Maximum Likelihood and Restricted Likelihood Solutions in Multiple-Method Studies

Andrew L. Rukhin

National Institute of Standards and Technology, Gaithersburg, MD 20899-8980

andrew.rukhin@nist.gov

A formulation of the problem of combining data from several sources is discussed in terms of random effects models. The unknown measurement precision is assumed not to be the same for all methods. We investigate maximum likelihood solutions in this model. By representing the likelihood equations as simultaneous polynomial equations, the exact form of the Groebner basis for their stationary points is derived when there are two methods. A parametrization of these solutions which allows their comparison is suggested. A numerical method for solving likelihood equations is outlined, and an alternative to the maximum likelihood method, the restricted maximum likelihood, is studied. In the situation when methods variances are considered to be known an upper bound on the between-method variance is obtained. The relationship between likelihood equations and moment-type equations is also discussed.

1. Introduction: Meta Analysis and Interlaboratory Studies

This article is concerned with mathematical aspects of an ubiquitous problem in applied statistics: how to combine measurement data nominally on the same property by methods, instruments, studies, medical centers, clusters or laboratories of different precision. One of the first approaches to this problem was suggested by Cochran [1], who investigated maximum likelihood (ML) estimates for one-way, unbalanced, heteroscedastic random-effects model. Cochran returned to this problem throughout his career [2, 3], and Ref. [4] reviews this work. Reference [5] discusses applications in metrology and gives more references.

The problem of combining data from several sources is central to the broad field of meta-analysis. It is most difficult when the unknown measurement precision varies among methods whose summary results may not seem to conform to the same measured property. Reference [6] provides some practical suggestions for dealing with the problem.

In this paper we investigate the ML solutions of the random effects model which is formulated in Sec. 2. By representing the likelihood equations as simultaneous polynomial equations, the so-called Groebner basis for them is derived when there are two sources. A parametrization of such solutions is suggested in Sec. 2.1. The maxima of the likelihood function are compared for positive and zero between-labs variance. A numerical method for solving likelihood equations by reducing them to an optimization problem for a homogeneous objective function is given in Sec. 2.2. An alternative to the ML method, the restricted maxi-

mum likelihood is considered in Sec. 3. An explicit formula for the restricted likelihood estimator is discovered in Sec. 3.2 in the case of two methods. Section 4 deals with the situation when methods variances are considered to be known, and an upper bound on the between-method variance is obtained. The Sec. 5 discusses the relationship between likelihood equations and moment-type equations, and Sec. 6 gives some conclusions. All auxiliary material related to an optimization problem and to elementary symmetric polynomials is collected in the Appendix.

2. ML Method and Polynomial Equations

To model interlaboratory testing situation, denote by n_i the number of observations made in laboratory i, $i = 1, ..., p$, whose observations x_{ij} have the form

$$x_{ij} = \mu + b_i + \varepsilon_{ij} . \tag{1}$$

Here μ is the true property value, b_i represents the method (or laboratory) effect, which is assumed to be normal with mean 0 and unknown variance σ^2, and ε_{ij} represent independent normal, zero mean random errors with unknown variances τ_i^2.

For a fixed i, the i-th sample mean $x_i = \sum_j x_{ij}/n_i$ is normally distributed with the mean μ and the variance $\sigma^2 + \sigma_i^2$, where $\sigma_i^2 = \tau_i^2/n_i$. If the σ's were known, then the best estimator of μ would be the weighted average of x_i with weights proportional to $1/(\sigma^2 + \sigma_i^2)$. Since these variances are unknown, the weights have to be estimated. Traditionally to evaluate σ_i^2 one uses the classical unbiased statistic $s_i^2 = \sum_j (x_{ij}-x_i)^2/(n_i v_i)$, $v_i = n_i - 1$, which has the distribution $\sigma_i^2 \chi^2(v_i)/v_i$. Since statistics x_i, s_i^2, $i = 1, ..., p$, are sufficient, we use the likelihood function based on them.

The ML solution minimizes in μ, σ^2, σ_i^2, $i = 1, ..., p$, the negative logarithm of this function which is proportional to

$$L = \sum_i \frac{(x_i - \mu)^2}{\sigma^2 + \sigma_i^2} + \sum_i \frac{v_i s_i^2}{\sigma_i^2} + \\ \sum_i \log(\sigma^2 + \sigma_i^2) + \sum_i v_i \log \sigma_i^2 . \tag{2}$$

It follows from (2) that if $\hat{\sigma}_i^2$ is the ML estimator of σ_i^2, then $\hat{\sigma}_i^2 > 0$. However, it is quite possible that $\hat{\sigma}_i^2 = 0$. In order to find these estimates one can replace μ in (2) by

$$\bar{\mu} = \frac{\sum_i x_i (\sigma^2 + \sigma_i^2)^{-1}}{\sum_i (\sigma^2 + \sigma_i^2)^{-1}} , \tag{3}$$

which reduces the number of parameters from $p + 2$ to $p + 1$.

Our goal is to represent the set of all stationary points of the likelihood equations as solutions to simultaneous polynomial equations. To that end, note that

$$\sum_i w_i (x_i - \bar{\mu})^2 = \sum_i w_i x_i^2 - (\sum_i w_i x_i)^2 /(\sum_i w_i) \\ = \sum_{1 \le i < j \le p} w_i w_j (x_i - x_j)^2 /(\sum_i w_i) . \tag{4}$$

This formula, which easily follows from the Lagrange identity [7, Sec. 1.3], will be used with $w_i = (\sigma^2 + \sigma_i^2)^{-1}$.

We introduce a polynomial P of degree p in σ^2,

$$P = P(\sigma^2, \sigma_1^2, ..., \sigma_p^2) = \prod_i (\sigma^2 + \sigma_i^2) = \sum_{\ell=0}^p \mathbb{S}_{p-\ell} \sigma^{2\ell} \tag{5}$$

with $\mathbb{S}_\ell = \mathbb{S}_\ell(\sigma_1^2, ..., \sigma_p^2), \ell = 0, 1, ..., p$, denoting the ℓ-th elementary symmetric polynomial. Another polynomial of interest is $Q = Q(\sigma^2, \sigma_1^2, ..., \sigma_p^2)$ having degree $p - 2$ in σ^2,

$$Q(\sigma^2, \sigma_1^2, ..., \sigma_p^2) = \sum_{1 \le i < j \le p} (x_i - x_j)^2 \prod_{k \ne i, j} (\sigma^2 + \sigma_k^2) \\ = \sum_{\ell=0}^{p-2} q_{p-2-\ell} \sigma^{2\ell} = \sum_{1 \le i < j \le p} (x_i - x_j)^2 P_{ij} , \tag{6}$$

where $P_{ij} = \dfrac{\partial^2}{\partial \sigma_i^2 \partial \sigma_j^2} P(\sigma^2, \sigma_1^2, ..., \sigma_p^2)$ is a polynomial in σ^2 of degree $p - 2$ which does not depend on σ_i^2, σ_j^2, and $q_\ell = q_\ell(\sigma_1^2, ..., \sigma_p^2)$ is a multilinear form in $\sigma_1^2, ..., \sigma_p^2$.

Since

$$\frac{P'}{P} = \frac{\partial}{\partial \sigma^2} \log P(\sigma^2, \sigma_1^2, ..., \sigma_p^2) = \sum_k \frac{1}{\sigma^2 + \sigma_k^2}, \tag{7}$$

the identity (4) can be written as

$$\sum_k \frac{(x_k - \bar{\mu})^2}{\sigma^2 + \sigma_k^2} = \frac{Q/P}{P'/P} = \frac{Q}{P'} . \tag{8}$$

The negative log-likelihood function (2) in this notation is

$$L = \frac{Q}{P'} + \log P + \sum_i \frac{v_i s_i^2}{\sigma_i^2} + \sum_i v_i \log \sigma_i^2 . \qquad (9)$$

Let $P_i = \Pi_{k \neq i}(\sigma^2 + \sigma_k^2)$ be the partial derivative of P with regard to σ_i^2; denote by Q_i the same partial derivative of Q and by $P_i' = \sum_{j:j \neq i} P_{ij}$ the derivative of $P'(\sigma^2, \sigma^2, \ldots, \sigma_p^2)$. By differentiating (9), we see that the stationary points of (2) satisfy polynomial equations,

$$\sigma_i^4(\sigma^2 + \sigma_i^2)(Q_i P' - Q P_i') + \sigma_i^4 (P')^2$$
$$+ v_i(\sigma_i^2 - s_i^2)(\sigma^2 + \sigma_i^2) P' = 0 . \qquad (10)$$

Each of these polynomials has degree 4 in σ_i^2, $i = 1, \ldots, p$. When $\sigma^2 = 0$, $P = \Pi \sigma_i^2$, and the equations (10) simplify to

$$\sigma_i^2 P(Q_i P' - Q P_i') + \sigma_i^2 P_i (P')^2$$
$$+ v_i(\sigma_i^2 - s_i^2) P_i (P')^2 = 0 . \qquad (11)$$

These polynomials have degree 3 in σ_i^2. If $\sigma^2 > 0$, in addition to (10) one has

$$F = (P')^3 + (Q'P' - QP'')P = 0 , \qquad (12)$$

and F has degree $3p-3$ in σ^2.

In both cases the collection of all stationary points forms an affine variety whose structure can be studied via the *Groebner basis* of the ideal of polynomials (11) or (10) and (12) which vanish on this variety. The Groebner basis allows for successive elimination of variables leading to a description of the points in the variety, i.e., to the characterization of all (complex) solutions. There are powerful numerical algorithms for evaluation of such bases [8]. Many polynomial likelihood equations are reviewed in Ref. [9]. We determine the Groebner basis for equations (11) when $p = 2$ in the next section.

2.1 ML Method: $p = 2$

When $p = 2$, $Q = (x_1 - x_2)^2$. If $\sigma^2 = 0$, the polynomial equations (11) take the form

$$(x_1 - x_2)^2 \sigma_i^4 - (n_i \sigma_i^2 - v_i s_i^2)(\sigma_1^2 + \sigma_2^2)^2 = 0, \ i = 1, 2 . \quad (13)$$

