Bernoulli **20**(4), 2014, 1979–1998
DOI: 10.3150/13-BEJ547

Restricted likelihood representation and decision-theoretic aspects of meta-analysis

ANDREW L. RUKHIN

National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899, USA.
E-mail: andrew.rukhin@nist.gov

In the random-effects model of meta-analysis a canonical representation of the restricted likelihood function is obtained. This representation relates the mean effect and the heterogeneity variance estimation problems. An explicit form of the variance of weighted means statistics determined by means of a quadratic form is found. The behavior of the mean squared error for large heterogeneity variance is elucidated. It is noted that the sample mean is not admissible nor minimax under a natural risk function for the number of studies exceeding three.

Keywords: DerSimonian–Laird estimator; Hedges estimator; Mandel–Paule procedure; minimaxity; quadratic forms; random-effects model; Stein phenomenon

1. Parameter estimation in meta-analysis: Random-effects model

In the simplest random-effects model of meta-analysis involving, say, n studies the data is supposed to consist of treatment effect estimators x_i, $i = 1, \ldots, n$, which have the form

$$x_i = \mu + b_i + \varepsilon_i.$$

Here μ is an unknown common mean, b_i is zero mean between-study effect with variance τ^2, $\tau^2 \geq 0$, and ε_i represents the measurement error of the ith study, with variance σ_i^2, $\sigma_i^2 > 0$. Then the variance of x_i is $\tau^2 + \sigma_i^2$. In practice σ_i is often treated as a given constant, s_i, which is the reported standard error or uncertainty of the ith study.

The considered here problem is that of estimation of the common mean μ and of the heterogeneity variance τ^2 from the statistical decision theory point of view under normality assumption. If τ^2 is known, then the best unbiased estimator of μ is the weighted means statistic, $\hat{\mu}_{\mathrm{opt}} = \sum \omega_i^0 x_i$, with the normalized weights,

$$\omega_i^0 = \frac{1}{\tau^2 + s_i^2} \left(\sum_k \frac{1}{\tau^2 + s_k^2} \right)^{-1}, \tag{1.1}$$

$\sum \omega_i^0 = 1$. Its variance has the form

$$\mathrm{Var}(\hat{\mu}_{\mathrm{opt}}) = \left[\sum_i \frac{1}{\tau^2 + s_i^2} \right]^{-1}.$$

1350-7265

When τ^2 is unknown, to estimate μ the common practice uses a plug-in version of $\hat{\mu}_{\mathrm{opt}}$,

$$\hat{\mu}_{\mathrm{plug}} = \sum_i \frac{x_i}{\hat{\tau}^2 + s_i^2} \left(\sum_i \frac{1}{\hat{\tau}^2 + s_i^2} \right)^{-1}, \tag{1.2}$$

so that an estimator $\hat{\tau}^2$ of τ^2 is required in the first place.

Usually such an estimator is obtained from a moment-type equation [15]. For example, the DerSimonian–Laird [3] estimator of τ^2 is

$$\hat{\tau}^2_{\mathrm{DL}} = \frac{\sum_i (x_i - \delta_{\mathrm{GD}})^2 s_i^{-2} - n + 1}{\sum_i s_i^{-2} - \sum_i s_i^{-4} / (\sum_i s_i^{-2})}$$

with $\delta_{\mathrm{GD}} = \sum_i s_i^{-2} x_i / \sum_i s_i^{-2}$ denoting the Graybill–Deal estimator of μ. The popular DerSimonian–Laird μ-estimator is obtained from (1.2) by using the positive part of $\hat{\tau}^2_{\mathrm{DL}}$.

Similarly the estimator of τ^2,

$$\hat{\tau}^2_{\mathrm{H}} = \frac{\sum_i (x_i - \bar{x})^2 - (n-1) \sum_i s_i^2 / n}{n-1},$$

leads to the Hedges estimator of μ.

The paper questions the wisdom of using under all circumstances the tradition of plugging in τ^2 estimators to get μ estimators. Indeed the routine of plug-in estimators may lead to poor procedures. For example, by replacing the unknown τ^2 by $\hat{\tau}^2$ in the above formula for $\mathrm{Var}(\hat{\mu}_{\mathrm{opt}})$, one can get a flagrantly biased estimator which leads to inadequate confidence intervals for μ.

A large class of weighted means statistics is motivated by the form of Bayes procedures derived in Section 2.2. These statistics which typically *do not* admit the representation (1.2) induce estimators of the weights (1.1) which shows the primary role of μ-estimation.

The main results of this work are based on a canonical representation of the restricted likelihood function in terms of independent normal random variables and possibly of some χ^2-random variables. An important relationship between the weighted means statistics with weights of the form (1.1) and linear combinations of x's, which are shift invariant and independent, follows from this fact. Our representation transforms the original problem to that of estimating curve-confined expected values of independent heterogeneous χ^2-random variables. This reduction makes it possible to describe the risk behavior of the weighted means statistics whose weights are determined by a quadratic form.

We make use of the concept of permissible estimators which cannot be uniformly improved in terms of the differential inequality in Section 2.3. This inequality shows that the sample mean exhibits the Stein-type phenomenon being an inadmissible estimator of μ under the quadratic loss when $n > 3$. A risk function for the weights in a weighted means statistic whose main purpose is μ-estimation is suggested in Section 2.4. It is shown there that under this risk the sample mean is not even minimax. Section 2.5 discusses the case of approximately equal uncertainties, and Section 3 gives an example. The derivation of the canonical representation of the likelihood function is given in the Appendix; the proof of Theorem 2.1 is delegated to the Electronic Supplement [16].

2. Estimating the common mean

2.1. Restricted likelihood, heterogeneity variance estimation and quadratic forms

The setting with the common mean μ and the heterogeneity variance τ^2 described in Section 1 is a special case of a mixed linear model where statistical inference is commonly based on the restricted (residual) likelihood function.

The (negative) restricted log-likelihood function ([17], Section 6.6) has the form

$$\mathcal{L} = \frac{1}{2}\left[\sum_i \frac{(x_i - \hat{\mu}_{\text{opt}})^2}{\tau^2 + s_i^2} + \sum_i \log\bigl(\tau^2 + s_i^2\bigr) + \log\left(\sum_i \frac{1}{\tau^2 + s_i^2}\right)\right].$$

It is possible that some of s_i^2 are equal; let s_i^2 have the multiplicity ν_i, $\nu_i \geq 1$, so that $\sum \nu_i = n$. Then with the index i now taking values from 1 to p,

$$\mathcal{L} = \frac{1}{2}\left[\sum_i \frac{\nu_i(\bar{x}_i - \hat{\mu}_{\text{opt}})^2}{\tau^2 + s_i^2} + \sum_i \nu_i \log\bigl(\tau^2 + s_i^2\bigr) + \log\left(\sum_i \frac{\nu_i}{\tau^2 + s_i^2}\right) + \sum_i \frac{(\nu_i - 1)u_i^2}{\tau^2 + s_i^2}\right].$$

Here, p denotes the number of pairwise different s_i^2, $\bar{x}_i = \sum_{k:s_k = s_i} x_k/\nu_i$ represents the average of ν_i x's corresponding to the particular $s_i^2, i = 1, \ldots, p$, and u_i^2 is their sample variance when $\nu_i \geq 2$. To simplify the notation, we write x_i for $\bar{x}_i$, so that $\hat{\mu}_{\text{opt}} = \sum_i \nu_i(\tau^2 + s_i^2)^{-1}x_i/\sum_i \nu_i(\tau^2 + s_i^2)^{-1}$. In our problem $x = (x_1, \ldots, x_p)$ and $u_i^2 = \sum_{k:s_k=s_i}(x_k - \bar{x}_i)^2/(\nu_i - 1)$, $\nu_i > 1$, form a sufficient statistic for μ and τ^2.

Throughout this paper, we assume that $p \geq 2$. Otherwise all μ-estimators reduce to the sample mean (but see Section 2.5 where τ^2-estimation for equal uncertainties is considered). The results in the Appendix relate the likelihood function $\mathcal{L}$ to the joint density of $p-1$ independent normal, zero mean random variables $y_1, \ldots, y_{p-1}$. The $(p-1)$-dimensional normal random vector $y = (y_1, \ldots, y_{p-1})^{\mathrm{T}}$ which is a linear transform of x has zero mean (no matter what μ is) and the covariance matrix, $\operatorname{diag}(\tau^2 + t_1^2, \ldots, \tau^2 + t_{p-1}^2)$, with $t_1^2, \ldots, t_{p-1}^2$ larger than $\min s_i^2$.

