^{(k)}, \dots, b_{p_k-1}^{(k)})$.

Using (7), we can see that the polynomial $(v + \sum_i \nu_{ik} s_{ik}/n) Q_k(v)/n - M(v)$ has degree $p_k - 2$. The values of this polynomial at $v = -t_{jk}$ coincide with $-M(-t_{jk})$. Therefore,

$$\begin{aligned}\sum_j \frac{b_j^{(k)}}{\xi_k + t_{jk}} &= -\sum_j \frac{M(-t_{jk})}{Q(-t_{jk})(\xi_k + t_{jk})} = \frac{1}{n}\left(\xi_k + \frac{\sum_i \nu_{ik} s_{ik}}{n}\right) - \frac{M(\xi_k)}{Q(\xi_k)} \\ &= Var(\bar{x}_k) - Var(\tilde{x}_k).\end{aligned} \tag{11}$$

In particular,

$$\sum_j b_j^{(k)} = \lim_{v\to\infty} v\left[\frac{1}{n}\left(v + \frac{\sum_i \nu_{ik}s_{ik}}{n}\right) - \frac{M(v)}{Q(v)}\right]$$
$$= \frac{\sum_i \nu_{ik}(s_{ik} - \sum_\ell \nu_{\ell k}s_{\ell k}/n)^2}{n^2}.$$

If

$$U_k = C_k - \xi_k N_k^{-1} = \operatorname{diag}\left(\frac{s_{1k}}{\nu_{1k}}, \dots, \frac{s_{p_k k}}{\nu_{p_k k}}\right),$$

then the matrix $A_k^T U_k A_k = \operatorname{diag}(b_1^{(k)} t_{1k}, \dots, b_{p_k-1}^{(k)} t_{p_k-1k})$ is also diagonal. Indeed for $j \neq m$,

$$\sum_i \frac{\nu_{ik}s_{ik}}{(s_{ik}-t_{jk})(s_{ik}-t_{mk})} = t_{jk}\sum_i \frac{\nu_{ik}}{(s_{ik}-t_{jk})(s_{ik}-t_{mk})} = 0,$$

implying that

$$\sum_i \frac{s_{ik}A_{ij}^{(k)}A_{im}^{(k)}}{\nu_{ik}} = 0.$$

For $j = m$,

$$\frac{1}{t_{jk}}\sum_i \frac{s_{ik}(A_{ij}^{(k)})^2}{\nu_{ik}} = -\sum_i \frac{\nu_{ik}s_{ik}}{(s_{ik}-t_{jk})^2} = \sum_i \frac{\nu_{ik}}{(s_{ik}-t_{jk})^2}$$
$$= \frac{M_k(-t_{jk})}{Q_k'(-t_{jk})} = b_j^{(k)}. \qquad (12)$$

Since the vector $\hat{X}_k$ has the diagonal $p_k \times p_k$ covariance matrix, $C_k = \xi_k N_k^{-1} + U_k$, the normal random vector $Y_k = (A_k^T N_k^{-1} A_k)^{-1/2} A_k^T \hat{X}_k$, has the $(p_k - 1)\times(p_k - 1)$ covariance matrix

$$\left(A_k^T N_k^{-1} A_k\right)^{-1/2} A_k^T C_k A_k \left(A_k^T N_k^{-1} A_k\right)^{-1/2} = \operatorname{diag}(\xi_k + t_{1k}, \dots, \xi_k + t_{p_k-1k})$$
$$= \xi_k I + T_k.$$

We prove next that

$$A_k\left(A_k^T N_k^{-1} A_k\right)^{-1/2}(\xi_k I + T_k)^{-1}\left(A_k^T N_k^{-1} A_k\right)^{-1/2} A_k^T$$
$$= C_k^{-1} - \frac{C_k^{-1} e e^T C_k^{-1}}{e^T C_k^{-1} e}. \qquad (13)$$

The (i,ℓ)-th element of the matrix in the right-hand side of (13) is of the form

$$\sum_j \frac{A_{ij}^{(k)} A_{\ell j}^{(k)}}{(\xi_k + t_{jk}) b_j^{(k)}} = -\nu_{ik} \sum_j \frac{A_{\ell j}^{(k)}}{(\xi_k + t_{jk})(t_{jk} - s_{ik})}$$
$$= \frac{\nu_{ik}}{\xi_k + s_{ik}} \left[\sum_j \frac{A_{\ell j}^{(k)}}{t_{jk} - s_{ik}} - \sum_j \frac{A_{\ell j}^{(k)}}{\xi_k + t_{jk}} \right].$$

