Journal of Statistical Planning and Inference 154 (2014) 156–165

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Bayes estimators of heterogeneity variance and T-systems

Andrew L. Rukhin

Statistical Engineering Division, National Institute of Standards and Technology, United States

ARTICLE INFO

Available online 8 January 2014

Keywords:
Curved exponential family
Dirichlet averages
Laguerre polynomials
Meta-analysis
Minimaxity
Quadrature formulas
Stieltjes transform

ABSTRACT

The considered problem concerns simultaneous inference for curve-confined natural parameters of independent, heterogeneous gamma random variables with known shape parameters. A loss function is suggested that is motivated by meta-analysis, and some properties of the minimax value of the corresponding risk are obtained. Bayes estimators and quadrature formulas for their numerical evaluation are provided, along with Monte Carlo simulations and comparisons of numerical implementations of several alternative estimators.

Published by Elsevier B.V.

1. Introduction and summary

In this work the entries of the observed data vector $y=(y_1,\ldots,y_n)$ are supposed to be realizations of independent, heterogeneous, gamma-random variables with given shape parameters and restricted scale parameters. More precisely, for fixed positive ν_i

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

where Γ_ν is a gamma random variable with the density

$$g_\nu(u)=\frac{e^{-u}u^{\nu-1}}{\Gamma(\nu)}, \quad 0<u<\infty,$$

whose cumulative distribution function will be denoted by G_ν. Thus y_i/ν_i is an unbiased estimator of $\tau^2+t_i^2$. The unknown parameter τ^2, $\tau^2\geq 0$, has the meaning of the heterogeneity variance (equivalently the between study effect variance) in meta analysis. The distinct constants t_i^2, $t_i^2>0$, $i=1,\ldots,n$, are supposed to be given. In practice they are determined from the reported uncertainties (Rukhin, 2014).

This model was introduced by Efron and Morris (1973, p. 128) in terms of chi-squared random variables for the empirical Bayes approach to multivariate normal mean estimation. The joint distribution of y_i's forms a curved exponential family whose natural parameters consist of $(\tau^2+t_i^2)^{-1}$, $i=1,\ldots,n$.

In this situation the sufficient statistic $y=(y_1,\ldots,y_n)$ is incomplete, and mathematically convenient conjugate prior distributions for τ^2 are not available. An unbiased estimator of $(\tau^2+t_i^2)^{-1}$ does not even exist if $\nu_i\leq 1$, which happens in the most interesting case, $\nu_i=1/2$, corresponding to a normal variance.

There is a body of literature on the estimation of the natural parameter vector for independent exponential families, in particular of gamma-distributions with a thorough investigation of the Stein inadmissibility phenomenon under various loss functions (Berger, 1980; DasGupta, 1986; Ghosh and Parsian, 1980). That work is largely based on solving differential

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

0378-3758/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jspi.2013.12.013

inequalities arising from an integration by parts technique. The estimation of positive powers of the natural parameter in exponential families is studied by Baringhaus (2003).

The (random) loss function $L = L(\delta, \tau^2)$ for estimation of the heterogeneity variance τ^2 was introduced in Rukhin (2012) when $n=1$; Rukhin (2014) deals with the general case. Motivated by meta-analysis applications, L has the form

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

Here $\delta = (\delta_1, \ldots, \delta_n)$ represents the vector estimator of the reciprocals of the scale parameters, $((\tau^2+t_1^2)^{-1}, \ldots, (\tau^2+t_n^2)^{-1})$. It is convenient to allow a possibility of y_i (containing some information about τ^2) being present, when $b_i=0$ and the estimate of $(\tau^2+t_i^2)^{-1}$ is not required as such. The Fisher information in y_i about τ^2 is $\nu_i/(\tau^2+t_i^2)^2$, so that the informational content of y_i with large t_i^2 and small ν_i is relatively small. Under our convention, $t_1^2 < \cdots < t_n^2$, so that when all ν's are equal, y_1 is the most informative and y_n is the least informative data point.

The traditional procedures are of the form, $\delta_j = (\tilde{\tau}^2 + t_j^2)^{-1}$, where $\tilde{\tau}^2 = \tilde{\tau}^2(y_1, \ldots, y_n)$ is an estimate of τ^2. However, under the loss function (1), $\tilde{\tau}^2$ is not designed to estimate τ^2 itself, but rather $(\tilde{\tau}^2+t_j^2)^{-1}$ estimates $(\tau^2+t_j^2)^{-1}$. Then an estimator $\delta_j = (\tilde{\tau}_j^2 + t_j^2)^{-1}$ may be more natural provided that $\tilde{\tau}_j^2$ is τ^2 estimator adjusted for the presence of y_j in the loss L.

A convenient normalization of the loss leads to the risk function

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

Some properties of the minimax value for this risk are given in Section 3. They are motivated by the case of equal t_j's discussed in the next section. The form of Bayes estimators and different methods for their numerical evaluation are provided in Section 4. The paper concludes with Section 5 which presents some comparisons and results of a Monte Carlo study.

One of the approaches to the described problem in Section 4 is based on the fact that the reciprocals of scale parameters form a T-system on the positive half-line. Many fundamental contributions to the study of these systems were made by the author's former colleague, Bill Studden, to whose memory this work is dedicated.

2. Risk for equal scale parameters

Let

$$m(y|\tau^2) = \prod_i \frac{1}{(\tau^2+t_i^2)^{\nu_i}} \exp\left\{-\frac{y_i}{\tau^2+t_i^2}\right\} \tag{3}$$

denote the density of the vector y with respect to the measure μ

$$d\mu(y) = \prod_i \frac{y_i^{\nu_i - 1}}{\Gamma(\nu_i)}\, dy.$$

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

$$m_j(y|\tau^2) = \frac{y_j m(y|\tau^2)}{\nu_j(\tau^2+t_j^2)} \tag{4}$$

is a probability density with respect to μ as well. This density represents independent gamma-random variables, $y_1', \ldots, y_n'$, with $y_j' \sim (\tau^2+t_j^2)G_{\nu_j+1}$ and $y_i' \sim (\tau^2+t_i^2)G_{\nu_i}, i \neq j$. It allows for the following expression of the expected loss, in (1):

$$EL(\delta, \tau^2) = \sum_j \frac{b_j}{\tau^2+t_j^2} \int [(\tau^2+t_j^2)\delta_j(y') - 1]^2 m_j(y'|\tau^2)\, d\mu(y').$$

The increased (by one) shape parameter for y_j also happens in evaluation of the relative savings loss in the mentioned empirical Bayes approach to multivariate normal mean estimation (Efron and Morris, 1973).

