Ann Inst Stat Math (2016) 68:571–588
DOI 10.1007/s10463-015-0505-1

Decision-theoretic issues in heterogeneity variance estimation

Andrew L. Rukhin

Received: 11 June 2014 / Revised: 8 November 2014 / Published online: 8 February 2015

Abstract Motivated by a heteroscedastic random effects setting of meta-analysis, a general model for the between-study variance is studied from the decision-theoretic point of view. This model leads to estimation of a linear in the variance reciprocals, random function or to simultaneous inference on curve-confined natural parameters of independent heterogeneous χ^2-random variables with given degrees of freedom. A form of the Stein phenomenon for the suggested loss functions is noted; the exact minimax value is determined, and minimax estimators are derived.

Keywords Bayes estimator · Loss function · Meta-analysis · Minimax value · Random effects model · Stein phenomenon

1 Introduction: random effects meta-analysis model and its linear transform

In this work the observed data vector $y = (y_1, \ldots, y_n)$ is supposed to consist of realizations of independent heterogeneous χ^2-random variables with given degrees of freedom ν_i,

$$y_i \sim (\tau^2 + t_i^2)\chi^2_{\nu_i}, \quad i = 1, \ldots, n. \tag{1}$$

The unknown parameter τ^2, $\tau^2 \geq 0$, has the meaning of the heterogeneity variance in the following meta-analysis formulation, while the distinct positive constants t_i^2, $i = 1, \ldots, n$, are supposed to be given. In the increasingly popular heteroscedastic random effects model of meta-analysis the treatment effect estimators x_k have the form

A. L. Rukhin (✉)
Statistical Engineering Division, National Institute of Standards and Technology, 100 Bureau Dr. Gaithersburg, MD 20899, USA
e-mail: andrew.rukhin@nist.gov

$$x_k = \mu + \lambda_k + \epsilon_k, \quad k = 1, \ldots, K. \tag{2}$$

Here μ is the unknown common mean (treatment effect), λ_k is zero mean between-study random effect with the variance τ^2, and ϵ_k represents the independent measurement error of the kth study with variance σ_k^2. By assumed independence the variance of x_k is $\tau^2 + \sigma_k^2$. In practice σ_k is often treated as a given constant, s_k, which is the reported standard error or the uncertainty provided by the kth study, and so does this work.

If τ^2 is known, the best unbiased estimator of μ is the weighted means statistic,

$$\tilde{\mu} = \sum_k \frac{x_k}{\tau^2 + s_k^2} \left(\sum_k \frac{1}{\tau^2 + s_k^2} \right)^{-1}. \tag{3}$$

Traditionally to estimate the principal parameter μ, one uses a plug-in version of (3), say γ, where the unknown τ^2 is replaced by an estimate $\tilde{\tau}^2$. All such estimators γ are unbiased if $\tilde{\tau}^2$ is a shift invariant function of $x_1, \ldots, x_K$. Then $\gamma = \sum_k \omega_k x_k$ with positive normalized weights ω_k, $\sum \omega_k = 1$, and the variance of γ (which does not depend on μ) can be decomposed in the sum of two terms,

$$\mathrm{Var}(\gamma) = \mathrm{Var}(\tilde{\mu}) + E(\gamma - \tilde{\mu})^2 = \left[\sum_k (\tau^2 + s_k^2)^{-1} \right]^{-1} + E(\gamma - \tilde{\mu})^2. \tag{4}$$

Thus this variance is completely determined by $E(\gamma - \tilde{\mu})^2$.

The model (2) is related to the setting (1) via a linear transformation (Rukhin 2014a) which reduces the data $\{x_k\}$ to independent normal statistics $z_j \sim N(0, \tau^2 + t_j^2)$, $j = 1, \ldots, p-1$. Here p is the number of distinct s's whose values determine positive t's. For any τ^2, $\tilde{\mu} = \bar{x} - \sum_j \sqrt{b_j} z_j (\tau^2 + t_j^2)^{-1}$, where b_j are positive constants also determined by $s_1, \ldots, s_p$, and $\bar{x} = \sum x_k / K$ is the sample mean. Any γ as above can be written in the form, $\gamma = \bar{x} - \sum_j \sqrt{b_j} \delta_j z_j$, where $\delta_j = (\tilde{\tau}^2 + t_j^2)^{-1}$ can be interpreted as an approximation to $(\tau^2 + t_j^2)^{-1}$. In particular $\delta_j \equiv 0$ for $\gamma = \bar{x}$. The quadratic discrepancy between γ and $\tilde{\mu}$ becomes $(\gamma - \tilde{\mu})^2 = \left[\sum_j \sqrt{b_j} z_j (\delta_j - (\tau^2 + t_j^2)^{-1}) \right]^2$, suggesting a new loss function $\sum_j b_j y_j (\delta_j - (\tau^2 + t_j^2)^{-1})^2$. Here $y_j = z_j^2 \sim (\tau^2 + t_j^2)\chi_1^2$, $j = 1, \ldots, p-1$ and δ's does not have to be of the form $\delta_j = (\tilde{\tau}^2 + t_j^2)^{-1}$.

In the general setting (1) this argument motivates the definition of the loss,

$$L(\delta, \tau^2) = \sum_i \frac{b_i y_i}{\nu_i} \left(\delta_i - \frac{1}{\tau^2 + t_i^2} \right)^2, \tag{5}$$

for fixed non-negative constants b's. Whereas there are n scaled χ^2-variables, $y_1, \ldots, y_n$, the number m of strictly positive b's in (5) is allowed to be smaller so that possibly a y_i (containing some information about the unknown τ^2) is present, but $b_i = 0$ (as it happens in the meta-analysis setting if the multiplicity of s_i exceeds one).

Even if $b_i = 0$, the corresponding y_i can be profitably used to decrease the risk as Theorem 1 in Sect. 3 shows.

The model (1) was considered by Efron and Morris (1973) sec. 8 in an application of the empirical Bayes approach to multivariate normal mean estimation. Welham and Thompson (1997) use a similar setting when studying the tests based on the restricted likelihood in homoscedastic mixed effects model. The joint distribution of y's forms a curved exponential family whose natural parameter consists of $(\tau^2 + t_i^2)^{-1}, i = 1, \ldots, n$. An unbiased estimator of $\tau^2 + t_i^2, i = 1, \ldots, n$, is y_i/ν_i, whereas an unbiased estimator of $(\tau^2 + t_i^2)^{-1}$ does not exist unless $\nu_i > 2$.

It is shown in Sect. 2 that under the loss L the estimator $\delta^0 = 0$ exhibits the Stein-type phenomenon being inadmissible when $\sum_i \nu_i > 2$. As was mentioned, in meta-analysis model (2) δ^0 corresponds to the sample mean $\bar{x}$, which therefore is inadmissible when the number of studies exceeds three. Section 2 also suggests some seemingly reasonable procedures, namely those which cannot be uniformly improved in the sense of the differential inequality considered there. The normalized risk function is defined in Sect. 3. where the case of large τ^2 is studied. It is shown there that δ^0 is not minimax.

The main contributions of this work are the surprisingly simple expression for the minimax value and the form of a minimax estimator. These results obtained in Sect. 5 are motivated by the case of approximately equal uncertainties considered in Sect. 4. The relationship between our estimation problem and that of positive normal mean as well as the shape of the risk functions are discussed there. The main technical tool to establish the minimax value is the multivariate hypergeometric functions (the Dirichlet averages) although the proof as such is based on a very classical approximation by proper Bayes procedures.

2 Differential inequality and Stein phenomenon

The loss function (5) relates our setting to the differential inequality of a statistical estimation problem. For notational convenience we assume here that $b_1 > 0, \ldots, b_m > 0, b_{m+1} = \cdots = b_n = 0, m \le n$.