The Groebner basis is useful for solving these equations as it allows elimination of one of the variables, say, σ_2^2. With $n = n_1 + n_2$, $f = n + n_2$, $u = \sigma_1^2/(x_1 - x_2)^2$, $v = \sigma_2^2/(x_1 - x_2)^2$, $z_1 = v_1 s_1^2/(x_1 - x_2)^2$ and $z_2 = v_2 s_2^2/(x_1 - x_2)^2$, under lexicographic order, $\sigma_1^2 > \sigma_2^2$, the Groebner basis for equations (13) written in the form

$$u^2 + (z_1 - n_1 u)(u + v)^2 = 0, \qquad (14)$$

$$v^2 + (z_2 - n_2 v)(u + v)^2 = 0, \qquad (15)$$

consists of two polynomials,

$$G_1(u, v) = c_0 u^2 + c_1 v^5 + c_2 v^4 + c_3 v^3 + c_4 v^2 , \qquad (16)$$

$$c_0 = n_1^2 z_2^3 [f(1 + z_1) + n_1 z_2] ,$$

$$c_1 = -n^2 n_2^3 f z_1 + n_1 n_2 (n_1 + 3n_2) n^3 z_2 - n^2 n_2^3 f ,$$

$$c_2 = -2 n_2^3 n f z_1^2 + 2 n_2 n (n_1^3 + 4 n_1^2 n_2 + 4 n_1 n_2^2 + 2 n_2^3) z_1 z_2$$
$$- n_1 n (n_1^2 n_2 + 4 n_1 n_2 + 6 n_2^2) z_2^2 + f n_1^2 n_2^2 z_1$$
$$- n(5 n_1^4 + 5 n_1^3 n_2 + 8 n_1^2 n_2^2 + 4 n_1^2 n_2^3 - 4 n_2^4) z_2$$
$$+ f n_2^2 (n_1^2 + 2 n_1 n_2 + 2 n_2^2) ,$$

$$c_3 = -f n_2^3 z_1^3 + n_2(n_1^3 + 4 n_1^2 n_2 + 5 n_1 n_2^2 + 4 n_2^3) z_1^2 z_2$$
$$- (2 n_1^4 + 8 n_1^3 n_2 + 12 n_1^2 n_2^2 + 7 n_1 n_2^3 + 2 n_2^4) z_1 z_2^2$$
$$+ 3 n_1 n_2 n^2 z_2^3 - 3 f n_2^3 z_1^2 + (3 n_1^3 n_2 + 10 n_1^2 n_2^2 + 6 n_1 n_2^3) z_1 z_2^2$$
$$+ n_2 n (5 n_1^2 + 7 n_1 n_2 - 2 n_2^2) z_2^2$$
$$+ 3 f n_2^3 z_1 + n_2 (2 n_1^3 + 6 n_1^2 n_2 + n_1 n_2^2 - 4 n_2^3) z_2 - f n_2^3 ,$$

$$c_4 = -n_1^2 z_1 z_2^2 [f(1 + z_1) + n_1 z_2] ,$$

and

$$G_2(u, v) = d_0 u v + d_1 v^5 + d_2 v^4 + d_3 v^3 + d_4 v^2 , \quad (17)$$

$$d_0 = 2 n_1 z_2^2 [f(1 + z_1) + n_1 z_2] ,$$

$$d_1 = -n_2 f^2 n^2 ,$$

$$d_2 = -2 n n_2 f^2 z_1 + n f (n_1^2 + 3 n_1 n_2 + 4 n_2^2) z_2$$
$$+ f^2 (n_1^2 + 2 n_1 n_2 + 2 n_2^2) ,$$

$$d_3 = -n_2 f^2 z_1^2 + 2 f(n_1^2 + 2 n_1 n_2 + 2 n_2^2) z_1 z_2$$
$$- n_2 (3 n_1^2 + 6 n_1 n_2 + 4 n_2^2) z_2^2 - 2 n_2 f^2 z_1 - 4 n n_2 f z_2 - f n_2^3 ,$$

$$d_4 = n_1 z_2 (1 + z_1 + z_2)[f(1 + z_1) + n_1 z_2] .$$

This fact can be derived from the definition of the Groebner basis and confirmed by existing computational algebra software.

It follows that for stationary points u, v, $G_1(u, v) = G_2(u, v) = 0$. All these points can be found by expressing u through v via (17),

$$u = -v(d_1 v^3 + d_2 v^2 + d_3 v + d_4)/d_0 , \qquad (18)$$

substituting this expression in (16), and solving then the resulting sextic equation for v,

$$c_0(d_1v^3 + d_2v^2 + d_3v + d_4)^2$$
$$+ d_0^2(c_1v^3 + c_2v^2 + c_3v + c_4) = 0. \tag{19}$$

Thus, there are $3 \times 6 = 18$ complex root pairs out of which positive roots u, v are to be chosen. Although in practice most all these roots are complex or negative and are not meaningful, sometimes the number of positive roots is fairly large. For example, $x_1 = -0.391$, $x_2 = 0.860$, $s_1^2 = 0.075$, $s_2^2 = 0.102$, $n_1 = n_2 = 3$, there are three positive solutions, ($u = 0.263, v = 0.053$), $u = 0.138$, $v = 0.106$), ($u = 0.035$, $v = 0.312$). In this particular case one has to compare the likelihood function evaluated at these solutions with the likelihood evaluated when $\hat{\sigma}^2 > 0$, in which case the stationary points are ($u = v = 0.065$, $z = 0.230$), ($u = 0.246$, $v = 0.055$, $z = 0.003$), ($u = 0.042$, $v = 0.234$, $z = 0.019$), ($u = 0.048$, $v = 0.065$, $z = 0.193$). (The likelihood is maximized at the last solution.)

If $\hat{\sigma}^2 = 0$, the estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ satisfy the equations,

$$\frac{(x_1 - x_2)^2}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{n_1}{\sigma_1^2} - \frac{v_1 s_1^2}{\sigma_1^4} = \frac{n_2}{\sigma_2^2} - \frac{v_2 s_2^2}{\sigma_2^4}, \tag{20}$$

which imply the inequalities $v_i s_i^2/n_i < \hat{\sigma}_i^2 < [(x_1 - x_2)^2 + v_i s_i^2]/n_i$. Evaluation of second derivatives shows that $2n\hat{\sigma}_1^2\hat{\sigma}_2^2(x_1 - x_2)^2 \leq n_1 n_2(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$.

These solutions are to be compared with the solutions when $\hat{\sigma}^2 > 0$, in which case with $y = \sigma^2/(x_1 - x_2)^2 > 0$,

$$(2y + u + v)^3 - 2(y + u)(y + v) = 0, \tag{21}$$

$$u^2(u - v) + 2(z_1 - v_1 u)(y + u)(y + v) = 0, \tag{22}$$

$$v^2(v - u) + 2(z_2 - v_2 v)(y + u)(y + v) = 0. \tag{23}$$

Notice that (21)-(23) imply that if $\hat{\sigma}^2 > 0$, then the inequalities $s_1^2 \geq \hat{\sigma}_1^2$, $s_2^2 \leq \hat{\sigma}_2^2$, and $\hat{\sigma}_1^2 \leq \hat{\sigma}_2^2$, are all equivalent. The fact excludes the possibility that $\hat{\sigma}_i^2 = s_i^2$, unless $s_1^2 = s_2^2$, in which case $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = s_1^2$, $\hat{\sigma}^2 = $

$[(x_1 - x_2)^2/4 - s_1^2]_+$, $x_+ = \max(0, x)$. In this and only in this situation (whose probability is zero), $\hat{\sigma}_i^2 = s_i^2$. The equation (21) implies that $2y + u + v \leq 1/2$, so that

$$\hat{\sigma}^2 \leq \frac{(x_1 - x_2)^2}{4}. \tag{24}$$

Unfortunately, the Groebner basis for equations (21)-(23) has a much more complicated structure: the form of the monomials entering into the basis polynomials depends on z_1, z_2, n_1 and n_2. To find solutions to (21)-(23) we start with conditions on u and v for which $y > 0$ in (21)-(23).

For fixed u and v the behavior of the derivative of (9) with respect to σ^2 is determined by that of the cubic polynomial (12) which now takes the form,

$$F(y) = (2y + u + v)^3 - 2(y + u)(y + v). \tag{25}$$

This derivative is positive if and only if $F(y) > 0$.

As is easy to check, the derivative of F has two roots: $-(u + v)/2$ and $y^* = 1/6 - (u + v)/2$. The polynomial F has no positive roots if and only if either $y^* \leq 0$, $F(0) \geq 0$ or if $y^* > 0$, $F(y^*) \geq 0$. We rewrite these conditions as follows: (21) has a positive root if and only if either

$$u + v \geq \frac{1}{3}, \quad (u + v)^3 < 2uv, \tag{26}$$

or

$$u + v < \frac{1}{3}, \quad |u - v| < \frac{1}{3\sqrt{3}}. \tag{27}$$

If (26) holds, there is a unique positive stationary point. When condition (27) is met, only the largest root of $F(y) = 0$ (i.e., the one exceeding y^*) can be the ML estimator $\hat{\sigma}^2$. Analysis of second derivatives shows that (y, u, v), $y > 0$, can provide the minimum only if $2y + u + v > \sqrt{3}|u - v|$ and $(y + u)(y + v)$ min $(v_1(y + u)/u, v_2(y + v)/v) \geq y|u - v|$.

Figure 1 shows the region where (21) has a positive root in the (u, v) plane. Its boundary is formed by two straight segments where $3\sqrt{3}|u - v| = 1$, $3(u + v) \leq 1$, and by a cubic curve when $u + v \geq 1/3$. The largest possible value of u or v such that $\hat{\sigma}^2 > 0$ is 8/27.
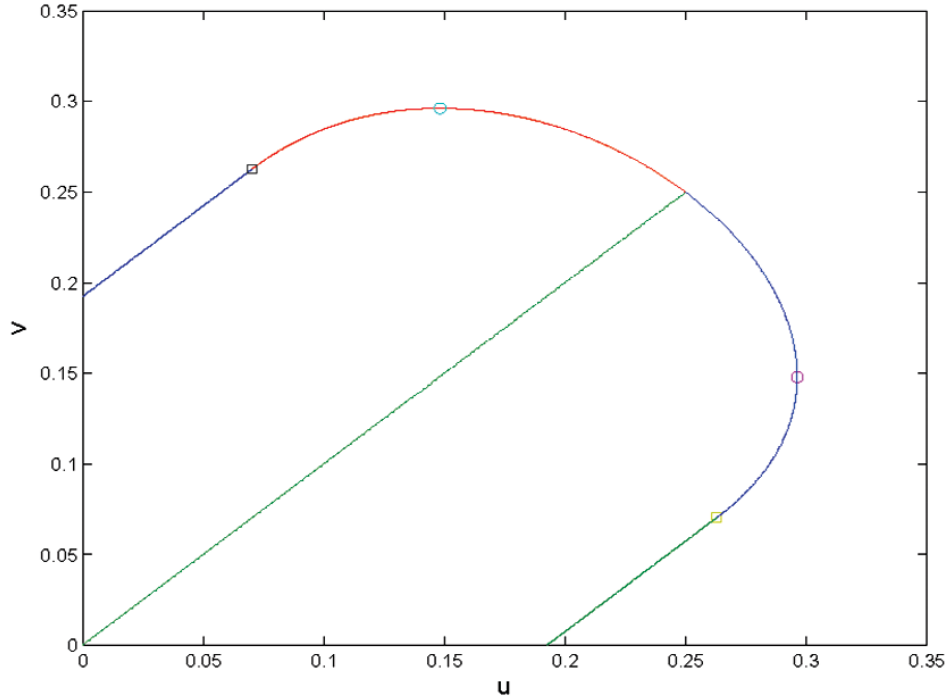
Fig. 1. The region where (21) has a positive root. The squares mark the points where the boundary changes from linear to cubic. Two points of this region at which u or v take the largest possible value (8/27) are marked by o.

