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 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.
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

$$ x_{ij} = \mu + b_i + \epsilon_{ij}. \quad (1) $$

Here $\mu$ 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 $\sigma^2_i$, and $\epsilon_{ij}$ represent independent normal, zero mean random errors with unknown variances $\tau^2_i$.

For a fixed $i$, the $i$-th sample mean $x_i = \sum_j x_{ij}/n_i$ is normally distributed with the mean $\mu$ and the variance $\sigma^2 + \tau^2$, where $\sigma^2 = \tau^2/n_i$. If the $\sigma^2$'s were known, then the best estimator of $\mu$ would be the weighted average of $x_i$ with weights proportional to $1/(\sigma^2 + \tau^2)$. Since these variances are unknown, the weights have to be estimated. Traditionally to evaluate $\sigma^2_i$ one uses the classical unbiased statistic $s^2_i = \sum_j (x_{ij} - x_i)^2/(n_i - 1)$, which has the distribution $\sigma^2_i \chi^2(n_i)$.

Statistics $x_i$, $s^2_i$, $i = 1, ..., p$, are sufficient, we use the likelihood function based on them.

The ML solution minimizes in $\mu$, $\sigma^2$, $\sigma^2_i$, $i = 1, ..., p$, the negative logarithm of this function which is proportional to

$$ L = \sum_i \frac{(x_i - \mu)^2}{\sigma^2 + \sigma^2_i} + \sum_i \frac{v_i s^2_i}{\sigma^2_i} + \sum_i \log(\sigma^2 + \sigma^2_i) + \sum_i v_i \log \sigma^2_i. \quad (2) $$

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

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

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

$$ \sum_j w_j (x_j - \bar{\mu})^2 = \sum_j w_j x_j^2 - \left(\sum_j w_j x_j\right)^2/\left(\sum_j w_j\right) $$

$$ = \sum_{1 \leq i < j \leq p} w_i w_j (x_i - x_j)^2/\left(\sum_j w_j\right) . \quad (4) $$

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

We introduce a polynomial $P$ of degree $p$ in $\sigma^2$,

$$ P = P(\sigma^2, \sigma^2_1, ..., \sigma^2_p) = \prod_i (\sigma^2 + \sigma^2_i) - \sum_{\ell=0}^p S_\ell \sigma^2\ell \quad (5) $$

with $S_\ell = S_\ell (\sigma^2_1, ..., \sigma^2_p)$, $\ell = 0, 1, ..., p$, denoting the $\ell$-th elementary symmetric polynomial. Another polynomial of interest is $Q = Q(\sigma^2, \sigma^2_1, ..., \sigma^2_p)$ having degree $p - 2$ in $\sigma^2$,

$$ Q(\sigma^2, \sigma^2_1, ..., \sigma^2_p) = \sum_{1 \leq i < j \leq p} (x_i - x_j)^2 \prod_{k \neq i, j} (\sigma^2 + \sigma^2_k) $$

$$ = \sum_{\ell=0}^p q_{p-2,\ell} \sigma^2\ell = \sum_{1 \leq i < j \leq p} (x_i - x_j)^2 P_\ell, \quad (6) $$

where $P_\ell = \frac{\partial^2}{\partial \sigma^2_j \partial \sigma^2_j} P(\sigma^2, \sigma^2_1, ..., \sigma^2_p)$ is a polynomial in $\sigma^2$ of degree $p - 2$ which does not depend on $\sigma^2_i, \sigma^2_p$, and $q_{p-2,\ell}(\sigma^2_1, ..., \sigma^2_p)$ is a multilinear form in $\sigma^2_1, ..., \sigma^2_p$.

Since

$$ \frac{P'}{P} = -\frac{\partial}{\partial \sigma^2} \log P(\sigma^2, \sigma^2_1, ..., \sigma^2_p) = \sum_{i=1} p \frac{1}{\sigma^2 + \sigma^2_i}, \quad (7) $$

the identity (4) can be written as

$$ \sum_i \frac{(x_i - \bar{\mu})^2}{\sigma^2 + \sigma^2_i} = \frac{Q}{P'} \frac{P}{P'} = \frac{Q}{P'} \quad (8) $$
The negative log-likelihood function (2) in this notation is
\[ L = \frac{Q}{P'} + \log P + \sum_{i} v_{i} x_{i}^{2} + \sum_{i} v_{i} \log \sigma_{i}^{2}. \]  
(9)

Let \( P_{i} = \Pi_{\text{var}}(\sigma^{2} + \sigma_{i}^{2}) \) be the partial derivative of \( P \) with regard to \( \sigma_{i}^{2} \); denote by \( Q \) the same partial derivative of \( Q \) and by \( P'_{i} = \sum_{\text{var}} P_{i} \) the derivative of \( P'(\sigma^{2}, \sigma_{1}, \ldots, \sigma_{p}) \). By differentiating (9), we see that the stationary points of (2) satisfy polynomial equations,
\[ \begin{align*}
\sigma_{i}^{2}(\sigma^{2} + \sigma_{i}^{2})(Q'P' - QP') + \sigma_{i}^{2}(P')^{2} \\
+ v_{i}(\sigma_{i}^{2} - \sigma_{i}^{2})(\sigma^{2} + \sigma_{i}^{2}) P' = 0.
\end{align*} \]  
(10)

Each of these polynomials has degree 4 in \( \sigma_{i}^{2} \), \( i = 1, \ldots, p \). When \( \sigma^{2} = 0 \), \( P = \Pi_{\text{var}} \), and the equations (10) simplify to
\[ \begin{align*}
\sigma_{i}^{2}(\sigma^{2} + \sigma_{i}^{2})(Q'P' - QP') + \sigma_{i}^{2}(P')^{2} \\
+ v_{i}(\sigma_{i}^{2} - \sigma_{i}^{2})(\sigma^{2} + \sigma_{i}^{2}) P' = 0.
\end{align*} \]  
(11)

These polynomials have degree 3 in \( \sigma_{i}^{2} \). If \( \sigma_{i}^{2} > 0 \), in addition to (10) one has
\[ \begin{align*}
F = (P')^{3} + (Q'P' - QP')P = 0, \\
\end{align*} \]  
(12)
and \( F \) has degree 3\( p-3 \) in \( \sigma^{2} \).