The formula (8) implies that

$$\sum_j \frac{A_{\ell j}^{(k)}}{\xi_k + t_{jk}} = \frac{\nu_{\ell k}}{n} - \frac{\nu_{\ell k}}{(\xi_k + s_{\ell k}) e^T C_k^{-1} e},$$

and

$$\sum_j \frac{A_{\ell j}^{(k)}}{t_{jk} - s_{ik}} = \frac{\nu_{\ell k}}{n} - \delta_{i\ell},$$

where $\delta_{i\ell}$ is the Kronecker symbol ($\delta_{i\ell} = 1$, if $i = \ell$; $\delta_{i\ell} = 0$ otherwise). Therefore, (13) is true.

To complete the proof of (4) we need to show that

$$(\hat{X}_k - \tilde{x}_k e)^T C_k^{-1} (\hat{X}_k - \tilde{x}_k e) = \hat{X}_k C_k^{-1} \hat{X}_k - \left(\tilde{x}^T C_k^{-1} e\right)^2 / \left(e^T C_k^{-1} e\right)$$
$$= Y_k^T (T_k + \xi_k I)^{-1} Y_k. \tag{14}$$

It suffices to prove that

$$C_k^{-1} - \frac{C_k^{-1} e e^T C_k^{-1}}{e^T C_k^{-1} e} = A_k \left(A_k^T N_k^{-1} A_k\right)^{-1/2} (T_k + \xi_k I)^{-1} \left(A_k^T N_k^{-1} A_k\right)^{-1/2} A_k^T,$$

which follows from (13). Combination of (6) and (14) completes the proof of Theorem 2.1. □

The numerical check-up on formulas (9) and (10) can be performed by verifying the equalities

$$\prod_{ij} A_{ij}^{(k)} = \frac{[R(M_k, Q_k) \prod \nu_{ik}]^{p_k - 1}}{[R(Q_k, Q_k')]^{p_k}},$$

and

$$\prod_j b_j^{(k)} = (-1)^{p_k - 1} \frac{R(M_k, Q_k)}{n^2 R(Q_k, Q_k')}.$$

Here $R(M_k, Q_k)$ denotes the resultant of the polynomials M_k and Q_k. Another useful identity is

$$\sum_i A_{ij}^{(k)} s_{ik} = n b_j^{(k)}.$$

Thus, the representation (4) of the restricted likelihood function corresponds to q independent zero mean, normal $(p_k - 1)$-dimensional random vectors Y_k with covariance matrices $\xi_k I + T_k$. In addition it includes independent v_{ik}, each being a scaled χ^2-random variable with $\nu_{ik} - 1$ degrees of freedom. When $\nu_{ik} > 1$, v_{ik} is an unbiased estimator of $\xi_k + s_{ik}$, $v_{ik} \sim (\xi_k + s_{ik})\chi^2_{\nu_{ik}-1}/(\nu_{ik} - 1)$. For $\nu_{ik} = 1$, with probability one, $v_{ik} = 0$.

According to the sufficiency principle, all statistical inference about Ξ involving the restricted likelihood can be based exclusively on Y_k and $\{v_{ik}\}$. Their joint distribution forms a curved exponential family whose natural parameter is formed by $(\xi_k + t_{jk})^{-1}$ (and possibly by some $(\xi_k + s_{ik})^{-1}$). The total dimension of sufficient statistics y_{jk} and v_{ik}, $j = 1, \dots, p_k - 1$, $k = 1, \dots, q$, is $\sum(p_k - 1) + \sum(\nu_{ik} - 1) = (n - 1)q$, so that the q degrees of freedom used for estimating θ are accounted for.