To find these numbers, we introduce the polynomial $P(v) = \prod_i (v + s_i^2)^{\nu_i}$ of degree n, and its minimal annihilating polynomial $M(v) = \prod_i (v + s_i^2)$ which has degree p. Define

$$Q(v) = M(v)\frac{P'(v)}{P(v)} = \sum_i \nu_i \prod_{k:k\neq i} \bigl(v + s_k^2\bigr). \tag{2.1}$$

Thus Q is a polynomial of degree $p-1$ which has only real (negative) roots, denoted by $-t_1^2, \ldots, -t_{p-1}^2$ (coinciding with the roots of P' different from $-s_1^2, \ldots, -s_p^2$). Thus $Q(v) = n\prod_j (v + t_j^2)$. Note that $M(-t_j^2) \neq 0$. When $\nu_i \equiv 1$, $M(v) = P(v)$, and $Q(v) = P'(v)$.

According to (2.1),

$$\sum_i \log\bigl(\tau^2 + s_i^2\bigr) + \log\left(\sum_i \frac{\nu_i}{\tau^2 + s_i^2}\right) = \sum_j \log\bigl(\tau^2 + t_j^2\bigr) + \log n,$$

so that by using (A.8) one gets

$$\begin{aligned}\mathcal{L} = \frac{1}{2}\Bigg[&\sum_j \frac{y_j^2}{\tau^2+t_j^2} + \sum_j \log\bigl(\tau^2+t_j^2\bigr)\\ &+ \sum_i \frac{(\nu_i-1)u_i^2}{\tau^2+s_i^2} + \sum_i (\nu_i-1)\log\bigl(\tau^2+s_i^2\bigr) + \log n\Bigg].\end{aligned} \tag{2.2}$$

The representation (2.2) of the restricted likelihood function very explicitly takes into account one degree of freedom used for estimating μ, as it corresponds to $p-1$ independent zero mean, normal random variables y_j with variances $\tau^2+t_j^2$, $j=1,\dots,p-1$. In addition, this likelihood includes independent u_i^2, each being a multiple of a χ^2-random variable with ν_i-1 degrees of freedom. When $\nu_i>1$, u_i^2 is an unbiased estimator of $\tau^2+s_i^2$, $u_i^2 \sim (\tau^2+s_i^2)\chi^2_{\nu_i-1}/(\nu_i-1)$. For $\nu_i=1$, $u_i^2=0$ with probability one. According to the sufficiency principle, all statistical inference about τ^2 involving the restricted likelihood can be based exclusively on $y_1^2,\dots,y_{p-1}^2$ and $u_1^2,\dots,u_p^2$. Their joint distribution forms a curved exponential family whose natural parameter is formed by $(\tau^2+t_j^2)^{-1}$ (and perhaps by some $(\tau^2+s_i^2)^{-1}$).

Evaluation of the restricted maximum likelihood estimator (REML) $\hat\tau^2$ is considerably facilitated by employing $y_1^2,\dots,y_{p-1}^2$ and $u_1^2,\dots,u_p^2$. Indeed (2.2) shows that this estimator can be determined by simple iterations as

$$\hat\tau^2 = \left(\sum_j \frac{y_j^2-t_j^2}{(\hat\tau^2+t_j^2)^2} + \sum_i \frac{(\nu_i-1)(u_i^2-s_i^2)}{(\hat\tau^2+s_i^2)^2}\right)\Big/\left(\sum_j \frac{1}{(\hat\tau^2+t_j^2)^2} + \sum_i \frac{\nu_i-1}{(\hat\tau^2+s_i^2)^2}\right)$$

with $\hat\tau^2_{\mathrm{DL}}$ as a good starting point, and truncation at zero if the iteration process converges to a negative number. Thus, $\hat\tau^2$ is related to a quadratic form whose coefficients are inversely proportional to the estimated variances of $y_j^2-t_j^2$ and of $u_i^2-s_i^2$ (cf. [4], Section 8).

The form of the likelihood function $\mathcal{L}$ also motivates the moment-type equations based on general quadratic forms, $\sum_j q_j y_j^2 + \sum_i (\nu_i-1) r_i u_i^2$ with positive constants q_j, r_i. The moment-type equation written in terms of random variables $y_1^2,\dots,y_{p-1}^2$ and $u_1^2,\dots,u_p^2$ is

$$E\left[\sum_j q_j y_j^2 + \sum_i (\nu_i-1) r_i u_i^2\right] = \left[\sum_j q_j + \sum_i (\nu_i-1) r_i\right]\tau^2 + \sum_j q_j t_j^2 + \sum_i (\nu_i-1) r_i s_i^2.$$

Then the estimator of τ^2 by the method of moments is

$$\hat\tau^2 = \frac{\sum_j q_j (y_j^2-t_j^2) + \sum_i (\nu_i-1) r_i (u_i^2-s_i^2)}{\sum_j q_j + \sum_i (\nu_i-1) r_i}.$$

Unless τ^2 is large, the probability that $\hat\tau^2$ takes negative values is non-negligible. Non-negative statistics $\hat\tau^2_+ = \max(\hat\tau^2, 0)$ are used to get μ-estimators of the form (1.2).

The representations of two traditional statistics in Section 1 easily follow,

$$\hat{\tau}^2_{\mathrm{DL}} = \frac{\sum_j t_j^{-2} y_j^2 + \sum_i (\nu_i - 1) s_i^{-2} u_i^2 - n + 1}{\sum_j t_j^{-2} + \sum_i (\nu_i - 1) s_i^{-2}}$$

and

$$\hat{\tau}^2_{\mathrm{H}} = \frac{\sum_j (y_j^2 - t_j^2) + \sum_i (\nu_i - 1)(u_i^2 - s_i^2)}{n-1}.$$

A different method-of-moments procedure suggested by Paule and Mandel [12] is based on solving the equation,

$$\sum_i \frac{\nu_i (x_i - \hat{\mu})^2}{\tau^2 + s_i^2} + \sum_i \frac{(\nu_i - 1) u_i^2}{\tau^2 + s_i^2} = n - 1,$$

which has a unique positive solution, $\tau^2 = \hat{\tau}^2_{\mathrm{MP}}$, provided that $\sum_i \nu_i (x_i - \hat{x}_{\mathrm{GD}})^2 s_i^{-2} + \sum_i (\nu_i - 1) u_i^2 s_i^{-2} > n - 1$. If this inequality does not hold, $\hat{\tau}^2_{\mathrm{MP}} = 0$. Because of (A.8), the equation can be rewritten in terms of y's and t's as

$$\sum_j \frac{y_j^2}{\tau^2 + t_j^2} + \sum_i \frac{(\nu_i - 1) u_i^2}{\tau^2 + s_i^2} = n - 1.$$

This representation allows for an explicit form of $\hat{\tau}^2_{\mathrm{MP}}$ in some cases.

Indeed, when $n = p = 2$, $\hat{\tau}^2_{\mathrm{MP}} = \hat{\tau}^2_{\mathrm{DL}} = \hat{\tau}^2_{\mathrm{H}} = \max[0, y_1^2 - t_1^2]$, which is also the REML. When $n = p = 3$, $y_1^2/t_1^2 + y_2^2/t_2^2 \geq 2$,

$$\hat{\tau}^2_{\mathrm{MP}} = \frac{y_1^2 + y_2^2}{4} - \frac{t_1^2 + t_2^2}{2} + \sqrt{\left(\frac{y_1^2 + y_2^2}{4}\right)^2 + \left(\frac{t_1^2 - t_2^2}{2}\right)^2 - \frac{(t_1^2 - t_2^2)(y_1^2 - y_2^2)}{4}}.$$

We conclude this section by noticing that the widely used heterogeneity index I^2 ([1], page 117) in terms of y's and u's takes the from

$$I^2 = \frac{\sum_j t_j^{-2} y_j^2 + \sum_i (\nu_i - 1) s_i^{-2} u_i^2 - n + 1}{\sum_j t_j^{-2} y_j^2 + \sum_i (\nu_i - 1) s_i^{-2} u_i^2}, \qquad 0 \leq I^2 < 1.$$

2.2. Weighted means statistics and suggested estimators

Let us look now at the generalized Bayes estimator of μ when Λ is a prior distribution for τ^2 while μ has the uniform (non-informative) prior. Under the quadratic loss this estimator has the form with $\mathcal{L}$ given in (2.2)

$$\delta = \frac{\int_0^\infty \hat{\mu}_{\mathrm{opt}} \exp\{-\mathcal{L}\}\, \mathrm{d}\Lambda(\tau^2)}{\int_0^\infty \exp\{-\mathcal{L}\}\, \mathrm{d}\Lambda(\tau^2)} = \sum_i \omega_i x_i. \tag{2.3}$$

Thus δ is a weighted means statistic with normalized weights, $\omega_i \propto \nu_i \int_0^\infty (\tau^2+s_i^2)^{-1}[\sum \nu_k(\tau^2+s_k^2)^{-1}]^{-1} \exp\{-\mathcal{L}\} \,\mathrm{d}\Lambda(\tau^2)$, $\sum_i \omega_i = 1$, which are shift invariant, $\omega_i(x_1+c,\ldots,x_p+c) = \omega_i(x_1,\ldots,x_p)$ for any real c. (Any function of $y_1,\ldots,y_{p-1}$ is shift invariant.) Indeed the use of restricted likelihood is tantamount to the practice of weighted means statistics with invariant weights as μ estimators. (cf. [17], Section 9.2).