To study further the form of solutions to (21)-(23), we use the fact that

$$\arg \min L \tag{28}$$
$$= \arg \min_{z_1/u + z_2/v + 1/(2y+u+v)=n} \left[v_1 \log u + v_2 \log v + \log(y+u) + \log(y+v) \right],$$

which can be shown by the direct calculation or by the argument from Sec. 2.2. Thus assuming the condition $z_1/u + z_2/v + 1/(2y+u+v) = n$, one can parametrize the solutions of Eqs. (21)-(23) as follows: for s in the unit interval, $u = (z_1 s + z_2 (1-s))/[s(n-1/K)]$, $v = (z_1 s + z_2(1-s))/[(1-s)(n-1/K)]$, $u+y = 2w(1-w)^2$, $v+y = 2w^2(1-w)$ with $K = 2y+u+v = 2w(1-w)>1/n$, where w, $0<w<1$, solves the cubic equation $u - v = 2(1-2w)w(1-w)$ or

$$(1-2w)^3 - \frac{(n-2)(1-2w)}{n} + \frac{2(1-2s)(z_1 s + z_2(1-s))}{ns(1-s)} = 0. \tag{29}$$

The conditions (26) and (27) reduce to the inequalities,

$$2s(1-s)\left(n - \frac{1}{K}\right) \geq \frac{z_1 s + z_2 (1-s)}{ns(1-s)}, \tag{30}$$

and

$$|1-2w| w(1-w) < \frac{1}{6\sqrt{3}}. \tag{31}$$

The discriminant of the cubic equation (29) reveals that when $s \neq 1/2, |1-2s|[z_1 s + z_2(1-s)] < (n-2)^{3/2}$ $s(1-s)/(3\sqrt{3n})$, and (29) has three real roots, two of them of the same sign as $1-2s$. If $s=1/2$, then we have $w=1/2$. Indeed the condition $y \geq 0$ leads to the following restriction on the range of $s: |1-2w| \leq |1-2s|$ with a strict inequality when $y > 0$.

The end points of this domain can be found as solutions to the equation,

$$s(1-s)(1-2s)^2 - \frac{(n-2)s(1-s)}{n} + \frac{2(z_1 s + z_2(1-s))}{n} = 0, \tag{32}$$

which is a quartic equation in $\xi = 1-2s$,

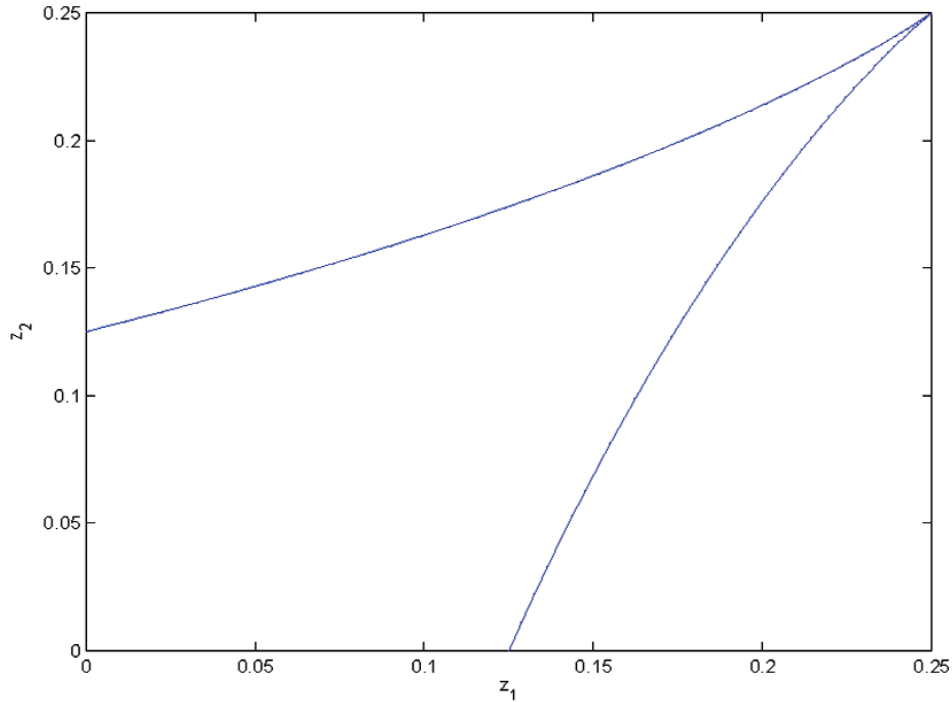$$\xi^4 - \frac{2(n-1)}{n}\xi^2 - \frac{8\Delta}{n}\xi - \frac{8\bar{z}-n+2}{n} = 0. \tag{33}$$

Fig. 2. The region where $\hat{\sigma}^2 > 0$ when $n_1 = 2$, $n_2 = 2$.

Here $\bar{z} = (z_1 + z_2)/2, \Delta = (z_2 - z_1)/2, |\Delta| < \bar{z}$.

If the equation (33) has two distinct roots $\underline{\xi}, \bar{\xi}$ in the interval $(-1, 1)$, then the range of s-values is the interval $[\underline{s}, \bar{s}]$, $0 < \underline{s} = (1 - \bar{\xi})/2 < \bar{s} = (1 - \underline{\xi})/2 < 1$, and for $s \in (\underline{s}, \bar{s})$ the polynomial in (32) is negative. Therefore, if there are two roots of (29) in the interval $(0, 1)$ which have the same sign as $1 - 2s$, the one with the smallest absolute value of $|1 - 2w|$, i.e., the one with

$|1 - 2w| < \sqrt{(n-2)/(3n)}$, is to be chosen. Thus, the solutions of equations (21)-(23) are parameterized by $s \in [\underline{s}, \bar{s}]$.

To determine conditions for the equation (33) to have two roots $\underline{\xi}, \bar{\xi}$, in the interval $(-1, 1)$, observe first that it cannot have more than two roots there. Indeed the polynomial in (33) assumes negative values at the end points, and it has exactly two roots if and only if there is a point in this region at which that polynomial is positive.

If the derivative of this polynomial, $4\xi^3 - 4(n-1)\xi/n - 8\Delta/n$, does not vanish in $(-1, 1)$, then (33) cannot have two roots there. This happens when this derivative has just one real root, or equivalently the discriminant of the cubic polynomial is negative which means that $|\Delta| > [(n-1)/(3n)]^{3/2}$ and which implies

that $|\Delta| > 1/2$. Then the derivative takes negative values at the end points of the interval $(-1, 1)$.

If ξ_0 is the point of maximum of the polynomial (33) in $(-1, 1)$, then $4\xi_0^3 - 4(n-1)\xi_0/n = 8\Delta/n$, and the roots $\underline{\xi}, \bar{\xi}$ exist if and only if

$$\xi_0^4 - \frac{2(n-1)}{n}\xi_0^2 - \frac{8\Delta}{n}\xi_0 - \frac{8\bar{z} - n + 2}{n} =$$

$$3\xi_0^4 + \frac{2(n-1)}{n}\xi_0^2 - \frac{8\bar{z} - n + 2}{n} > 0. \qquad (34)$$

According to inequality (34),

$$\left|\xi_0^2 - \frac{n-1}{3n}\right| \leq \sqrt{\frac{4n^2 - 8n + 1 - 24n\bar{z}}{9n^2}}, \qquad (35)$$

and (35) shows that $\bar{z} \leq (4n^2 - 8n + 1)/(24n) \leq (n-1)^{3/2}/(3\sqrt{3n})$. One must have $|\xi_0| \leq \sqrt{(n-1)/(3n)}$, as the derivative decreases only in this sub-interval of $(-1, 1)$. Thus

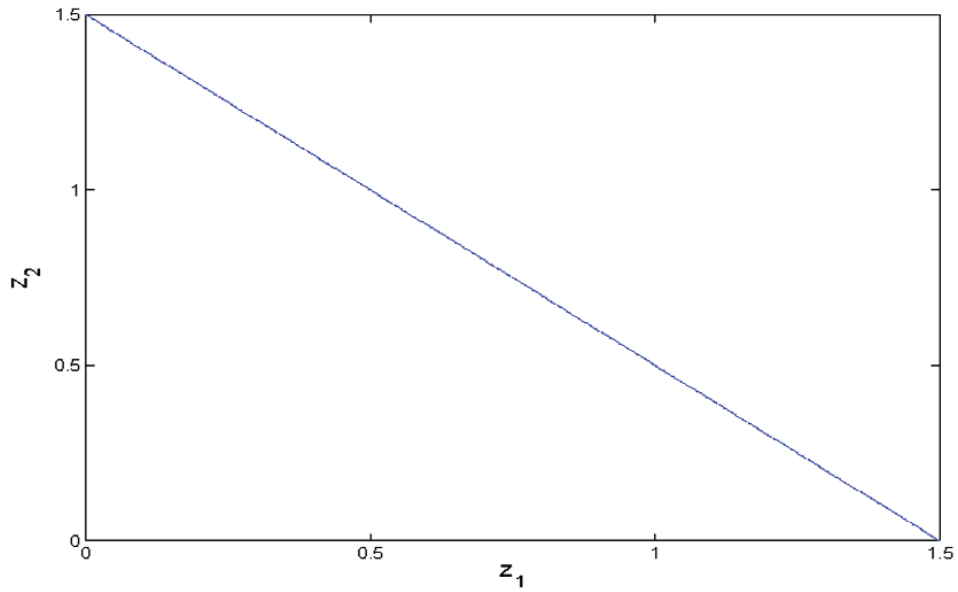$$\xi_0^2 \geq \frac{n-1}{3n}\left[1 - \sqrt{\frac{4n^2 - 8n + 1 - 24n\bar{z}}{(n-1)^2}}\right]. \qquad (36)$$

Fig. 3. The region where $\hat{\sigma}^2 > 0$ when $n_1 = 4$, $n_2 = 4$.
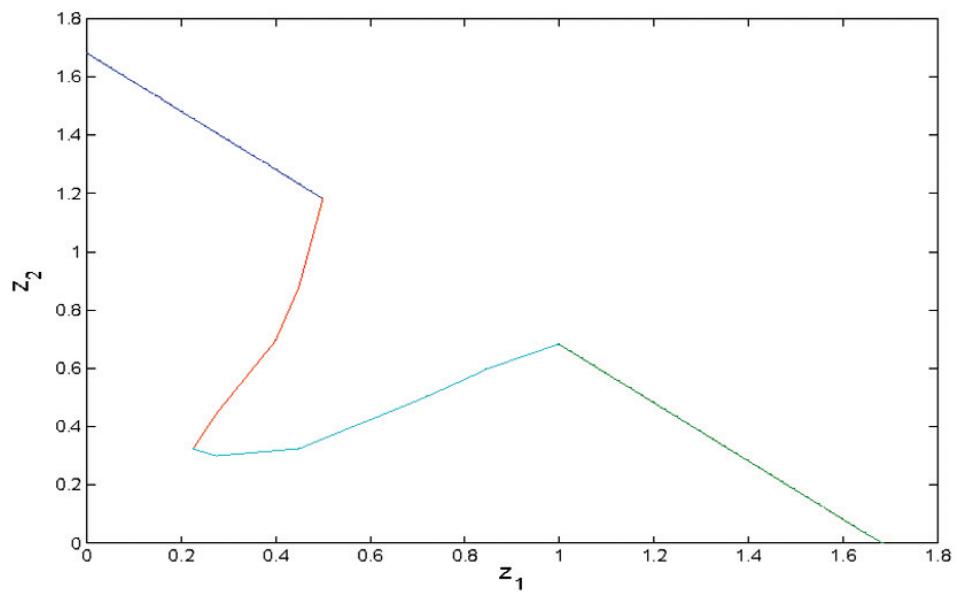

Fig. 4. The region where $\hat{\sigma}^2 > 0$ when $n_1 = 8$, $n_2 = 3$.