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
\[ (x_{1} - x_{2})^{2} \sigma_{1}^{2} - (n_{1} \sigma_{1}^{2} - v_{1} x_{1}) (\sigma_{1}^{2} + \sigma_{2}^{2})^{2} = 0, \]
\( i = 1, 2 \). \( \sigma_{1}^{2}. \) With \( n = n_{1} + n_{2}, \) \( f = n + n_{2}, \) \( u = \sigma_{2}^{2}/(x_{1} - x_{2})^{2}, \) \( v = \sigma_{2}^{2}/(x_{1} - x_{2})^{2}, \) \( z_{1} = v x_{1} s_{1}^{2}/(x_{1} - x_{2})^{2} \)
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
\[ u^{2} + (z_{1} - n_{1} u)(u + v)^{2} = 0, \]  
(14)
\[ \]  
\[ v^{2} + (z_{2} - n_{2} v)(u + v)^{2} = 0, \]  
(15)

consists of two polynomials,
\[ G_{1}(u, v) = c_{1} u^{2} + c_{2} u v + c_{3} u^{3} + c_{4} v^{3} + c_{5} v^{2}, \]  
(16)
\[ c_{1} = n_{1} z_{1}^{2} f(1 + z_{1}) + n_{2} z_{1}, \]
\[ c_{2} = -n_{2} n_{1} f z_{1} + n_{1} n_{2} n_{3} n_{4} z_{1} - n_{2} n_{1} f, \]
\[ c_{3} = -n_{1} n_{2} n_{3} n_{4} z_{1} + n_{1} n_{2} n_{3} n_{4} n_{5} + n_{1} n_{2} n_{3} + n_{2} n_{1} f z_{1} \]
\[ -n_{1} n_{2} n_{3} n_{4} z_{1} + 2 n_{1} n_{2} n_{3} n_{4} n_{5} + n_{1} n_{2} n_{3} + n_{2} n_{1} f z_{1} \]
\[ + f n_{1} n_{2} n_{3} + 2 n_{1} n_{2} n_{3} f, \]
\[ c_{4} = -n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} n_{4} n_{5} + n_{1} n_{2} n_{3} + n_{2} n_{1} f z_{1} \]
\[ -2 n_{1} n_{2} n_{3} n_{4} z_{1} + 2 n_{1} n_{2} n_{3} n_{4} n_{5} + n_{1} n_{2} n_{3} + n_{2} n_{1} f z_{1} \]
\[ + 3 n_{2} n_{1} n_{2} n_{3} n_{4} n_{5} z_{1} + 3 n_{2} n_{1} n_{2} n_{3} n_{4} n_{5} + n_{1} n_{2} n_{3} + n_{2} n_{1} f z_{1} \]
\[ + n_{1} n_{2} n_{3} n_{4} z_{1} + n_{1} n_{2} n_{3} n_{4} + n_{2} n_{1} f z_{1} - f n_{1} n_{2} z_{1}, \]
\[ c_{4} = -n_{1} n_{2} n_{3} z_{1}^{2} f(1 + z_{1}) + n_{2} z_{1}, \]

and
\[ G_{2}(u, v) = d_{1} u v + d_{2} v^{3} + d_{3} u^{3} + d_{4} u v + d_{5} v^{2}, \]  
(17)
\[ d_{0} = 2 n_{1} n_{2} z_{1} f(1 + z_{1}) + n_{2} z_{1}, \]
\[ d_{1} = -n_{1} n_{2} f z_{1} n^{2} n_{1} + n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} z_{1}, \]
\[ d_{2} = -2 n_{1} n_{2} f z_{1} n^{2} n_{1} + n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} z_{1}, \]
\[ d_{3} = -n_{1} n_{2} f z_{1} n^{2} n_{1} + 2 n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} z_{1} + n_{1} n_{2} n_{3} z_{1}, \]
\[ d_{4} = n_{1} n_{2} (1 + z_{1}) f(1 + z_{1}) + n_{2} z_{1}, \]

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),
\[ u = -v(d_{1} u v + d_{2} v^{3} + d_{3} u v + d_{4} v + d_{5})/d_{0}, \]  
(18)
substituting this expression in (16), and solving then the resulting sextic equation for \( u \),

\[
c_0 (d_1 u^3 + d_2 u^2 + d_3 u + d_4)^2 + d_0^2 (c_1 u^2 + c_2 v^2 + c_3 v + c_4) = 0. \tag{19}
\]

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, \( x_i = -0.391, x_2 = 0.860, s_1^2 = 0.075, s_2^2 = 0.102, n_1 = n_2 = 3 \), there are three positive solutions, \( u = 0.263, u = 0.106, u = 0.035, \nu = 0.312 \). In this particular case one has to compare the likelihood function evaluated at these solutions with the likelihood evaluated when \( \sigma^2 > 0 \), in which case the stationary points are \( (u = 0.065, z = 0.230), (u = 0.246, \nu = 0.055, z = 0.003), (u = 0.042, u = 0.234, z = 0.019), (u = 0.048, \nu = 0.065, z = 0.193) \). (The likelihood is maximized at the last solution.)