Formula (A.10) in the Appendix gives

$$\delta = \bar{x} - \sum_j \frac{\int_0^\infty (\tau^2+t_j^2)^{-1} \exp\{-\mathcal{L}\} \,\mathrm{d}\Lambda(\tau^2)}{\int_0^\infty \exp\{-\mathcal{L}\} \,\mathrm{d}\Lambda(\tau^2)} \sqrt{b_j} y_j = \bar{x} - \sum_j w_j \sqrt{b_j} y_j$$

with y_j discussed in Section 2.1. Positive coefficients b_j (the diagonal elements of the diagonal matrix $A^{\mathrm{T}} J^{-1} A$ defined in Lemma A.1) can be found from (A.3) or rather from (A.11); w_j is the posterior mean of $(\tau^2+t_j^2)^{-1}$,

$$w_j = -2\frac{\partial}{\partial y_j^2} \log \lambda\big(y_1^2,\ldots,y_{p-1}^2,u_1^2,\ldots,u_p^2\big)$$

with $\lambda = \int_0^\infty \exp\{-\mathcal{L}\} \,\mathrm{d}\Lambda(\tau^2)$. Thus positive w_j is designed to estimate $(\tau^2+t_j^2)^{-1}$, $w_j \le t_j^{-2}$, and as a function of y_ℓ^2, w_j decreases. The inequalities, $t_j^2 < t_\ell^2$, and $w_j > w_\ell$, are equivalent.

If $p > 2$ and the support of Λ has at least two points, δ does not admit representation (1.2) which suggests a more general class of μ-estimators. Namely, we propose to use weighted means statistics $\delta = \sum_i \omega_i x_i$ with weights $\omega_i = 1/p - \sum_j w_j A_{ij}$. The Bayes weights belong to a smaller part of this polyhedron, namely to the convex hull of the vectors with coordinates $(\tau^2+t_1^2)^{-1},\ldots,(\tau^2+t_{p-1}^2)^{-1}$ for $\tau^2 \ge 0$. If $\hat{\tau}^2$ is an estimate of τ^2, the weights corresponding to (1.2),

$$w_j = \frac{1}{\hat{\tau}^2 + t_j^2}, \tag{2.4}$$

lie on the boundary of this convex hull. A corner point, $(t_1^{-2},\ldots,t_{p-1}^{-2})$, of the convex hull always is an inner point of the polyhedron.

Thus the focus in this paper is on estimators δ of μ, which admit the representation,

$$\delta = \sum_i \omega_i x_i = \bar{x} - \sum_j \sqrt{b_j} w_j y_j \tag{2.5}$$

with $w_j, 0 \le w_j \le t_j^{-2}$, y_j and b_j as defined above. The last term in the right-hand side of (2.5) can be viewed as an arguably necessary heterogeneity correction to $\bar{x}$.

Notice that (2.5) does not need an estimate of τ^2 as a prerequisite. Since w_j is an approximation to $(\tau^2+t_j^2)^{-1}$, when $n = p$, the form of the REML $\hat{\tau}^2$ in Section 2.1 suggests such an estimator: $[\sum w_j^2(y_j - t_j^2)]_+/\sum w_j^2$. If some of the multiplicities exceed one, an estimator of $(\tau^2+s_i^2)^{-1}$ can be derived from w_j by replacing t_j^2 by s_i^2. According to (A.10), $\hat{\mu}_{\mathrm{opt}}$ as well as $\bar{x}$, has the form (2.5). In fact, all traditional statistics (1.2) admit this representation.

2.3. Estimation of multivariate normal mean and permissible procedures

We look now at the quadratic risk behavior of μ-estimators of the form (2.5). If $\delta = \sum_i \omega_i x_i$ is such an estimator with positive normalized weights ω_i which are shift invariant functions of $x_1, \dots, x_p$, then it is unbiased. Its variance does not depend on μ and can be written as

$$\mathrm{Var}(\delta) = \mathrm{Var}(\hat{\mu}_{\mathrm{opt}}) + E(\delta - \hat{\mu}_{\mathrm{opt}})^2 = \left[\sum_i \frac{\nu_i}{\tau^2 + s_i^2}\right]^{-1} + E(\delta - \hat{\mu}_{\mathrm{opt}})^2 \tag{2.6}$$

by independence of $\hat{\mu}_{\mathrm{opt}}$ and $\delta - \hat{\mu}_{\mathrm{opt}}$. This and more general decompositions of the mean squared error are discussed by Harville [5]. The second term in the right-hand side of this identity is an important variance component which shows how well δ approximates the optimal but unavailable $\hat{\mu}_{\mathrm{opt}}$, and which relates our setting to the classical estimation problem of the multivariate $(p-1)$-dimensional normal mean.

Proposition 2.1. *If the coefficients $w_j = w_j(y_1^2, \dots, y_{p-1}^2, u_1^2, \dots, u_p^2)$ defining the estimator (2.5) are piecewise differentiable in y's, then*

$$\begin{aligned}\mathrm{Var}(\delta) &= \mathrm{Var}(\hat{\mu}_{\mathrm{opt}}) + \sum_j b_j E y_j^2 \left(w_j - \frac{1}{\tau^2 + t_j^2}\right)^2 \\ &= \mathrm{Var}(\bar{x}) + E \sum_j b_j \left(f_j^2 - 2\frac{\partial}{\partial y_j} f_j\right),\end{aligned}$$

where $f_j = y_j w_j$. When $p > 3$, $\bar{x}$ is an inadmissible estimator of μ under the quadratic loss.

The omitted proof of Proposition 2.1 is based on (A.10), (A.11), and on familiar integration by parts technique. It demonstrates linkage of our situation to the differential inequality of a statistical estimation problem [2]. Namely, if for some vector θ, $Y \sim N_{p-1}(\theta, I)$, then $\sum_j b_j(f_j^2 - 2\,\partial f_j/\partial y_j)$ is an unbiased estimate of $\sum_j b_j E(Y_j + f_j(Y) - \theta_j)^2 - \sum_j b_j \theta_j^2$. Therefore $Y + g(Y)$, $g = (g_1, \dots, g_{p-1})^{\mathrm{T}}$, improves on $Y + f(Y)$, $f = (f_1, \dots, f_{p-1})^{\mathrm{T}}$, as a θ-estimator provided that for all values $Y_1, \dots, Y_{p-1}$,

$$\sum_j b_j \left(f_j^2 - 2\frac{\partial}{\partial Y_j} f_j\right) \geq \sum_j b_j \left(g_j^2 - 2\frac{\partial}{\partial Y_j} g_j\right). \tag{2.7}$$

Following [13], let us call a (piecewise differentiable) vector function f *permissible* if (2.7) does not have any solutions g providing a strict inequality at some point. Thus, $Y + f$ is a permissible estimator of the vector normal mean θ if and only if the corresponding scalar μ-estimator, $\bar{x} - \sum_j \sqrt{b}_j f_j$, cannot be improved upon in the sense of differential inequality (2.7). Since for $p > 3$, $f \equiv 0$ is not a permissible function, the sample mean $\bar{x}$ is inadmissible in the original setting. Indeed the left-hand side of (2.7) is negative for $f_j^S = y_j w_j^S$, $w_j^S = (p-3)/(b_j \sum y_\ell^2 / b_\ell)$ proving this statement.

The differential operator in (2.7) does not involve t's or s's, but in our problem only functions f_j such that $|f_j| \le |y_j| t_j^{-2}$ and $f_j/y_j \ge 0$ are of interest. Since $(\tau^2 + t_j^2)^{-1}$ is positive and cannot exceed t_j^{-2}, according to the first equality in Proposition 2.1, w_j can be improved by $\max[0, \min(w_j, t_j^{-2})]$.