If $8\bar{z} < n-2$ (i.e., $4n^2 - 8n + 1 - 24n\bar{z} > (n-1)^2$), Eq. (33) must have two roots ξ, $\bar{\xi}$ as it takes a positive value when $\xi = 0$. Since $\text{sign}(\xi_0) = -\text{sign}(\Delta)$, assuming that $8\bar{z} \geq n-2$, we get

$$\frac{2|\Delta|}{n} = (\frac{n-1}{n} - \xi_0^2)|\xi_0|$$
$$\geq \left(\frac{n-1}{3n}\right)^{3/2} \left[2 + \sqrt{\frac{4n^2 - 8n + 1 - 24n\bar{z}}{(n-1)^2}}\right] \quad (37)$$
$$\left[1 - \sqrt{\frac{4n^2 - 8n + 1 - 24n\bar{z}}{(n-1)^2}}\right]^{1/2},$$

which means that

$$(4n^2 - 8n + 1 - 24n\bar{z})^{3/2}$$
$$+ 3(n-1)(4n^2 - 8n + 1 - 24n\bar{z}) \quad (38)$$
$$\geq 4[(n-1)^3 - 27n\Delta^2] \geq 0.$$

The condition (38) can also be shown to be sufficient for the existence of ξ, $\bar{\xi}$.

This parametrization and the Groebner basis for (13) allow for given z_1, z_2 to compare the minimal values of (9) when $y > 0$ with that when $y = 0$. Figures 2-4 show bounded regions where $\hat{\sigma}^2 > 0$ in the space (z_1, z_2) when $n_1 = n_2 = 2$, $n_1 = n_2 = 4$, or $n_1 = 8$, $n_2 = 3$. When $n_1 = n_2 \geq 3$, this region is a triangle, $z_1 + z_2 < c$; for $n_1 \neq n_2$ it is more complicated.

We summarize the obtained results.

Theorem 2.1. *When $p = 2$, the Groebner basis for (13) is given by (16), (17). The solutions of (21)-(23) satisfy conditions (26) and (27). These solutions can be parametrized via (29) by $s \in (\underline{s}, \bar{s})$ where the end points can be found from (32) given that (38) holds.*

We conclude by noticing the relationship of the problem discussed here to the likelihood solutions to the classical Behrens-Fisher problem [10] which assumes that $\sigma^2 = 0$.

2.2 Solving Likelihood Equations Numerically

In view of the difficulty in evaluating the Groebner basis for $p > 2$, an iterative method to solve the optimization problem in (2) is of interest. Notice that the sum of the first two terms in (2) is a homogeneous function of σ^2, σ_1^2, ..., σ_p^2, of degree -1. Therefore, with $n = \sum_i n_i$,

$$\min_{\sigma^2, \sigma_1^2, ..., \sigma_p^2} L = \min_{\lambda\sigma^2, \lambda\sigma_1^2, ..., \lambda\sigma_p^2} L$$

$$= \min_{\lambda, \sigma^2, \sigma_1^2, ..., \sigma_p^2} \left[\sum_i \frac{(x_i - \bar{\mu})^2}{\lambda(\sigma^2 + \sigma_i^2)} + \sum_i \frac{v_i s_i^2}{\lambda \sigma_i^2}\right.$$

$$+ \sum_i \log(\lambda(\sigma^2 + \sigma_i^2)) + \sum_i v_i \log(\lambda\sigma_i^2)\bigg]$$

$$= \min_{\sigma^2, \sigma_1^2, ..., \sigma_p^2} \left[\min_\lambda [\frac{1}{\lambda}(\sum_i \frac{(x_i - \bar{\mu})^2}{\sigma^2 + \sigma_i^2}\right.$$

$$+ \sum_i \frac{v_i s_i^2}{\sigma_i^2}) + n\log\lambda] + \sum_i \log(\sigma^2 + \sigma_i^2) \quad (39)$$

$$+ \sum_i v_i \log\sigma_i^2\bigg]$$

$$= \min_{y_0, y_1, ..., y_p} \left[n\log\left(\sum_i \frac{(x_i - \bar{\mu})^2}{y_0 + y_i} + \sum_i \frac{v_i s_i^2}{y_i}\right)\right.$$

$$+ \sum_i \log(y_0 + y_i) + \sum_i v_i \log y_i\bigg]$$

$$+ n - n\log n.$$

The minimizing value of λ is $\lambda_0 = (\sum_i (x_i - \bar{\mu})^2 / (y_0 + y_i) + \sum_i v_i s_i^2 / y_i)/n$. Thus the objective function in (39) becomes homogeneous of degree 0 in $y_0, y_1, ..., y_p$ which reduces the problem to minimization of such a function. If $y_0, y_1, ..., y_p$ is a solution to the problem (39), then $\hat{\sigma}^2 = \lambda_0 y_0, \hat{\sigma}_1^2 = \lambda_0 y_1, ..., \hat{\sigma}_p^2 = \lambda_0 y_p$ is a ML solution.

To minimize the objective function in (39)

$$F(y_0, y_1, ..., y_p)$$
$$= n\log\left(\sum_i \frac{(x_i - \bar{\mu})^2}{y_0 + y_i} + \sum_i \frac{v_i s_i^2}{y_i}\right) \quad (40)$$
$$+ \sum_i \log(y_0 + y_i) + \sum_i v_i \log y_i,$$

one can impose a restriction on $y_0, y_1, ..., y_p$ such as $\sum y_i = 1$ or $\lambda_0 = 1$. Note that F tends to $+\infty$ when $y_i \to 0$ or $y_i \to +\infty$ for any fixed $y_0 > 0$. Since $\sum_i (x_i - \bar{\mu}) / (y_0 + y_i) = 0$, one sees that

$$\frac{\partial}{\partial y_i} F = -n \frac{\frac{(x_i - \bar{\mu})^2}{(y_0 + y_i)^2} + \frac{v_i s_i^2}{y_i^2}}{\sum_j \left[\frac{(x_j - \bar{\mu})^2}{y_0 + y_j} + \frac{v_j s_j^2}{y_j}\right]} + \frac{1}{y_0 + y_i} + \frac{v_i}{y_i}, \quad (41)$$

which must be positive for large y_i and negative for sufficiently small y_i.

If $y_0 > 0$, by equating this derivative to zero, we obtain

$$\frac{s_i^2}{y_i^2} - \frac{\lambda_0}{y_i} + \frac{\lambda_i - n_i \lambda_0}{v_i y_0} = 0, \quad (42)$$

where $\lambda_i = (x_i - \bar{\mu})^2/(y_0 + y_i) + v_i s_i^2/y_i$, so that $\Sigma \lambda_i = n\lambda_0$. The equation (42) must have a positive root, which means that for λ_i evaluated at a stationary point,

$$\frac{v_i y_0}{4 s_i^2} \geq \frac{\lambda_i - n_i \lambda_0}{\lambda_0^2}, \quad (43)$$

and

$$\frac{s_i}{y_i} = \frac{\lambda_0}{2 s_i} + \sqrt{\frac{\lambda_0^2}{4 s_i^2} + \frac{n_i \lambda_0 - \lambda_i}{v_i y_0}} < \frac{\lambda_0}{2 s_i} + \sqrt{\frac{\lambda_0^2}{4 s_i^2} + \frac{n_i \lambda_0}{v_i y_0}}. \quad (44)$$

It follows that for any stationary point, $2 s_i^2 > \lambda_0 y_i$, i.e.,

$$\frac{2 s_i^2}{1 + \sqrt{1 + \dfrac{4 n_i s_i^2}{v_i \hat{\sigma}^2}}} < \hat{\sigma}_i^2 < 2 s_i^2. \quad (45)$$

For a fixed value of λ_0, say, $\lambda_0 = 1$, the Eq. (42) leads to an iteration scheme, with specified initial values of $y_0^{(0)}$ and $\bar{\mu}_0$. We take these to be the estimates arrived at by the method of moments as described later in Sec. 5, but once they are given, one can solve the cubic equation for $v_i = y_0/y_i$,

$$\begin{aligned} v_i (1 + v_i)^2 v_i s_i^2 - (1 + v_i)^2 v_i y_0 \\ - (1 + v_i) y_0 + v_i (x_i - \bar{\mu})^2 = 0. \end{aligned} \quad (46)$$

It is easy to see that each of these equations has either one or three positive roots. If there is just one root, then it uniquely defines v_i. In case of three positive roots, the root which minimizes (40) is chosen. Under this agreement, at stage k in (46) $v_i = v_i^{(k)}$, $y_0 = y_0^{(k)}$, $\bar{\mu} = \bar{\mu}_k$, and after solving (46) we put

$$\bar{\mu}_{k+1} = \frac{\displaystyle\sum_j \frac{v_j^{(k+1)} x_j}{1 + v_j^{(k+1)}}}{\displaystyle\sum_j \frac{v_j^{(k+1)}}{1 + v_j^{(k+1)}}}, \quad (47)$$

and

$$y_0^{(k+1)} = \frac{1}{n} \left[\sum_i \frac{v_i^{(k+1)} (x_i - \bar{\mu}_{k+1})^2}{1 + v_i^{(k+1)}} + \sum_i v_i s_i^2 v_i^{(k+1)} \right], \quad (48)$$

$k = 0, 1, \ldots.$

As in [11], (46) defines a sequence converging to a stationary point, and at each step the value of (40) is decreased.

Theorem 2.2 *Successive-substitution iteration defined by equations (46), (47) and (48) converges to a stationary point of (40) so that at each step this function decreases. At any stationary point inequalities (43) and (44) hold and (45) is valid.*

Notice that Vangel and Rukhin tacitly assume in [11] that $y_0 > 0$. The case of the global minimum attained at the boundary, i.e., when $y_0 = 0$, can be handled as follows. Equating (41) to zero gives

$$y_i \propto \frac{(x_i - \bar{\mu})^2 + v_i s_i^2}{n_i}, \quad (49)$$

and a simpler version of an iterative scheme

$$\begin{aligned} y_i^{(k+1)} &= \frac{\dfrac{(x_i - \bar{\mu}_k)^2 + v_i s_i^2}{n_i}}{\displaystyle\sum_j \frac{(x_j - \bar{\mu}_k)^2 + v_j s_j^2}{y_j^{(k)}}}, \quad \bar{\mu}_{k+1} \\ &= \sum_j \frac{x_j}{y_j^{(k+1)}} \left[\sum_j \frac{1}{y_j^{(k+1)}} \right]^{-1}, \end{aligned} \quad (50)$$

$k = 0, 1, \ldots, y_j^{(0)} = s_j^2, y_0 = 0$, converges fast. However the solutions obtained via (46), (47), (48) and (50) are to be compared by evaluating L. To assure that a global minimum has been found, several starting values should be tried.

The maximum likelihood estimator and the restricted maximum likelihood estimator discussed in the next section can be computed via their *R*-language implementation through the *lme* function from *nlme* library [12]. However this routine has a potential (false or singular) convergence problem occurring in some practical situations.

3. Restricted Maximum Likelihood

The possible drawback of the maximum likelihood method is that it may lead to biased estimators. Indeed the maximum likelihood estimators of σ's do not take into account the loss in degrees of freedom that results from estimating μ. This undesirable feature is eliminated in the restricted maximum likelihood (REML) method which maximizes the likelihood corresponding to the distribution associated with linear combinations of observations whose expectation is zero. This method discussed in detail by Harville [13] has gained considerable popularity and is employed now as a tool of choice by many statistical software packages.