3. Traditional estimators and Bayes procedures

According to (8), the coordinates of $\tilde{x}$ in (3) admit the following representation,

$$\begin{aligned}\tilde{x}_k &= \frac{e^T \hat{X}_k}{n} - \sum_{i,j} \frac{A_{ij}^{(k)} \hat{x}_{ik}}{\xi_k + t_{jk}} = \bar{x}_k - \sum_j \frac{\sqrt{b_j^{(k)}}\, y_{jk}}{\xi_k + t_{jk}} \\ &= \bar{x}_k - e^T (T_k + \xi_k I)^{-1} \left(A_k^T N_k^{-1} A_k\right)^{1/2} Y_k, \quad k = 1, \dots, q.\end{aligned}$$

Here $y_{jk} = \sum_i A_{ij}^{(k)} \hat{x}_{ik} / \sqrt{b_{jk}}$, the coordinates of Y_k, are independent normal, zero mean random variables with the variances $\xi_k + t_{jk}$, $\bar{x}_k = \sum_i \nu_{ik} x_{ik}/n$.

Let Y denote the vector stacked by $Y_1, \dots, Y_q$ (of dimension $p = p_1 + \dots + p_q - q$), and let $\bar{X} = n^{-1} \sum_i X_i$ represent the sample mean, i.e., the vector with coordinates $\bar{x}_k$. Then

$$\tilde{X} = \bar{X} - M(\Xi) Y, \tag{15}$$

where $M(\Xi)$ is $q \times p$ block-diagonal matrix formed by the blocks $e^T (A_k^T N_k^{-1} A_k)^{1/2} (T_k + \xi_k I)^{-1}$ of the sizes $1 \times (p_k - 1)$. It follows that any traditional estimator of θ, say,

$$\delta = \left[\sum_i (S_i + \tilde{\Xi})^{-1}\right]^{-1} \sum_i (S_i + \tilde{\Xi})^{-1} X_i,$$

which uses an estimator $\tilde{\Xi}$ has the form (15) with $M = M(\tilde{\Xi})$.

However the Bayes estimators of θ as a rule do not admit such a representation. Indeed assume a prior distribution Λ for Ξ which has more than two points of support while θ has the uniform (non-informative) prior. Mainly for notational convenience let $(\delta-\theta)^T(\delta-\theta)$ be a quadratic loss for θ estimation. Under this loss with $\mathcal{L}$ defined by (4) the (generalized) Bayes estimator,

$$\delta = \frac{\int_0^\infty \cdots \int_0^\infty \tilde{X} \exp\{-\mathcal{L}\}\, d\Lambda(\Xi)}{\int_0^\infty \cdots \int_0^\infty \exp\{-\mathcal{L}\}\, d\Lambda(\Xi)},$$

can be written as

$$\delta = \bar{X} - \mathbf{M}Y. \tag{16}$$

According to (15), the block-diagonal matrix $\mathbf{M}$ has the same structure as $M(\Xi)$ with blocks of the form $e^T(A_k^T N_k^{-1} A_k)^{1/2}\Omega_k$, $k = 1, \dots, q$, and the diagonal matrix

$$\begin{aligned}\Omega_k &= \Omega_k\big(Y_1, \dots, Y_q, \{v_{ik}\}\big)\\ &= \frac{\int_0^\infty \cdots \int_0^\infty (T_k + \xi_k I)^{-1} \exp\{-\mathcal{L}\}\, d\Lambda(\xi_1, \dots, \xi_q)}{\int_0^\infty \cdots \int_0^\infty \exp\{-\mathcal{L}\}\, d\Lambda(\xi_1, \dots, \xi_q)},\end{aligned}$$

can be thought of as an estimator $(T_k + \xi_k I)^{-1}$.

Any δ of the form (16) is an unbiased estimator of θ whose covariance matrix, $Var(\delta)$, does not depend on θ. Since $\tilde{X}$ and $\delta - \tilde{X}$ are independent as the random vectors $\tilde{X}$ and Y_k, $k = 1, \dots, q$, are independent,

$$Var(\delta) = Var(\tilde{X}) + E(\delta - \tilde{X})(\delta - \tilde{X})^T = \left[\sum_i (S_i + \Xi)^{-1}\right]^{-1} + R, \tag{17}$$

where the (k, ℓ)-th element of the matrix R has the form,

$$\begin{aligned}&e^T\big(A_k^T N_k^{-1} A_k\big)^{1/2} E\big[\big[\Omega_k - (T_k + \xi_k I)^{-1}\big] Y_k Y_\ell^T \big[\Omega_\ell - (T_\ell + \xi_\ell I)^{-1}\big]\big]\\ &\quad \times \big(A_\ell^T N_\ell^{-1} A_\ell\big)^{1/2} e.\end{aligned}$$