If all t_i^2 are equal, $t_i^2 \equiv t^2$, then the sum $y_1' + \cdots + y_n'$ forms a sufficient statistic. Its distribution is $(\tau^2+t^2)\Gamma_{N+1}$ with $N = \sum_i \nu_i$. In this situation, $\delta_j \equiv \delta$ and

$$EL(\delta, \tau^2) = \frac{\sum_j b_j}{\tau^2+t^2} \int [(\tau^2+t^2)\delta(y') - 1]^2 m_j(y'|\tau^2)\, d\mu(y').$$

Then R in (2) has the form

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

Thus when $t_i^2 \equiv t^2$, $\sigma = \tau^2 + t^2$ is the scale parameter, and our estimation problem is that of its reciprocal under the restriction, $\sigma \geq t^2 > 0$. The "data" u in this situation has a gamma distribution, $u \sim \sigma\Gamma_{N+1}$. The invariant quadratic loss function, $\sigma^2(\delta - \sigma^{-1})^2 = (\sigma\delta - 1)^2$, corresponds to our R.

If $N \le 1$, the minimax value, $V = \inf_\delta \sup_{\sigma \ge t^2} E(\sigma\delta - 1)^2$ in this problem is one. For $N > 1$, $V = N^{-1}$, so that the minimax value is the same as in the non-restricted parameter case, i.e., it does not depend on t^2. See Efron and Morris (1973, Theorem 2), Gajek and Kaluszka (1995, Corollary 4.4), and Marchand and Strawderman (2005, Remark 12). The generalized prior, $d\sigma/\sigma$, $\sigma \ge t^2$, or $d\tau^2/(\tau^2+t^2)$, provides a least favorable distribution. A similar prior distribution, $d\sigma/\sigma$, $0 < \sigma \le t^2$, is least favorable in the estimation problem of upper-bounded σ under the quadratic loss $\sigma^{-2}(\delta-\sigma)^2 = (\delta/\sigma - 1)^2$. The minimax value remains the same N^{-1}.

The generalized Bayes estimator of $\tau^2 + t^2$ when $N > 1$ is

$$\delta^B(u) = \frac{\int_0^\infty \exp\{-u/(\tau^2+t^2)\}(\tau^2+t^2)^{-N-1}\, d\tau^2}{\int_0^\infty \exp\{-u/(\tau^2+t^2)\}(\tau^2+t^2)^{-N}\, d\tau^2}$$
$$= \frac{\int_0^{u/t^2} e^{-z} z^{N-1}\, dz}{u \int_0^{u/t^2} e^{-z} z^{N-2}\, dz} = \frac{(N-1)G_N(u/t^2)}{uG_{N-1}(u/t^2)}, \tag{5}$$

($\delta^B(u) = 0$ if $N \le 1$).

Mathematically these facts mean that

$$\inf_\delta \sup_\pi \int R(\delta, \tau^2)\pi(\tau^2)\, d\tau^2 = \sup_{\tau^2} R(\delta^B, \tau^2) = \min\left(1, \frac{1}{N}\right).$$

The R-risk of δ^B is a bowl-shaped function taking equal values N^{-1} at $\tau^2 = \infty$ and at $\tau^2 = 0$ (Rukhin, 2014).

Actually the R-risk of δ^B is well defined for all $\tau^2 > -t^2$. When $\tau^2 \downarrow -t^2$, $R(\delta^B, \tau^2) \to 1$, while $R(\delta^B, 0) = N^{-1}$, no matter how small is t^2. More generally, if $\tau^2 + t_1^2 \to 0$, then $R(\delta, \tau^2) \sim (\tau^2 + t_1^2)^{-1} E y_1 (\delta_1(\tau^2 + t_1^2) - 1)^2 \to 1$ for any bounded positive δ_1. These facts present a warning against the interpretation of τ^2 as a possibly negative variance component which is practiced in some studies (Prysley et al., 2011). Indeed the risk of positive τ^2-estimators is the largest possible for such, arguably, irrelevant values. The use of possibly negative τ^2 estimators ignores the information, $\sigma > t^2$, but leads to the same minimax value. The (unique) minimax estimator $(N-1)/u$ has a constant risk which is uniformly larger than that of (5) for all positive τ^2.

In the next section we will see that from the minimax value point of view the problem is the easiest when $t_i^2 \equiv t^2$.

3. Bayes procedures and minimax value

Here we look at the Bayes estimators for the loss L in (1) when Π is a (generalized) prior distribution for τ^2. With $m(y|\tau^2)$ given by (3), let

$$\mathcal{M}(y) = \int_0^\infty \left[\sum_j \frac{b_j}{\tau^2 + t_j^2}\right]^{-1} m\left(y|\tau^2\right) d\Pi(\tau^2) = \int_0^\infty m(y|\tau^2)\, d\Lambda(\tau^2),$$

where

$$d\Lambda(\tau^2) = \left[\sum_j \frac{b_j}{\tau^2 + t_j^2}\right]^{-1} d\Pi(\tau^2).$$