The negative restricted likelihood function has the form

$$RL = L + \log(\sum_i (\sigma^2 + \sigma_i^2)^{-1}) =$$

$$\sum_{1 \le i < j \le p} \frac{(x_i - x_j)^2}{(\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2)} \left[\sum_i (\sigma^2 + \sigma_i^2)^{-1} \right]^{-1}$$

$$+ \log(\sum_i (\sigma^2 + \sigma_i^2)^{-1}) + \sum_i \log(\sigma_i^2 + \sigma^2)$$

$$+ \sum_i \frac{v_i s_i^2}{\sigma_i^2} + \sum_i v_i \log \sigma_i^2 . \tag{51}$$

By using notation (5) and (6), we can rewrite

$$RL = \frac{Q}{P'} + \log P' + \sum_i \frac{v_i s_i^2}{\sigma_i^2} + \sum_i v_i \log \sigma_i^2 . \tag{52}$$

To estimate σ^2 and σ_i^2, $i = 1, \ldots, p$ one has to find the minimal value of $Q/P' + \log P'$. The derivative of this function in σ^2 is proportional to the polynomial H of degree $2p{-}3$ in this variable,

$$H = P'P'' + Q'P' - QP'', \tag{53}$$

as opposed to $3p{-}3$ which is the degree of the corresponding polynomial F in (12) under the ML scenario.

For fixed σ_i^2, $i = 1, \ldots, p$, all p roots of the polynomial $P = P(\sigma^2)$ are real numbers, so that

$$PP'' \le \frac{p-1}{p}[P']^2 \le [P']^2 \tag{54}$$

[14, 3.3.29].

An application of (54) shows that

$$F \ge P(P'P'' + Q'P' - QP'') = PH . \tag{55}$$

Since $P(\sigma^2) > 0$, if F has a positive root σ^2, then H also must have a positive root τ^2. Moreover, $\tau^2 \le \sigma^2$, so that the ML estimator of σ^2 is always smaller than the REML estimator of the same parameter for the same data.

The polynomial equations for the restricted likelihood method have the form,

$$\sigma_i^4(Q_i' \, P' - QP_i' + P_i'P') + v_i(\sigma_i^2 - s_i^2)(P')^2 = 0 , \tag{56}$$

with each of these polynomials being of degree 3 in σ_i^2, $i = 1, \ldots, p$. When $\sigma^2 > 0$, Eqs. (56) have to be augmented by the equation $H = 0$.

3.1 Solving REML Equations

To solve the optimization problem in (51) in practice one can use a method similar to that in Sec. 2.2. Indeed,

$$\min_{\sigma^2, \sigma_1^2, \ldots, \sigma_p^2} RL = \min_{\lambda, \sigma^2, \sigma_1^2, \ldots, \sigma_p^2} \left[\sum_i \frac{(x_i - \bar{\mu})^2}{\lambda(\sigma^2 + \sigma_i^2)} + \sum_i \frac{v_i s_i^2}{\lambda \sigma_i^2} \right.$$

$$+ \log(\sum_i \frac{1}{\lambda(\sigma^2 + \sigma_i^2)}) + \sum_i \log(\lambda(\sigma^2 + \sigma_i^2))$$

$$\left. + \sum_i v_i \log(\lambda \sigma_i^2) \right]$$

$$= \min_{\sigma^2, \sigma_1^2, \ldots, \sigma_p^2} \left[\min_\lambda [\frac{1}{\lambda}(\sum_i \frac{(x_i - \bar{\mu})^2}{\sigma^2 + \sigma_i^2} + \sum_i \frac{v_i s_i^2}{\sigma_i^2}) \right.$$

$$+ (n-1)\log \lambda] + \log \left(\sum_i \frac{1}{\sigma^2 + \sigma_i^2} \right) + \sum_i \log \left(\sigma^2 + \sigma_i^2 \right)$$

$$\left. + \sum_i v_i \log \sigma_i^2 \right]$$

$$= \min_{y_0, y_1, \ldots, y_p} \left[(n-1)\log \left(\sum_i \frac{(x_i - \bar{\mu})^2}{y_0 + y_i} + \sum_i \frac{v_i s_i^2}{y_i} \right) \right.$$

$$+ \log \left(\sum_i \frac{1}{y_0 + y_i} \right) + \sum_i \log(y_0 + y_i) + \sum_i v_i \log y_i \right]$$

$$+ n - 1 - (n-1)\log(n-1) . \tag{57}$$

The minimizing value of λ is $\lambda_0 = (\sum_i (x_i - \bar{\mu})^2 / (y_0 + y_i) + \sum_i v_i s_i^2 / y_i) / (n-1)$, and the objective function in (57) is homogeneous of degree 0 in y's. If $y_0, y_1, \ldots, y_p$ is a solution to the problem (57), then $\tilde{\sigma}^2 = \lambda_0 y_0$, $\tilde{\sigma}_1^2 = \lambda_0 y_1$, $\ldots$, $\tilde{\sigma}_p^2 = \lambda_0 y_p$ is a REML solution.

When $y_0 > 0$, the function in (57) takes the form

$$(n-1)\log\left(\sum_i \frac{(x_i - \bar{\mu})^2}{y_0 + y_i} + \sum_i \frac{v_i s_i^2}{y_i}\right) \tag{58}$$

$$+\log\left(\sum_i \frac{1}{y_0 + y_i}\right) - \sum_i \log\left(\frac{1}{y_0 + y_i}\right) + \sum_i v_i \log y_i .$$

Its derivative with respect to y_i is

$$-(n-1)\frac{\dfrac{(x_i - \bar{\mu})^2}{(y_0 + y_i)^2} + \dfrac{v_i s_i^2}{y_i^2}}{\sum_j \left[\dfrac{(x_j - \bar{\mu})^2}{y_0 + y_j} + \dfrac{v_i s_j^2}{y_j}\right]}$$

$$-\frac{\dfrac{1}{(y_0 + y_i)^2}}{\sum_j \dfrac{1}{y_0 + y_j}} + \frac{1}{y_0 + y_i} + \frac{v_i}{y_i} . \tag{59}$$

As in Sec. 2.2, for a fixed value $\lambda_0 = 1$ we can use an iterative scheme which is based on (59). However, now one has to specify initial values of $y_0^{(0)}$, $\bar{\mu}_0$ and of $A^{(0)} = \sum(y_0 + y_j)^{-1}$ after which the cubic equations for $v_i = y_0 / y_i$,

$$v_i (1 + v_i)^2 v_i s_i^2 - (1 + v_i)^2 v_i y_0 - (1 + v_i) y_0$$
$$+ v_i / A + v_i (x_i - \bar{\mu})^2 = 0 \tag{60}$$

are solved for $i = 1, \ldots, p$, with updating as in (47) and (48). Each of these equations has either one or three positive roots. If there is just one root, then it defines v_i, out of three roots the largest is taken.

If $y = 0$, then with the vectors $e = (1, \ldots, 1)^T$, $z = (z_1, \ldots, (z_p)^T$, $z_i = 1/\sigma_i^2$, and $p \times p$ symmetric matrix $\mathbb{R}$ defined by positive elements $((x_i - x_j)^2 + v_i s_i^2 + v_j s_j^2)/2$, (58) takes the form

$$\min_{z_1, \ldots, z_p > 0} \left[(n-1)\log(z^T \mathbb{R} z) - (n-2)\log(z^T e) - \sum_i n_i \log z_i\right]. \tag{61}$$

The gradient of the function in (61)

$$\frac{2(n-1)\mathbb{R}z}{z^T \mathbb{R} z} - (n-2)\frac{e}{z^T e} - \frac{\vec{n}}{z}, \tag{62}$$

and its Hessian is

$$\nabla^2 F = \frac{2(n-1)\mathbb{R}}{z^T \mathbb{R} z} - \frac{4(n-1)\mathbb{R}z^T z \mathbb{R}}{(z^T \mathbb{R} z)^2} + \frac{(n-2)ee^T}{(z^T e)^2}$$
$$+ diag\left(\frac{\vec{n}}{z^2}\right). \tag{63}$$

where $\vec{n}/z$ is the vector with coordinates n_i / z_i, $i = 1, \ldots, p$, leads to the iteration scheme,

$$\frac{\vec{n}}{z^{(k+1)}} \propto \frac{2(n-1)\mathbb{R}z^{(k)}}{(z^{(k)})^T \mathbb{R} z^{(k)}} - (n-2)e,$$
$$z_i^{(0)} = \frac{s_i^{-2}}{\sum_k s_k^{-2}}, \quad e^T z^{(k)} = 1, \tag{64}$$

$k = 0, 1, \ldots$, which converges fast.

The solutions obtained via (60) and (64) are to be compared by evaluating RL in (57).

3.2 REML $p = 2$

To find the REML solutions $\tilde{\sigma}_i^2$ and $\tilde{\sigma}^2$ when $p = 2$ notice that (51) takes the form

$$RL = \left(\frac{(x_1 - x_2)^2}{2y + \sigma_1^2 + \sigma_2^2} + \log(2y + \sigma_1^2 + \sigma_2^2)\right)$$
$$+ v_1\left(\frac{s_1^2}{\sigma_1^2} + \log \sigma_1^2\right) + v_2\left(\frac{s_2^2}{\sigma_2^2} + \log \sigma_2^2\right), \tag{65}$$

which shows that if $(x_1 - x_2)^2 \geq s_1^2 + s_2^2$,

$$\tilde{\sigma}_i^2 = s_i^2, i = 1, 2, \tag{66}$$

$$\tilde{\sigma}^2 = \frac{1}{2}\left[(x_1 - x_2)^2 - s_1^2 - s_2^2\right]. \tag{67}$$

Indeed this choice simultaneously minimizes each of three bracketed terms in (65) guaranteeing the global minimum at this point. When $(x_1 - x_2)^2 < s_1^2 + s_2^2$, as we will see, $\sigma^2 = 0$. To find σ_i^2, $i = 1, 2$, one can employ the Groebner basis of the polynomial equations (56) which in the notation of Sec. 2.1 takes the form

$$u^2(u+v-1)+(v_1 u-z_1)(u+v)^2 = 0 \qquad (68)$$

$$v^2(u+v-1)+(v_2 v-z_2)(u+v)^2 = 0. \qquad (69)$$

This Groebner basis consists of three polynomials

$$G_3(u,v) = a_0 u^2 + a_1 v^5 + a_2 v^4 + a_3 v^3 + a_4 v^2, \qquad (70)$$

whose coefficients are not needed here,

$$G_4(u,v) = b_0 uv + b_1 v^5 + b_2 v^4 + b_3 v^3 + b_4 v^2,$$

$$b_0 = -2n_1 z_2^2 [n_1 z_1 z_2 + 2(n_2 -1)z_1 + (n_1 -2)z_2 + n_1 -2],$$

$$b_1 = -n_2(n-1)(n-2)(n_1 +2n_2 -2)^2,$$

$$b_2 = -(n_1 +2n_2 -2)^2 (2n_1 n_2 +2n_2^2 -n_1 -4n_2 +2)z_1$$

$$+(n_1 +2n_2 -2)(n_1^3 +4n_1^2 n_2 +7n_1 n_2^2 +4n_2^3 -5n_1^2$$

$$\qquad -14n_1 n_2 -12n_2^2 +8n_1 +12n_2 -4)z_2$$

$$+(n_1 +2n_2 -2)^2 (n_1^2 +2n_1 n_2 +2n_2^2 -3n_1 -4n_2 +2),$$

$$b_3 = -(n_1 +2n_2 -2)^2 v_2 z_1^2$$

$$+2(n_1 +2n_2 -2)(n_1^2 +2n_1 n_2 +2n_2^2 -3n_1 -4n_2 +2)z_1 z_2$$

$$-(3n_1^2 n_2 +6n_1 n_2^2 +4n_2^3 -n_1^2 -12n_1 n_2 -12n_2^2 +4n_1$$

$$\qquad +12n_2 -4)z_2^2$$

$$-2v_2(n_1 +2n_2 -2)^2 z_1 -2(n_1 +2n_2 -2)$$

$$\quad (2n_1 n_2 +2n_2^2 -n_1 -4n_2 +2)z_2$$

$$-2v_2(n_1 +2n_2 -2)^2,$$

$$b_4 = n_1 z_2(1+z_1 +z_2)[(n_1 +2n_2 -2)z_1 +(n_1 -2)z_2$$

$$+n_1 +2n_2 -2], \qquad (71)$$

and

$$G_5(u,v) = c_0 v^6 + c_1 v^5 + c_2 v^4 + c_3 v^3, \qquad (72)$$

with

$$c_0 = n_2(n-1)(n-2),$$

$$c_1 = (2n_2^2 +2n_1 n_2 -n_1 -4n_2 +2)z_1$$

$$-(n_1^2 +3n_2^2 +4n_1 n_2 -3n_1 -6n_2 +2)z_2$$

$$-n_1^2 -2n_2^2 -2n_1 n_2 +3n_1 +4n_2 -2,$$

$$c_2 = v_2 z_1^2 -(2n_1 +4n_2 -4)z_1 z_2$$

$$+(2n_1 +3n_2 -3)z_2^2 +2v_2 z_1 +(2n_1 +4n_2 -4)z_2 +v_2,$$

$$c_3 = -z_2[(z_1 -z_2)^2 +2(z_1 +z_2)+1].$$

All stationary points of the restricted likelihood equations can be found by solving the cubic equation $G_5(v) = 0$ for $v > 0$, and substituting this solution in (71), allowing one to express u through v,

$$u = -v(b_1 v^3 + b_2 v^2 + b_3 v + b_4)/b_0.$$

Thus, in this case there are 3 complex non-zero roots pairs u, v.