This matrix shows how well δ approximates the optimal but unavailable $\tilde{X}$, so that it can be used to define a risk function for θ estimators. Provided that $\Omega_k = \Omega_k(\{y_{jk}^2\}, \{v_{ik}\})$, the matrix R is diagonal with the non-zero elements of the form,

$$\sum_j b_j^{(k)} E y_{jk}^2 \left(\omega_{jk} - \frac{1}{\xi_k + t_{jk}}\right)^2, \quad k = 1, \dots, q,$$

which is being used in Theorem 4.1 in the next section.

Before that we briefly review the alternative estimating procedures for Ξ. The transformed expression (4) of the likelihood function $\mathcal{L}$ motivates the moment-type equations based on a general quadratic form, $\sum_j q_{jk} y_{jk}^2 + \sum_i (\nu_{ik} - 1) r_{ik} v_{ik}$ with positive coefficients q_{jk}, r_{ik}. The moment-type equation for ξ_k written in terms of random variables y_{jk} and v_{ik} is

$$E\left[\sum_j q_{jk} y_{jk} + \sum_i (\nu_{ik} - 1) r_{ik} v_{ik}\right]$$

$$= \left[\sum_j q_{jk} + \sum_i (\nu_{ik} - 1) r_{ik}\right] \xi_k + \sum_j q_{jk} t_{jk} + \sum_i (\nu_{ik} - 1) r_{ik} s_{ik}.$$

Its estimator by the method of moments has the form

$$\tilde{\xi}_k = \frac{\sum_j q_j (y_{jk}^2 - t_{jk}) + \sum_i (\nu_{ik} - 1) r_{ik} (v_{ik} - s_{ik})}{\sum_j q_{jk} + \sum_i (\nu_{ik} - 1) r_{ik}}.$$

Commonly the probability that such an estimator takes negative values is non-negligible. Non-negative statistics, $\tilde{\xi}_{k+} = \max(\tilde{\xi}_k, 0)$, are used to get traditional θ estimators.

The representations of the popular DerSimonian–Laird statistic,

$$\tilde{\xi}_k^{DL} = \frac{\sum_j t_{jk}^{-1} y_{jk}^2 + \sum_i (\nu_{ik} - 1) s_{ik}^{-1} v_{ik} - n + 1}{\sum_j t_{jk}^{-1} + \sum_i (\nu_{ik} - 1) s_{ik}^{-1}},$$

and of the Hedges statistic,

$$\tilde{\xi}_k^{H} = \frac{\sum_j (y_{jk}^2 - t_{jk}) + \sum_i (\nu_{ik} - 1)(v_{ik} - s_{ik})}{n - 1},$$

follow easily.

The expression of these statistics in terms of y_{jk} and v_{ik} facilitate numerical implementation of these procedures as well as of the restricted maximum likelihood estimator (REML) $\hat{\xi}$. Indeed (4) shows that this estimator can be determined by simple iterations as

$$\hat{\xi}_k = \frac{\sum_j \frac{y_{jk}^2 - t_{jk}}{(\hat{\xi}_k + t_{jk})^2} + \sum_i \frac{(\nu_{ik} - 1)(v_{ik} - s_{ik})}{(\hat{\xi}_k + s_{ik})^2}}{\sum_j \frac{1}{(\hat{\xi}_k + t_{jk})^2} + \sum_i \frac{\nu_{ik} - 1}{(\hat{\xi}_k + s_{ik})^2}},$$

with truncation at zero if the iteration process converges to a negative number.

A multivariate extension of the DerSimonian–Laird statistic in the non-commutative case was suggested in [9, Sec 3.1] and [10, Sec 3] where results of numerical comparisons are reported. See also [5] and references there for the use of method of moments multivariate estimators in random effects model in biostatistics. However the latter references deal with the estimators which are not invariant under linear transforms or are based on non-symmetric matrices leading to ambiguity about the choice among a matrix and its transpose.