Then the Bayes estimator can be written in the form

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

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

If Π is a probability distribution, the Bayes risk has the form

$$\int R(\delta^\Pi, \tau^2)\, d\Pi(\tau^2) = \sum_j \frac{b_j}{\nu_j} \int \cdots \int y_j \left[\frac{\int (\tau^2 + t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)}{\int m(y|\tau^2)\, d\Lambda(\tau^2)} - \frac{1}{\tau^2 + t_j^2}\right]^2 m(y|\tau^2)\, d\mu(y)\, d\Lambda(\tau^2)$$
$$= \sum_j \frac{b_j}{\nu_j} \int \cdots \int \left[\frac{\int (\tau^2 + t_j^2)^{-2} m(y|\tau^2)\, d\Lambda(\tau^2)}{\int m(y|\tau^2)\, d\Lambda(\tau^2)} - \left(\frac{\int (\tau^2 + t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)}{\int m(y|\tau^2)\, d\Lambda(\tau^2)}\right)^2\right]$$
$$\times y_j m(y|\tau^2)\, d\mu(y)\, d\Lambda(\tau^2)$$
$$= \sum_j \frac{b_j}{\nu_j} \int \cdots \int y_j \left[\int \frac{m(y|\tau^2)\, d\Lambda(\tau^2)}{(\tau^2 + t_j^2)^2} - \frac{[\int (\tau^2 + t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)]^2}{\int m(y|\tau^2)\, d\Lambda(\tau^2)}\right] d\mu(y)$$
$$= 1 - \sum_j \frac{b_j}{\nu_j} \int \cdots \int y_j \frac{[\int (\tau^2 + t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)]^2}{\int m(y|\tau^2)\, d\Lambda(\tau^2)}\, d\mu(y). \tag{6}$$

Concavity of the Bayes risk as a function of the prior distribution Π can be obtained directly from the inequality (9) in Vidakovic and DasGupta (1995), although it also follows from a more general result, e.g. DeGroot (1970, Section 8.4).

The minimax value, $V=\inf_\delta \sup_{\tau^2} R(\delta,\tau^2)$, is of interest. According to (6), its expression in terms of the Bayes R-risk has the form

$$V=V(t_1^2,\ldots,t_n^2)=\sup_\Pi \int R(\delta^\Pi,\tau^2)\, d\Pi(\tau^2)$$
$$=1-\inf_\Lambda \sum_j \frac{b_j}{\nu_j}\int\cdots\int y_j \frac{[\int_0^\infty (\tau^2+t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)]^2}{\int_0^\infty m(y|\tau^2)\, d\Lambda(\tau^2)}\, d\mu(y). \tag{7}$$

Since Π is a probability distribution, for Λ in (7)

$$\sum_j b_j \int \frac{d\Lambda(\tau^2)}{\tau^2+t_j^2}=1. \tag{8}$$

The next result gives some properties of $V(t_1^2,\ldots,t_n^2)$.

Proposition 1. *The function $V(t_1^2,\ldots,t_n^2)$, $0<t_1^2<\cdots<t_n^2$, is homogeneous of degree zero, i.e. for any $\zeta>0$*

$$V(\zeta t_1^2,\ldots,\zeta t_n^2)=V(t_1^2,\ldots,t_n^2). \tag{9}$$

For any fixed positive $t_1^2,\ldots,t_{n-1}^2$

$$\lim_{t_n^2\to\infty} V(t_1^2,\ldots,t_n^2)=V(t_1^2,\ldots,t_{n-1}^2). \tag{10}$$

If $N=\sum_i \nu_i>1$

$$1>V(t_1^2,\ldots,t_n^2)\ge \frac{1}{N}. \tag{11}$$

Proof. It follows from (8) that with $d\tilde\Lambda(\tau^2)=\zeta\, d\Lambda(\tau^2/\zeta)$

$$\sum_j b_j \int (\tau^2+\zeta t_j^2)^{-1}\, d\tilde\Lambda(\tau^2)=\sum_j b_j \int (\tau^2+t_j^2)^{-1}\, d\Lambda(\tau^2)=1,$$

so that with

$$m_\zeta(y|\tau)=\prod_i \frac{1}{(\tau^2+\zeta t_i^2)^{\nu_i}} \exp\left\{-\frac{y_i}{\tau^2+\zeta t_i^2}\right\},$$

one gets

$$1-V(\zeta t_1^2,\ldots,\zeta t_n^2)=\inf_{\tilde\Lambda} \sum \frac{b_j}{\nu_j}\int\cdots\int y_j \frac{[\int_0^\infty (\tau^2+\zeta t_j^2)^{-1} m_\zeta(y|\tau^2)\, d\tilde\Lambda(\tau^2)]^2}{\int_0^\infty m_\zeta(y|\tau^2)\, d\tilde\Lambda(\tau^2)}\, d\mu(y)$$
$$=\inf_\Lambda \sum \frac{b_j}{\nu_j}\int\cdots\int y_j \frac{[\int_0^\infty (\tau^2+t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)^2}{\int_0^\infty m(y|\tau^2)\, d\Lambda(\tau^2)}\, d\mu(y)=1-V(t_1^2,\ldots,t_n^2).$$

To prove (10), notice that for any prior distribution Λ satisfying (8) as $t_n^2\to\infty$

$$\int m(y|\tau^2)\, d\Lambda(\tau^2)\sim \int m(y_1,\ldots,y_{n-1}|\tau^2)\, d\Lambda(\tau^2)\frac{e^{-y_n/t_n^2}}{t_n^{2\nu_n}}.$$

For $j<n$

$$\int_0^\infty \frac{m(y|\tau^2)\, d\Lambda(\tau^2)}{\tau^2+t_j^2}\sim \int_0^\infty \frac{m(y_1,\ldots,y_{n-1}|\tau^2)\, d\Lambda(\tau^2)}{\tau^2+t_j^2}\frac{e^{-y_n/t_n^2}}{t_n^{2\nu_n}},$$

and

$$\int_0^\infty \frac{m(y|\tau^2)\, d\Lambda(\tau^2)}{\tau^2+t_n^2}\sim \int m(y_1,\ldots,y_{n-1}|\tau^2)\, d\Lambda(\tau^2)\frac{e^{-y_n/t_n^2}}{t_n^{2\nu_n+2}}.$$

Therefore the transformation of variables, $y_n\to t_n^2 y_n$, shows that when $b_n=0$

$$\lim_{t_n^2\to\infty}\sum_{j=1}^{n-1} b_j\int\cdots\int y_j \frac{[\int_0^\infty (\tau^2+t_j^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)]^2}{\int_0^\infty m(y|\tau^2)\, d\Lambda(\tau^2)}\, d\mu(y)$$
$$=\sum_{j=1}^{n-1} b_j\int\cdots\int y_j \frac{[\int_0^\infty (\tau^2+t_j^2)^{-1} m(y_1,\ldots,y_{n-1}|\tau^2)\, d\Lambda(\tau^2)]^2}{\int_0^\infty m(y_1,\ldots,y_{n-1}|\tau^2)\, d\Lambda(\tau^2)}\, d\mu(y_1,\ldots,y_{n-1}). \tag{12}$$