Another approach is to use the argument in Sec. 3.1 which for a fixed sum $2y_0 + y_1 + y_2 = 1$, leads to the optimization problem,

$$\min_{\substack{y_1+y_2 \le 1 \\ y_1>0, y_2>0}} \left[(n-1)\log\left(1+\frac{z_1}{y_1}+\frac{z_2}{y_2}\right) + \sum v_i \log y_i \right], \quad (73)$$

If $(x_1 - x_2)^2 < s_1^2 + s_2^2$, then according to Lemma 1 in the Appendix, for a fixed sum $y_1 + y_2 = y$, the minimum in (73) monotonically decreases when $0 < y \le 1$, so that this minimum is attained at the boundary, $y_1 + y_2 = 1$, in which case indeed $\tilde{\sigma}^2 = 0$. Thus, one can solve the cubic equation for $w = 1 - 2y_1$,

$$(n-2)w^3 -(4\Delta+2\delta)w^2 -(n+8\Delta\delta -4n\bar{z}-2)w$$

$$-4\Delta(n-1)-2\delta(1+4\bar{z}) = 0, \qquad (74)$$

where $\delta = (n_2 - n_1)/2$, and as before $\bar{z} = (z_1 + z_2)/2$, $\Delta = (z_2 - z_1)/2$.

An alternative method is to use the iteration scheme (64) to solve (57).

4. Known Variances

4.1 Conditions and Bounds for Strictly Positive Variance Estimates

In view of the rather complicated nature of likelihood equations, in many situations it is desirable to have a simpler estimating method for the common mean μ. The most straightforward way is to assume that the variances σ_i^2, $i = 1, \ldots, p$, are known. In this case, essentially suggested in Ref. [15], but also pursued in Refs. [16, 17], these variances are taken to be s_i^2 (or a multiple thereof). Because of (3), the only parameter to estimate is $y = \sigma^2$.

We give here upper bounds on the ML estimator $\hat{\sigma}^2$ and on the REML estimator $\tilde{\sigma}^2$.

Theorem 4.1 Assume that the variances σ_i^2, $i = 1, \ldots, p$, are known. Then

$$\hat{\sigma}^2 \leq \frac{1}{2} \max_{i \neq j} \left[\frac{(p-1)^2 (x_i - x_j)^2}{p} - (\sigma_i^2 + \sigma_j^2) \right]_+, \quad (75)$$

and

$$\tilde{\sigma}^2 \leq \frac{1}{2} \max_{i \neq j} \left[(p-1)(x_i - x_j)^2 - (\sigma_i^2 + \sigma_j^2) \right]_+. \quad (76)$$

In particular, if

$$\max_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} \leq \frac{p}{(p-1)^2}, \quad (77)$$

then $\hat{\sigma}^2 = 0$. *If*

$$\max_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} \leq \frac{1}{p-1}, \quad (78)$$

$\tilde{\sigma}^2 = 0$.
Proof: We prove first that

$$|QP'' - Q'P'| \leq (p-1) \max_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} P'P''. \quad (79)$$

Indeed, one can write

$$QP'' - Q'P' = \sum_{k,\ell} k(k-1-\ell) q_{p-2-\ell} \mathbb{S}_{p-k} y^{k+\ell-2}, \quad (80)$$

and Lemma 3 in the Appendix shows that since $(p-1-\ell)|k-1-\ell| \leq (p-1)(k-1)$, (79) holds.

One gets for the polynomial H in (53),

$$H \geq P'P'' \left[1 - (p-1) \max_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} \right], \quad (81)$$

which is positive under condition (78). Then $\tilde{\sigma}^2 = 0$.

Similarly (77) and (54) imply that

$$F \geq P' \left[(P')^2 - (p-1) \max_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} PP'' \right] \geq 0. \quad (82)$$

To prove (76) observe that with y replaced by $y + a$ a formula like (79) holds for any positive a. The only modification is that $\mathbb{S}_{p-\ell}(\sigma_1^2, \ldots, \sigma_p^2)$ is replaced by $\mathbb{S}_{p-\ell}(a + \sigma_1^2, \ldots, a + \sigma_p^2)$ and the $H^{(k)}(a)/k!$ represent polynomial coefficients. Then a version of Lemma 3 in the Appendix states that

$$\frac{Q^{(\ell)}(a)}{\ell!} \leq (\ell+1)(p-\ell-1)$$

$$\max_{1 \leq i < j \leq p} \left[\frac{(x_i - x_j)^2}{2a + \sigma_i^2 + \sigma_j^2} \right] \frac{P^{(\ell+1)}(a)}{(\ell+1)!} \quad (83)$$

$$= (p-\ell-1) \max_{1 \leq i < j \leq p} \left[\frac{(x_i - x_j)^2}{2a + \sigma_i^2 + \sigma_j^2} \right] \frac{P^{(\ell+1)}(a)}{\ell!}.$$

The argument, as before, shows now that if $\max_{1 \leq i < j \leq p} (x_i - x_j)^2 / (2a + \sigma_i^2 + \sigma_j^2) \leq (p-1)^{-1}$, then $H^{(k)}(a) \geq 0$, $k = 0, \ldots, 2p-3$. It follows that $\tilde{\sigma}^2 \leq a$ so that (76) is valid with $a = 2^{-1} \max[(p-1)(x_i - x_j)^2 - (\sigma_i^2 + \sigma_j^2)]_+$. The proof of (75) is similar. $_+$

When $p = 2$, the bounds (75) and (76) are sharp as (24) and (67) show. When p increases, their accuracy decreases. Section 4.2 gives necessary and sufficient conditions for $\tilde{\sigma}^2 = 0$ when $p = 3$. It is possible to get better estimates under additional conditions. For example, if the ordering of sequences $\mathbb{S}_{p-\ell}(\bar{\sigma}_{ij}^2)(\sigma_i^2 + \sigma_j^2)$ (notation of Lemma 2) and $(x_i - x_j)^2(\sigma_i^2 + \sigma_j^2)^{-1}$ is the same, then the maximum of $(x_i - x_j)^2(\sigma_i^2 + \sigma_j^2)$, $1 \leq i < j \leq p$, in (77) and (78) can be replaced by their average.

4.2 Example: Restricted Maximum Likelihood for $p = 3$

When $p = 3$, $Q(y) = q_0 y + q_1$, is a polynomial of degree one, and

$$H(y) = 18 y^3 + 3(6\mathbb{S}_1 - q_0) y^2$$
$$+ (6\mathbb{S}_2 + 4\mathbb{S}_1^2 - 6q_1) y + 2\mathbb{S}_2 \mathbb{S}_1 + q_0 \mathbb{S}_2 - 2q_1 \mathbb{S}_1 \quad (84)$$
$$= \sum_0^3 h_{3-k} y^k$$

is a polynomial of degree three. If $H(0) = h_3 < 0$, it has a positive root, which means that $\tilde{\sigma}^2 > 0$. Otherwise the existence of a positive root is related to the sign of the discriminant

$$D = h_1^2 h_2^2 - 4h_0 h_2^3 - 4h_1^3 h_3$$
$$- 27 h_0^2 h_3^2 + 18 h_0 h_1 h_2 h_3. \quad (85)$$

If $D < 0$ and $h_3 \geq 0$, there is just one real root which must be negative, and then $\tilde{\sigma}^2 = 0$. If $D \geq 0$, then there are three real roots. The condition $h_3 \geq 0$ implies that either two of them are positive and one is negative (so that at least one of the coefficients h_1 or h_2 is negative) and $\tilde{\sigma}^2 > 0$, or all three roots are negative (so that $\tilde{\sigma}^2 > 0$.)

Separation between these cases depends on the ratio $\lambda = q_1/q_0^2$. Let $E_1 = \mathbb{S}_1/q_0$, $E_2 = \mathbb{S}_2/q_0^2$. Then $E_2 \leq E_1^2/3$, $h_0 = 18$, $h_1/q_0 = 18(E_1 - 1/6)$, $h_2/q_0^2 = 6(E_2 + 2E_1^2/3 - \lambda)$, $h_3/q_0^3 = 2E_1 E_2 + E_2 - 2\lambda E_1$. Actually, as follows from the proof of Theorem 4.1, $E_1 \geq \lambda$, but we do not use this fact here.

Algebra shows that when $E_2 = E_1^2/3$, one has $D = 544(E_1 - 3\lambda)^2(3\lambda + 1/16 - E_1)$, when $h_3 = 0$, i.e., if $E_2 = \lambda E_1/(E_1 + 0.5)$, then $D = 36(4E_1^3 + E_1^2 - 3\lambda)^2 (8E_1^3 - 20E_1^2 - 10E_1 + 1 + 48\lambda)(1 + 2E_1)^{-3}$, and on the curve $h_2 = 0$, i.e., when $E_2 = \lambda - 2E_1^2/3$, D factorizes as follows, $D = 36(4E_1^3 + 2E_1^2 - 3\lambda)(108E_1^3 - 162E_1^2 - 18E_1 + 81\lambda)$. Let $\hat{E}_1$ be the smallest positive root of the cubic equation

$$8E_1^3 - 20E_1^2 - 10E_1 + 1 + 48\lambda = 0,\qquad(86)$$

which is defined for $\lambda < 0.5137\ldots$, and let $\bar{E}_1$ be the (only) positive root of the equation

$$4E_1^3 + 2E_1^2 - 3\lambda = 0.\qquad(87)$$

The discriminant of (87) is negative, and according to Descartes's rule there are always two complex roots, and one positive root. It is easy to check that $\bar{E}_1 = \bar{E}_1(\lambda)$ is monotonically increasing in $\lambda < 0$ and $\bar{E}_1(1/16) = 1/4$.