Proposition 1 *Let $\delta = (\delta_1, \ldots, \delta_m)$ be a piecewise differentiable estimator of $(\tau^2 + t_1^2, \ldots, \tau^2 + t_m^2)$.*

Then for the loss L defined by (5),

$$EL(\delta, \tau^2) - \sum_j \frac{b_j}{\tau^2 + t_j^2} = \sum_j \frac{b_j}{\nu_j} E y_j \left(\delta_j^2 - \frac{2\delta_j}{\tau^2 + t_j^2} \right)$$

$$= E \sum_j \frac{b_j y_j}{\nu_j} \left(\delta_j^2 - \frac{2\nu_j \delta_j}{y_j} - 4 \frac{\partial}{\partial y_j} \delta_j \right).$$

When $M = \sum_{j=1}^m \nu_j > 2$, $\delta^0 \equiv 0$ is an inadmissible estimator under this loss function.

We do not give the proof as it is based on the familiar integration by parts formula,

$$\frac{1}{\tau^2+t_j^2} E y_j h = E\left(\nu_j h + 2y_j \frac{\partial h}{\partial y_j}\right), \quad y_j \sim (\tau^2+t_j^2)\chi^2_{\nu_j}.$$

Under the loss L, $\delta = (\delta_1, \ldots, \delta_m)$ is better than $\xi = (\xi_1, \ldots, \xi_m)$ provided that for all values $y_1, \ldots, y_n$,

$$\sum_j \frac{b_j y_j}{\nu_j}\left(\delta_j^2 - \frac{2\nu_j\delta_j}{y_j} - 4\frac{\partial}{\partial y_j}\delta_j\right) \le \sum_j \frac{b_j y_j}{\nu_j}\left(\xi_j^2 - \frac{2\nu_j\xi_j}{y_j} - 4\frac{\partial}{\partial y_j}\xi_j\right). \quad (6)$$

When $M > 2$ and $\xi_j \equiv 0$, there are non-trivial solutions of (6), so that $\delta^0 = 0$ cannot be admissible for the loss L. Indeed by taking $\xi_j = \alpha\nu_j/\left(b_j \sum \nu_\ell y_\ell/b_\ell\right)$, we see that the left-hand side of (6) is $\alpha(\alpha - 2M + 4)/\left(\sum \nu_\ell y_\ell/b_\ell\right)$, which is negative when $0 < \alpha < 2(M-2)$. The case $\alpha = M - 2$ corresponds to the crude Stein estimator.

The differential operator in (6) can be written as $2D\delta + \delta^T B\delta$ with

$$D\delta = -\sum_j \frac{b_j}{\nu_j}\left(\nu_j\delta_j + 2y_j\frac{\partial\delta_j}{\partial y_j}\right),$$

and $B = \text{diag}(b_1 y_1/\nu_1, \ldots, b_m y_m/\nu_m)$. Brown (1988) gave a necessary and sufficient condition for (6) not to have non-trivial solutions. If $D^\star$ is the conjugate (dual) operator of D acting on scalar functions $H = H(y)$:

$$D^\star H = -\left(b_1\left[\frac{(\nu_1-2)H}{\nu_1} - \frac{2y_1}{\nu_1}\frac{\partial H}{\partial y_1}\right], \ldots, b_m\left[\frac{(\nu_m-2)H}{\nu_m} - \frac{2y_m}{\nu_m}\frac{\partial H}{\partial y_m}\right]\right),$$

for any δ which cannot be improved via (6) there is a differentiable function $H(y)$ such that $\delta = -B^{-1}D^\star H/H$, i.e. with $G(y) = H^2(y)\prod y_j^{2-\nu_j}$

$$\delta_j = \frac{\nu_j - 2}{\nu_j y_j} - \frac{2\partial}{\nu_j \partial y_j}\log H(y) = -\frac{\partial}{\nu_j\partial y_j}\log G(y).$$

Since in our problem only functions δ_j such that $0 \le \delta_j \le t_j^{-2}$ are viable, and positive $(\tau^2+t_j^2)^{-1}$ cannot exceed t_j^{-2}, the estimator $\max[0, \min(\delta_j, t_j^{-2})]$ is better than δ_j. This fact shows that any δ which cannot be improved in the sense of (6) must have the form

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

with some differentiable positive function $G = G(y)$. For example, when $H = H(Q)$ with a positive linear in $y_1, \ldots, y_n$ form $Q = \sum_i q_i y_i$, one gets $\delta_j = -q_j H'(Q)/H(Q)$.

The function, $H(Q) = Q^{-\alpha}$, $\alpha > 0$, leads to estimators

$$\delta_j = \min\left[\frac{\alpha q_j}{\sum_\ell q_\ell y_\ell}, \frac{1}{t_j^2}\right], \tag{7}$$

which are considered later. When $\alpha = M - 2$, $q_j \propto b_j^{-1}$, the statistic $\delta_j^+ = \min\left[(M-2)\nu_j/\left(b_j \sum \nu_\ell y_\ell / b_\ell\right), t_j^{-2}\right]$, is similar to the positive part of Stein's normal vector mean estimator.

3 *R*-risk and asymptotic optimality

For the loss L in (5) introduce the normalized risk of an estimator δ as

$$R(\delta, \tau^2) = \frac{EL(\delta, \tau^2)}{\sum_j b_j(\tau^2 + t_j^2)^{-1}}.$$

Since $EL(\delta^0, \tau^2) = \sum_j b_j(\tau^2 + t_j^2)^{-1}$, $R(\delta^0, \tau^2) \equiv 1$. According to Proposition 1, δ^0 typically is inadmissible. A natural query is R-minimaxity of this rule, i.e., if the minimax value, $V = \inf_\delta \sup_{\tau^2} R(\delta, \tau^2)$, which is the object of interest in this work, equals to one. The answer to this question is in negative. The next result gives a large class of estimators whose largest R-risk is smaller than that of δ^0 leading to the minimaxity theorem in Sect. 5.

Theorem 1 *Assume that*

$$N = \sum_{i=1}^n \nu_i > 2, \tag{8}$$

and let $Q = \sum_i q_i y_i$ *with non-negative coefficients* q_i. *If for* $\kappa \to \infty$, $\kappa \delta_j(\kappa y_1, \ldots, \kappa y_n) \to \alpha_j / Q, 0 < \alpha_j < \infty, j = 1, \ldots, m$, *then*

$$\lim_{\tau^2 \to \infty} R(\delta, \tau^2) = 1 - \frac{1}{\sum_j b_j} \sum_j \frac{b_j}{\nu_j}\left[E\frac{2\alpha_j \chi^2_{\nu_j}}{\sum_i q_i \chi^2_{\nu_i}} - E\frac{\alpha_j^2 \chi^2_{\nu_j}}{\left(\sum_i q_i \chi^2_{\nu_i}\right)^2}\right] \geq \frac{2}{N}, \tag{9}$$

with independent $\chi^2_{\nu_1}, \ldots, \chi^2_{\nu_n}$. *Equal coefficients* q_i *(and only they) provide the asymptotically optimal form* Q *for which* (9) *is an equality. If* $q_i \equiv 1$, *the optimal choice is* $\alpha_j \equiv N - 2$.