For any fixed Ξ, according to (14), $\sum(\hat{X}_k - \tilde{x}_k e)^T C_k^{-1}(\hat{X}_k - \tilde{x}_k e) = \sum_k Y_k^T(T_k + \xi_k I)^{-1} Y_k$ has χ_p^2-distribution, while $\tilde{X}$ is a normal vector with mean θ and the covariance matrix $[\sum_i (S_i + \Xi)^{-1}]^{-1}$. Therefore the ratio

$$\frac{p(\tilde{X} - \theta)^T[\sum_i (S_i + \Xi)^{-1}](\tilde{X} - \theta)}{q\sum(\hat{X}_k - \tilde{x}_k e)^T C_k^{-1}(\hat{X}_k - \tilde{x}_k e)}$$

has $F(q, p)$ distribution.

Similar facts are commonly used to get confidence ellipsoids for the unknown vector parameter. In our case when Ξ is replaced by one of its estimates $\tilde{\Xi}$ as above, or by using $\sum_k Y_k^T \Omega_k Y_k$, where Ω_k define the estimator (16), instead of the sum $\sum_k(\hat{X}_k - \tilde{x}_k e)^T C_k^{-1}(\hat{X}_k - \tilde{x}_k e)$. As Ξ has q unknown parameters, it makes sense to diminish the degrees of freedom by this number leading to the approximate $(1-\alpha)$-confidence ellipsoid,

$$(\delta - \theta)^T \left[\sum_i (S_i + \tilde{\Xi})^{-1}\right](\delta - \theta) \leqslant \frac{pF_\alpha(q, p-q)}{q\sum_k(\hat{X}_k - \tilde{x}_k e)^T C_k^{-1}(\hat{X}_k - \tilde{x}_k e)}, \tag{18}$$

cf. [6, Sec 7.3.4] where this degree of freedom is not adjusted and $q = 1$.

4. Estimation of multivariate normal mean

The decomposition (17) relates our setting to the classical estimation problem of the multivariate normal mean.

Theorem 4.1. *If the coefficients $\omega_{jk} = \omega_{jk}(\{y_{jk}^2\}, \{v_{ik}\})$, the elements of defining the estimator (16) are piecewise differentiable for $k = 1, \ldots, q$, then*

$$\begin{aligned} Var(\delta_k) &= Var(\tilde{x}_k) + \sum_j b_j^{(k)} E y_{jk}^2 \left(\omega_{jk} - \frac{1}{\xi_k + t_{jk}}\right)^2 \\ &= Var(\bar{x}_k) + E\sum_j b_j^{(k)} \left(f_{jk}^2 - 2\frac{\partial}{\partial y_{jk}} f_{jk}\right), \end{aligned} \tag{19}$$

where $f_{jk} = y_{jk}\omega_{jk}$. When $p_k > 3$, the estimator $\bar{x}_k$ is an inadmissible estimator of θ_k under the quadratic loss, $(\delta_k - \theta_k)^2$. The vector estimator $\bar{X}$ of θ is inadmissible under the quadratic loss, $\sum_k(\delta_k - \theta_k)^2$, when $p = \sum_k p_k - q > 2$.

Proof. To establish (19) we use (11) and the familiar integration by parts formula,

$$\frac{1}{\xi_k + t_{jk}} E y_{jk}^2 w = E\big(w + 2y_{jk}^2 w_j'\big), \quad y_{jk} \sim N(0, \xi_k + t_{jk}),$$

which holds for any piecewise differentiable function $w = w(y)$ (cf. [7, Ch 1, Lemma 5.15]).

To prove inadmissibility of $\bar{X}_k$ for $p_k > 3$, we put

$$\omega_{jk} = A_k \left[b_j^{(k)} \sum_\ell \frac{y_{\ell k}^2}{b_\ell^{(k)}} \right]^{-1}.$$

A straightforward calculation shows that then

$$E \sum_j b_j^{(k)} \left(f_{jk}^2 - 2 \frac{\partial}{\partial y_{jk}} f_{jk} \right) = A_k \left[A_k - 2(p_k - 3) \right] E \left[\sum_j \frac{y_{jk}^2}{b_j^{(k)}} \right]^{-1}.$$

The choice $A_k = p_k - 3$, which makes the left-hand side negative, corresponds to the James–Stein estimator.