If $b_n > 0$

$$\lim_{t_n^2\to\infty} b_n \int\cdots\int y_n \frac{[\int_0^\infty (\tau^2+t_n^2)^{-1} m(y|\tau^2)\, d\Lambda(\tau^2)]^2}{\int_0^\infty m(y|\tau^2)\, d\Lambda(\tau^2)}\, d\mu(y)$$
$$\le b_n \lim_{t_n^2\to\infty} \int\cdots\int m(y_1,\dots,y_{n-1}|\tau^2)\, d\mu(y_1,\dots,y_{n-1}) \frac{d\Lambda(\tau^2)}{\tau^2+t_n^2} = 0,$$

so that (12) holds in this case as well. Thus

$$\lim_{t_n^2\to\infty} V(t_1^2,\dots,t_n^2) \ge V(t_1^2,\dots,t_{n-1}^2).$$

The reverse inequality in (10) follows from the fact that

$$V(t_1^2,\dots,t_n^2) = \inf_{\delta=\delta(y)} \sup_{\tau^2} R(\delta,\tau^2)$$
$$\le \inf_{\delta=\delta(y_1,\dots,y_{n-1})} \sup_{\tau^2} R(\delta,\tau^2) = V(t_1^2,\dots,t_{n-1}^2).$$

To establish (11), we look at a sequence of probability prior densities $\pi_T(\tau^2) = T^{-1}\pi(\tau^2/T),\ 0 \le \tau^2 \le T$, where π is such a density on the unit interval. As $T\to\infty$

$$\mathcal{M}(Ty) = \int_0^T m(Ty|\tau^2)\pi_T(\tau^2)\left[\sum_j b_j(\tau^2+t_j^2)^{-1}\right]^{-1} d\tau^2$$
$$= \int_0^1 m(Ty|T\sigma)\pi(\sigma)\left[\sum_j b_j(T\sigma+t_j^2)^{-1}\right]^{-1} d\sigma$$
$$\sim \frac{1}{T^{N-1}\sum b_j}\int_0^1 \frac{e^{-\sum y_i/\sigma}\pi(\sigma)\, d\sigma}{\sigma^{N-1}}.$$

Similarly,

$$\int_0^T (\tau^2+t_j^2)^{-1} m(Ty|\tau^2)\pi_T(\tau^2)\left[\sum_j b_j(\tau^2+t_j^2)^{-1}\right]^{-1} d\tau^2$$
$$\sim \frac{1}{T^N\sum b_j}\int_0^1 \frac{e^{-\sum y_i/\sigma}\pi(\sigma)\, d\sigma}{\sigma^N}.$$

Therefore, by changing the integration variable from Ty to y, we see that

$$1 - V(t_1^2,\dots,t_n^2) \le \left[\sum_j b_j\right]^{-1} \sum_j \frac{b_j}{\nu_j}\int y_j \frac{[\int_0^1 e^{-\sum y_i/\sigma}\pi(\sigma)\sigma^{-N}\, d\sigma]^2}{\int_0^1 e^{-\sum y_i/\sigma}\pi(\sigma)\sigma^{1-N}\, d\sigma}\, d\mu(y)$$
$$= \int_0^\infty \frac{[\int_0^1 e^{-u/\sigma}\pi(\sigma)\sigma^{-N}\, d\sigma]^2}{\int_0^1 e^{-u/\sigma}\pi(\sigma)\sigma^{1-N}\, d\sigma}\frac{u^N\, du}{\Gamma(N+1)}.$$

The last integral does not depend on $t_1^2,\dots,t_n^2$. Its value is unity minus the Bayes risk in the estimation problem of an upper-bounded scale parameter σ, $\sigma \le 1$, for gamma family with the shape parameter $N+1,\ N>1$, under the quadratic loss when π is the prior density. According to the results discussed in Section 2, the smallest value of this integral taken over all probability densities π is $\Gamma^2(N)/[\Gamma(N-1)\Gamma(N+1)] = (N-1)/N$. For example, the prior density, $\varepsilon/\sigma^{1+\varepsilon}, 0<\sigma\le 1,\ \varepsilon>0$, makes these values ε-close.

The comparison of V and of the R-risk for the trivial estimator, $\delta_j \equiv 0$, demonstrates the remaining inequality in (11). □

In the proof of (9), the coefficients b_j were assumed to be fixed constants which do not depend on $t_1^2,\dots,t_n^2$. The proof remains valid if all functions $b_j = b_j(t_1^2,\dots,t_n^2),\ j=1,\dots,n$, are homogeneous of the same degree. This degree is two in the mentioned meta-analysis applications (Rukhin, 2014).

We conjecture that for all positive $t_1^2,\dots,t_n^2$, $V=\min(1,N^{-1})$.

4. Quadrature formulas and *T*-systems

Section 2 suggests the form of the least favorable prior distribution, or rather of the sequence

$$d\Pi_\varepsilon(\tau^2) \propto \sum_j \frac{b_j\pi_\varepsilon(\tau^2)\, d\tau^2}{\tau^2+t_j^2}, \tag{13}$$

where as $\varepsilon\to 0,\ \pi_\varepsilon(\tau^2)\to 1$, so that $\int(\tau^2+t^2)^{-1}\pi_\varepsilon(\tau^2)\, d\tau^2 < \infty$ if $t^2>0$. For example, one can take $\pi_\varepsilon(\tau^2) \propto (\tau^2+t^2)^{-\varepsilon}$ with a positive t^2. Then $d\Lambda(\tau^2) = d\Lambda_\varepsilon(\tau^2) = [\sum_j b_j \int \pi_\varepsilon(\tau^2)(\tau^2+t_j^2)^{-1}\, d\tau^2]^{-1}\pi_\varepsilon(\tau^2)\, d\tau^2$.