If \( \sigma^2 = 0 \), the estimators \( \hat{\sigma}^2, \hat{\sigma}^2 \) satisfy the equations,

\[
\frac{(x_i - x_j)^2}{(\hat{\sigma}^2 + \hat{\sigma}^2)^2} = \frac{n_i}{\hat{\sigma}^2} - \frac{V_i \hat{\sigma}^2}{\hat{\sigma}^2} = \frac{n_j}{\hat{\sigma}^2} - \frac{V_j \hat{\sigma}^2}{\hat{\sigma}^2}, \tag{20}
\]

which imply the inequalities \( V_i \hat{\sigma}^2/n_i < \hat{\sigma}^2 < [(x_i - x_j)^2 + V_i \hat{\sigma}^2]/n_i \). Evaluation of second derivatives shows that \( 2n_i \hat{\sigma}^2 (x_i - x_j)^2 \leq n_i (\hat{\sigma}^2 + \hat{\sigma}^2) \).

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

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

\[
u^2 (u - v) + 2(z_i - V_i) y (y + u)(y + v) = 0, \tag{22}
\]

\[
u^2 (u - v) + 2(z_i - V_i) y (y + u)(y + v) = 0. \tag{23}
\]

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

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

\[
\hat{\sigma}^2 \leq \frac{(x_i - x_j)^2}{4}. \tag{24}
\]

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 \( z_i, z_i, n_i \) and \( n_i \). To find solutions to (21)-(23) we start with conditions on \( u, v \) for which \( y > 0 \) in (21)-(23).

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

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

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

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

or

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

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 \( (y + u)(y + v) \min \left(V_i(y + u)/u, V_i(y + v)/v\right) \geq y|u - v| \).

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

\[ (28) \]

which can be shown by the direct calculation or by the argument from Sec. 2.2. Thus assuming the condition

\[ z_1 / u + z_2 / \nu + 1 / (2 \nu + u + \nu) = n, \]

one can parametrize the solutions of Eqs. (21)-(23) as follows: for \( s \) in the unit interval, \( u = w, \ 0 < w < 1, \) solves the cubic equation

\[ u - \nu = -(2w)^2 \] (29)

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

\[ 2s(1-s) \left( n - \frac{1}{K} \right) \geq \frac{z_is + z_j(1-s)}{ns(1-s)}, \]

and

\[ |1-2w| \nu(1-w) < \frac{1}{6\sqrt{3}}. \] (31)

The discriminant of the cubic equation (29) reveals that when \( s \neq 1/2, |1-2s| \left[ z_is + z_j(1-s) < (n-2) \right] 3s(1-s)/(3\sqrt{3}n) \), 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,

\[ s(1-s)(1-2s)^3 - \frac{(n-2)s(1-s)}{n} + \frac{2(z_is + z_j(1-s))}{ns(1-s)} = 0. \] (32)

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

\[ \xi^4 - \frac{2(n-1)}{n} \xi^2 - \frac{8\Delta}{n} \xi - \frac{8\xi^2 - n+2}{n} = 0. \] (33)
If the equation (33) has two distinct roots \( \xi_-, \xi_+ \) in the interval \((-1, 1)\), then the range of \( s \)-values is the interval \([s_-, s_+]\), \(0 < s_- = (1 - \xi_+)/2 < s_+ = (1 - \xi_-)/2 < 1\), and for \( s \in (s_-, s_+)\) 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 \(|1 - 2\omega|\), 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 \( s \in [s_-, s_+]\).

To determine conditions for the equation (33) to have two roots \( \xi_-, \xi_+ \), 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 \(\xi_0\) 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 \(\xi_-, \xi_+\) exist if and only if

\[
3\xi_0^4 + 2(n-1)\xi_0 - 8\xi^n + 2n > 0.
\]

According to inequality (34),

\[
|\xi_0^2 - \frac{n-1}{3n}| \leq \frac{4n^2 - 8n + 1 - 24n\tau}{9n^2},
\]

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

\[
\xi_0^2 \geq \frac{n-1}{3n} \left[ 1 - \frac{4n^2 - 8n + 1 - 24n\tau}{(n-1)^2} \right].
\]
Fig. 3. The region where $\sigma^2 > 0$ when $n = 4, n \neq 4$.

Fig. 4. The region where $\sigma^2 > 0$ when $n = 8, n \neq 3$. 
If \( 8 \xi < n - 2 \) (i.e., \( 4 n^2 - 8n + 1 - 24 n \xi > (n - 1)^2 \)), Eq. (33) must have two roots \( \xi_1, \xi_2 \) as it takes a positive value when \( \xi = 0 \). Since \( \text{sign}(\xi) = -\text{sign}(\Delta) \), assuming that \( 8 \xi \geq n - 2 \), we get

\[
\frac{2|\Delta|}{n} = (\frac{n - 1}{n})^2 |\xi_0| \geq \left( \frac{n - 1}{3n} \right)^2 \left[ 2 + \sqrt{\frac{4n^2 - 8n + 1 - 24 n \xi}{(n - 1)^2}} \right] \frac{n^2}{2}
\]

which means that

\[
(4n^2 - 8n + 1 - 24 n \xi)^{\frac{3}{2}} + 3(n-1)(4n^2 - 8n + 1 - 24 n \xi) \geq 4((n-1)^2 - 27 n \Delta^2) \geq 0 .
\]

The condition (38) can also be shown to be sufficient for the existence of \( \xi_1, \xi_2 \).

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 \( \sigma^2 > 0 \) 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 \( s \in (\xi, \bar{\xi}) \) 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 \( \sigma^2, \sigma_1^2, \ldots, \sigma_p^2 \) of degree \(-1\). Therefore, with \( n = \sum n_i \),