The procedure δ^0 *is not* $R-$*minimax. If* $N = M$, *it is improved by the estimator* (7) *provided that*

$$0 < \alpha \leq 2(N-2)\frac{\min b_j q_j / \nu_j}{\max b_j q_j}. \tag{10}$$

Proof One has

$$R(\delta,\tau^2)=\left(\sum_j \frac{b_j}{\tau^2+t_j^2}\right)^{-1}\times\sum_j b_j(\tau^2+t_j^2)E\frac{\chi^2_{\nu_j}}{\nu_j}$$
$$\times\left[\delta_j\left((\tau^2+t_1^2)\chi^2_{\nu_1},\ldots,(\tau^2+t_n^2)\chi^2_{\nu_n}\right)-\frac{1}{\tau^2+t_j^2}\right]^2,$$

with independent random variables $\chi^2_{\nu_i}, i=1,\ldots,n$. Because of our assumptions if $\tau^2\to\infty$,

$$R(\delta,\tau^2)$$
$$\sim\frac{1}{\sum_j b_j}\sum_j\frac{b_j}{\nu_j}E\chi^2_{\nu_j}\left[(\tau^2+t_j^2)\delta_j\left(\frac{(\tau^2+t_1^2)\chi^2_{\nu_1}}{\nu_1},\ldots,\frac{(\tau^2+t_n^2)\chi^2_{\nu_n}}{\nu_n}\right)-1\right]^2$$
$$\to\frac{1}{\sum_j b_j}\sum_j\frac{b_j}{\nu_j}E\chi^2_{\nu_j}\left(\frac{\alpha_j}{\sum_i q_i\chi^2_{\nu_i}}-1\right)^2,$$

so that the first formula in (9) holds.

The optimum for large τ^2 is achieved if $\alpha_j=E\chi^2_{\nu_j}(\sum_i q_i\chi^2_{\nu_i})^{-1}/E\chi^2_{\nu_j}(\sum_i q_i\chi^2_{\nu_i})^{-2}$, and then

$$\lim_{\tau^2\to\infty}R(\delta,\tau^2)=1-\frac{1}{\sum_j b_j}\sum_j\frac{b_j}{\nu_j}\left(E\frac{\chi^2_{\nu_j}}{\sum_i q_i\chi^2_{\nu_i}}\right)^2\left[E\frac{\chi^2_{\nu_j}}{\left(\sum_i q_i\chi^2_{\nu_i}\right)^2}\right]^{-1}.$$

To prove the inequality in (9) we demonstrate that for $N>2$ and any j,

$$\left(E\frac{\chi^2_{\nu_j}}{\sum_i q_i\chi^2_{\nu_i}}\right)^2\le\frac{\nu_j(N-2)}{N}E\frac{\chi^2_{\nu_j}}{\left(\sum_i q_i\chi^2_{\nu_i}\right)^2},\tag{11}$$

with the equality only when all q's are equal.

Indeed using the notation and the results in Carlson (1977) Ch. 6, one gets

$$E\frac{\chi^2_{\nu_j}}{\sum_i q_i\chi^2_{\nu_i}}=\frac{1}{2}\int_0^\infty E\chi^2_{\nu_j}\exp\left\{-u\sum_i q_i\chi^2_{\nu_i}/2\right\}\,\mathrm{d}u$$
$$=\frac{\nu_j}{2}\int_0^\infty\frac{\mathrm{d}u}{(1+q_ju)\prod_i(1+q_iu)^{\nu_i/2}}=\frac{\nu_jB(1,N/2)}{2}R_{-1}(a_j;\tilde{q}).\tag{12}$$

Here B is the beta-function, $\tilde{q} = (q_1, \ldots, q_n)$, and $a_j = (\nu_1/2, \ldots, \nu_{j-1}/2, \nu_j/2 + 1, \nu_{j+1}/2, \ldots, \nu_n/2)$. The function R_{-1} is the average of the function $\left(\sum q_i \omega_i\right)^{-1} = (\tilde{q}\omega)^{-1}, \omega = (\omega_1, \ldots, \omega_n)^T$, with respect to the Dirichlet distribution M_a corresponding to the parameter vector $a = a_j$, $R_{-1}(a; \tilde{q}) = \int \left(\sum q_i \omega_i\right)^{-1} \, dM_a(\omega)$, where integration is over the unit simplex.

Similarly, because of (8),

$$E \frac{\chi^2_{\nu_j}}{\left(\sum_i q_i \chi^2_{\nu_i}\right)^2} = \frac{\nu_j B(2, N/2 - 1)}{4} R_{-2}(a_j; \tilde{q}),$$

with

$$\begin{aligned} R_{-2}(a_j; \tilde{q}) &= \int (\tilde{q}\omega)^{-2} \, \mathrm{d}M_{a_j}(\omega) \\ &= \frac{1}{B(2, N/2 - 1)} \int_0^\infty \frac{u \, \mathrm{d}u}{(1 + q_j u) \prod_i (1 + q_i u)^{\nu_i/2}} \geq \left[R_{-1}(a_j; \tilde{q})\right]^2, \end{aligned} \tag{13}$$

which proves (11).

The estimator (7) is better than δ^0 if for all τ^2

$$\sum_j \frac{b_j}{\nu_j} E y_j \left(\frac{\alpha q_j}{\sum_i q_i \chi^2_{\nu_i}} - \frac{1}{\tau^2 + t_j^2} \right)^2 < \sum_j \frac{b_j}{\tau^2 + t_j^2}.$$

To prove this inequality when $N = M$ it suffices to demonstrate that

$$\alpha \sum_j \frac{b_j q_j^2}{\nu_j} E \frac{\chi^2_{\nu_j}}{\left(\sum_i q_i \chi^2_{\nu_i}\right)^2} < 2 \sum_j \frac{b_j q_j}{\nu_j (\tau^2 + t_j^2)} E \frac{\chi^2_{\nu_j}}{\sum_i q_i \chi^2_{\nu_i}}.$$

By using (12) and (13) now for $\tilde{q} = (q_1(\tau^2 + t_1^2), \ldots, q_n(\tau^2 + t_n^2))$, we see that the latter fact is equivalent to

$$\alpha \sum_j b_j q_j^2 (\tau^2 + t_j^2) R_{-2}(a_j; \tilde{q}) < 2(N - 2) \sum_j b_j q_j R_{-1}(a_j; \tilde{q}). \tag{14}$$

With $a = (\nu_1/2, \ldots, \nu_n/2)$,

$$\sum_j \nu_j q_j (\tau^2 + t_j^2) R_{-2}(a_j; \tilde{q}) = \sum_j \nu_j R_{-1}(a_j; \tilde{q}) = N R_{-1}(a; \tilde{q}),$$

see Relations 5.9-5 in Carlson (1977). Therefore,

$$\alpha \sum_j b_j q_j^2 (\tau^2 + t_j^2) R_{-2}(a_j; \tilde{q}) \leq \alpha \max_j \frac{b_j q_j}{\nu_j} \sum_j \nu_j R_{-1}(a_j; \tilde{q}).$$

Thus (10) implies (14), so that (7) improves upon δ^0. □

The proof of (9) shows that for an estimator (7) satisfying (10),

$$1 - \lim_{\tau^2 \to \infty} R(\delta, \tau^2) > 0,$$

which establishes lack of minimaxity on the part of δ^0 because of the risk continuity.

Theorem 1 demonstrates non-optimality of the traditional estimators (7) with $\tilde{\tau}^2 = Q/\alpha$ for large τ^2 unless Q coincides (up to a positive factor) with $Q^\infty = \sum_i y_i$, in which case $\alpha = N - 2$. For example, the statistic

$$\delta_j^1 = \min\left(\frac{N-2}{\sum y_i}, \frac{1}{t_j^2}\right), \ j = 1, \ldots, m, \tag{15}$$

is asymptotically optimal in this sense.