The proper Bayes estimator, $\int_0^\infty (\tau^2+t_j^2)^{-1} m(y|\tau^2)\pi_\varepsilon(\tau^2)\, d\tau^2 / \int_0^\infty m(y|\tau^2)\pi_\varepsilon(\tau^2)\, d\tau^2$, as $\varepsilon \to 0$ tends to the generalized Bayes estimator

$$\delta_j^B = \delta_j^B(y; t_1^2, \ldots, t_n^2)) = \frac{\int_0^\infty (\tau^2+t_j^2)^{-1} m(y|\tau^2)\, d\tau^2}{\int_0^\infty m(y|\tau^2)\, d\tau^2} = \frac{\nu_j}{y_j} \frac{\int_0^\infty m_j(y|\tau^2)\, d\tau^2}{\int_0^\infty m(y|\tau^2)\, d\tau^2}, \tag{14}$$

with $m(y|\tau^2)$ and $m_j(y|\tau^2)$ defined by (3) and (4) respectively. Strictly speaking, δ_j must be evaluated only when $b_j > 0$. However to check numerical accuracy it is convenient to have δ_j^B available for all $j = 1, \ldots, n$.

Admissibility of δ^B under the risk (2) can be proven by the standard (Blyth) method via approximating its Bayes risk for the prior $\pi_\varepsilon(\tau^2)$ as above. See van Eeden (1995) for a similar proof when $t_i^2 \equiv t^2$ and δ^B coincides with (5).

For large values of y's, say, if $\sum_i y_i = TY$ with a fixed positive Y and $T \to \infty$, one has

$$\lim_{T\to\infty} T\delta_j^B(y) = \int_0^\infty \frac{e^{-Y/v}\, dv}{v^{N+1}} \left[\int_0^\infty \frac{e^{-Y/v}\, dv}{v^N} \right]^{-1} = \frac{N-1}{Y},$$

so that $\delta_j^B(y) \sim (N-1)/\sum_i y_i$. It is known that under the risk R, $N-1$ is the optimal multiple of $(\sum_i y_i)^{-1}$ for large τ^2, and $R(\delta^B, \tau^2) \to N^{-1}$ as $\tau^2 \to \infty$ (Rukhin, 2014).

The values of δ^B at the origin, $y=0$, can be obtained from the multivariate hypergeometric functions (NIST Digital Library of Mathematical Functions, Section 19.16.9) which are defined as Dirichlet averages. More precisely, if the Dirichlet distribution D_ν over the unit simplex has the parameter vector $\nu = (\nu_1, \ldots, \nu_n)$, let for $z = (z_1, \ldots, z_n)$

$$R_{-1}(\nu; z) = \int \left(\sum z_i \omega_i\right)^{-1} dD_\nu(\omega),$$

be such an average (of the function x^{-1}). Here integration is over the unit simplex whose points are denoted by $\omega = (\omega_1, \ldots, \omega_n)$. Numerical evaluation of Dirichlet averages and their relationship with elliptic integrals is discussed in NIST Digital Library of Mathematical Functions (Section 19.36).

Under this notation

$$\int_0^\infty \frac{du}{\prod_i (u+t_i^2)^{\nu_i}} = \frac{R_{-1}(\nu; t_1^{-2}, \ldots, t_n^{-2})}{(N-1)\prod_i t_i^{2\nu_i}},$$

and for any j

$$\int_0^\infty \frac{du}{(u+t_j^2)\prod_i (u+t_i^2)^{\nu_i}} = \frac{R_{-1}(\nu^{(j)}; t_1^{-2}, \ldots, t_n^{-2})}{N t_j^2 \prod_i t_i^{2\nu_i}} \tag{15}$$

with $\nu^{(j)} = (\nu_1, \ldots, \nu_{j-1}, \nu_j+1, \nu_{j+1}, \ldots, \nu_n)$. Therefore,

$$\delta_j^B(0) = \frac{\int_0^\infty \frac{du}{(u+t_j^2)\prod_i (u+t_i^2)^{\nu_i}}}{\int_0^\infty \frac{du}{\prod_i (u+t_i^2)^{\nu_i}}} = \frac{(N-1)R_{-1}(\nu^{(j)}; t_1^{-2}, \ldots, t_n^{-2})}{N t_j^2 R_{-1}(\nu; t_1^{-2}, \ldots, t_n^{-2})}. \tag{16}$$

The known relations between associated R-functions (NIST Digital Library of Mathematical Functions, Section 19.18) show that

$$\sum \nu_j \delta_j^B(0) = \frac{N-1}{R_{-1}(\nu; t_1^{-2}, \ldots, t_n^{-2})}, \tag{17}$$

and

$$\sum \nu_j t_j^2 \delta_j^B(0) = N-1. \tag{18}$$

To evaluate (14) numerically observe that n functions $(v+t_j^2)^{-1}$ of v, $v \ge 0$, form a T-system (Tchebycheff or Chebyshev system) (Karlin and Studden, 1966). For any distribution W, the point with coordinates $(\int_0^\infty (\tau^2+t_1^2)^{-1}\, dW(\tau^2), \ldots, \int_0^\infty (\tau^2+t_n^2)^{-1}\, dW(\tau^2))$ belongs to the closed cone in the n-dimensional space which is a convex hull of the curve whose coordinates are $((v+t_1^2)^{-1}, \ldots, (v+t_n^2)^{-1})$, $v \ge 0$. Therefore,

$$\left(\int_0^\infty (\tau^2+t_1^2)^{-1}\, dW(\tau^2), \ldots, \int_0^\infty (\tau^2+t_n^2)^{-1}\, dW(\tau^2) \right) = \sum_k A_k((v_k+t_1^2)^{-1}, \ldots, (v_k+t_n^2)^{-1})$$

with some positive coefficients A_k and non-negative distinct nodes v_k, $k = 1, \ldots, K$, $K \le n+1$.

Karlin and Studden (1966, Chapter V, Theorem 7.2) discuss the most important case (the so-called principal representation) when the number K is $\lfloor (n+1)/2 \rfloor$ ($K = n/2$, if n is even; $K = (n+1)/2$ if n is odd, in which case $v_1 = 0$). For small n one may want to expand the original T-system by including additional values of t's.

