Ann. Inst. Statist. Math.
Vol. 50, No. 4, 773–788 (1998)

ESTIMATION OF THE COEFFICIENT OF MULTIPLE DETERMINATION

NABENDU PAL[1] AND WOOI K. LIM[2]

[1] *Department of Mathematics, University of Southwestern Louisiana, Lafayette, LA 70504, U.S.A.*
[2] *Department of Mathematics, University of New Orleans, New Orleans, LA 70148, U.S.A.*

(Received March 17, 1997; revised August 7, 1997)

Abstract. Assume that we have *iid* observations on the random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ following a multivariate normal distribution $N_p(\mu, \Sigma)$ where both $\mu \in \mathcal{R}^p$ and $\Sigma(p.d.)$ are unknown. Let $\rho_{1\cdot 23\cdots p}$ denote the multiple correlation coefficient between X_1 and $(X_2, \ldots, X_p)'$. The parameter $\lambda = \rho^2_{1\cdot 23\cdots p}$, called the multiple coefficient of determination, indicates the proportion of variability in X_1 explained by its best linear fit based on $(X_2, \ldots, X_p)'$. In this paper we consider the point estimation of λ under the ordinary squared error loss function. The usual estimators (MLE, UMVUE) have complicated risk expressions and hence it is quite difficult to get exact decision theoretic results. We therefore follow the asymptotic decision theoretic approach (as done by Ghosh and Sinha (1981, Ann. Statist., **9**, 1334–1338)) and study 'Second Order Admissibility' of various estimators including the usual ones.

Key words and phrases: Multiple correlation coefficient, loss function, risk function, second order admissibility.

1. Introduction

Consider a p-variate random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ following a multivariate distribution $N_p(\mu, \Sigma)$ where both $\mu \in \mathcal{R}^p$ and Σ (p.d.) are unknown. Let $\boldsymbol{X}_{(2)} = (X_2, \ldots, X_p)'$ and the regression of $\boldsymbol{X}_1$ on $X_{(2)}$ is defined as $E(X_1 | \boldsymbol{X}_{(2)})$. The multiple correlation coefficient between X_1 and $\boldsymbol{X}_{(2)}$, denoted by $\rho_{1\cdot 23\ldots p}$, is the simple correlation coefficient between X_1 and its best linear fit $X_{1\cdot 23\ldots p} = \beta_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ by $\boldsymbol{X}_{(2)}$ where β_i's are true regression coefficients. Define

$$\lambda = \rho^2_{1\cdot 23\cdots p}. \tag{1.1}$$

The parameter λ is called the multiple coefficient of determination and it is the true proportion of the mean regression sum of squares (SSR) to the mean total sum of squares (SSTO), i.e., $\lambda = E[E(X_1|\boldsymbol{X}_{(2)}) - E(X_1)]^2/E(X_1 - E(X_1))^2$. Note that λ is always between 0 and 1, and it indicates the proportion of variability in X_1 explained by its best linear fit $\hat{X}_1$. In this paper we consider the point estimation of λ in a decision theoretic setup

Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ be independent observations on $\boldsymbol{X}$ and define the sample dispersion matrix $\boldsymbol{A}$ as

$$A = \sum_{i=1}^{N} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'. \tag{1.2}$$

Through out this paper we assume that $N > p$ so that A is *p.d.* almost surely. Partition A as

$$A = \begin{bmatrix} a_{11} & \mathbf{a}_{12}' \\ \mathbf{a}_{12} & A_{22} \end{bmatrix}, \tag{1.3}$$

where A_{22} is $(p-1) \times (p-1)$. The sample multiple coefficient of determination is defined as

$$\hat{\lambda}_{o,o} = R = \left| \frac{\mathbf{a}_{12}' A_{22}^{-1} \mathbf{a}_{12}}{a_{11}} \right| \tag{1.4}$$

and the positive square root of R, say $r_{1\cdot 23\cdots p}$, is called the sample multiple correlation coefficient. The usual estimator, which is also the MLE, of $\lambda(\rho_{1\cdot 23\cdots p})$ is $R(r_{1\cdot 23\cdots p})$. The sampling distribution of R has been studied by many authors (see Gurland (1968), Muirhead (1982)) and the *pdf* of R is given as

$$\begin{aligned} (1.5)\quad & f(R|\lambda, n) \\ &= \frac{\Gamma(n/2)}{\Gamma((p-1)/2)\Gamma((n-p+1)/2)} R^{(p-3)/2}(1-R)^{(n-p-1)/2}(1-\lambda)^{n/2} \\ &\quad \times {}_2F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{(p-1)}{2}; \lambda R\right), \end{aligned}$$

where $0 < R < 1, n = N - 1$ and ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ is a Gaussian hypergeometric function given as

$${}_2F_1(a, b; c; z) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)\Gamma(c)z^j}{\Gamma(a)\Gamma(b)\Gamma(c+j)j!}.$$

Let k be any real number such that $(p + 2k) > 1$. The k-th moment of R is

$$E(R)^k = E_J \left\{ \frac{\Gamma\left(\frac{n}{2} + J\right) \Gamma\left(\frac{p + 2k - 1}{2} + J\right)}{\Gamma\left(\frac{n + 2k}{2} + J\right) \Gamma\left(\frac{p-1}{2} + J\right)} \right\}, \tag{1.6}$$

where the random variable J follows *Negative Binomial* $(n/2, (1-\lambda))$ distribution. Also see the expression (29) in Muirhead (1982) for details.