4 Equal variances and positive normal mean estimation problem

If all t_i^2 tend to t^2 and $\delta_j \to \delta$ (or when $n = 1$, $N = \nu_1$),

$$R(\delta, \tau^2) \to (\tau^2 + t^2)^2 \int_0^\infty \left[\delta(u) - \frac{1}{\tau^2 + t^2}\right]^2 \mathrm{d}G_{N+2}\left(\frac{u}{\tau^2 + t^2}\right)$$
$$= \int_0^\infty \left[(\tau^2 + t^2)\delta((\tau^2 + t^2)z) - 1\right]^2 \mathrm{d}G_{N+2}(z).$$

Here and further G_K is the distribution function of χ^2-law on K degrees of freedom with density g_K. Thus if $t_i \equiv t$, our estimation problem is that of the reciprocal of the scale parameter $\sigma = \tau^2 + t^2$ under the restriction, $\sigma \geq t^2$. The "data" u in this situation is χ^2-distributed, $u \sim \sigma\chi^2_{N+2}$. The invariant loss function, $\sigma^2(\delta - \sigma^{-1})^2 = (\delta\sigma - 1)^2$, corresponds then to (5).

If $N \leq 2$, the minimax value is 1, and for $N > 2$ the minimax value, $2N^{-1}$, is the same as in the non-restricted ($t = 0$) parameter case: Efron and Morris (1973), Theorem 2, Gajek and Kaluszka (1995), Corollary 4.4, Marchand and Strawderman (2005), Remark 12. Under the invariant quadratic loss the risk of the minimax estimator of lower-bounded scale parameter σ at $\sigma = t^2$ equals to the minimax value, $2N^{-1}$ as well, van Eeden (1995). See Kubokawa (2004) for non-quadratic loss function results.

The generalized prior, $\mathrm{d}\sigma/\sigma$, $\sigma \geq t^2$, or $\mathrm{d}\tau^2/(\tau^2 + t^2)$, provides a least favorable distribution, and the corresponding Bayes estimator for $N > 2$ is

$$\delta^B(u) = \frac{\int_0^\infty \exp\{-u/[2(\tau^2 + t^2)]\}(\tau^2 + t^2)^{-N/2-1}\,\mathrm{d}\tau^2}{\int_0^\infty \exp\{-u/[2(\tau^2 + t^2)]\}(\tau^2 + t^2)^{-N/2}\,\mathrm{d}\tau^2}$$
$$= \frac{\int_0^{u/t^2} \mathrm{e}^{-z/2} z^{N/2-1}\,\mathrm{d}z}{u\int_0^{u/t^2} \mathrm{e}^{-z/2} z^{N/2-2}\,\mathrm{d}z} = \frac{(N-2)G_N(u/t^2)}{uG_{N-2}(u/t^2)} = \frac{N-2}{u} - \frac{2g_{N-2}(u/t^2)}{t^2 G_{N-2}(u/t^2)}. \tag{16}$$

The R-risk of δ^B is a bowl-shaped function taking equal values $2N^{-1}$ at $\tau^2 = \infty$ and at $\tau^2 = 0$. Indeed putting $z = v/(\tau^2+t^2)$, we see that with $\zeta = (\tau^2+t^2)/t^2 \geq 1$,

$$R(\delta^B, \tau^2) = \frac{2}{N} - 4\zeta \int_0^\infty \left(\frac{N-2}{z} - 1\right) \frac{g_{N-2}(\zeta z)}{G_{N-2}(\zeta z)}\, \mathrm{d}G_{N+2}(z) + 4\zeta^2 \int_0^\infty \left[\frac{g_{N-2}(\zeta z)}{G_{N-2}(\zeta z)}\right]^2 \mathrm{d}G_{N+2}(z).$$

Integration by parts in the latter integral gives

$$\begin{aligned} 4\zeta^2 \int_0^\infty \left[\frac{g_{N-2}(\zeta z)}{G_{N-2}(\zeta z)}\right]^2 g_{N+2}(z)\, \mathrm{d}z &= 4\zeta \int_0^\infty g_{N-2}(\zeta z) g_{N+2}(z)\, \mathrm{d}\left(-\frac{1}{G_{N-2}(\zeta z)}\right) \\ &= 4\zeta \int_0^\infty \frac{[g_{N-2}(\zeta z) g'_{N+2}(z) + \zeta g'_{N-2}(z) g_{N+2}(z)]\, \mathrm{d}z}{G_{N-2}(\zeta z)} \\ &= 4\zeta \int_0^\infty \frac{g_{N-2}(\zeta z) g_{N+2}(z)}{G_{N-2}(\zeta z)} \left(\frac{N-2}{z} - \frac{\zeta+1}{2}\right) \mathrm{d}z. \end{aligned}$$

Substituting this expression in the formula above, one gets

$$\begin{aligned} R(\delta^B, \tau^2) &= \frac{2}{N} - 2(\zeta-1)\zeta \int_0^\infty \frac{g_{N-2}(\zeta z) g_{N+2}(z)\, \mathrm{d}z}{G_{N-2}(\zeta z)} \\ &= \frac{2}{N} - 2(\zeta-1) \int_0^\infty \frac{g_{N-2}(z) g_{N+2}(z/\zeta)\, \mathrm{d}z}{G_{N-2}(z)}, \end{aligned} \tag{17}$$

which is indeed a bowl-shaped function of τ^2 assuming the value $2N^{-1}$ at $\zeta = 1$ or $\tau^2 = 0$. The same value obtains when $\tau^2 \to \infty$.

According to the Central Limit Theorem, when $N \to \infty$ for any real y, $G_N(N + \sqrt{2N}y) \to \Phi(y)$ and $\sqrt{2N} g_N(N + \sqrt{2N}y) \to \varphi(y)$. Therefore if $\zeta = 1 + \sqrt{2}\theta/\sqrt{N}$ with $\theta \geq 0$,

$$\frac{N}{2} R(\delta^B, \tau^2) \to 1 - \theta \int_{-\infty}^\infty \frac{\varphi(y-\theta)\varphi(y)}{\Phi(y)}\, \mathrm{d}y. \tag{18}$$

The function in the right-hand side of (18) is bowl-shaped attaining its maximum at $\theta = 0$ and $\theta = \infty$, with the minimal value 0.58.. at $\theta = 1.074$... Thus, $\min_{\tau^2} R(\delta^B, \tau^2) = 1.1678N^{-1} + O(N^{-3/2})$, is taken on when

$$\tau_\star^2 \approx \frac{1.5189}{\sqrt{N}} t^2. \tag{19}$$

We turn now to the estimators $\delta(u) = \min(\alpha u^{-1}, t^{-2})$, for which

$$R(\delta, \tau^2) = 1 - \left(1 - \frac{\tau^4}{t^4}\right) G_{N+2}\left(\frac{\alpha}{\zeta}\right) - \frac{2\alpha}{N}\left[1 - G_N\left(\frac{\alpha}{\zeta}\right)\right] + \frac{\alpha^2}{N(N-2)}\left[1 - G_{N-2}\left(\frac{\alpha}{\zeta}\right)\right]. \tag{20}$$

The estimator (15) (with $\alpha = N - 2$) is minimax when $N > 2$. Indeed since $\zeta \geq 1$,

$$R(\delta_1, \tau^2) = \int \left[\min\left(\frac{N-2}{z}, \zeta \right) - 1 \right]^2 \, \mathrm{d}G_{N+2}(z) \leq \frac{2}{N}.$$

The minimal value of $R(\delta_1, \tau^2)$ is seen to be attained at $\tau^2 = 0$. To elucidate the asymptotic behavior of the risk for large N, observe that with $\theta = \sqrt{N}\tau^2/(\sqrt{2}t^2)$,

$$\frac{N}{2} R(\delta, \tau^2) \to \int_{-\infty}^{\infty} [\min(\theta, -y)]^2 \varphi(y) \, \mathrm{d}y = [1 - \Phi(\theta)]\theta^2 + \Phi(\theta) - \theta\varphi(\theta). \tag{21}$$

The comparison of (18) and (21) shows that the limiting behavior of the R-risk is that of the quadratic risk corresponding to two different estimators of the positive normal mean, namely the maximum likelihood estimator and the minimax generalized Bayes estimator against the uniform prior distribution over $\theta > 0$. The first estimator corresponding to δ_1 (or to $\hat{\delta}$) is not admissible. However it is minimax, and has its smallest risk value (0.5 at $\theta = 0$) below the smallest risk value of admissible (and minimax) Bayes estimator δ^B. Maruyama and Iwasaki (2005) discuss the stability of these properties when the variance is misspecified. Rukhin (1992) related this problem to the multivariate variance estimation, and Kubokawa (1999) gave a wide class of minimax estimators.