Suppose that for a positive weight function $w(\tau^2)$ the points v_k and the coefficients A_k are determined so that the (finite) "moments" f_j, $j=1,\dots,n$, have the form

$$f_j = \int_0^\infty \frac{w(\tau^2)\,d\tau^2}{\tau^2+t_j^2} = \sum_k \frac{A_k}{v_k+t_j^2}.$$

Then the numerical integration rule

$$\int_0^\infty f(\tau^2)w(\tau^2)\,d\tau^2 \approx \sum_k A_k f(v_k) \tag{19}$$

is *exact* for all functions f which are linear combination of our T-system, cf. Davis and Rabinowitz (1984, Section 2.7.7) and Karlin and Studden (1966, Chapter IV, Section 8). The moments f_j represent the values of the *Stieltjes* transform of w at $t_1^2,\dots,t_n^2$. As such they determine an orthogonal polynomial sequence obtained from the numerators of the continued fraction expansion of this transform (Dette and Studden, 1997, Section 3.3).

To determine A_k and v_k, $k=1,\dots,K$, say, when $n=2K$, notice that with $P(\tau^2)=\prod_i(\tau^2+t_i^2)$, (19) implies that

$$\int_0^\infty \frac{Q(\tau^2)w(\tau^2)\,d\tau^2}{P(\tau^2)} = \sum_k \frac{A_k Q(v_k)}{P(v_k)}$$

for any polynomial Q of degree $n-1$.

Let $p_0^\star(v),\dots,p_K^\star(v)$ denote the orthogonal polynomials for the weight function, $w(v)/P(v)$, $0\le v<\infty$, with zeros of $p_K^\star(v)$ denoted by $v_1<\cdots<v_K$. According to the well known Gauss quadrature formula applied to this weight function for any polynomial Q of degree $2K-1$

$$\int_0^\infty \frac{Q(\tau^2)w(\tau^2)\,d\tau^2}{P(\tau^2)} = \sum_k a_k Q(v_k),$$

with a_k explicitly given in terms of these polynomials for example in Davis and Rabinowitz (1984, (2.7.8)). The comparison of the last two formulas shows that the nodes v_k in (19) must coincide with the zeros of $p_K^\star(v)$, and $A_k=a_kP(v_k)$.

In particular, applying (19) to the numerator and to the denominator of δ_j^B, one gets an approximate formula for δ^B in (14)

$$\delta_j^B(y) = \frac{\sum_k \frac{a_k}{(v_k+t_j^2)w(v_k)} \prod_i \frac{\exp\{-y_i(v_k+t_i^2)^{-1}\}}{(v_k+t_i^2)^{\nu_i-1}}}{\sum_k \frac{a_k}{w(v_k)} \prod_i \frac{\exp\{-y_i(v_k+t_i^2)^{-1}\}}{(v_k+t_i^2)^{\nu_i-1}}}. \tag{20}$$

This formula has a clear Bayes interpretation, namely (20) represents the Bayes estimator against the discrete prior distribution supported by v_k, $k=1,\dots,K$, whose prior probabilities are proportional to $a_kP(v_k)/w(v_k)$.

Several choices of the weight function $w(v)$ suggest themselves. The most natural may be $w(v)=e^{-v}v^\alpha P(v)$, $\alpha>-1$. Then with ℓ_k, $k=1,\dots,K$, denoting the roots of the K-th associated *Laguerre* polynomial, $p_K^\star(v)=L_K^{(\alpha)}(v)$, $v_k=\ell_k$, and $a_k=\Gamma(K+1)\Gamma(K+\alpha+1)\ell_k/L_{K+1}^{(\alpha)}(\ell_k)$, $k=1,\dots,K$ (Davis and Rabinowitz, 1984, (3.6.6)).

Another choice of the weight function is $w(v)=1/P(v)$, in which case f_i coincides with (15). However then it is more awkward to find the roots v_i of polynomials $p_K^\star(v)$ which are defined only for $K<2N-1$. Solving the simultaneous equations, $f_i=\sum A_k(v_k+t_i^2)^{-1}$, $i=1,\dots,n$, for A_k and v_k may lead to an ill conditioned (Cauchy-type) system. For these reasons this choice is not included in the Monte Carlo results reported in the next section.

As a matter of fact, for numerical evaluation of the Bayes estimator δ^B, it is beneficial by using a monotone transformation to replace the original T-system by a T-system of functions defined on a finite interval, say, $[0,1]$. For example, one can employ a transformation, $t_1^2/(\tau^2+t_1^2)=x$, with $(\tau^2+t_i^2)^{-1}=x/(t_1^2+(t_i^2-t_1^2)x)$, $i=2,\dots,n$.

Then

$$\delta_j^B(y) = \frac{\int_0^1 x^{N-1}[t_1^2+(t_j^2-t_1^2)x]^{-1}\prod_i \exp\left\{-\frac{y_ix}{t_1^2+(t_i^2-t_1^2)x}\right\}[t_1^2+(t_i^2-t_1^2)x]^{-\nu_i}\,dx}{\int_0^1 x^{N-2}\prod_i \exp\left\{-\frac{y_ix}{t_1^2+(t_i^2-t_1^2)x}\right\}[t_1^2+(t_i^2-t_1^2)x]^{-\nu_i}\,dx}$$

$$= \frac{\sum_k b_kx_k[t_1^2+(t_j^2-t_1^2)x_k]^{-1}\prod_i \exp\left\{-\frac{y_ix_k}{t_1^2+(t_i^2-t_1^2)x_k}\right\}[t_1^2+(t_i^2-t_1^2)x_k]^{-\nu_i}}{\sum_k b_k\prod_i \exp\left\{-\frac{y_ix_k}{t_1^2+(t_i^2-t_1^2)x_k}\right\}[t_1^2+(t_i^2-t_1^2)x_k]^{-\nu_i}}, \tag{21}$$

Here b_k and x_k are determined from the orthogonal polynomials on the interval $[0,1]$ corresponding to the weight function x^{N-2}. Then the quadrature formula

$$\int_0^1 x^{N-2} f(x)\,dx = \sum_{k=1}^{K} b_k f(x_k)$$

is exact for all polynomials f of degree $2K$.

After stressing that the procedures (20) and (21) are Bayes rules in their own right, we summarize now the main results of this section.