The proof of Theorem 1 in [2] shows that any permissible w_j in our situation is of the form

$$w_j = \max\left[0, \min\left(-\frac{\partial}{\partial y_j^2} \log \lambda, \frac{1}{t_j^2}\right)\right]$$

with some piecewise differentiable positive function $\lambda = \lambda(y_1^2, \dots, y_{p-1}^2, u_1^2, \dots, u_p^2)$. When $n = p$ and $\lambda = \lambda(q)$ for a positive quadratic form $q = \sum_j q_j y_j^2, q_j > 0$, one gets $w_j = \min[-q_j (\log \lambda)'(q), t_j^{-2}]$. If there are multiplicities exceeding one, the quadratic form q is to be replaced by $q = \sum_j q_j y_j^2 + \sum_i (\nu_i - 1) r_i u_i^2$. For example, the function, $\lambda(q) = q^{-\alpha}, \alpha > 0$, leads to the estimator (2.5) with

$$w_j = \min\left[\frac{\alpha q_j}{\sum_\ell q_\ell y_\ell^2 + \sum_i (\nu_i - 1) r_i u_i^2}, \frac{1}{t_j^2}\right]. \tag{2.8}$$

The statistic $w_j^{JS} = \min(w_j^S, t_j^{-2})$, corresponding to $q_j = b_j^{-1}, \alpha = p - 3$, when $n = p$ is similar to the positive part of the Stein estimator of the vector normal mean which improves over Y. However, in the meta-analysis context it is desirable having the coefficients q_j of the same ordering as t_j^{-2}, and this condition may not hold for $q_j \propto b_j^{-1}$. As a matter of fact, despite doing better than $w_j \equiv 0$ or w_j^S, the weights w_j^{JS} do not produce a good estimator of μ. The same is true for many other procedures (2.8) satisfying condition (2.10) of Theorem 2.1 in the next section. This theorem shows that if $p \le 3 < n$, $\bar{x}$ is an inadmissible estimator of μ although the function $f \equiv 0$ is permissible then.

2.4. *R*-risk and asymptotic optimality

According to (2.6) the variance of estimator (2.5) is completely determined by the term, $E(\delta - \hat{\mu}_{\mathrm{opt}})^2$, which can be interpreted as a cost of not knowing τ^2 when estimating μ, or as a new risk of δ viewed as a procedure providing approximations to $(\tau^2 + t_j^2)^{-1}$, $j = 1, \dots, p-1$. More conveniently, with $s^2 = \sum_i \nu_i s_i^2 / n$, define

$$R(\delta, \tau^2) = \frac{E(\delta - \hat{\mu}_{\mathrm{opt}})^2}{\mathrm{Var}(\bar{x}) - \mathrm{Var}(\hat{\mu}_{\mathrm{opt}})} = \frac{E[\sum_i (\omega_i - \omega_i^0) x_i]^2}{(\tau^2 + s^2)/n - [\sum_i \nu_i / (\tau^2 + s_i^2)]^{-1}}$$

to be the R-risk of δ. Because of (A.10) and (2.5), the ensuing random loss function has the form,

$$L(\delta, \tau^2) = \frac{(\delta - \hat{\mu}_{\mathrm{opt}})^2}{(\tau^2 + s^2)/n - [\sum_i \nu_i / (\tau^2 + s_i^2)]^{-1}} = \frac{\sum_j (w_j - 1/(\tau^2 + t_j^2))^2 b_j y_j^2}{\sum_j b_j / (\tau^2 + t_j^2)}.$$

This loss is invariant under a scale change of y_j, τ, t_j (or of x_i, τ, s_i). For $\tau^2 \to \infty$,

$$\frac{\tau^2 + s^2}{n} - \left[\sum_i \frac{\nu_i}{\tau^2 + s_i^2}\right]^{-1} \sim \frac{\sum_i \nu_i (s_i^2 - s^2)^2}{n^2 \tau^2},$$

so that the normalizing factor in the definition of L amplifies the error in approximating $\hat{\mu}_{\text{opt}}$ when τ^2 is large. The results of this section show that for estimators δ satisfying conditions of the following Theorem 2.1,

$$\operatorname{Var}(\delta) = \operatorname{Var}(\bar{x}) + \left[\operatorname{Var}(\bar{x}) - \operatorname{Var}(\hat{\mu}_{\text{opt}})\right]\left[R(\delta, \tau^2) - 1\right] = \frac{\tau^2 + s^2}{n} + \mathrm{O}\left(\frac{1}{\tau^2}\right)$$

when $\tau^2 \to \infty$. Thus, $\operatorname{Var}(\bar{x}) = (\tau^2 + s^2)/n$ is the dominating contribution to the variance of δ when τ^2 is large. The R-risk is a useful tool for the comparison of estimators (2.5), as unlike the normalized quadratic risk, $E(\delta - \mu)^2 / \operatorname{Var}(\bar{x})$, it removes this linear in τ^2 term.

If $\delta = \hat{\mu}_{\text{plug}}$ with an invariant $\hat{\tau}^2$, then $R(\hat{\mu}_{\text{plug}}, \tau^2)$ can be interpreted as a conventional risk of the estimator $\hat{\tau}^2$. However under this risk large values of $\hat{\tau}^2$ are not penalized very much no matter what τ^2 is. Indeed $\hat{\tau}^2$ is not designed to estimate τ^2 itself, but rather $(\hat{\tau}^2 + t_j^2)^{-1}$ estimates $(\tau^2 + t_j^2)^{-1}$ (cf. [11], page 329). When $n = p = 2$, the estimator $\bar{x}$, which corresponds to $\hat{\tau}^2 = \infty$, is even admissible which of course cannot happen for any unbounded loss function. This circumstance explains why an estimator $\hat{\tau}^2$ may have a large quadratic risk, while the associated estimator $\hat{\mu}_{\text{plug}}$ in (1.2) has a small variance. That phenomenon is known to happen in the case of the DerSimonian–Laird procedure [6].

The estimator $\bar{x}$ has a constant risk, $R(\bar{x}, \tau^2) \equiv 1$, which raises the question of its R-minimaxity, i.e., if $\inf_\delta \sup_{\tau^2} R(\delta, \tau^2) = 1$. In contrast, for the Graybill–Deal estimator, $R(\delta_{\text{GD}}, \tau^2) = \tau^4[\sum b_j (\tau^2 + t_j^2)^{-1} t_j^{-4}]/[\sum b_j (\tau^2 + t_j^2)^{-1}]$, so that its R-risk, which vanishes when $\tau^2 = 0$, grows quadratically in τ^2. The next result gives a large class of estimators with bounded R-risk improving on $\bar{x}$ when $n > 3$.

Theorem 2.1. *Under notation of Section 2.1, let for $n > 3$, $q = \sum_j q_j y_j^2 + \sum_i (\nu_i - 1) r_i u_i^2$ be a quadratic form with positive coefficients q_j, r_i. If δ has the form (2.5) such that for $q \to \infty$, $w_j \sim \alpha_j / q$, $0 < \alpha_j < \infty$, then*

$$\begin{aligned}\lim_{\tau^2 \to \infty} R(\delta, \tau^2) = 1 - \frac{1}{\sum_j b_j} \sum_j b_j \\ \times \left[2\alpha_j E \frac{z_j^2}{\sum_\ell q_\ell z_\ell^2 + \sum_i r_i \chi^2_{\nu_i - 1}} - \alpha_j^2 E \frac{z_j^2}{(\sum_\ell q_\ell z_\ell^2 + \sum_i r_i \chi^2_{\nu_i - 1})^2}\right] \qquad (2.9) \\ \geq \frac{2}{n-1},\end{aligned}$$

where independent standard normal $z_1, \ldots, z_{p-1}$ are independent of $\chi^2_{\nu_1 - 1}, \ldots, \chi^2_{\nu_p - 1}$. Equal coefficients $q_j = r_i$ (and only they) provide the asymptotically optimal quadratic form. If $q_j =$

$r_i = 1$, *the optimal choice is* $\alpha_j \equiv n-3$. *The sample mean* $\bar{x}$ *is not* R-*minimax*, *any estimator* (2.5) *with weights* (2.8) *improves on it if*

$$0 < \alpha \le 2(n-3)\frac{\min[\min_j q_j^2 t_j^4, \min_{i:\nu_i \ge 2} r_i^2 s_i^4]\sum_j b_j q_j}{\max[\max_j q_j^2 t_j^4, \max_{i:\nu_i \ge 2} r_i^2 s_i^4]\sum_j b_j q_j^2}. \tag{2.10}$$

Theorem 2.1 shows that the traditional weights (2.4) with $\hat{\tau}^2 = q/\alpha$ are not asymptotically optimal unless the quadratic form q coincides (up to a positive factor) with $q^\infty = \sum_j y_j^2 + \sum_i (\nu_i - 1)u_i^2 = \sum_i \nu_i (x_i - \bar{x})^2 + \sum_i (\nu_i - 1)u_i^2$, and $\alpha = n-3$. Only then (2.9) is an equality. Thus, the Hedges estimator for which $\alpha = n-1$ and $R(\delta_{\mathrm{H}}, \tau^2) \sim 2(n-3)^{-1}$, is not asymptotically optimal albeit its performance is the best when τ^2 is large. For the Mandel–Paule estimator from Section 2.1, as well as for the REML, (2.9) also holds with the same quadratic form and the same α. The DerSimonian–Laird estimator is defined by the quadratic form $q^0 = \sum_j y_j^2/t_j^2 + \sum_i (\nu_i - 1)u_i^2/s_i^2$ with $\alpha = \sum_j t_j^{-2} + \sum_i (\nu_i - 1)s_i^{-2}$. Therefore, these three statistics are not optimal for large τ^2 either.