The MLE $\hat{\lambda}_{o,o} = R$ has mean and variance given as

$$E(R) = \lambda + \frac{p-1}{n}(1-\lambda) - \frac{2}{n+2}\lambda(1-\lambda) + O(n^{-2}); \quad \text{and} \tag{1.7}$$
$$\text{Var}(R) = \frac{4\lambda(1-\lambda)^2(n-p+1)^2}{n(n+2)(n+4)} + O(n^{-2}).$$

Obviously the MLE is biased and its bias function is

$$\text{Bias } (\hat{\lambda}_{o,o}) = \frac{(p-1-2\lambda)(1-\lambda)}{n} + O(n^{-2}). \tag{1.8}$$

Olkin and Pratt (1958) have shown that the UMVUE of λ is

$$\begin{aligned} \hat{\lambda}_u &= 1 - \frac{(n-2)}{(n-p+1)}(1-R)\,{}_2F_1\left(1,1;\frac{(n-p+3)}{2};1-R\right) \\ &= R - \frac{(p-3)}{(n-p+1)}(1-R) \\ &\quad - \frac{2(n-2)}{(n-p+1)(n-p+3)}(1-R)^2 + O(n^{-2}). \end{aligned} \tag{1.9}$$

Hence a first order approximation of $\hat{\lambda}_u$ is

$$\begin{aligned} \hat{\lambda}_{u(1)} &= R - \frac{(p-3)}{(n-p+1)}(1-R) - \frac{2(n-2)}{(n-p+1)(n-p+3)}(1-R)^2 \\ &\approx R - \frac{1}{n}(1-R)\{(p-3) + 2(1-R)\}. \end{aligned} \tag{1.10}$$

Using the result (see Muirhead (1982)) that

$${}_2F_1(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \tag{1.11}$$

for $c \neq 0, -1, -2, \ldots$, and $c > (a+b)$, one can show that at $R = 0$, $\hat{\lambda}_u = -(p-1)/(n-p-1)$; i.e., the unbiased estimator $\hat{\lambda}_u$ can take a negative value even though λ is always nonnegative. Therefore, the truncated version of $\hat{\lambda}_u$, say $\hat{\lambda}_u^+$, given as $\hat{\lambda}_u^+ = \max(0, \hat{\lambda}_u)$ is more desirable than $\hat{\lambda}_u$. Similarly, $\hat{\lambda}_{u(1)}^+$ is preferable to $\hat{\lambda}_{u(1)}$.

It appears that not much work has been done so far on estimation of λ from a decision theoretic point of view. The main objective of this paper is to investigate the properties of various estimators of λ in a decision theoretic setup under the squared error loss function.

$$L(\hat{\lambda}, \lambda) = (\hat{\lambda} - \lambda)^2. \tag{1.12}$$

The performance of an estimator $\hat{\lambda}$ is evaluated based on its average loss, or risk, which is defined as $R(\hat{\lambda}, \lambda) = EL(\hat{\lambda}, \lambda)$. It is found that the risk expressions of the usual estimators (MLE, first order UMVUE etc.) are extremely complicated and hence it is nearly impossible to apply the standard decision theoretic tools (for example, Blyth's (1951) limiting Bayes method, Karlin's (1958) generalized Bayes approach, etc.) to check admissibility and/or minimaxity of these estimators under the above mentioned loss. To avoid this difficulty we use the asymptotic decision theoretic approach of Ghosh and Sinha (1981) and study second order admissibility/inadmissibility of various estimators of λ.

For two given estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ of the parameter λ, $\hat{\lambda}_1$ is called second order better than $\hat{\lambda}_2$ under the squared error loss function provided

$$E(\hat{\lambda}_1 - \lambda)^2 \leq E(\hat{\lambda}_2 - \lambda)^2 \quad \text{up to} \quad o(n^{-2}) \quad \forall \quad \lambda, \tag{1.13}$$

and strict inequality for at least one parameter point (i.e., we judge an estimator by its approximated risk after neglecting the $o(n^{-2})$ terms of the risk). An estimator $\hat{\lambda}_2$ is called second order admissible (SOA) if there doesn't exist any other estimator $\hat{\lambda}_1$ satisfying (1.13).

For notational simplicity we will use $R_2(\hat{\lambda}, \lambda)$ to denote the second order risk of the estimator $\hat{\lambda}$, i.e., $R(\hat{\lambda}, \lambda) = R_2(\hat{\lambda}, \lambda) + o(n^{-2})$.