Proposition 2. *The generalized Bayes estimator (14) is admissible under the risk R in (2). Its values at $y=0$ are given by the formula (16); identities (17) and (18) provide numerical accuracy checks. In the integration rule (19) which is exact for linear combinations of n functions $(\tau^2+t_j^2)^{-1}, j=1,\dots,n$, the nodes v_k can be found as the zeros of $p_K^\star(v)$, where $p_0^\star(v),\dots,p_K^\star(v)$ are the orthogonal polynomials for the weight function, $w(v)/P(v)=w(v)/\prod(v+t_j^2)$, while $A_k=a_kP(v_k)$, $k=1,\dots,K$, $K=\lfloor (n+1)/2\rfloor$, with a_k determined from Davis and Rabinowitz (1984, (2.7.8)). Numerical integration formulas (20) and (21) can be used for evaluation of (14).*

There are other approaches to numerical evaluation of (14) which avoid integration rules altogether. One of them appeals to the generating function for the already mentioned Laguerre polynomials

$$\frac{e^{-xz/(1-z)}}{(1-z)^{\alpha+1}} = \sum_k L_k^{(\alpha)}(x) z^k, \quad |z|<1.$$

After the transformation, $t_n^2/(\tau^2+t_n^2)=z$, $\tau^2+t_i^2=t_n^2/z-(t_n^2-t_i^2)$, $i=1,\dots,n-1$, one obtains by using this formula

$$\int_0^\infty m(y|\tau^2)\,d\tau^2 = \frac{1}{t_n^{2(\nu_n-1)}} \sum_{k_1,\dots,k_{n-1}} \prod_{i=1}^{n-1} \left(1-\frac{t_i^2}{t_n^2}\right)^{k_i} L_{k_i}^{(\nu_i-1)}\left(\frac{y_i}{t_n^2-t_i^2}\right) \times \int_0^1 e^{-zy_n/t_n^2} z^{N-2+\sum k_i}\,dz$$

with a similar expression for $\int_0^\infty m_j(y|\tau^2)\,d\tau^2$. Thus an exact formula

$$\delta_j^B(y) = \frac{\sum_{k_1,\dots,k_{n-1}} \prod_{i=1}^{n-1} \left(1-\frac{t_i^2}{t_n^2}\right)^{k_i} L_{k_i}^{(\nu_i^{(j)}-1)}\left(\frac{y_i}{t_n^2-t_i^2}\right) \int_0^1 e^{-zy_n/t_n^2} z^{N-1+\sum k_i}\,dz}{t_n^2 \sum_{k_1,\dots,k_{n-1}} \prod_{i=1}^{n-1} \left(1-\frac{t_i^2}{t_n^2}\right)^{k_i} L_{k_i}^{(\nu_i-1)}\left(\frac{y_i}{t_n^2-t_i^2}\right) \int_0^1 e^{-zy_n/t_n^2} z^{N-2+\sum k_i}\,dz} \tag{22}$$

is derived. Here $\nu_i^{(j)}$ have the same meaning as in (15). However, the resulting expression (22) is practical only for small values of n, like $n\le 4$. Then it gives good results even in the case when t_1^2 is small which presents numerical instabilities in the original formulation. The simulation results show that to be accurate for larger n values, (22) demands too many terms to be included in the sums. For this reason this method is not considered in the next section.

Another approach which does not use quadratures is the plug-in procedure which in this situation can also be justified by the asymptotic Laplace method. Each of these methods suggests the following approximation:

$$\delta_j^B(y) \approx \frac{1}{\hat\tau^2+t_j^2}, \tag{23}$$

where $\hat\tau^2$ is the maximum likelihood rule, $\hat\tau^2=\arg\max m(y|\tau^2)$. This estimator can be determined by simple iterations as

$$\hat\tau^2 = \frac{\sum_j \frac{y_j-\nu_j t_j^2}{(\hat\tau^2+t_j^2)^2}}{\sum_j \frac{\nu_j}{(\hat\tau^2+t_j^2)^2}}, \tag{24}$$

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

The Laplace method also provides a different procedure

$$\delta_j^B(y) \approx \frac{\nu_j m_j(y|\hat\tau_j^2)}{y_j m(y|\hat\tau^2})\sqrt{\frac{(\log m)''|_{\tau^2=\hat\tau^2}}{(\log m_j)''|_{\tau^2=\hat\tau_j^2}}}$$

Here $\hat\tau_j^2$ is the maximum likelihood rule for the density $m_j(y|\tau^2)$ defined in (4). This estimator can be found from an iteration process like in (24) with ν_j replaced by ν_{j+1}. Of course this approximation tacitly assumes that all maximums are attained at strictly positive τ^2's so that the first derivatives vanish at these points and the second derivatives are negative. A more serious difficulty is that even for moderately large n (say, $n\ge 8$) the likelihood functions $m(y|\hat\tau^2)$ and/or $m_j(y|\hat\tau_j^2)$ can become very small creating numerical instability. Thus this approximation to δ^B is excluded from the simulation results in the next section.

It is more practical to use analogy with (23) and to consider the maximum likelihood estimator which takes into account the form of the loss (1)

$$\delta_j^A(y) = \frac{1}{\hat{\tau}_j^2 + t_j^2}. \tag{25}$$

5. Monte Carlo results

Here we compare several numerical implementations of the Bayes estimator (14), namely, (20) with $w(v) = e^{-v}P(v)$, (21), (23), and (25). All these estimators when written as $\delta(y) = \delta(y; t_1^2, \ldots, t_n^2)$ exhibit the *equivariance* property: $\delta(\zeta y; \zeta t_1^2, \ldots, \zeta t_n^2) = \zeta^{-1}\delta(y; t_1^2, \ldots, t_n^2)$ for any $\zeta > 0$. For such procedures, the risk R is invariant under simultaneous scale change in t_j^2's and τ^2

$$R(\delta(y; \zeta^{-1}t_1^2, \ldots, \zeta^{-1}t_n^2), \zeta\tau^2) = R(\delta(y; t_1^2, \ldots, t_n^2), \tau^2).$$

This invariance allows to assume that $t_1^2 = 1$, and this condition was imposed in the reported simulation results. In addition the choice of uncertainties $t_1^2 = 1, t_2^2, \ldots, t_n^2$ was motivated by a metrology example (Rukhin 2013).