A more accurate asymptotic formula for $R(\delta, 0) = 1 - G_{N+2}(\alpha) - 2\alpha[1 - G_N(\alpha)]/N + \alpha^2[1 - G_{N-2}(\alpha)]/[N(N-2)]$, can be derived from the Edgeworth expansion, e.g. Johnson et al. (1994) Ch 18, Sec 4. For a fixed real number a, $G_{N-2a}(N) = 1/2 + (1-3a)/[3\sqrt{\pi N}] + O(N^{-1})$.

Thus as $N \to \infty$, the risk of estimator (15) admits the following representation,

$$R(\delta_1, 0) \sim \frac{1}{N} + \frac{10}{3\sqrt{\pi} N^{3/2}}. \tag{22}$$

For the maximum likelihood estimator $\hat{\tau}^2$, $\alpha = N + 2$, so that $R(\hat{\tau}^2, 0) \sim N^{-1} - 26/[3\sqrt{\pi}N^{3/2}]$. Therefore $\hat{\tau}^2$ has a smaller risk at the origin than δ_1 or δ^B. However this risk increases as $\tau^2 \to \infty$ to $2(N+6)/[N(N-2)]$ (which is about twice as large as its risk at zero).

We summarize now the main results of this section.

Proposition 2 *When $t_j^2 \equiv t^2$ and $N > 2$, the R-risk of the minimax Bayes estimator* (16) *is a bowl-shaped function which takes its largest value $2N^{-1}$ at $\tau^2 = \infty$ and at $\tau^2 = 0$, while the smallest value is attained at* (19). *The risk of estimators $\delta(u) = \min(\alpha u^{-1}, t^{-2})$ can be determined from* (20). *The estimator* (15) *($\alpha = N - 2$) is minimax. The asymptotic behavior of the risks of δ^B and of $\hat{\tau}^2$ ($\alpha = N + 2$) is described by* (18) *and* (21) *respectively. The minimal value of $R(\delta_1, \tau^2)$ attained at the origin satisfies* (22).

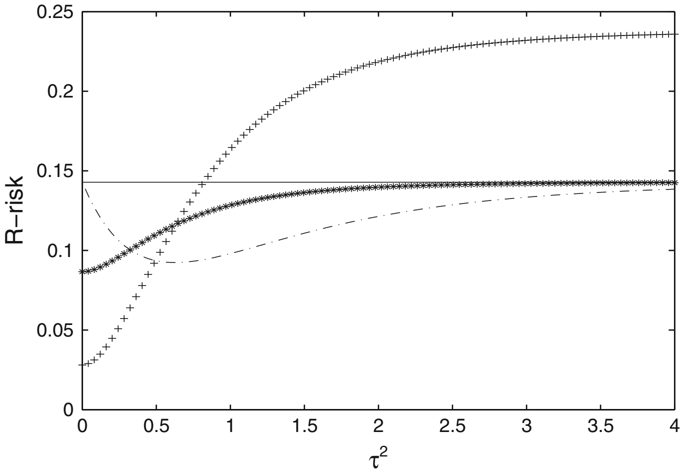

Fig. 1 Plots of R-risks of estimators δ^B (*dash-dotted line*), δ^1 from (15) (*line marked by asterisks*), and of $\hat{\delta}$ (*line* marked by +) when $N = 14$, $t^2 = 1$. The *solid line* portrays the value $2N^{-1}$

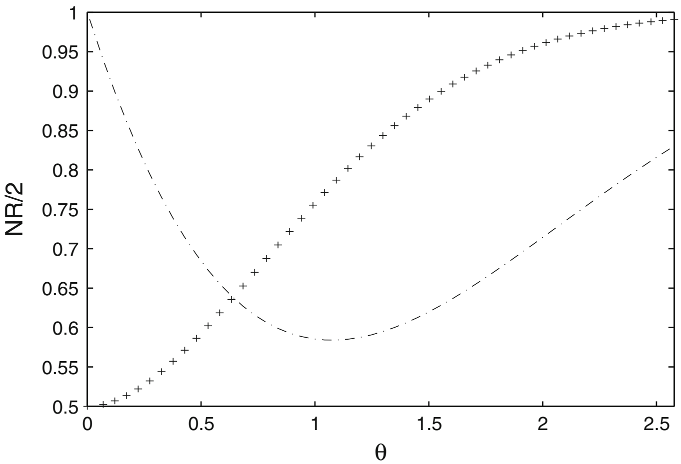

Fig. 2 Plots of the limiting R-risks $\lim NR(\hat{\delta}, \theta)/2$ (*line* marked by +) and $\lim NR(\delta^B, \theta)/2$ (*dash-dotted line*)

Figure 1 shows for $N = 14$ the plots of R-risks of the minimax estimators (15), (16), and that of $\hat{\tau}^2$. The latter rule is not minimax but has a smaller risk at the origin although the interval where it dominates δ^B shrinks as N increases. Better minimax generalized Bayes estimators of τ^2 can be found in Maruyama and Strawderman (2006).

Figure 2 portrays the plots of limiting normalized R-risk for two estimators δ^B and $\hat{\delta}$, i.e. $\lim NR(\hat{\delta}, \theta)/2$ and $\lim NR(\delta^B, \theta)/2$. There is no clear domination, but $\hat{\delta}$ gives a noticeable gain over δ^B only for fairly small values of τ^2 (e..g., $\tau^2 \le t^2/\sqrt{N}$).

5 Bayes procedures and minimax value

Here we look at the Bayes estimators when Π is a (generalized) prior distribution for τ^2. Let

$$p(y|\tau^2) = \prod_{i=1}^{n} \frac{1}{(\tau^2 + t_i^2)^{\nu_i/2}} \exp\left\{ - \frac{y_i}{2(\tau^2 + t_i^2)} \right\} \tag{23}$$

denote the density of the vector y with respect to the measure $\mathcal{M}$, $\mathrm{d}\mathcal{M}(y) = \prod_i y_i^{\nu_i/2-1}/[2^{\nu_i/2}\Gamma(\nu_i/2)]$.

Under the risk R, the Bayes estimator $\delta^\Pi = (\delta_1^\Pi, \ldots, \delta_n^\Pi)$ has the form

$$\delta_j^\Pi = \frac{\int_0^\infty (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\, \mathrm{d}\Lambda(\tau^2)}{\int_0^\infty p(y|\tau^2)\, \mathrm{d}\Lambda(\tau^2)} = -\frac{2\partial}{\partial y_j} \log \mathcal{P}(y),$$

where $\mathrm{d}\Lambda(\tau^2) = \left[\sum_j b_j(\tau^2 + t_j^2)^{-1}\right]^{-1} \mathrm{d}\Pi(\tau^2)$, and

$$\mathcal{P}(y) = \mathcal{P}(y; t_1^2, \ldots, t_n^2) = \int_0^\infty p(y|\tau^2)\, \mathrm{d}\Lambda(\tau^2). \tag{24}$$

Thus δ_j^Π is merely the posterior mean of $(\tau^2 + t_j^2)^{-1}$.