Similarly when $p > 2$, one can use

$$\omega_{jk} = \frac{A}{b_j^{(k)}} \left[\sum_{\ell\, k} \frac{y_{\ell k}^2}{b_\ell^{(k)}} \right]^{-1},$$

in which case

$$E \sum_{j\, k} b_j^{(k)} \left(f_{jk}^2 - 2 \frac{\partial}{\partial y_{jk}} f_{jk} \right) = A \left[A - 2(p-2) E \left(\sum_{j\, k} \frac{y_{jk}^2}{b_j^{(k)}} \right)^{-1} \right],$$

which establishes the last statement of Theorem 4.1. □

Thus the sample mean $\bar{X}$ cannot be recommended as an estimator of the common mean θ when $(n-1)q > 2$. Indeed $p_k \leqslant n$, so that $p = \sum_k p_k - q \leqslant (n-1)q$. As a matter of fact, the sample mean suffers from a more serious drawback than inadmissibility: it is not even minimax under the loss function normalized by $\sum_k [Var(\bar{x}_k) - Var(\tilde{x}_k)]$, and under this loss the risk of $\bar{X}$ is constant. This result (as well as the form of the Bayes estimators in Section 3) holds for any positive-definite matrix W with $(\delta - \theta)^T W (\delta - \theta)$ providing a quadratic loss for θ estimation.

5. Example: prognostic test study

Our example is a meta-analysis study that summarizes the existing evidence concerning mutant p53 tumor suppressor gene as a prognostic factor for patients with squamous cell carcinoma [13]. Estimates of log hazard ratios of mutant p53 to normal p53 both for disease-free survival (θ_1) and for overall survival (θ_2) were obtained in three observational studies, $X_1 = (-0.58, -0.18)^T$, $X_2 = (-1.02, -0.63)^T$, $X_3 = (-0.69, -0.64)^T$. The within-study correlations were not given, but the estimates of standard deviations are available: $\sqrt{s_{11}} = \sqrt{s_{12}} = 0.56$, $\sqrt{s_{21}} = 0.39$, $\sqrt{s_{22}} = 0.29$, $\sqrt{s_{31}} = \sqrt{s_{32}} = 0.40$.

Please cite this article in press as: A.L. Rukhin, Restricted likelihood representation for multivariate heterogeneous random effects models, Linear Algebra Appl. (2014), http://dx.doi.org/10.1016/j.laa.2014.01.008

Additional three studies provided only the estimates of overall survival log hazard ratio, θ_2, so that with $\star$ denoting the missing entries, $X_4 = (\star, 0.79)^T$, $X_5 = (\star, 0.21)^T$, $X_6 = (\star, 1.01)^T$; $\sqrt{s_{42}} = 0.24$, $\sqrt{s_{52}} = 0.66$, $\sqrt{s_{62}} = 0.48$. While this data can be used to get a better estimate of θ_2, one may hope for better inference about θ_1 as well. Indeed the independence assumption of disease-free survival and of overall survival is not viable.

For this reason we looked at the commuting matrices $D_i R D_i$, $i = 1, \ldots, 6$ where D_i's are scalar matrices, $D_i = \sqrt{s(i)}I$, with $s(1) = s_{11} = s_{12}$, $s(2) = (s_{21} + s_{22})/2$, $s(3) = s_{31} = s_{32}$, $s(4) = s_{42}$, $s(5) = s_{52}$, $s(6) = s_{62}$, and R is the 2×2 correlation matrix (i.e., R has the unit diagonal and the off-diagonal elements equal to ρ). Since R diagonalizes to $\operatorname{diag}(1-\rho, 1+\rho)$, after an orthogonal transform diagonal matrices $S_i = s(i)\operatorname{diag}(1-\rho, 1+\rho)$ are obtained. Following [4], we took $\rho = 0.7$ or $\rho = 0.95$.

In the first case when $\rho = 0.7$, ξ_2 can be estimated by using the calculated values, $t_{12} = 0.13, t_{22} = 0.23, t_{32} = 0.34, t_{42} = 0.48, t_{52} = 0.68$. These values indicate presence of the between-studies effect, as $\hat{\xi}_2 = 0.94$. For the ξ_1 the REML estimator using three data points gives $\hat{\xi}_1 = 0$. The matrix $\hat{\Xi}$ is singular with all elements being 0.47. The REML estimator of θ is $(-0.12, 0.15)^T$ (to be contrasted with $(-0.32, 0.09)$ in [4, Table VII]). An approximate confidence ellipsoid can be derived from (18).