The case when $n = p = 2$ was studied in [14]. Then $\bar{x}$ is admissible (so that it is automatically minimax under R). Any estimator (2.5) has the form (1.2) with some $\hat{\tau}^2 = \hat{\tau}^2(y_1^2)$, and its R-risk grows linearly in τ,

$$R(\delta, \tau^2) \sim \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty \frac{y^2\,\mathrm{d}y}{(\hat{\tau}^2 + s^2)^2}\tau.$$

For $n = p = 3$, as $\tau^2 \to \infty$, $R(\delta, \tau^2) \sim C \log \tau^2$ (see Electronic Supplement). By analogy with the Stein phenomenon, admissibility of the sample mean when $n = 3$ is expected.

2.5. Equal uncertainties and minimax value

When $n > 3$, the minimax value, $\inf_\delta \sup_{\tau^2} R(\delta, \tau^2)$, (which does not exceed one since $R(\bar{x}, \tau^2) \equiv 1$) cannot be smaller than $2(n-1)^{-1}$. Indeed for any estimator δ,

$$\sup_{\tau^2} R(\delta, \tau^2) \ge \limsup_{\tau^2 \to \infty} R(\delta, \tau^2) \ge \frac{2}{n-1}.$$

This fact can be proven by constructing a sequence of proper prior densities for τ^2 such that the corresponding sequence of the Bayes R-risks converges to $2(n-1)^{-1}$.

Thus for large τ^2, the estimators (2.5) with q^∞, $\alpha = n-3$, cannot be improved upon. The most natural of these statistics, say, δ_1 has the form (2.5) with

$$w_j^1 = \min\left(\frac{n-3}{q^\infty}, \frac{1}{t_j^2}\right). \tag{2.11}$$

Another modified Hedges estimator, δ_{mH}, has the form (1.2) with $\hat{\tau}^2 = (n-3)^{-1}[q^\infty - \sum_j t_j^2 - \sum_i (\nu_i - 1)s_i^2]_+$ and also is asymptotically optimal although in general its performance is worse than that of (2.11).

If $\sum_i \nu_i (s_i^2 - s^2)^2 \to 0$, so that all t_j^2 and s_i^2 tend to $s^2 = \sum_i \nu_i s_i^2/n$, $w_j(v) \to w(v)$,

$$R(\delta, \tau^2) \to (\tau^2 + s^2)^2 \int_0^\infty \left[w(v) - \frac{1}{\tau^2 + s^2} \right]^2 \mathrm{d}G_{n+1}\left(\frac{v}{\tau^2 + s^2} \right).$$

Here and further G_k is the distribution function of χ^2-law with k degrees of freedom. Thus if $s_i^2 \approx s^2$, our problem is that of estimation of the reciprocal of the scale parameter $\sigma = \tau^2 + s^2$ under the restriction, $\sigma \geq s^2$. The "data" v in this problem is χ^2-distributed, $v \sim \sigma \chi^2_{n+1}$, and the invariant loss function, $\sigma^2(w - \sigma^{-1})^2$, corresponds to the R-risk. Then the minimax value, $2(n-1)^{-1}$, is the same as in the non-restricted ($s = 0$) parameter case [8]. As in unrestricted scale parameter estimation, the generalized prior, $\mathrm{d}\sigma/\sigma$, $\sigma \geq s^2$, or $\mathrm{d}\tau^2/(\tau^2 + s^2)$, provides a least favorable distribution. See also [9] for more general results.

Thus in meta-analysis problems with s_i^2 exhibiting little variation, the minimax value is expected to stay close to $2(n-1)^{-1}$. Indeed when $w(v) = \min(\alpha v^{-1}, s^{-2})$, $\xi = \alpha s^2/(\tau^2 + s^2)$,

$$R(\delta, \tau^2) \to 1 - \left(1 - \frac{\tau^4}{s^4}\right) G_{n+1}(\xi) - \frac{2\alpha(1 - G_{n-1}(\xi))}{n-1} + \frac{\alpha^2(1 - G_{n-3}(\xi))}{(n-1)(n-3)}. \tag{2.12}$$

The formula (2.12) shows that the estimator (2.11) is minimax unlike δ_{mH} for which $w(v) = \{[v - (n-1)s^2]_+/(n-3) + s^2\}^{-1}$.

The DerSimonian–Laird rule, $w_{\mathrm{DL}}(v) = \{[v - (n-1)s^2]_+/(n-1) + s^2\}^{-1}, \alpha = n-1$, coincides in this situation with the REML and the Hedges estimator. For the proper maximum likelihood estimator of $(\tau^2 + s^2)^{-1}$, $\alpha = n+1$. None of these procedures is minimax which indicates that their good properties in meta-analysis may be attributable to a large number of individual studies (large n) or to lack of interest in high heterogeneity (small τ^2).

Figure 1 shows the graphs of the R-risk in (2.12) when $s^2 = 1$. It suggests that the estimator δ_1 performs quite well for small/medium n's. Indeed this estimator is better than other procedures except for small τ^2 in which case δ_{DL} dominates δ_{mH} (at the price of higher risk for larger values of τ^2).

3. Example: $p = 2$

When there are only two different values s_1^2 and s_2^2 with multiplicities ν_1 and ν_2, $n = \nu_1 + \nu_2 > 3$,

$$t_1^2 = t^2 = \frac{\nu_2 s_1^2 + \nu_1 s_2^2}{n},$$

$$A_{11} = -A_{21} = \frac{\nu_1 \nu_2 (s_1^2 - s_2^2)}{n^2},$$

$$b_1 = b = \frac{\nu_1 \nu_2 (s_1^2 - s_2^2)^2}{n^3},$$

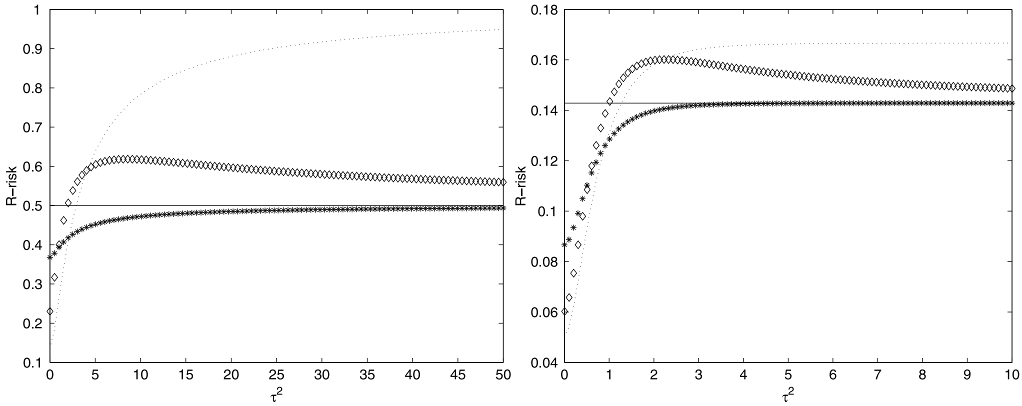

Figure 1. Plots of R-risks of estimators corresponding to δ_{DL} (dash-dotted line), δ_{mH} (line marked by diamonds) and δ_1 (line marked by $*$) when $n=5$ (left panel), and of the same risks when $n=15$ (right panel). The straight line depicts the minimax value $2(n-1)^{-1}$.

and if $s_1^2 > s_2^2$,

$$y_1 = y = \frac{\sqrt{\nu_1\nu_2}(x_1-x_2)}{\sqrt{n}}.$$

Any estimator (2.5) has the form (1.2) for some $\hat{\tau}^2$,

$$\delta = \frac{x_1+x_2}{2} - \frac{\nu_1\nu_2(s_1^2-s_2^2)(x_1-x_2)}{n^2(\hat{\tau}^2+t^2)}.$$

For δ_1, $\hat{\tau}^2 = [(n-3)^{-1}q^\infty - t^2]_+$, $q^\infty = y^2 + (\nu_1-1)u_1^2 + (\nu_2-1)u_2^2$. The modified Hedges estimator δ_{mH} with $\hat{\tau}^2 = (n-3)^{-1}[q^\infty - t^2 - (\nu_1-1)s_1^2 - (\nu_2-1)s_2^2]_+$ typically has its R-risk at $\tau^2=0$ larger than that of δ_1. (The exact condition for δ_{mH} to have a smaller R-risk at $\tau^2=0$ than δ_1 is: $n \ge 5$, and if $\nu_1 \le \nu_2$, then $\nu_1 \le n(n-4)/(2n-5)$, $[n(n-1)-\nu_1(2n-5)]s_2^2 \ge [n(n-4)-\nu_1(2n-5)]s_1^2$.)