\[
\begin{align*}
\min_{\sigma^2, \sigma_1^2, \ldots, \sigma_p^2} L &= \min_{\sigma^2, \sigma_1^2, \ldots, \sigma_p^2} L \\
&= \min_{\sigma^2, \sigma_1^2, \ldots, \sigma_p^2} \left[ \sum_{i} \left( \frac{x_i - \bar{y}}{\sigma_1} \right)^2 + \sum_{i} \frac{y_i}{\lambda \sigma_i^2} \right] \\
&+ \sum_{i} \log \left( \frac{\lambda \sigma_i^2}{\sigma_i^2} \right) + \sum_{i} y_i \log (\lambda \sigma_i^2) \\
&= \min_{\lambda, \sigma^2, \sigma_1^2, \ldots, \sigma_p^2} \left[ \sum_{i} \left( \frac{x_i - \bar{y}}{\sigma_1} \right)^2 + \sum_{i} \frac{y_i}{\lambda \sigma_i^2} \right] \\
&+ \sum_{i} \log (\lambda \sigma_i^2) + n \log \lambda + \sum_{i} \log (\sigma_i^2 + \sigma_i^2) \\
&\geq \min_{\lambda, \sigma^2, \sigma_1^2, \ldots, \sigma_p^2} \left[ n \log \left( \frac{\sum_{i} (x_i - \bar{y})^2}{\sum_{i} y_i \log (\lambda \sigma_i^2)} + \sum_{i} \frac{y_i}{\lambda \sigma_i^2} \right) \\
&+ \sum_{i} \log (\lambda \sigma_i^2) + n \log \lambda + \sum_{i} \log (\sigma_i^2 + \sigma_i^2) \right] \\
&+ n \log n .
\end{align*}
\]

The minimizing value of \( \lambda \) is \( \lambda_0 = (\sum_{i} (x_i - \bar{y})^2/(\sum_{i} y_i \log y_i)) + \sum_{i} y_i \sigma_i^2 y_i \)/n. Thus the objective function in (39) becomes homogeneous of degree 0 in \( y_0, y_1, \ldots, y_p \) which reduces the problem to minimization of such a function. If \( y_0, y_1, \ldots, y_p \) is a solution to the problem (39), then \( \sigma^2 = \lambda_0 y_0, \sigma_1^2 = \lambda_0 y_1, \ldots, \sigma_p^2 = \lambda_0 y_p \) is a ML solution.

To minimize the objective function in (39)

\[
F(y_0, y_1, \ldots, y_p) = n \log \left( \sum_{i} \left( \frac{x_i - \bar{y}}{\sigma_1} \right)^2 + \sum_{i} \frac{y_i}{\lambda \sigma_i^2} \right) + \sum_{i} \log (\lambda \sigma_i^2) + \sum_{i} \log (\sigma_i^2 + \sigma_i^2) + n \log n
\]

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

\[
\frac{\partial}{\partial y_i} F = -n \frac{\left( \frac{x_i - \bar{y}}{\sigma_1} \right)^2 (y_0 + y_i)^2 + y_i y_j}{(y_0 + y_i)^2 (y_0 + y_j)} + \frac{1}{y_0 + y_i} + \frac{y_j}{y_j},
\]

which must be positive for large \( y_i \) and negative for sufficiently small \( y_j \).
If \( y_0 > 0 \), by equating this derivative to zero, we obtain
\[
\frac{s_i^2 - \lambda_0 y_i + \lambda_i - n_i \lambda_i}{y_i^2} = 0, \quad (42)
\]
where \( \lambda_i = (x_i - \bar{\mu})^2/(y_0 + y_i) + v_i s_i^2/y_i \), so that \( \Sigma \lambda = n \lambda_0 \).

The equation (42) must have a positive root, which means that for \( \lambda_i \) evaluated at a stationary point,
\[
\frac{v_i y_i}{4 s_i^2} \geq \frac{\lambda_i - n_i \lambda_0}{\lambda_0}, \quad (43)
\]
and
\[
\frac{s_i}{y_i} = \frac{\lambda_0}{2 s_i} \left( \frac{\lambda_0^2}{4 s_i^2} \frac{n_i \lambda_0 - \lambda_i}{v_i y_i} \right) < \frac{\lambda_0}{2 s_i} \left( \frac{\lambda_0^2}{4 s_i^2} \frac{n_i \lambda_0}{v_i y_i} \right), \quad (44)
\]
It follows that for any stationary point, \( 2 s_i^2 > \lambda_0 y_i \), i.e.,
\[
2 s_i^2 < \frac{\sigma_i^2}{1 + \left( \frac{4 n_i s_i^2}{v_i \sigma_i^2} \right)}, \quad (45)
\]

For a fixed value of \( \lambda_0 \), say, \( \lambda_0 = 1 \), the Eq. (42) leads to an iteration scheme, with specified initial values of \( y_0^{(0)} \) and \( \bar{\mu} \). 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 \( \upsilon_i = y_0/y_i \),
\[
\upsilon_i (1 + \upsilon_i) y_i v_i s_i^2 - (1 + \upsilon_i)^2 v_i y_i = 0. \quad (46)
\]

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 \( \upsilon_i \). In case of three positive roots, the root which minimizes (40) is chosen. Under this agreement, at stage \( k \) in (46) \( \upsilon_i = \upsilon_i^{(0)} \), \( y_0 = y_0^{(0)} \), \( \bar{\mu} = \bar{\mu}^{(k)} \), and after solving (46) we put
\[
\bar{\mu}^{(k+1)} = \frac{\sum_j u_j^{(k+1)} x_j}{\sum_j 1 + u_j^{(k+1)}} \quad (47)
\]
and
\[
y_0^{(k+1)} = \frac{1}{n} \sum_j u_j^{(k+1)} (x_j - \bar{\mu}^{(k+1)})^2 + \sum_j v_j s_j y_j u_j^{(k+1)}, \quad (48)
\]
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
\[
y_i = \frac{(x_i - \bar{\mu})^2 + v_i s_i^2}{n_i} \quad (49)
\]
and a simpler version of an iterative scheme
\[
y_i^{(k+1)} = \frac{n_i}{\sum_j (x_j - \bar{\mu}^{(k)})^2 + v_i s_j^2} \quad (50)
\]

\[
y_0 = \bar{\mu}^{(k+1)} \quad (51)
\]

\[
k = 0, 1, \ldots, y_0^{(0)} = s_j^2, y_0 = 0, \] 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 $\sigma$'s do not take into account the loss in degrees of freedom that results from estimating $\mu$. 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

$$RL = L + \log \left( \sum_i (\sigma^2 + \sigma_i^2)^{-1} \right) = \\
\sum_{i \neq j} \frac{(x_i - x_j)^2}{(\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2)} \left[ \sum_i (\sigma^2 + \sigma_i^2)^{-1} \right]^{-1} + \log \left( \sum_i (\sigma^2 + \sigma_i^2)^{-1} \right) + \sum_i \log (\sigma_i^2 + \sigma^2) + \sum_i \frac{v_i s_i^2}{\sigma_i^2} + \sum_i v_i \log \sigma_i^2. \quad (51)$$