The estimator $\delta^0 \equiv 0$ is generalized Bayes against any prior density $\pi(\tau^2) = \lambda(\tau^2)\left[\sum_j b_j(\tau^2 + t_j^2)^{-1}\right]$ such that $\int_0^\infty p(y|\tau^2)\ \lambda(\tau^2)\ \mathrm{d}\tau^2 = \infty$, but $\int_0^\infty (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\lambda(\tau^2)\ \mathrm{d}\tau^2 < \infty$. For example, the prior density $\pi(\tau^2) = (\tau^2 + t^2)^a$ with $t^2 > 0$, makes δ^0 the generalized Bayes when $N/2 - 2 \le a < N/2 - 1$. If $N > 2$, such densities cannot be well approximated by proper prior densities implying non-minimaxity of δ^0 and its inadmissibility.

If Π is a probability distribution, then with the measure $\mathcal{M}$ defined above, the Bayes risk can be written as:

$$\begin{aligned}
&\int R(\delta^\Pi, \tau^2)\, \mathrm{d}\Pi(\tau^2) \\
&= \sum_j \frac{b_j}{\nu_j} \int \cdots \int y_j \left(\delta_j^\Pi - \frac{1}{\tau^2 + t_j^2} \right)^2 p(y|\tau^2)\, \mathrm{d}\mathcal{M}(y) \mathrm{d}\Lambda(\tau^2) \\
&= \sum_j \frac{b_j}{\nu_j} \int \cdots \int y_j \left[\int \frac{p(y|\tau^2)\mathrm{d}\Lambda(\tau^2)}{(\tau^2 + t_j^2)^2} - \frac{\left[\int (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\mathrm{d}\Lambda(\tau^2)\right]^2}{\int p(y|\tau^2)\mathrm{d}\Lambda(\tau^2)} \right] \mathrm{d}\mathcal{M}(y) \\
&= 1 - \sum_j \frac{b_j}{\nu_j} \int \cdots \int \frac{y_j \left[\int (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\mathrm{d}\Lambda(\tau^2)\right]^2}{\int p(y|\tau^2)\mathrm{d}\Lambda(\tau^2)}\, \mathrm{d}\mathcal{M}(y).
\end{aligned} \tag{25}$$

Since Π is a probability distribution,

$$\sum_j b_j \int \frac{\mathrm{d}\Lambda(\tau^2}{\tau^2 + t_j^2} = 1. \tag{26}$$

Our goal is to determine the minimax value, $V = \inf_\delta \sup_{\tau^2} R(\delta, \tau^2)$ for fixed positive $\nu_1, \ldots \nu_n$. According to (25), V can be expressed in terms of the Bayes R-risk,

$$V = \sup_\Pi \int R(\delta^\Pi, \tau^2)\, \mathrm{d}\Pi(\tau^2) = 1 - \inf_{\mathcal{P}} \sum \frac{4b_j}{\nu_j} \int \cdots \int \frac{y_j \left[\frac{\partial}{\partial y_j}\mathcal{P}\right]^2}{\mathcal{P}(y)}\, \mathrm{d}\mathcal{M}(y). \tag{27}$$

Here $\mathcal{P}(y)$ is defined in (24) with a probability distribution Π or Λ satisfying (26).

Section 4 suggests the form of the anticipated least favorable prior distribution, or rather of the sequence

$$\mathrm{d}\Pi_\epsilon(\tau^2) \propto \sum_j \frac{b_j \lambda_\epsilon(\tau^2)\, \mathrm{d}\tau^2}{\tau^2 + t_j^2}, \tag{28}$$

where as $\epsilon \to 0, \lambda_\epsilon(\tau^2) \to 1$, so that $\int (\tau^2 + t^2)^{-1} \pi_\epsilon(\tau^2)\, \mathrm{d}\tau^2 < \infty$ if $t^2 > 0$. In the proof of the next Theorem we take for simplicity $\lambda_\epsilon(\tau^2) = (\tau^2 + t_1^2)^{-\epsilon}$ with $t_1^2 = \min t_j^2$. Then as $\epsilon \to 0$, $\sum_j b_j \int \lambda_\epsilon(\tau^2)(\tau^2 + t_j^2)^{-1}\, \mathrm{d}\tau^2 \sim \sum_j b_j \epsilon^{-1}$, and $\mathrm{d}\Lambda(\tau^2) = \mathrm{d}\Lambda_\epsilon(\tau^2) = C(\epsilon)\lambda_\epsilon(\tau^2)\, \mathrm{d}\tau^2$ with $C(\epsilon) \sim \epsilon / \left(\sum_j b_j\right)$.

For any j, $\delta_j^\epsilon = \int_0^\infty (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\lambda_\epsilon(\tau^2)\, \mathrm{d}\tau^2 / \int_0^\infty p(y|\tau^2)\lambda_\epsilon(\tau^2)\, \mathrm{d}\tau^2$ as $\epsilon \to 0$ tends to

$$\delta_j^B(y) = \frac{\int_0^\infty (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\, \mathrm{d}\tau^2}{\int_0^\infty p(y|\tau^2)\, \mathrm{d}\tau^2}, \tag{29}$$

which corresponds to the generalized prior density $\sum_j b_j(\tau^2 + t_j^2)^{-1}$. When all t's coincide this is the least favorable prior density discussed in Sect. 4 and then δ^B coincides with (16). The generalized Bayes estimator δ^B satisfies the conditions of Theorem 1 with $q_j \equiv 1$ and $\alpha_j \equiv N - 2$, so that it is optimal for large τ^2,, i.e. $\lim_{\tau^2 \to \infty} R(\delta^B, \tau^2) = 2N^{-1}$. Its numerical evaluation is discussed by Rukhin (2014b).

The main result shows that V does not depend on $t_1^2, \ldots, t_n^2$, and that (29) is indeed minimax.

Theorem 2 *Under condition (8) for any positive $t_1^2, \ldots, t_n^2$, $V = 2N^{-1}$. The sequence (28) is least favorable, i.e.,*

$$\lim_{\epsilon \to 0} \int R(\delta^\epsilon, \tau^2)\, \mathrm{d}\Pi_\epsilon(\tau^2) = \frac{2}{N},$$

and the estimator (29) *is minimax.*

Proof We start by proving the inequality $V \geq 2N^{-1}$. To this end it suffices to show that

$$\lim_{\epsilon\to 0} \frac{\epsilon}{\sum_j b_j} \sum_j \frac{b_j}{\nu_j} \int \cdots \int \frac{y_j \left[\int_0^\infty (\tau^2+t_j^2)^{-1} p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2\right]^2}{\int_0^\infty p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2} \times \mathrm{d}\mathcal{M}(y) = \frac{N-2}{N}. \quad (30)$$

To establish (30), let U_ϵ be a positive sequence such that $U_\epsilon \to \infty$, but $\epsilon U_\epsilon \to 0$. Then

$$\sum_j \frac{b_j}{\nu_j} \int \cdots \int_{\sum y_i \leq U_\epsilon} y_j \frac{\left[\int_0^\infty (\tau^2+t_j^2)^{-1} p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2\right]^2}{\int_0^\infty p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2} \mathrm{d}\mathcal{M}(y)$$
$$\leq \max_j \frac{b_j U_\epsilon}{\nu_j} \int \cdots \int \int_0^\infty \frac{p(y|\tau^2)\mathrm{d}\tau^2\, \mathrm{d}\mathcal{M}(y)}{(\tau^2+t_1^2)^{2+\epsilon}}$$
$$\leq \max_j \frac{b_j U_\epsilon}{\nu_j} \int_0^\infty \frac{\mathrm{d}\tau^2}{(\tau^2+t_1^2)^{2+\epsilon}},$$

so that the contribution of smaller values of $\sum y_i$ to V becomes negligible as $\epsilon \to 0$, and the behavior of the integrals in (30) when $\epsilon \to 0$ is determined by integrand's asymptotics for large $\sum y_i$,