The R-risk of δ_1 at $\tau^2=0$ can be larger than $2/(n-1)$. Indeed

$$R(\delta_1,0) = \int_{(n-3)}^{\infty}\left(\frac{n-3}{v}-1\right)^2 \mathrm{d}F\big(t^2 v\big),$$

where $F(v)$ is the distribution function of $t^2\chi_3^2 + s_1^2\chi_{\nu_1-1}^2 + s_2^2\chi_{\nu_2-1}^2$. With $a = t^2/[t^6 s_1^{2(\nu_1-1)} \times s_2^{2(\nu_2-1)}]^{1/(n+1)}$, according to the Okamoto inequality [10], $F(t^2v) \le G_{n+1}(av)$. Thus, since $[(n-3)v^{-1}-1]^2$ is an increasing function of v, $v \ge n-3$,

$$\begin{aligned} R(\delta_1,0) &> \int_{(n-3)}^{\infty}\left(\frac{n-3}{v}-1\right)^2 \mathrm{d}G_{n+1}(av) \\ &= 1 - G_{n+1}\big(a(n-3)\big) \end{aligned}$$

$$- \frac{2(n-3)a[1-G_{n-1}(a(n-3))]}{n-1} + \frac{(n-3)a^2[1-G_{n-3}(a(n-3))]}{n-1}.$$

This inequality shows that $R(\delta_1, 0) \geq 2(n-1)^{-1}$, if $a < a_0 < 1$, where $a_0 = a_0(n)$ is monotonically increasing to 1 in $n, a_0(4) = 0.637\ldots, a_0(10) = 0.798\ldots$ For small a, δ_1 cannot have its risk at the origin smaller than $2(n-1)^{-1}$. For example, when $n = 4, \nu_1 = 1, \nu_2 = 3$, $R(\delta_1, 0) \leq 2(n-1)^{-1}$ if and only if $s_1^2/s_2^2 \geq 0.173\ldots$, i.e., iff $a \geq 0.679\ldots$.

The DerSimonian–Laird estimator δ_{DL} with $\hat{\tau}_{\mathrm{DL}}^2 = (q^0 - n + 1)_+/(1/t^2 + (\nu_1 - 1)/s_1^2 + (\nu_2 - 1)/s_2^2)$, $q^0 = y^2/t^2 + (\nu_1 - 1)u_1^2/s_1^2 + (\nu_2 - 1)u_2^2/s_2^2$, has its R-risk at $\tau^2 = 0$ of the form

$$R(\delta_{\mathrm{DL}}, 0) = \int_{(n-1)}^{\infty} \left[\frac{1}{1 + \kappa^{-1}(v - n + 1)} - 1 \right]^2 \mathrm{d}G_{n+1}(v)$$

with $\kappa = 1 + t^2[(\nu_1 - 1)/s_1^2 + (\nu_2 - 1)/s_2^2]$.

For the estimator δ_0 defined by (2.5),

$$w_j^0 = \min\left(\frac{n-1}{q^0}, 1\right)\frac{1}{t_j^2},$$

so that $\hat{\tau}^2 = t^2[q^0/(n-1) - 1]_+$. Its risk at $\tau^2 = 0$,

$$R(\delta_0, 0) = \int_{(n-1)}^{\infty} \left(\frac{n-1}{v} - 1\right)^2 \mathrm{d}G_{n+1}(v),$$

is always smaller than that of δ_1.

But δ_0 is also competitive against δ_{DL}. Indeed $R(\delta_0, 0) < R(\delta_{\mathrm{DL}}, 0)$ if and only if $\kappa < n - 1$, that is, iff

$$\frac{(\nu_2 - 1)s_1^2}{\nu_1 s_2^2} + \frac{(\nu_1 - 1)s_2^2}{\nu_2 s_1^2} < \frac{\nu_1\nu_2(n-2)}{n} - 2 + \frac{n}{\nu_1\nu_2}.$$

Thus provided that $\nu_1, \nu_2 > 1, s_1^2/s_2^2 \approx \sqrt{(\nu_1 - 1)\nu_1/[(\nu_2 - 1)\nu_2]}$, δ_0 improves upon δ_{DL} for small τ^2. If, say, $\nu_1 = 1$, this domination means that $s_1^2 < s_2^2$. Thus, when one study reports a smaller uncertainty than all other studies whose standard errors are approximately equal, δ_0 improves upon the DerSimonian–Laird estimator for small τ^2.

However, there is no uniform domination as the condition, $\kappa < n - 1$, means that for large τ^2, $R(\delta_0, \tau^2) > R(\delta_{\mathrm{DL}}, \tau^2)$.

4. Conclusions

Author's attempt was to give a perspective of a meta-analysis setting from the point of view of the statistical decision theory. Although concepts like admissibility or minimaxity have not so far generated much interest among meta-analysts, there is a realization that different desirable qualities of the employed procedures call for different loss functions. The quadratic loss for the mean effect estimators from a wide class leads in a natural way to the R-risk suggested and studied in this paper. This risk strongly recommends against the use of the sample mean as a consensus estimate which still happens in some collaborative studies.

Moreover, the R-risk questions well recognized excellent properties of the DerSimonian–Laird estimator δ_{DL} in the situation when s_i are almost equal, or when one study claims a high precision while all other studies report larger uncertainties which are about the same. The unsatisfactory performance of the Graybill–Deal estimator δ_{GD} is well known in the latter case. It is of interest that δ_0 improves on the DerSimonian–Laird estimator for moderate/small τ^2. Inference on the overall effect can be obtained before the heterogeneity variance is estimated, but even in the simplest cases considered here there is no unique rule dominating all others.

This paper is dedicated to the memory of George Casella who was always interested in implications of the statistical decision theory results to practical estimation problems [7].

Appendix

A.1. Partial fraction decomposition and weighted means

Let e denote unit coordinates vector whose dimension is clear from the context, and put $J = \mathrm{diag}(\nu_1, \ldots, \nu_p)$, $S = \mathrm{diag}(s_1^2/\nu_1, \ldots, s_p^2/\nu_p)$. In the used here notation of Section 2.1 the vector x has the diagonal covariance matrix, $C = \tau^2 J^{-1} + S$.

Lemma A.1. *For any v different from $-t_j^2$, $j = 1, \ldots, p-1$, and for any $i = 1, \ldots, p$,*

$$\frac{\nu_i}{v + s_i^2}\left[\sum_k \frac{\nu_k}{v + s_k^2}\right]^{-1} = \frac{\nu_i}{n} - \sum_j \frac{A_{ij}}{v + t_j^2}, \tag{A.1}$$

where

$$A_{ij} = \frac{\nu_i M(-t_j^2)}{Q'(-t_j^2)(t_j^2 - s_i^2)}. \tag{A.2}$$

For any j, $j = 1, \ldots, p-1$,

$$b_j = \sum_i \frac{A_{ij}^2}{\nu_i} = \frac{1}{t_j^2} \sum_i \frac{s_i^2 A_{ij}^2}{\nu_i} = -\frac{M(-t_j^2)}{Q'(-t_j^2)}. \tag{A.3}$$

If the $p \times (p-1)$ matrix A is determined by its elements A_{ij} in (A.2), then

$$A^{\mathrm{T}} e = 0, \tag{A.4}$$

and

$$Ae = \frac{1}{n} J\big(S - s^2 I\big)e, \qquad s^2 = \frac{\sum_i \nu_i s_i^2}{n}. \tag{A.5}$$

The matrices $A^{\mathrm{T}} J^{-1} A = \operatorname{diag}(b_1, \dots, b_{p-1})$ *and* $A^{\mathrm{T}} S A = \operatorname{diag}(b_1 t_1^2, \dots, b_{p-1} t_{p-1}^2)$ *are diagonal*, *and*

$$A\big(A^{\mathrm{T}} J^{-1} A\big)^{-1} A^{\mathrm{T}} = J - \frac{J e e^{\mathrm{T}} J}{e^{\mathrm{T}} J e}. \tag{A.6}$$

With $\rho = ((\tau^2 + t_1^2)^{-1}, \dots, (\tau^2 + t_{p-1}^2)^{-1})^{\mathrm{T}}$,

$$A^{\mathrm{T}} J^{-1} C^{-1} J^{-1} A = \operatorname{diag}\big(A^{\mathrm{T}} J^{-1} A\rho\big) + \left(\sum_i \frac{\nu_i}{\tau^2 + s_i^2}\right)\big(A^{\mathrm{T}} J^{-1} A\rho\big)\big(A^{\mathrm{T}} J^{-1} A\rho\big)^{\mathrm{T}}, \tag{A.7}$$

and

$$\sum_i \frac{\nu_i (x_i - \hat{\mu}_{\mathrm{opt}})^2}{\tau^2 + s_i^2} = \sum_j \frac{y_j^2}{\tau^2 + t_j^2}. \tag{A.8}$$

Proof. By the definition of the polynomial Q in Section 2.1,

$$\begin{aligned} \frac{\nu_i}{n} - \frac{\nu_i}{v + s_i^2}\left[\sum_k \frac{\nu_k}{v + s_k^2}\right]^{-1} &= \frac{\nu_i}{n} - \frac{\nu_i P(v)}{(v + s_i^2) P'(v)} \\ &= \frac{\nu_i}{n} - \frac{\nu_i M(v)}{(v + s_i^2) Q(v)} = \frac{\nu_i [\prod_j (v + t_j^2) - \prod_{k \neq i} (v + s_k^2)]}{Q(v)} \end{aligned}$$

with the right-hand side of this identity being the ratio of two polynomials of degree $p-2$ and $p-1$, respectively. The formulas (A.1) and (A.2) easily follow from the classical result on partial fraction decomposition for such ratios.