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

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

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

$$H = P'P'' + Q'P' - QP'', \quad (53)$$

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

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

$$PP'' \leq \frac{p-1}{p} [P']^2 \leq [P']^2 \quad (54)$$

An application of (54) shows that

$$F \geq P(P'P'' + Q'P' - QP'') = PH. \quad (55)$$

Since $P(\sigma^2) > 0$, if $F$ has a positive root $\sigma^2$, then $H$ also must have a positive root $\tau^2$. Moreover, $\tau^2 \leq \sigma^2$, so that the ML estimator of $\sigma^2$ 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,

$$\sigma_i^2 (Q' P' - QP' + P'P') + v_i (\sigma_i^2 - \bar{s}^2)(P')^2 = 0, \quad (56)$$

with each of these polynomials being of degree 3 in $\sigma_i^2$, $i = 1, \ldots, p$. When $\sigma_i^2 > 0$, 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,

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

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

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

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

$$+ n - 1 - (n-1) \log (n-1). \quad (57)$$
The minimizing value of \( \lambda \) is \( \lambda_0 = \frac{1}{n} \sum_i (x_i - \overline{x})^2 / (y_i + y_j) + \sum_i v_i \sigma_i^2 / y_j / (n-1) \), and the objective function in (57) takes the form

\[
\kappa \left[ (n-1) \log(x^T \mathbf{R} x) - (n-2) \log(z^T \mathbf{e}) - \sum_i n_i \log z_i \right],
\]

(61)

The gradient of the function in (61)

\[
\frac{2(n-1) \mathbf{R} z}{z^T \mathbf{R} z} - (n-2) \frac{e - \mathbf{n}}{z^T \mathbf{e} - z},
\]

(62)

and its Hessian is

\[
\mathbf{V}^2 \mathbf{F} = \frac{2(n-1) \mathbf{R} z^T \mathbf{R} z}{z^T \mathbf{R} z^2} - \frac{4(n-1) \mathbf{R} z^T \mathbf{R} e}{z^T \mathbf{R} e^2} + \frac{(n-2) e^2}{(z^T \mathbf{e})^2}
\]


\[+ \text{diag} \left( \frac{n}{z^2} \right).
\]

(63)

where \( n/z \) is the vector with coordinates \( n_i/z_i, i = 1, \ldots, p \), leads to the iteration scheme,

\[
\frac{n_i}{z^{(k+1)}} \approx \frac{2(n-1) \mathbf{R} z^{(k)}}{(z^{(k)})^2} - (n-2)e,
\]

(64)

\[
z_i^{(0)} = \frac{s_i^2}{\sum_k s_k^2}, \quad e^{(k)} = 1,
\]

\[k = 0, 1, \ldots, \text{which converges fast.}
\]

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 \( \sigma_i^2 \) and \( \sigma^2 \) when \( p = 2 \) notice that (51) takes the form

\[
RL = \left\{ \frac{(x_i - x_j)^2}{2y + \sigma_i^2 + \sigma_j^2} + \log(2y + \sigma_i^2 + \sigma_j^2) \right\}
\]

\[+ \kappa \left( \frac{s_i^2}{\sigma_i^2} + \log \sigma_i^2 \right) + \kappa \left( \frac{s_j^2}{\sigma_j^2} + \log \sigma_j^2 \right),
\]

(65)

which shows that if \( (x_i - x_j)^2 \geq s_i^2 + s_j^2 \),

\[
\sigma_i^2 = s_i^2, \quad i = 1, 2,
\]

(66)

\[
\sigma^2 = \frac{1}{2} \left[ (x_i - x_j)^2 - s_i^2 - s_j^2 \right].
\]

(67)

Indeed this choice simultaneously minimizes each of three bracketed terms in (65) guaranteeing the global minimum at this point. When \( (x_i - x_j)^2 < s_i^2 + s_j^2 \), as we will see, \( \sigma^2 = 0 \). To find \( \sigma_i^2, i = 1, 2 \), one can employ the Groebner basis of the polynomial equations (56) which in the notation of Sec. 2.1 takes the form
This Groebner basis consists of three polynomials

\[ G_i(u, v) = a_0 u v^3 + a_1 u v^2 + a_2 u v + a_3 v^2, \]  

whose coefficients are not needed here,

\[ G_i(u, v) = b_0 u v + b_1 u + b_2 v + b_3 u^2, \]

\[ b_0 = -2n_z z _1 [z_1 z_2 + 2(n_2 - 1)z_2 + (n_1 - 2)z_1 + n_1 - 2], \]

\[ b_1 = -n_2 (n - 1)(n - 2)(n_1 + 2n_2 - 2), \]

\[ b_2 = -(n_1 + 2n_2 - 2)^2(2n_1 n_2 + 2n_1^2 - n - 4n_2 + 2)z_1 + (n_1 + 2n_2 - 2)(n_1 + 4n_2 n_1 + 7n_1 n_2 + 4n_1^2 - 5n_1^2) \]

\[ -14n_1 z_1 + 8n_1 z_1 + 12n_2 - 4)z_1 + (n_1 + 2n_2 - 2)^2(n_1 + 4n_2 n_1 + 7n_1 n_2 + 4n_1^2 - 5n_1^2), \]