In this notation, when $\lambda \leq 1/16$ (so that $h_2 \leq 0$ implies that $h_1 \leq 0$), the region $\{(E_1, E_2) : E_2 \leq E_1^2/3,$ $h_3 \geq 0$, $\sigma_2 > 0\}$, is formed by three curves: (i) $h_3 = 0$ between the point $M_1 = ((\sqrt{1+48\lambda} - 1)/4, (1 + 24\lambda - \sqrt{1+48\lambda})/24)$, where $h_3 = 0$ intersects the curve $E_2 = E_1^2/3$, and the point $(\bar{E}_1, \lambda - 2\bar{E}_1^2/3)$ on $h_2 = 0$, (ii) $E_2 = E_1^2/3$ between M_1 and the point $M_2 = (E_1 = 3\lambda + 1/6, E_1^2/3)$ where $D = 0$ intersects $E_2 = E_1^2/3$, and by (iii) the cubic in E_2 curve corresponding to the equation $D = 0$, which connects the point M_2 and the point $M_3 = (3\lambda + 1/6, E_1^2/3)$. See Fig. 5. If $\lambda < 2/81$, this region also includes the part of the curve $h_3 = 0$ where $\bar{E}_1 \leq \hat{E}_1$, but the probability of having the likelihood solution there is zero.

When $\lambda > 1/16$, one has $E_1 > 1/4 = \bar{E}_1(1/16)$, so that $h_1 > 0$. Also then $E_2 > \lambda/3$, $\bar{E}_1 > \sqrt{\lambda}$, so that $h_2 > 0$. Thus in this situation $\tilde{\sigma}^2 > 0$ if and only if $h_3 \geq 0$.

5. Moment-Type Equations

5.1 Weighted Means Statistics

When the within-lab and between-lab variances σ_i^2 and σ^2 are known, the best (in terms of the mean squared error) unbiased estimator of the treatment effect μ in the model (1) is a weighted means statistic (3) with $w_i = w_i^0 = 1/(\sigma^2 + \sigma_i^2)$. Even without the normality assumption for these optimal weights, $E\sum_i w_i^0 (x_i - \bar{\mu})^2 = p - 1$, and

$$E(\bar{\mu} - \mu)^2 = \frac{1}{\sum_i w_i^0} = \frac{1}{\sum_i (\sigma^2 + \sigma_i^2)^{-1}}.\qquad(88)$$

If $Var(x_i) = \sigma_i^2 + \sigma^2$, but the weights w_i are arbitrary,

$$E\sum_i w_i(x_i - \bar{\mu})^2 = \sum_1^p (\sigma^2 + \sigma_i^2)w_i - \frac{\sum_1^p (\sigma^2 + \sigma_i^2)w_i^2}{\sum_1^p w_i}\qquad(89)$$

$$= \sigma^2\left[\sum_1^p w_i - \frac{\sum_1^p w_i^2}{\sum_1^p w_i}\right] + \sum_1^p \sigma_i^2 w_i - \frac{\sum_1^p \sigma_i^2 w_i^2}{\sum_1^p w_i}.$$

In particular, when $w_i = 1/\sigma_i^2$

$$E\sum_i \frac{(x_i - \bar{\mu})^2}{\sigma_i^2} = p - 1 + \sigma^2\left[\sum_1^p \frac{1}{\sigma_i^2} - \frac{\sum_1^p \frac{1}{\sigma_i^4}}{\sum_1^p \frac{1}{\sigma_i^2}}\right].\qquad(90)$$

The simplest estimate of the within-trials variances σ_i^2 is by the available s_i^2, but the problem of estimating the between-trial component of variance σ^2 remains.
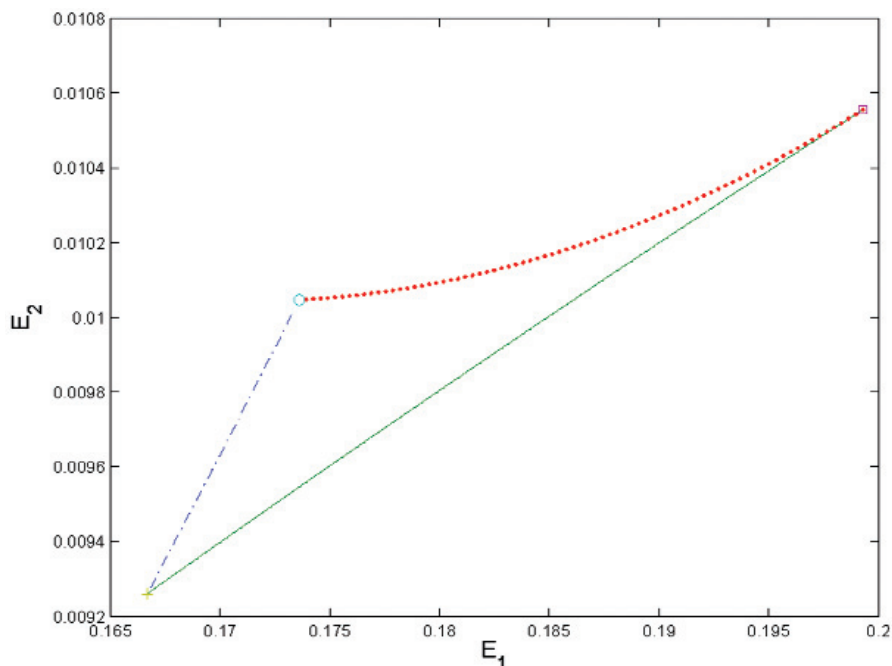
Fig. 5. The region where $h_3 \geq 0$ and $\bar{\sigma}^2 > 0$ when $\lambda = 1/27$. The solid line is $h_3 \geq 0$, the dash-dotted line is $E_2 = E_1^2/3$, and the dotted line is $D = 0$. The point M_1 is marked by $+$, the point M_2 is marked by a square, the point M_3 by o.

5.2 DerSimonian-Laird Procedure

By employing the idea behind the method of moments, DerSimonian and Laird [18] made use of the identity (90) as an estimating equation for μ and σ^2, provided that σ_i^2 are estimated by s_i^2, in the following way. For weights of the form

$$w_i = \frac{1}{y + s_i^2}, \qquad (91)$$

determine a non-negative $y = y_{DL}$ from the formula,

$$y_{DL} = \max\left[0, \frac{\sum_i s_i^{-2}(x_i - \tilde{x}^0)^2 - p + 1}{\sum_{i=1}^{p} s_i^{-2} - \sum_{i=1}^{p} s_i^{-4}\left(\sum_{i=1}^{p} s_i^{-2}\right)^{-1}}\right] \qquad (92)$$

and put $\tilde{x}_{DL} = (\sum_{i=1}^{p} x_i(y_{DL} + s_i^2)^{-2})/\sum_{i=1}^{p}(y_{DL} + s_i^2)^{-2}$. Here $\tilde{x}^0 = (\sum_{i=1}^{p} x_i s_i^{-2})/\sum_{i=1}^{p} s_i^{-2}$, is one of traditional estimators for the common mean (the Graybill-Deal estimator). In other words, the statistic $\tilde{x}^0$ and the weights $w_i = s_i^{-2}$ corresponding to $\sigma^2 = 0$, are used to evaluate the sum $\sum_i s_i^{-2}(x_i - \tilde{x}^0)^2$ which is then employed to estimate σ^2 via (90).

5.3 Mandel-Paule Method

The Mandel-Paule algorithm uses weights of the form (91) as well, but now $y = y_{MP}$, which is designed to approximate σ^2, is found from the moment-type estimating equation,

$$\sum_{i=1}^{p} \frac{(x_i - \tilde{x})^2}{y_{MP} + s_i^2} = p - 1. \qquad (93)$$

See Refs. [15], [19]. The *modified Mandel-Paule* procedure with $y = y_{MMP}$ is defined as above, but $p - 1$ in the right-hand side of (93) is replaced by p, i.e.,

$$\sum_{i=1}^{p} \frac{(x_i - \tilde{x})^2}{y_{MMP} + s_i^2} = p. \qquad (94)$$

Notice that when $p = 2$, the DerSimonian-Laird estimator coincides with the Mandel-Paule rule, as, in this case,

$$y_{MP} = y_{DL} = \frac{1}{2}\max\left[0, (x_1 - x_2)^2 - s_1^2 - s_2^2\right], \qquad (95)$$

so that this estimator is similar to the REML estimates estimates (66) and (67). In the general case, both of these rules set $y = 0$, when

$$\sum_{i=1}^{p} \frac{(x_i - \tilde{x})^2}{s_i^2} \leq p - 1. \qquad (96)$$

It was shown in Ref. [20] that the modified Mandel-Paule is characterized by the following fact: the ML estimator $\hat{\sigma}^2$ of σ^2 coincides with y_{MMP}, if in the reparametrized version of the likelihood equation the weights w_i admit representation (91). As a consequence, the corresponding weighted means statistic (3) must be close to the ML estimator, so that the modified Mandel-Paule estimator approximates its ML counterpart. Thus, the modified Mandel-Paule estimator can be interpreted as a procedure which uses weights of the form $1/(y + s_i^2)$, instead of solutions of the likelihood equation that are more difficult to find, and still maintains the same estimate of σ^2 as ML.

A similar interpretation holds for the original Mandel-Paule rule and the REML function [21]. For this reason both Mandel-Paule rules are quite natural. It is also suggestive to use the weights (91) with $y = y_{MP}$ determined from the Mandel-Paule equation (93) as a first approximation when solving the likelihood equations via the iterative scheme (47) and (48).

5.4 Uncertainty Assessment

It is tempting to use formula (88) to obtain an estimator of the variance of $\tilde{x}$. For example, DerSimonian and Laird [18] suggested an approximate formula for the estimate of the variance of their estimator,

$$\widehat{Var}(\tilde{x}_{DL}) = \frac{1}{\sum_i (y_{DL} + s_i^2)^{-1}}. \qquad (97)$$

Similarly motivated by (88), Paule and Mandel [19] suggested to use $[\sum_i (y_{MP} + s_i^2)^{-1}]^{-1}$ to estimate the variance of $\tilde{x}_{MP}$. However these estimators typically underestimate the true variance, Ref. [5]. They are not GUM consistent [22] in the sense that the variance estimate is not representable as a quadratic form in deviations $x_i - \tilde{x}$.

Horn, Horn and Duncan [23] in the more general context of linear models have suggested the following GUM consistent estimate of $Var(\tilde{x})$, which has the form $\sum_i \omega_i^2 (x_i - \tilde{x})^2 / (1 - \omega_i)$, with $\omega_i = w_i / \sum_k w_k$. When the plug-in weights are $\omega_i = (y + s_i^2)^{-1} / \sum_k (y + s_k^2)^{-1}$, $i = 1, \ldots, p$, this estimate is

$$\delta = \frac{\sum_i \frac{(x_i - \tilde{x})^2}{(y + s_i^2)^2} \left[\sum_{k:k \neq i} \frac{1}{y + s_k^2} \right]^{-1}}{\sum_i \frac{1}{y + s_i^2}}. \qquad (98)$$

Simulations show that (98) gives good confidence intervals of the form, $\tilde{x} \pm t_{\alpha/2}(p-1)\sqrt{\delta_2}$. For the Mandel-Paule rule or the DerSimonian-Laird procedure they outperform the intervals, $\tilde{x} \pm t_{\alpha/2}(p-1)\sqrt{\widehat{Var}(\tilde{x}_{DL})}$. Still when all sample means are close $x_i \approx \tilde{x}$, then $y_{DL} = 0$ and the uncertainty estimate $(\sum_i s_i^{-2})^{-1}$ may be a more satisfactory answer than $\delta \approx 0$.