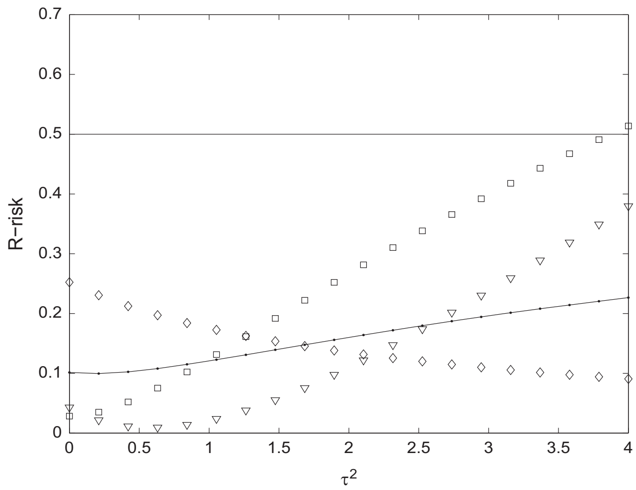

Fig. 1. Plots of R-risks of estimators (20) (line marked by diamonds), (21) (line marked by triangles), (23) (continuous line), and (25) (line marked by squares), when $n=4$, $t_1^2=1$, $t_2^2=3$, $t_3^2=5$, $t_4^2=7.5$. The solid line portrays the value $N^{-1}=0.5$.

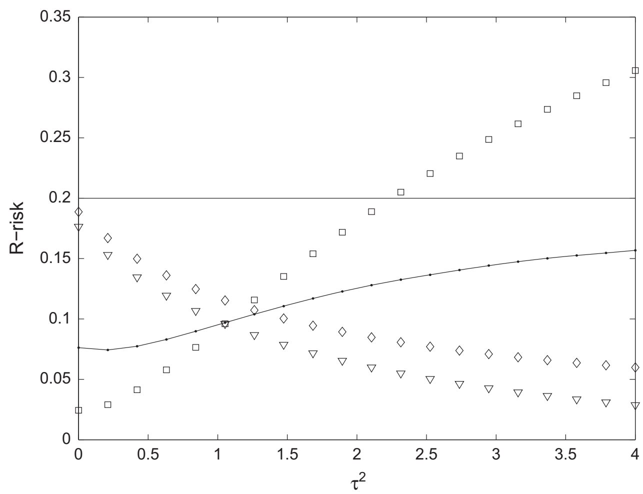

Fig. 2. Plots of R-risks of estimators (20) (line marked by diamonds), (21) (line marked by triangles), (23) (continuous line), and (25) (line marked by squares), when $n=10$, $t_1^2=1$, $t_2^2=3$, $t_3^2=5$, $t_4^2=7.5$, $t_5^2=9$, $t_6^2=11$, $t_7^2=14$, $t_8^2=16$, $t_9^2=19$, $t_{10}^2=22$. The solid line portrays the value $N^{-1}=0.2$.

More precisely, we took $n=4$ and $n=10$ with $\nu_i \equiv 1/2$. In the first case $t_2^2=3$, $t_3^2=5$, $t_4^2=7.5$, in the second the additional values were $t_5^2=9$, $t_6^2=11$, $t_7^2=14$, $t_8^2=16$, $t_9^2=19$, $t_{10}^2=22$, with $b_j \equiv 1$ in both cases. The number of Monte Carlo simulations was 50,000 for $\tau^2=0:0.2:4$. Independent $\Gamma_{1/2}$ random variables were generated and then used for all τ^2 values to obtain $y_j=(\tau^2+t_j^2)\Gamma_{1/2}$ following the remark in Robert and Casella (2004, p. 141).

As Figs. 1 and 2 show, there is no clear winner among these estimators although for large n (21) is somewhat better than (20). The estimator δ^A in (25) exhibits worse performance than (21), (20) or (23) for large τ^2. However for smaller τ^2 it outperforms other estimators. The same pattern was observed under other heterogeneity scenarios and n values.

References

Baringhaus, L., 2003. On unbiased estimation of positive integral powers of the natural parameter in exponential families. Scand. J. Statist. 30, 597–608.
Berger, J.O., 1980. Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. Ann. Statist. 8, 545–571.
DasGupta, A., 1986. Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. Ann. Statist. 14, 206–219.
Davis, P.J., Rabinowitz, P., 1984. Methods of Numerical Integration, 2nd ed. Academic Press, New York.
DeGroot, M.H., 1970. Optimal Statistical Decisions. McGraw Hill, New York.
Dette, H., Studden, W., 1997. The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis. Wiley, New York.
Efron, B., Morris, C., 1973. Stein's estimation rule and its competitors: an empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117–130.
Gajek, L., Kaluszka, M., 1995. Nonexponential applications of a global Cramér–Rao inequality. Statistics 26, 111–122.
Ghosh, M., Parsian, A., 1980. Admissible and minimax multiparameter estimation in exponential families. J. Multivariate Anal. 10, 551–564.
Karlin, S., Studden, W., 1966. Tchebycheff Systems: With Applications in Statistics and Analysis. Interscience, New York.
Marchand, E., Strawderman, W., 2005. On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint. J. Stat. Plan. Inference 134, 90–101.
NIST Digital Library of Mathematical Functions, ⟨http://dlmf.nist.gov⟩.
Prysley, A., Tchonlafi, C., Verbeke, G., Molenberghs, G., 2011. Estimating negative variance components from Gaussian and non-Gaussian data: a mixed models approach. Computat. Statist. Data Anal. 55, 1071–1085.
Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods, 2nd ed. Springer, New York.
Rukhin, A.L., 2012. Estimating common mean and heterogeneity variance in two study case meta-analysis. Statist. Probab. Lett. 82, 1318–1325.
Rukhin, A.L., 2013. Estimating heterogeneity variance in meta-analysis studies. J. R. Statist. Soc., Ser. B 75, 451–469.
Rukhin, A.L., 2014. Restricted likelihood representation and decision-theoretic aspects of meta-analysis. Bernoulli 20. http://dx.doi.org/10.3150/13-BEJ547.
van Eeden, C., 1995. Minimax estimation of a lower-bounded scale parameter of a gamma-distribution for scale-invariant squared-error loss. Canad. J. Statist. 23, 245–256.
Vidakovic, B., DasGupta, A., 1995. Lower bounds on Bayes risks for estimating a normal variance: with applications. Canad. J. Statist. 23, 269–282.