\[ b_3 = -(n_1 + 2n_2 - 2)^2v_z z_1 + (n_1 + 2n_2 - 2)(n_1 + 4n_2 n_1 + 7n_1 n_2 + 4n_1^2 - 5n_1^2), \]

\[ b_4 = -(n_1 + 2n_2 - 2)^2v_z z_1 - 2(n_1 + 2n_2 - 2)(2n_1 n_2 + 2n_1^2 - n - 4n_2 + 2)z_1, \]

\[ + 2(n_1 + 2n_2 - 2)(n_1 + 2n_1^2 - n - 4n_2 + 2)z_1, \]

\[ b_5 = n_1 z_1 (1 + z_1 + z_2) [(n_1 + 2n_2 - 2)z_1 + (n_1 - 2)z_1 + n_1 + 2n_2 - 2], \]

and

\[ G_i(u, v) = c_0 u^6 + c_1 u v^3 + c_2 u v + c_3 v^2, \]

with

\[ c_0 = n_2 (n - 1)(n - 2), \]

\[ c_1 = (2n_1 + 2n_1 n_1 - n - 4n_2 + 2)z_1 \]

\[ - (n_1 + 3n_1 + 4n_2 n_1 - 3n_2 - 6n_2 + 2) z_1 \]

\[ - 2n_1^2 - 2n_1 n_2 + 3n_2 + 4n_2 - 2, \]

\[ c_2 = v_z z_1^2 - (2n_1 + 4n_2 - 4)z_2 z_1 \]

\[ + (2n_1 + 3n_1 - 3)z_2^2 + 2v_z z_2 + (2n_1 + 4n_2 - 4)z_2 + v_z, \]

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

All stationary points of the restricted likelihood equations can be found by solving the cubic equation

\[ G_i(v) = 0 \]

for \( v > 0 \), and substituting this solution in (71), allowing one to express \( u \) through \( v \),

\[ u = -v(b_1 v^2 + b_2 v^2 + b_3 v + b_4)/b_0. \]

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

\[ \min_{y_1, y_2 \leq 0} \left( n - 1 \right) \log \left( 1 + \frac{z_1}{y_1} + \frac{z_2}{y_2} \right) + \sum y_i \log y_i \right), \]

(73)

If \( (x_i - x_j)^2 < s_i^2 + s_j^2 \), then according to Lemma 1 in the Appendix, for a fixed sum \( y_1 + y_2 = y \), the minimum in (73) monotonically decreases when \( 0 < y \leq 1 \), so that this minimum is attained at the boundary, \( y_1 + y_2 = 1 \), in which case indeed \( \sigma^2 = 0 \). Thus, one can solve the cubic equation for \( w = 1 - 2y_1 \),

\[ (n - 2)w^3 - (4\Delta + 2\delta) w^2 - (n + 8\Delta \delta - 4n\pi - 2) w \]

\[ - 4\Delta(n - 1) - 2\delta(1 + 4\pi) = 0, \]

(74)

where \( \delta = (n_2 - n_1)/2 \), and as before \( \pi = (z_1 + z_2)/2, \Delta = (z_1 - z_2)/2 \).

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 \( \mu \). The most straightforward way is to assume that the variances \( \sigma_i^2, i = 1, \ldots, p \), 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 $\hat{\sigma}^2$.

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

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

and

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

In particular, if

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

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

$$\frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} \leq \frac{1}{p-1},$$

$\hat{\sigma}^2 = 0$.

**Proof:** We prove first that

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

Indeed, one can write

$$QP'' - Q'P' = \sum_{k \leq \ell} k (k-1-\ell) q_{p-2-k} S_{p-k} y^{k+\ell-2},$$

and Lemma 3 in the Appendix shows that since $(p-1-\ell) k \leq (p-1)(k-1)$, (79) holds. One gets for the polynomial $H$ in (53),

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

which is positive under condition (78). Then $\hat{\sigma}^2 = 0$. Similarly (77) and (54) imply that

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

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 $S_{p-1} (\sigma_1^2, \ldots, \sigma_p^2)$ is replaced by $S_{p-1} (a + \sigma_1^2, \ldots, a + \sigma_p^2)$ and the $H^0(a)/k!$ represent polynomial coefficients. Then a version of Lemma 3 in the Appendix states that

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

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

The argument, as before, shows now that if

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

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 $\hat{\sigma}^2 = 0$ when $p = 3$. It is possible to get better estimates under additional conditions. For example, the ordering of sequences $S_{p-1} (\sigma_i^2)$ (notation of Lemma 2) and $(x_i - x_j)^2 (\sigma_i^2 + \sigma_j^2)$ 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 $p = 3$, $Q(y) = q_3 y + q_1$, is a polynomial of degree one, and

$$H(y) = 18y^3 + 3(6S_1 - q_3)y^2 + (6S_2 + 4S_1^2 - 6q_3)y + 2S_3S_1 + q_3S_2 - 2q_1S_1,$$

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

$$D = h_3^2h_2 - 4h_1h_2h_3 - 4h_1^2h_3 - 27h_2^2h_3^2 + 18h_3h_1h_2h_3.$$  

If $D < 0$ and $h_3 \geq 0$, there is just one real root which must be negative, and then $\hat{\sigma}^2 = 0$. 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 $\hat{\sigma}^2 > 0$, or all three roots are negative (so that $\hat{\sigma}^2 > 0$).
Separation between these cases depends on the ratio
\[ \lambda = q_i/q_0^i. \]
Let \( E_i = S_i/q_i, E_0 = S_0/q_0^i. \) Then \( E_i \leq E_0^i/3, \)
hence \( h_0 = 18, h_0/q_0 = 18(E_0 - 1/6), h_0/\lambda = (6E_0^i + 2E_0^i/3 - \lambda), \)
hence \( h_0 = 2E_0^i + 2E_0^i - 2\lambda E_0. \) Actually, as follows from the
proof of Theorem 4.1, \( E_i \geq \lambda, \) but we do not use this fact here.