For $\sum_i y_i \geq U_\epsilon$, the change of variables, $x = (\tau^2+t_1^2)^{-1}$, leads to the following estimates,

$$\int_0^\infty p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2$$
$$\geq \int_0^{t_1^{-2}} x^{N/2+\epsilon-2} \mathrm{e}^{-x\sum y_i/2} \max\left[1 - \frac{x\sum \nu_i(t_i^2-t_1^2)}{2}, \prod_j \left(\frac{t_1^2}{t_j^2}\right)^{\nu_j/2}\right] \mathrm{d}x$$
$$\geq \frac{\Gamma(N/2+\epsilon-1)}{\left(\sum y_i/2\right)^{N/2+\epsilon-1}} \left(1 - \frac{C_1}{U_\epsilon}\right),$$

and

$$\int_0^\infty (\tau^2+t_j^2)^{-2} p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2$$
$$\leq \int_0^{t_1^{-2}} x^{N/2+\epsilon-1} \mathrm{e}^{-x\sum y_i/2} \left[1 + C_2 \sum y_i(t_n^2-t_1^2)x^2\right] \mathrm{d}x$$
$$\leq \frac{\Gamma(N/2+\epsilon)}{\left(\sum y_i/2\right)^{N/2+\epsilon}} \left[1 + \frac{C_3}{U_\epsilon}\right].$$

Here and below positive constants $C_k, k = 1, \ldots, 4$, do not depend on ϵ, x or y's. Thus if $\sum_i y_i \geq U_\epsilon$,

$$\frac{\left[\int_0^\infty (\tau^2 + t_j^2)^{-2} p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2\right]^2}{\int_0^\infty p(y|\tau^2)\lambda_\epsilon(\tau^2)\mathrm{d}\tau^2} \leq \frac{\Gamma(N/2+\epsilon)(N/2+\epsilon-1)}{(\sum y_i/2)^{N/2+\epsilon+1}}\left(1+\frac{C_4}{U_\epsilon}\right),$$

so that

$$1 - V \leq \lim_{\epsilon\to 0}\left(\sum_j b_j\right)^{-1} \Gamma\left(\frac{N}{2}+\epsilon\right)\left(\frac{N}{2}+\epsilon-1\right)\left(1+\frac{C_4}{U_\epsilon}\right)$$
$$\times \sum_j b_j \int\cdots\int_{\sum y_i \geq U_\epsilon} \frac{\epsilon y_j \mathrm{d}\mu(y)}{\left(\sum y_i/2\right)^{N/2+\epsilon+1}}.$$

Since for any $j = 1, \ldots, n$,

$$\lim_{\epsilon\to 0}\epsilon\int\cdots\int_{\sum y_i\geq U_\epsilon}\frac{y_j \mathrm{d}\mathcal{M}(y)}{\left(\sum y_i/2\right)^{N/2+\epsilon+1}} = \lim_{\epsilon\to 0}\epsilon\int_{U_\epsilon/2}^\infty \frac{\mathrm{d}z}{z^{1+\epsilon}}\left[\Gamma\left(\frac{N}{2}+1\right)\right]^{-1}$$
$$= \left[\Gamma\left(\frac{N}{2}+1\right)\right]^{-1},$$

(30) is proven.

To demonstrate the reverse inequality, $V \leq 2N^{-1}$, let $f_j = f_j(y)$ be a function of y such that the expected values $g_j(\tau^2) = \int\cdots\int f_j(y)p(y|\tau^2)\, \mathrm{d}\mathcal{M}(y)$ and $h_j(\tau^2) = \int\cdots\int y_j^{-1} f_j^2(y)p(y|\tau^2)\, \mathrm{d}\mathcal{M}(y)$ exist for any τ^2. Then

$$\sum \frac{2b_j}{\nu_j}\int\cdots\int \frac{\partial\mathcal{P}}{\partial y_j} f_j\, \mathrm{d}\mathcal{M}(y) = -\sum \frac{b_j}{\nu_j}\int_0^\infty \frac{g_j(\tau^2)\,\mathrm{d}\Lambda(\tau^2)}{\tau^2+t_j^2},$$

and

$$\sum \frac{b_j}{\nu_j}\int\cdots\int \frac{\mathcal{P}f_j^2}{y_j}\,\mathrm{d}\mathcal{M}(y) = \sum\frac{b_j}{\nu_j}\int_0^\infty h_j(\tau^2)\,\mathrm{d}\Lambda(\tau^2).$$

The inequality

$$\sum\frac{4b_j}{\nu_j}\int\cdots\int y_j \frac{\left[\frac{\partial}{\partial y_j}\mathcal{P}\right]^2}{\mathcal{P}(y)}\,\mathrm{d}\mathcal{M}(y) \geq \frac{\left[\sum\frac{2b_j}{\nu_j}\int\cdots\int f_j\frac{\partial}{\partial y_j}\mathcal{P}\,\mathrm{d}\mathcal{M}(y)\right]^2}{\sum b_j\nu_j^{-1}\int\cdots\int f_j^2 y_j^{-1}\mathcal{P}\,\mathrm{d}\mathcal{M}(y)},$$

implies that

$$1 - V \geq \inf_{\Lambda} \frac{\left[\sum \frac{b_j}{\nu_j} \int_0^{\infty} (\tau^2 + t_j^2)^{-1} g_j(\tau^2)\, \mathrm{d}\Lambda(\tau^2)\right]^2}{\sum \frac{b_j}{\nu_j} \int_0^{\infty} h_j(\tau^2)\, \mathrm{d}\Lambda(\tau^2)}.$$

By choosing $f_j = y_j \left(\sum q_i y_i\right)^{-1}$, one gets in the notation of Sect. 3,

$$g_j(\tau^2) = (\tau^2 + t_j^2) E \frac{\chi^2_{\nu_j}}{\sum_i q_i (\tau^2 + t_i^2) \chi^2_{\nu_i}} = \frac{\nu_j (\tau^2 + t_j^2) B(1, N/2)}{2} R_{-1}(a_j; \tilde{q}),$$

and

$$h_j(\tau^2) = (\tau^2 + t_j^2) E \frac{\chi^2_{\nu_j}}{\left[\sum_i q_i (\tau^2 + t_i^2) \chi^2_{\nu_i}\right]^2} = \frac{\nu_j (\tau^2 + t_j^2) B(2, N/2 - 1)}{4} R_{-2}(a_j; \tilde{q}).$$

Here $\tilde{q}$ is the vector whose coordinates are $q_i(\tau^2 + t_i^2)$; R_{-1}, R_{-2}, and the vectors a_j are as in the proof of Theorem 1.