When $p = 2$,

$$\delta = \frac{(x_1 - x_2)^2 (y + s_1^2)(y + s_2^2)}{(2y + s_1^2 + s_2^2)^2}. \qquad (99)$$

In this case δ is an increasing function of y with the largest value $(x_1 - x_2)^2 / 4$ attained when $y \to \infty$.

An alternative method of obtaining confidence intervals for μ on the basis of REML estimators was suggested in Ref. [24] for an adjusted estimator of $\widehat{Var}(\tilde{\mu})$ based on the inverse of the Fisher information matrix. This (not GUM consistent) estimator is more complicated.

6. Conclusions

The original motivation for this work was an attempt to employ modern computational algebra techniques by evaluating the Groebner bases of likelihood polynomial equations for the random effects model. While this attempt leads to an explicit answer when there are two labs with no between-lab effect, it was not successful in a more general situation. The classical iterative algorithms appear to be more efficient in this application although they do not guarantee the global optimum. Simplified, method of moments based procedures, especially the DerSimonian-Laird method, deserve much wider use in interlaboratory studies.

7.   Appendix

7.1   Lemma 1

Lemma 1 *If* $\dfrac{z_1}{v_1} + \dfrac{z_2}{v_2} > 1,$ *then the following function*

of y,

$$\min_{y_1 + y_2 = y} \left[(n-1)\log\left(1 + \frac{z_1}{y_1} + \frac{z_2}{y_2}\right) + \sum v_i \log y_i \right], \quad (100)$$

is monotonically nondecreasing in the interval $(0, 1)$

Proof: With $B = (1 + z_1/y_1 + z_2/y_2)/(n-1)$, any stationary point (y_1, y_2) in (100) satisfies the simultaneous equations,

$$\frac{z_1}{y_1^2 B} - \frac{v_1}{y_1} = \frac{z_2}{y_2^2 B} - \frac{v_1}{y_2} = \zeta , \quad (101)$$

where ζ is the Lagrange multiplier. By multiplying these equations by y_1 and y_2 respectively and adding them, we see that $\zeta y = 1 - 1/B$. Notice that $B \leq 1$ if and only if $\zeta \leq 0$, which means that $B \geq z_i /(v_i y_i)$ for $i = 1, 2$. Then $z_1/v_1 + z_2/v_2 \leq 1$, which contradicts the condition of the Lemma 1. Therefore, $B > 1$.

To prove Lemma 1 it suffices to show that the derivative of (100) is negative. This fact follows from the inequality

$$\frac{(n-1)B'}{B} + \frac{v_1 y_1'}{y_1} + \frac{v_2 y_2'}{y_2} = \frac{B-1}{By} < 0, \quad (102)$$

which can be proven by differentiation of (101). Indeed for $i = 1, 2$,

$$\frac{v_i y_1'}{y_i} + \frac{z_i B'}{y_i B^2} = \frac{y_i(B-1)}{y^2 B} + \frac{y_i B'}{y B^2} - \frac{2y_i'(B-1)}{yB} . \quad (103)$$

Since $y_1 + y_2 = y$, and $y_1' + y_2' = 1$, by adding these identities we get

$$\frac{v_1 y_1'}{y_1} + \frac{v_2 y_2'}{y_2} = -\frac{B'}{B^2}\left(1 + \frac{z_1}{y_1} + \frac{z_2}{y_2}\right) - \frac{B-1}{By} . \quad (104)$$

7.2   Elementary Symmetric Polynomials: Lemmas 2 and 3

We note the following identities for elementary symmetric polynomials. For $1 \leq i < j \leq p$, define $\bar{\sigma}_{ij}^2 = (\sigma_1^2, ..., \sigma_{i-1}^2, \sigma_{i+1}^2, ..., \sigma_{j-1}^2, \sigma_{j+1}^2, ..., \sigma_p^2)$. Since $P(y) = (y^2 + (\sigma_i^2 + \sigma_j^2)y + \sigma_i^2 \sigma_j^2)\prod_{k \neq i,j}(y + \sigma_k^2)$, one has

$$\mathbb{S}_{p-k} = \mathbb{S}_{p-k}(\bar{\sigma}_{ij}^2) + (\sigma_i^2 + \sigma_j^2)\mathbb{S}_{p-k-1}(\bar{\sigma}_{ij}^2) + \sigma_i^2 \sigma_j^2 \mathbb{S}_{p-k-2}(\bar{\sigma}_{ij}^2) . \quad (105)$$

Comparison of the coefficients of the polynomial $P'' = \sum_{\ell=2}^{p} \ell(\ell-1)\mathbb{S}_{p-\ell}y^{\ell-2}$ with its other expression, $2\sum_{1 \leq i < j \leq p}\prod_{k \neq i,j}(y + \sigma_k^2)$, shows that

$$\sum_{1 \leq i < j \leq p} \mathbb{S}_{p-\ell}(\bar{\sigma}_{ij}^2) = \binom{\ell}{2}\mathbb{S}_{p-\ell} . \quad (106)$$

By differentiating the identity, $\mathbb{S}_{p-\ell}(\lambda\sigma_1^2, ..., \lambda\sigma_p^2) = \lambda^{p-\ell}\mathbb{S}_{p-\ell}(\sigma_1^2, ..., \sigma_p^2)$ twice with respect to λ, we get

$$\sum_{1 \leq i < j \leq p} \sigma_i^2 \sigma_j^2 \mathbb{S}_{p-2-\ell}(\bar{\sigma}_{ij}^2) = \binom{p-\ell}{2}\mathbb{S}_{p-\ell} . \quad (107)$$

Combined with (105), identities (106) and (107) demonstrate the following formula (108) (which the author was unable to find in the literature.)

Lemma 2 *Under the notation above, for* $1 \leq \ell \leq p-1$,

$$\sum_{1 \leq i < j \leq p} (\sigma_i^2 + \sigma_j^2)\mathbb{S}_{p-1-\ell}(\bar{\sigma}_{ij}^2) = \ell(p-\ell)\mathbb{S}_{p-\ell} . \quad (108)$$

The following result Lemma 2 gives an upper bound on the coefficients of polynomial Q.

Lemma 3 *For* $\ell = 0, 1, , ..., p - 2$,

$$q_{p-2-\ell} \leq (\ell+1)(p-\ell-1)\mathbb{S}_{p-1-\ell}\max_{1 \leq i < j \leq p}\frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} . \quad (109)$$

Indeed,

$$
\begin{aligned}
q_{p-2-\ell} &= \sum_{1 \le i < j \le p} (x_i - x_j)^2 \, \mathbb{S}_{p-2-\ell}(\bar{\sigma}_{ij}^2) \\
&= \sum_{1 \le i < j \le p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} (\sigma_i^2 + \sigma_j^2) \, \mathbb{S}_{p-2-\ell}(\bar{\sigma}_{ij}^2).
\end{aligned}
\tag{110}
$$

8. References

[1] W. G. Cochran, Problems arising in the analysis of a series of similar experiments, J. Royal Statist. Soc. Supplement **4**, 102-118 (1937).

[2] W. G. Cochran, The combination of estimates from different experiments, Biometrics **10**, 101-129 (1954).

[3] P. S. R. S. Rao, J. Kaplan, and W. G. Cochran, Estimators for the one-way random effects model with unequal error variances, J. Am. Stat. Assoc. **76**, 89-97 (1981).

[4] P. S. R. S. Rao, Cochran's contributions to variance component models for combining estimates, in P. Rao and J. Sedransk, edis., W. G. Cochran's Impact on Statistics, Wiley, New York, (1981), pp. 203-222.

[5] A. L. Rukhin, Weighted means statistics in interlaboratory studies, Metrologia **46**, 323-331 (2009).

[6] D. R. Cox, Combination of data, in Encyclopedia of Statistical Sciences, Vol 2, S. Kotz, C. B. Read, N. Balakrishnan, and B. Vidakovic, eds., Wiley, New York (2006), pp. 1074-1081.

[7] E. F. Beckenbach and R. Bellman, Inequalities, Springer, Berlin, (1961), p. 3.

[8] D. Cox, J. Little, and D. O'Shea, Ideals, Varieties and Algorithms, 2nd ed, Springer, New York (1997), pp. 96-97.

[9] M. Drton, B. Sturmfels, and S. Sullivant, Lectures on Algebraic Statistics, Birkhauser, (2009), pp. 15-55.

[10] A. Belloni and G. Didier, On the Behrens-Fisher problem: a globally convergent algorithm and a finite-sample study of the Wald, LR and LM tests, Ann. Statist. **36**, 2377-2408 (2008).

[11] M. G. Vangel and A. L. Rukhin, Maximum likelihood analysis for heteroscedastic one-way random effects ANOVA in interlaboratory studies, Biometrics **55**, 129-136 (1999).

[12] R Development Core Team R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (2009) ISBN 3-900051-07-0, URL http://www.R-project.org.

[13] D. Harville, Maximum likelihood approaches to variance components estimation and related problems, J. Am. Stat. Assoc. **72,** 320-340 (1977).

[14] D. S. Mitrinovich, Analytic Inequalities, Springer, Berlin, (1970) p 227.

[15] J. Mandel and R. C. Paule, Interlaboratory evaluation of a material with unequal number of replicates, Anal. Chem. **42**, 1194-1197 (1970).

[16] R. Willink, Statistical determination of a comparison reference value using hidden errors, Metrologia **39**, 343-354 (2002).

[17] E. Demidenko, Mixed Models, Wiley, New York, (2004), pp. 253-255.

[18] R. DerSimonian and N. Laird, Meta-analysis in clinical trials, Control. Clin. Trials **7**, 177-188 (1986).

[19] R. C. Paule and J. Mandel, Consensus values and weighting factors, J. Res. Natl. Bureau Stand. **87**, 377-385 (1982).

[20] A. L. Rukhin and M. G. Vangel, Estimation of a common mean and weighted means statistics, J. Am. Stat. Assoc. **93**, 303-308 (1998).

[21] A. L. Rukhin, B. Biggerstaff, and M. G. Vangel, Restricted maximum likelihood estimation of a common mean and Mandel-Paule algorithm, J. Stat. Plann. Inf. **83**, 319-330 (2000).

[22] International Organization for Standardization, Guide to the Expression of Uncertainty in Measurement. Geneva, Switzerland, (1993).

[23] S. A. Horn, R. A. Horn, and D. B. Duncan, Estimating heteroscedastic variance in linear models. J. Am. Stat. Assoc. **70**, 380-385 (1975).

[24] M. G. Kenward and J. H. Roger, Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, **53**, 983-997 (1997).

***About the author*:** *Andrew L. Rukhin is a Mathematical Statistician in the Statistical Engineering Division of the NIST Information Technology Laboratory. He got his M.S. in Mathematics (Honors) from the Leningrad State University (Russia) in 1967 and his Ph.D. from the Steklov Mathematical Institute in 1970. After emigrating from the USSR he worked as a professor at Purdue University (1977-1987), at University of Massachusetts, Amherst (1987-1989) and at the University of Maryland at Baltimore County (1989-2008). Now Rukhin is a Professor Emeritus at this university. Rukhin is a Fellow of the Institute of Mathematical Statistics, and a Fellow of the American Statistical Association. He won Senior Distinguished Scientist Award By Alexander von Humboldt-Foundation in 1990, and the Youden Prize for Interlaboratory Studies in 1998 and in 2008. The National Institute of Standards and Technology is an agency of the U.S. Department of Commerce.*