Algebra shows that when \( E_0 = E_0^i/3, \) one has
\[ D = 544(E_i - 3\lambda)^2(3\lambda + 1/16 - E_0), \]
when \( h_1 = 0, \) i.e., if \( E_2 = \lambda E_i/(E_i + 0.5), \) then
\[ D = 36(4E_i^2 + E_0^2 - 3\lambda^2)(8E_0^3 - 20E_0^2 - 10E_0 + 1 + 48\lambda)(1 + 2E_0^2), \]
and on the curve \( h_2 = 0, \) i.e., when \( E_2 = \lambda - 2E_0^2/3, \) \( D \) factorizes as follows,
\[ D = 36(4E_i^2 + 2E_0^2 - 3\lambda)((108E_i^3 - 162E_0^2 + 18E_0 + 81\lambda). \]
Let \( \hat{E}_i \) be the smallest positive root of the
the cubic equation
\[ 8E_i^3 - 20E_i^2 - 10E_0 + 1 + 48\lambda = 0, \]  
which is defined for \( \lambda < 0.5137, \ldots, \) and let \( \hat{E}_0 \) be the
the only positive root of the equation
\[ 4E_0^3 + 2E_0^2 - 3\lambda = 0. \]  
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 \( \hat{E}_0 = E_0(\lambda) \)
is monotonically increasing in \( \lambda < 0 \) and \( E_0(1/16) = 1/4. \)

In this notation, when \( \lambda \leq 1/16 \) (so that \( h_0 \leq 0 \)
implies that \( h_0 \leq 0 \), the region \( \{ (E_1, E_0) : E_2 \leq E_0^i/3, \)
hence \( E_0 \geq 0, \) \( \sigma > 0 \}, \) is formed by three curves: (i) \( h_1 = 0 \)
below the point \( M_1 = (\sqrt{1 + 48\lambda} - 1)/4, (1 + 24\lambda - \sqrt{1 + 48\lambda})/24, \) where \( h_1 = 0 \) intersects the curve \( E_2 = \)
\( E_0^i/3, \) and the point \( \{ \hat{E}_0, \lambda = 2E_0^i/3 \} \) when \( h_0 = 0, \) (ii) \( E_2 = \)
\( E_0^i/3 \) between \( M_0 \) and the point \( M_2 = (E_2 = 3\lambda + 1/6, \)
\( E_0^i/3 \) where \( D = 0 \) intersects \( E_2 = E_0^i/3, \) and by (iii)
the cubic in \( E_2 \) curve corresponding to the equation
\( D = 0, \) which connects the point \( M_2 \) and the point \( M_3 = (3\lambda + 1/6, E_0^i/3). \) See Fig. 5. If \( \lambda < 2/81, \) this region also
includes the part of the curve \( h_1 = 0 \) where \( \hat{E}_0 \leq \hat{E}_i, \)
but the probability of having the likelihood solution
there is zero.

When \( \lambda > 1/16, \) one has \( E_i > 1/4 = E_0(1/16), \) so that \( h_0 > 0. \) Also then \( E_0 > \lambda/3, \) \( \hat{E}_i > \sqrt{\lambda}, \) so that \( h_0 > 0. \)
Thus in this situation \( \sigma^2 > 0 \) if and only if \( h_1 \geq 0. \)

5. Moment-Type Equations

5.1 Weighted Means Statistics

When the within-lab and between-lab variances \( \sigma^2 \)
and \( \sigma^2 \) are known, the best (in terms of the mean
squared error) unbiased estimator of the treatment
effect \( \mu \) in the model (1) is a weighted means statistic
\( \mu, \) when \( \lambda < 0. \) Even without the normality assumption for these optimal weights,
\( E \sum w_i (x_i - \mu)^2 = p - 1, \) and
\[ E(\overline{\mu} - \mu)^2 = \frac{1}{\sum w_i} \sum \frac{1}{(\sigma^2 + \sigma^2)^2}. \]  
If \( V ar(x_i) = \sigma^2 + \sigma^2, \) but the weights \( w_i \) are arbitrary,
\[ E \sum w_i (x_i - \overline{\mu})^2 = \sum \frac{1}{(\sigma^2 + \sigma^2)^2} w_i \]
\[ = \sigma^2 \left[ \sum w_i \frac{1}{\sum w_i} \right] \left[ \sum \frac{1}{\sum w_i} \right] \left[ \sum \frac{1}{\sum w_i} \right]. \]  
In particular, when \( w_i = 1/\sigma^2 \)
\[ E \sum (x_i - \overline{\mu})^2 = p - 1 + \sigma^2 \left[ \sum \frac{1}{\sigma^2} \left( \sum \frac{1}{\sigma^2} \right) - \sum \frac{1}{\sigma^2} \right]. \]  
The simplest estimate of the within-trials variances \( \sigma^2 \)
is by the available \( s^2, \) but the problem of estimating the between-trial component of variance
\( \sigma^2 \) remains.
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 $\mu$ and $\sigma^2$, provided that $\sigma_i^2$ are estimated by $s_i^2$, in the following way. For weights of the form (91)

$$w_i = \frac{1}{y + s_i^2},$$

(91)

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

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

(92)

and put

$$\bar{x}_{DL} = \left( \sum_{i=1}^{p} x_i (y_{DL} + s_i^2) \right) / \sum_{i=1}^{p} s_i^2,$$

Here $\bar{x}^0 = \left( \sum_{i=1}^{p} x_i s_i^2 \right) / \sum_{i=1}^{p} s_i^2$, is one of traditional estimators for the common mean (the Graybill-Deal estimator). In other words, the statistic $\bar{x}^0$ and the weights $w_i = s_i^{-2}$ corresponding to $\sigma^2 = 0$, are used to evaluate the sum $\sum_{i=1}^{p} s_i^{-2} (x_i - \bar{x})^{-2}$ which is then employed to estimate $\sigma^2$ 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 $\sigma^2$, is found from the moment-type estimating equation,

$$\sum_{i=1}^{p} (x_i - \bar{x})^2 = p - 1.$$  

(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.,

$$\sum_{i=1}^{p} (x_i - \bar{x})^2 = p.$$  

(94)