With this choice

$$1 - V \geq \frac{N-2}{N} \inf_{\Lambda} \frac{\left[\sum b_j \int R_{-1}(a_j; \tilde{q})\, \mathrm{d}\Lambda(\tau^2)\right]^2}{\sum b_j \int (\tau^2 + t_j^2) R_{-2}(a_j; \tilde{q})\, \mathrm{d}\Lambda(\tau^2)}.$$

To complete the proof it suffices to notice that for any Λ under condition (26),

$$\left[\sum b_j \int R_{-1}(a_j; \tilde{q})\, \mathrm{d}\Lambda(\tau^2)\right]^2 \leq \sum b_j \int (\tau^2 + t_j^2) R_{-2}(a_j; \tilde{q})\, \mathrm{d}\Lambda(\tau^2). \quad (31)$$

Take $\mathrm{d}\Lambda(\tau^2) = \mathrm{d}\Lambda_\epsilon\left(\tau^2\right) \propto \epsilon \left(\sum_j b_j\right)^{-1} \lambda_\epsilon(\tau^2) \mathrm{d}\tau^2$, with $\lambda_\epsilon(\tau^2)$ as above. By using the properties of the Dirichlet averages R_{-1} and R_{-2} we see that

$$\lim_{\epsilon \to 0} \int R_{-1}(a_j; \tilde{q})\, \mathrm{d}\Lambda_\epsilon(\tau^2) = R_{-1}\left(a_j; q_1, \ldots, q_n\right) \left(\sum_j b_j\right)^{-1},$$

and similarly

$$\begin{aligned} \lim_{\epsilon \to 0} \int (\tau^2 + t_j^2) R_{-2}\left(a_j; \tilde{q}\right)\, \mathrm{d}\Lambda_\epsilon(\tau^2) &= \lim_{\tau^2 \to \infty} (\tau^2 + t_j^2)^2 R_{-2}\left(a_j; \tilde{q}\right) \left(\sum_j b_j\right)^{-1} \\ &= R_{-2}(a_j; q_1, \ldots, q_n) \left(\sum_j b_j\right)^{-1}. \end{aligned}$$

Thus for this sequence, (31) means that

$$\left[\sum_j b_j R_{-1}\left(a_j; q_1, \ldots, q_n\right)\right]^2 \le \sum_j b_j R_{-2}\left(a_j; q_1, \ldots, q_n\right)\left(\sum_j b_j\right),$$

which [as the inequality (11)] becomes an equality when q_i are all equal. Therefore, (28) indeed provides the least favorable prior. □

The estimator (29) is admissible under the loss (5) which can be proven by the traditional method via approximating its Bayes risk for the prior $\pi_\epsilon(\tau^2)$. van Eeden (1995) proves a similar result when $t_i \equiv t$ so that δ^B coincides with (16).

6 Second order minimaxity and concluding remarks

The proof of Theorem 2 indicates the special role played by large values of τ^2 in the minimaxity issue. Since extremely heterogeneous models are hardly useful in applications, the importance of these τ^2 values can be questioned. When τ^2 is supposed to be bounded from above, say $0 \le \tau^2 \le T$, the following result holds.

Proposition 3 *The Bayes estimator,* $\delta^T = (\delta_1^T, \ldots, \delta_n^T)$ *with*

$$\delta_j^T = \frac{\int_0^T (\tau^2 + t_j^2)^{-1} p(y|\tau^2)\, \mathrm{d}\tau^2}{\int_0^T p(y|\tau^2)\, \mathrm{d}\tau^2}, \quad j = 1, \ldots, n,$$

corresponding to the prior π *proportional to* $\sum_j b_j(\tau^2 + t_j^2)^{-1}$ *on the interval,* $0 \le \tau^2 \le T$*, is second order minimax, i.e. it minimizes*

$$\limsup_{T\to\infty}\left[\inf_\delta \sup_{0\le\tau^2<\infty} R(\delta, \tau^2) - \inf_\delta \sup_{0\le\tau^2\le T} R(\delta, \tau^2)\right].$$

If τ^2 is supposed to belong to the interval $[\mathrm{e}^{-T}, \mathrm{e}^T]$ for large T, one obtains another sequence of second order minimax estimators, which minimize $\limsup_{T\to\infty} T^2[2N^{-1} - \inf_\delta \sup_{|\log\tau^2|\le T} R(\delta, \tau^2)]$. As in the bounded normal mean problem these procedures are related to a very different prior probability density π, proportional to $\cos^2(\pi \log\tau^2/T)/\tau^2$, which minimizes the Fisher information on the interval $|\log\tau^2| \le T$ (Bickel 1981). Gajek and Kaluszka (1994) have shown that if $t_i \equiv t$, and $T \to \infty$, these densities form a sequence of the least favorable priors. See Kubokawa (1999), Marchand and Strawderman (2004) for reviews.

In most of meta-analysis studies small parametric values should not be excluded by the prior distribution. On the contrary, the priors which give $\tau^2 = 0$ a positive weight may be quite important in applications although they do not lead to minimax rules.

The key issue in a given practical situation is to decide how wide the possible range of heterogeneity variance can be. If this range cannot be reliably specified, the use of minimax estimators such as δ^{Π} can be recommended. If τ^2 is deemed to be fairly small, arguably the maximum likelihood estimator may be preferable.

The author is grateful to Bill Strawderman and to two anonymous referees for helpful comments.

References

Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Annals of Statistics*, *9*, 1301–1309.

Brown, L. D. (1988). The differential inequality of a statistical estimation problem. In S. S. Gupta, J. O. Berger (Eds.), Statistical Decision Theory and Related Topics IV (pp 299–324). NY: Springer.

Carlson, B. C. (1977). *Special Functions of Applied Mathematics*. New York: Academic Press.

Efron, B., Morris, C. (1973). Stein's estimation rule and its competitors: an empirical Bayes approach. *Journal of the American Statistical Association*, *68*, 117–130.

Gajek, L., Kaluszka, M. (1994). Lower bounds for the asymptotic Bayes risk in the scale model (with an application to the second-order minimax estimation). *Annals of Statistics*, *22*, 1831–1839.

Gajek, L., Kaluszka, M. (1995). Nonexponential applications of a global Cramér-Rao inequality. *Statistics*, *26*, 111–122.

Johnson, N., Kotz, S., Balakrishnan, N. (1994). *Univariate Continuous Distributions* (2nd ed.). New York: J. Wiley.

Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: a review. *Communications in Statistics-Theory and Methods*, *28*, 613–650.

Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters. *Journal of Japan Statistical Society*, *34*, 229–253.

Marchand, E., Strawderman, W. (2004). Estimation in restricted parameter spaces: a review. In H. Rubin, A. DasGupta (Eds.), *Festschrift for IMS Lecture Notes*, vol 45, (pp 21–44).

Marchand, E., Strawderman, W. (2005). On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint. *Journal of Statistical Planning and Inference*, *134*, 90–101.

Maruyama, Y., Iwasaki, K. (2005). Sensitivity of minimaxity and admissibility in the estimation of positive normal mean. *Annals of Institute of Statistical Mathematics*, *57*, 145–156.

Maruyama, Y., Strawderman, W. (2006). A new class of minimax generalized Bayes estimators of a normal variance. *Journal of Statistical Planning and Inference*, *136*, 3822–3836.

Rukhin, A. L. (1992). Asymptotic risk behavior of mean vector and variance estimators and the problem of positive normal mean. *Annals of Institute of Statistical Mathematics*, *44*, 299–311.

Rukhin, A. L. (2014a). Restricted likelihood representation and decision-theoretic aspects of meta-analysis. *Bernoulli*, *20*, 1979–1998.

Rukhin, A. L. (2014b). Bayes estimators of heterogeneity variance and T-systems. *Journal of Statistical Planning and Inference*, *154*, 156–165.

van Eeden, C. (1995). Minimax estimation of a lower-bounded scale parameter of a gamma-distribution for scale-invariant squared-error loss. *Canadian Journal of Statistics*, *23*, 245–256.

Welham, S. J., Thompson, R. (1997). Likelihood ratio tests for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Ser B.*, *59*, 701–714.