When $\rho = 0.95$, results turned out to be quite similar, giving $\hat{\xi}_1 = 0$, $\hat{\xi}_2 = 0.95$. The REML estimator of θ is $(-0.11, 0.15)^T$.

In [2] a concept of quantifying the amount of information contained in correlated data is discussed. The defined characteristic relates the size of an independent sample and the amount of information in the actual sample. In principle this approach also can be used to replace the covariance matrix S_i by an available $\operatorname{diag}(S_i)$ after the sample size adjustment. However this approach did not lead to practical answers in the situation described in this section.

6. Conclusions

The linear transformation suggested here offers a new perspective on the classical by now restricted maximum likelihood estimation. Not only it exhibits remarkable algebraic properties demonstrated in Theorem 2.1, it also facilitates numerical implementation of the restricted maximum likelihood estimator and of the method of moments based procedures Although the mentioned properties hold in the special case when all covariance matrices commute, lack of data may justify this restriction at least in some applications.

One of our findings is that the vector sample mean cannot be recommended as an estimator of the treatment effect if the number of studies is not very small (at least four when $q = 1$, any number larger than one if $q \geqslant 2$). Indeed individual studies can borrow strength from the accompanying data to get better estimators via the estimated weights matrix $\mathbf{M}$ in (16). This "correction to $\bar{X}$" term takes into account intrinsic heterogeneity. The author advocates to consider the class of θ estimators (16) which directly estimate the diagonal matrices $(T_k + \xi_k I)^{-1}$, $k = 1, \ldots, q$, rather then the commonly used plug-in estimators (15) which cannot have a Bayes origin.

References

[1] M. Drton, B. Sturmfels, S. Sullivant, Lectures on Algebraic Statistics, Birkhäuser, Basel, 2009.
[2] C. Faes, C. Molenberghs, M. Aerts, G. Verbeke, M. Kenward, Effective sample size and an alternative small sample degrees of freedom method, Amer. Statist. 63 (2009) 389–399.
[3] E. Gross, M. Drton, S. Petrovic, Maximum likelihood degree of variance component models, Electron. J. Stat. 6 (2012) 993–1016.
[4] D. Jackson, R. Riley, I. White, Multivariate meta-analysis: potential and promise, Stat. Med. 30 (2011) 2481–2498.
[5] D. Jackson, I. White, R. Riley, A matrix based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression, Biom. J. 55 (2013) 231–245.
[6] J. Hartung, G. Knapp, B.K. Sinha, Statistical Meta-Analysis with Applications, Wiley, New York, 2008.
[7] E. Lehmann, G. Casella, Theory of Point Estimation, second ed., Springer, New York, 1998.
[8] I.-S. Nam, K. Mengerson, P. Garthwaite, Multivariate meta-analysis, Stat. Med. 22 (2003) 2309–2333.
[9] A.L. Rukhin, Estimating common vector mean in interlaboratory studies, J. Multivariate Anal. 98 (2007) 435–454.
[10] A.L. Rukhin, Estimating common parameters in heterogeneous mixed effects model, J. Statist. Plann. Inference 141 (2011) 3181–3192.
[11] A.L. Rukhin, Maximum likelihood and restricted maximum likelihood solutions in multiple-method studies, J. Res. Natl. Inst. Stand. Technol. 116 (2011) 539–556.
[12] S.R. Searle, G. Casella, C.E. McCulloch, Variance Components, Wiley, New York, 1992.
[13] S. Tandon, C. Tudur-Smith, R.D. Riley, M.T. Boyd, T.M. Jones, A systematic review of p53 as a prognostic factor of survival in squamous cell carcinoma of the four main anatomical subsites of the head and neck, Cancer Epidemiol. Biomark. Prev. 19 (2010) 574–587.

Please cite this article in press as: A.L. Rukhin, Restricted likelihood representation for multivariate heterogeneous random effects models, Linear Algebra Appl. (2014), http://dx.doi.org/10.1016/j.laa.2014.01.008