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

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

(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

$$\sum_{i=1}^{p} s_i^2 \leq p - 1.$$  

(96)
It was shown in Ref. [20] that the modified Mandel-Paule is characterized by the following fact: the ML estimator \( \sigma^2 \) coincides with \( y_{\text{MMP}} \), 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 \( \sigma^2 \) 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 \( y = y_{\text{MP}} \) 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 \( \hat{x} \). For example, DerSimonian and Laird [18] suggested an approximate formula for the estimate of the variance of their estimator,

\[
\hat{\text{Var}}(\hat{x}_{\text{DL}}) = \frac{1}{y_{\text{DL}} + s_j^2}.
\]  

(97)

Similarly motivated by (88), Paule and Mandel [19] suggested to use \( (\sum (y_{\text{MP}} + s_i^2)^{-1})^{-1} \) to estimate the variance of \( \hat{x}_{\text{MP}} \). 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 \( x_i - \bar{x} \).

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

\[
\delta = \frac{\sum (x_i - \bar{x})^2 \left[ \sum \frac{1}{y + s_i^2} \right]^{-1}}{\sum \frac{1}{y + s_i^2}}. \tag{98}
\]

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

When \( p = 2 \),

\[
\delta = \frac{(x_1 - x_2)^2 (y + s_1^2)(y + s_2^2)}{(2y + s_1^2 + s_2^2)^2}. \tag{99}
\]

In this case \( \delta \) 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 \( \mu \) on the basis of REML estimators was suggested in Ref. [24] for an adjusted estimator of \( \hat{\text{Var}}(\mu) \) 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 \( \frac{z_1}{y_1} + \frac{z_2}{y_2} > 1 \), then the following function of \( y \),

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

(100)

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

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

\[
\frac{z_1}{y_1} B = \frac{z_2}{y_2} B = \zeta,
\]

(101)

where \( \zeta \) is the Lagrange multiplier. By multiplying these equations by \( y_1 \) and \( y_2 \) respectively and adding them, we see that \( \zeta y = 1 - \frac{1}{B} \).

Notice that \( B \leq 1 \) if and only if \( \zeta \leq 0 \), which means that \( B \geq \frac{z_i}{(y_i y_i)} \) for \( i = 1, 2 \). Then \( z_i/y_1 + z_2/y_2 \leq 1 \), 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

\[
\frac{(n-1) B'}{B} + \frac{v_1 y_1'}{y_1} + \frac{v_2 y_2'}{y_2} = \frac{B-1}{B} < 0,
\]

(102)

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

\[
\frac{v_i y_i'}{y_i} + \frac{z_i B'}{y_i B^2} = \frac{y_i (B-1)}{y^2 B} + \frac{y_i B'}{y B^2} + \frac{2 y_i' (B-1)}{y B}.
\]

(103)

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

\[
\frac{v_1 y_1'}{y_1} + \frac{v_2 y_2'}{y_2} = \frac{B'}{B} \left( 1 + \frac{z_1}{y_1} + \frac{z_2}{y_2} \right) \frac{B-1}{B}.
\]

(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 \( \sigma_i^{(j)} = (\sigma_1^i, \ldots, \sigma_{i-1}^i, \sigma_{i+1}^j, \ldots, \sigma_p^j) \). Since \( P(y) = (y^2 + (\sigma_1^i + \sigma_2^j) y + \sigma_1^i \sigma_2^j) \prod_{i 
eq j} (y + \sigma_i^j) \), one has

\[
S_{p-k} = S_{p-k}^{(i)} + (\sigma_1^i + \sigma_2^j) S_{p-k-1}^{(i)} \sigma_1^i \sigma_2^j
\]

(105)

Comparison of the coefficients of the polynomial \( P'' = \sum_{i=2}^{p} (\ell - 1) S_{p-i} y^{i-2} \) with its other expression,

\[
2 \sum_{i=1}^{p} \sigma_1^i \sigma_2^j S_{p-i} (\sigma_1^i)^{i-2} y_{i-2} \]

shows that

\[
\sum_{i=1}^{p} S_{p-i}^{(i)} (\sigma_1^i)^{i-2} y_{i-2} = \frac{p-1}{2} S_{p-1}.
\]

(106)

By differentiating the identity, \( S_{p-i}^{(i)} (\lambda \sigma_1^i, \ldots, \lambda \sigma_p^i) = \lambda^{p-i} S_{p-i}^{(i)} (\sigma_1^i, \ldots, \sigma_p^i) \) twice with respect to \( \lambda \), we get

\[
\sum_{i=1}^{p} \sigma_1^i \sigma_2^j S_{p-i}^{(i)} (\sigma_1^i)^{i-2} y_{i-2} = \frac{p-1}{2} S_{p-1}.
\]

(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 \( 1 \leq \ell \leq p - 1 \),

\[
\sum_{i=1}^{p} (\sigma_1^i + \sigma_2^j) S_{p-\ell} = \ell (p - \ell) S_{p-\ell}.
\]

(108)

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

**Lemma 3** For \( \ell = 0, 1, \ldots, p - 2 \),

\[
q_{p-\ell} \leq (\ell + 1) (p - \ell - 1) S_{p-\ell-1} \max_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_1^i + \sigma_2^j}.
\]

(109)
Indeed,

\[ q_{p-2-t} = \sum_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} S_{p-2,t}(\sigma_i^2) \]

\[ = \sum_{1 \leq i < j \leq p} \frac{(x_i - x_j)^2}{\sigma_i^2 + \sigma_j^2} S_{p-2,t}(\sigma_i^2) \]  

(110)

8. References


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 1990 and in 2008. The National Institute of Standards and Technology is an agency of the U.S. Department of Commerce.