For any fixed j,

$$\sum_i \frac{\nu_i}{s_i^2 - t_j^2} = \frac{P'(-t_j^2)}{P(-t_j^2)} = 0,$$

so that (A.4) follows from (A.2),

$$\sum_i A_{ij} = \frac{M(-t_j^2)}{Q'(-t_j^2)} \sum_i \frac{\nu_i}{s_i^2 - t_j^2} = 0.$$

By equating coefficients at v^{n+p-2} of QP and MP', one gets $\sum t_j^2 = \sum s_i^2 - s^2$. The comparison of coefficients at v^{p-2} of two equal polynomials, $n \sum_j A_{ij} \prod_{\ell \neq j} (v + t_\ell^2)$ and

$\nu_i[\prod_j(v+t_j^2)-\prod_{k\neq i}(v+s_k^2)]$, shows that

$$n\sum_j A_{ij}=\nu_i\left(\sum_j t_j^2-\sum_{k\neq i}s_k^2\right)=\nu_i\left(s_i^2-\sum_k \nu_k s_k^2/n\right),$$

which implies (A.5).

For any j

$$\sum_i \frac{\nu_i}{(s_i^2-t_j^2)^2}=-\frac{Q'(-t_j^2)}{M(-t_j^2)},$$

and

$$\sum_i \frac{\nu_i s_i^2}{(s_i^2-t_j^2)^2}=t_j^2\sum_i \frac{\nu_i}{(s_i^2-t_j^2)^2}=-t_j^2\frac{Q'(-t_j^2)}{M(-t_j^2)},$$

so that (A.3) is established by substituting (A.2) for A_{ij}.

For any different j and ℓ

$$0=\sum_i \frac{\nu_i}{s_i^2-t_j^2}-\sum_i \frac{\nu_i}{s_i^2-t_\ell^2}=\left(t_j^2-t_\ell^2\right)\sum_i \frac{\nu_i}{(s_i^2-t_j^2)(s_i^2-t_\ell^2)},$$

which implies that $\sum_i \nu_i^{-1}A_{ij}A_{i\ell}=0$, or that $A^{\mathrm{T}}J^{-1}A$ is a diagonal matrix.

This argument also shows that

$$\sum_i \frac{s_i^2 A_{ij}A_{i\ell}}{\nu_i}=0,$$

as

$$\sum_i \frac{\nu_i s_i^2}{(s_i^2-t_j^2)(s_i^2-t_\ell^2)}=t_j^2\sum_i \frac{\nu_i}{(s_i^2-t_j^2)(s_i^2-t_\ell^2)}=0.$$

To prove (A.6), observe that for $i\neq k$, the (i,k)th element of the matrix $A(A^{\mathrm{T}}J^{-1}A)^{-1}A^{\mathrm{T}}$ has the form,

$$\begin{aligned}\sum_j \frac{A_{ij}A_{kj}}{b_j}&=\nu_i\nu_k\sum_j \frac{b_j}{(t_j^2-s_i^2)(t_j^2-s_k^2)}\\&=\frac{\nu_i\nu_k}{s_i^2-s_k^2}\left[\sum_j \frac{b_j}{t_j^2-s_i^2}-\sum_j \frac{b_j}{t_j^2-s_k^2}\right]=-\frac{\nu_i\nu_k}{n}.\end{aligned}$$

Here we used the facts that $A_{ij}=-\nu_i b_j/(t_j^2-s_i^2)$, and $\sum_j b_j(t_j^2-s_i^2)^{-1}=(s^2-s_i^2)/n$.

To determine the diagonal elements of $A(A^{\mathrm{T}}J^{-1}A)^{-1}A^{\mathrm{T}}$, observe that according to the definition of Q, $Q(-s_i^2)/M'(-s_i^2) = \nu_i$. Therefore for any i,

$$\sum_j \frac{A_{ij}^2}{b_j} = -\nu_i \sum_j \frac{A_{ij}}{t_j^2 - s_i^2} = \nu_i^2 \left[\frac{M'(-s_i^2)}{Q(-s_i^2)} - \frac{1}{n} \right] = \nu_i - \frac{\nu_i^2}{n}.$$

Thus, (A.6) holds.

Because of (A.2),

$$A^{\mathrm{T}}J^{-1}C^{-1}e = -\left(\sum_i \frac{\nu_i}{\tau^2 + s_i^2} \right) \left(A^{\mathrm{T}}J^{-1}A\right)\rho. \tag{A.9}$$

To prove (A.7) for fixed j, ℓ, multiply (A.1) by A_{ij}, $A_{i\ell}$, divide by ν_i^2, and sum up over i to get the following expression for the (j,ℓ)th element of the matrix $A^{\mathrm{T}}J^{-1}C^{-1}J^{-1}A$,

$$\left(\sum_i \frac{\nu_i}{\tau^2 + s_i^2} \right) \left[\frac{\delta_{j\ell} b_j}{n} - \sum_{i,m} \frac{A_{ij} A_{i\ell} A_{im}}{\nu_i^2 (\tau^2 + t_m^2)} \right],$$

where δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1$, if $i = j$; $= 0$ otherwise). It is easy to see that $\sum_i A_{ij} A_{i\ell} A_{im} \nu_i^{-2} = 0$, unless there are at least two equal indices among j, ℓ, m. When all three of these indices coincide,

$$\begin{aligned}
\sum_i \frac{A_{ij}^3}{\nu_i^2} &= -\frac{M^3(-t_j^2)}{[Q'(-t_j^2)]^3} \sum_i \frac{\nu_i}{(s_i^2 - t_j^2)^3} \\
&= -\frac{M(-t_j^2)[Q''(-t_j^2)M(-t_j^2) - 2Q'(-t_j^2)M'(-t_j^2)]}{2[Q'(-t_j^2)]^3} \\
&= \frac{b_j[Q''(-t_j^2)M(-t_j^2) - 2Q'(-t_j^2)M'(-t_j^2)]}{2[Q'(-t_j^2)]^2} = -\frac{b_j^2 Q''(-t_j^2) + 2b_j M'(-t_j^2)}{2Q'(-t_j^2)}.
\end{aligned}$$

If, say, $m = j \neq \ell$,

$$\begin{aligned}
\sum_i \frac{A_{ij}^2 A_{i\ell}}{\nu_i^2} &= -\frac{M^2(-t_j^2)M(-t_\ell^2)}{[Q'(-t_j^2)]^2 Q'(-t_\ell^2)} \sum_i \frac{\nu_i}{(s_i^2 - t_j^2)^2 (s_i^2 - t_\ell^2)} \\
&= -\frac{M^2(-t_j^2)M(-t_\ell^2)}{[Q'(-t_j^2)]^2 Q'(-t_\ell^2)(t_j^2 - t_\ell^2)} \sum_i \frac{\nu_i}{(s_i^2 - t_j^2)^2} \\
&= \frac{M(-t_j^2)M(-t_\ell^2)}{Q'(-t_j^2)Q'(-t_\ell^2)(t_j^2 - t_\ell^2)} = \frac{b_j b_\ell}{t_j^2 - t_\ell^2}.
\end{aligned}$$

The last formula shows that off-diagonal elements of $A^{\mathrm{T}}J^{-1}C^{-1}J^{-1}A$, for $j \neq \ell$ have the form

$$-\left(\sum_i \frac{\nu_i}{\tau^2+s_i^2}\right)\left[\frac{b_j b_\ell}{(t_j^2-t_\ell^2)(\tau^2+t_j^2)}+\frac{b_j b_\ell}{(t_\ell^2-t_j^2)(\tau^2+t_\ell^2)}\right]$$
$$=\left(\sum_i \frac{\nu_i}{\tau^2+s_i^2}\right)\frac{b_j b_\ell}{(\tau^2+t_j^2)(\tau^2+t_\ell^2)},$$

that is, (A.7) holds for the off-diagonal elements.

We demonstrate now the equality of the diagonal elements of matrices in (A.7). These elements for the matrix $A^{\mathrm{T}}J^{-1}C^{-1}J^{-1}A$ are

$$b_j\left(\sum_i \frac{\nu_i}{\tau^2+s_i^2}\right)\left[\frac{1}{n}-\sum_{\ell\neq j}\frac{b_\ell}{(t_j^2-t_\ell^2)(\tau^2+t_\ell^2)}+\frac{b_j Q''(-t_j^2)+2M'(-t_j^2)}{2Q'(-t_j^2)(\tau^2+t_j^2)}\right].$$

Define the polynomial Q_j by the formula,

$$\frac{Q_j(\tau^2)}{Q(\tau^2)}=\sum_{\ell\neq j}\frac{b_\ell}{(t_\ell^2-t_j^2)(\tau^2+t_\ell^2)}+\frac{b_j Q''(-t_j^2)+2M'(-t_j^2)}{2Q'(-t_j^2)(\tau^2+t_j^2)}.$$

Then the degree of Q_j is $p-2$, and this polynomial is determined by its values at $-t_\ell^2, \ell = 1,\dots,p-1$: $Q_j(-t_\ell^2)=b_\ell Q'(-t_\ell^2)/(t_\ell^2-t_j^2)=-M(-t_\ell^2)/(t_\ell^2-t_j^2), \ell\neq j$, and $Q_j(-t_j^2)=b_j Q''(-t_j^2)/2+M'(-t_j^2)$. It follows that

$$Q_j(\tau^2)=\frac{M(\tau^2)}{\tau^2+t_j^2}+\frac{b_j Q(\tau^2)}{(\tau^2+t_j^2)^2}-\frac{Q(\tau^2)}{n}.$$

Indeed, the polynomial in the right-hand side has degree $p-2$. Since

$$\lim_{\tau^2\to -t_j^2}\frac{M(\tau^2)(\tau^2+t_j^2)+b_j Q(\tau^2)}{(\tau^2+t_j^2)^2}=M'(-t_j^2)+\frac{b_j Q''(-t_j^2)}{2},$$

it assumes the same values as Q_j at $-t_\ell^2, \ell=1,\dots,p-1$, which establishes (A.7).

Because of (A.6) and (A.2), $x-\hat{\mu}_{\mathrm{opt}}e = x-\bar{x}e+(\bar{x}-\hat{\mu}_{\mathrm{opt}})e=[I-ee^{\mathrm{T}}J/(e^{\mathrm{T}}Je)]x+(\rho^{\mathrm{T}}A^{\mathrm{T}}x)e$. Thus the quadratic form in the left-hand side of (A.8) can be written as

$$\begin{aligned}
&\left[J^{-1}A\left(A^{\mathrm{T}}J^{-1}A\right)^{-1}A^{\mathrm{T}}x+e\rho^{\mathrm{T}}A^{\mathrm{T}}x\right]^{\mathrm{T}}C^{-1}\left[J^{-1}A\left(A^{\mathrm{T}}J^{-1}A\right)^{-1}A^{\mathrm{T}}x+e\rho^{\mathrm{T}}A^{\mathrm{T}}x\right]\\
&\quad=y^{\mathrm{T}}\left[\left(A^{\mathrm{T}}J^{-1}A\right)^{-1/2}A^{\mathrm{T}}J^{-1}+\left(A^{\mathrm{T}}J^{-1}A\right)^{1/2}\rho e^{\mathrm{T}}\right]\\
&\qquad\times C^{-1}\left[J^{-1}A\left(A^{\mathrm{T}}J^{-1}A\right)^{-1/2}+e\rho^{\mathrm{T}}\left(A^{\mathrm{T}}J^{-1}A\right)^{1/2}\right]y\\
&\quad=y^{\mathrm{T}}\operatorname{diag}(\rho)y,
\end{aligned}$$

where the second equality follows from (A.7) and (A.9). □

The following important representation for $\hat{\mu}_{\mathrm{opt}}$

$$\hat{\mu}_{\mathrm{opt}} = \bar{x} - \sum_{i,j} \frac{A_{ij} x_i}{\tau^2 + t_j^2} = \bar{x} - \sum_j \frac{\sqrt{b_j}\, y_j}{\tau^2 + t_j^2} \tag{A.10}$$

is a consequence of Lemma A.1. Here $y_j = \sum_i A_{ij} x_i / \sqrt{b_j}$ are independent normal, zero mean random variables with the variances $\tau^2 + t_j^2$. Indeed the normal random vector $y = (A^{\mathrm{T}} J^{-1} A)^{-1/2} A^{\mathrm{T}} x$ has the covariance matrix $(A^{\mathrm{T}} J^{-1} A)^{-1/2} A^{\mathrm{T}} C A (A^{\mathrm{T}} J^{-1} A)^{-1/2} = \mathrm{diag}(\tau^2 + t_1^2, \ldots, \tau^2 + t_{p-1}^2)$. Since $E y_j (\hat{\mu}_{\mathrm{opt}} - \mu) = 0$, $\hat{\mu}_{\mathrm{opt}}$ and y_j are independent implying independence of $\hat{\mu}_{\mathrm{opt}}$ and $\delta - \hat{\mu}_{\mathrm{opt}}$ in Section 2.3.

The coefficients A_{ij} provide a simple expression for $\mathrm{Var}(\bar{x}) - \mathrm{Var}(\hat{\mu}_{\mathrm{opt}})$. Indeed, by dividing (A.1) by ν_i and multiplying it by $A_{i\ell}$, one gets after summing up over all i and ℓ and using (A.4), (A.5),

$$\begin{aligned}
\frac{P(\tau^2)}{P'(\tau^2)} \sum_{i,\ell} \frac{A_{i\ell}}{\tau^2 + s_i^2} &= \frac{M(\tau^2)}{n Q(\tau^2)} \sum_i \frac{\nu_i (s_i^2 - s^2)}{\tau^2 + s_i^2} \\
&= \left[\sum_i \frac{\nu_i}{\tau^2 + s_i^2} \right]^{-1} - \frac{\tau^2 + s^2}{n} = - \sum_{i,j} \frac{A_{ij}^2}{\nu_i (\tau^2 + t_j^2)}.
\end{aligned}$$

This formula can be written in the form,

$$\sum_j \frac{b_j}{\tau^2 + t_j^2} = \mathrm{Var}(\bar{x}) - \mathrm{Var}(\hat{\mu}_{\mathrm{opt}}) = \frac{\sum_i \nu_i (s^2 - s_i^2) \prod_{k \neq i} (\tau^2 + s_k^2)}{n Q(\tau^2)}, \tag{A.11}$$

which provides the representation of the left-hand side of (A.11) as a ratio of two polynomials of degree $p - 2$ and $p - 1$, respectively and which allows numerical evaluation of b's without calculating A_{ij}.

Supplementary Material

Restricted likelihood representation and decision-theoretic aspects of meta-analysis: Electronic supplement (DOI: 10.3150/13-BEJ547SUPP; .pdf). The supplement contains the proof of Theorem 2.1.

References

[1] Borenstein, M., Hedges, L., Higgins, J. and Rothstein, H. (2009). *Introduction to Meta-Analysis*. New York: Wiley.

[2] Brown, L.D. (1988). The differential inequality of a statistical estimation problem. In *Statistical Decision Theory and Related Topics*, *IV*, *Vol.* 1 (*West Lafayette*, *Ind.*, 1986) (S.S. Gupta and J.O. Berger, eds.) 299–324. New York: Springer. MR0927109

[3] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Control. Clin. Trials* **7** 177–188.
[4] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors – an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597
[5] Harville, D.A. (1985). Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80** 132–138. MR0786599
[6] Jackson, D., Bowden, J. and Baker, R. (2010). How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *J. Statist. Plann. Inference* **140** 961–970. MR2574658
[7] Maatta, J.M. and Casella, G. (1990). Developments in decision-theoretic variance estimation. *Statist. Sci.* **5** 90–120. With comments and a rejoinder by the authors. MR1054858
[8] Marchand, É. and Strawderman, W.E. (2005). On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint. *J. Statist. Plann. Inference* **134** 90–101. MR2146087
[9] Marchand, É. and Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli* **18** 635–643. MR2922464
[10] Marshall, A.W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. *Mathematics in Science and Engineering* **143**. New York: Academic Press. MR0552278
[11] Morris, C.N. and Normand, S.L. (1992). Hierarchical models for combining information and for meta-analyses. In *Bayesian Statistics, Vol.* 4 (*Peñíscola*, 1991) (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.) 321–344. New York: Oxford Univ. Press. MR1380284
[12] Paule, R.C. and Mandel, J. (1982). Consensus values and weighting factors. *J. Res. Natl. Bur. Stand.* **87** 377–385.
[13] Rukhin, A.L. (1995). Admissibility: Survey of a concept in progress. *Int. Stat. Rev.* **63** 95–115.
[14] Rukhin, A.L. (2012). Estimating common mean and heterogeneity variance in two study case meta-analysis. *Statist. Probab. Lett.* **82** 1318–1325. MR2929781
[15] Rukhin, A.L. (2013). Estimating heterogeneity variance in meta-analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 451–469. MR3065475
[16] Rukhin, A.L. (2014). Supplement to "Restricted likelihood representation and decision-theoretic aspects of meta-analysis." DOI:10.3150/13-BEJ543SUPP.
[17] Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. New York: Wiley. MR1190470

Received January 2013 and revised April 2013