In the case of one parameter (say, θ) exponential families Ghosh and Sinha (1981) considered estimators of the form $\hat{\theta}_c = \hat{\theta}_{\text{MLE}} + c(\hat{\theta}_{\text{MLE}})/n$ for the parameter θ (where $c(\cdot)$ admits Taylor series expansion at θ) and gave necessary and sufficient conditions under which $\hat{\theta}_c$ could be SOA. These conditions are similar in nature to Karlin's (1958) sufficient conditions for admissibility of generalized Bayes estimators. To prove admissibility of a given generalized Bayes estimator Karlin's (1958) result essentially puts restrictions on the tail behavior of the corresponding improper prior, whereas to prove second order admissibility Ghosh and Sinha's (1981) result puts restrictions on the bias function of the estimator under consideration. Ghosh and Sinha (1981) also indicated how to construct a new improved SOA estimator if a given estimator turns out to be second order inadmissible.

In the next section (Section 2) we consider second order point estimation of λ under the loss function (1.12). The main result here is that the MLE turns out to be second order inadmissible when p, the dimension of the underlying multivariate normal distribution, is greater than 7. For $2 \leq p \leq 7$, the MLE is SOA. In Section 3 we have produced two classes of simple estimators which are uniformly better than the MLE for suitable dimensions and these improved estimators are obtained without using the technique suggested by Ghosh and Sinha (1981). Interestingly, the first order unbiased estimator $\hat{\lambda}_{u(1)}$ is SOA always. We have also presented some numerical results to compare various improved estimators.

2. Estimation of λ

Motivated by the structure of the first order unbiased estimator we consider a class of estimators of the form

$$\hat{\lambda}_{a,b} = R - \frac{1}{n}(1 - R)\{a + b(1 - R)\}, \tag{2.1}$$

where a and b are real numbers (of order $O(1)$ as $n \to \infty$). Note that $a = b = 0$ gives the MLE and $a = (p - 3), b = 2$ gives a slight variation of the first order unbiased estimator of λ. In Subsection 2.1 we compute the asymptotic bias and variance of $\hat{\lambda}_{a,b}$. In Subsection 2.2, we address the second order admissibility of $\hat{\lambda}_{a,b}$.

2.1 *Asymptotic bias and variance of* $\hat{\lambda}_{a,b}$

First we derive the asymptotic bias of $\hat{\lambda}_{a,b}$. The bias of $\hat{\lambda}_{a,b}$ is given as

$$\text{Bias of } \hat{\lambda}_{a,b} = E(R-\lambda) - \frac{a}{n}E(1-R) - \frac{b}{n}E(1-R)^2. \tag{2.2}$$

It can be shown that

$$E(R-\lambda) = 1 - \frac{(n-p+1)}{n}(1-\lambda)_2F_1(1,1,(n/2)+1;\lambda) - \lambda. \tag{2.3}$$

Expanding the Gaussian hypergeometric function as an infinite series, and after some simplifications we get

$$E(R-\lambda) = \frac{(1-\lambda)(p-1-2\lambda)}{n} + \frac{2\lambda(1-\lambda)}{n^2}((p+1)-8\lambda) + o(n^{-2}). \tag{2.4}$$

Thus the bias of $\lambda_{a,b}$ is

$$\begin{aligned} \text{Bias of } \hat{\lambda}_{a,b} = {} & \frac{(1-\lambda)(p-1-2\lambda)}{n} \\ & - \frac{a}{n}\frac{(n-p+1)}{n}(1-\lambda)\ {}_2F_1(1,1;(n/2)+1;\lambda) \\ & - \frac{b}{n}\frac{(n-p+1)(n-p+3)}{n(n+2)} \\ & \quad \cdot (1-\lambda)^2\ {}_2F_1(2,2;(n/2)+2;\lambda) + O(n^{-2}). \end{aligned} \tag{2.5}$$

Again expanding the Gaussian hypergeometric functions and after some simplifications we get

$$\begin{aligned} \text{Bias of } \hat{\lambda}_{a,b} &= \frac{(1-\lambda)(p-1-2\lambda)}{n} \\ & \quad - \frac{a}{n}\frac{(n-p+1)}{n}(1-\lambda)\{1+O(n^{-1})\} \\ & \quad - \frac{b}{n}\frac{(n-p+1)(n-p+3)}{n(n+2)}(1-\lambda)^2\{1+O(n^{-1})\} \\ & \quad + O(n^{-2}) \\ &= \frac{(1-\lambda)}{n}(p-1-a-b) + \frac{(1-\lambda)}{n}\lambda(b-2) + O(n^{-2}). \end{aligned} \tag{2.6}$$

Note that (1.8) follows immediately from (2.6) by using $a = b = 0$. Next we compute $E(R-\lambda)^2$. Observe that

$$E(R-\lambda)^2 = E(R^2) - 2\lambda E(R) + \lambda^2.$$

We can write

$$(2.7)\quad E(R-\lambda)^2 = \frac{(n-p+1)(n-p+3)}{n(n+2)}(1-\lambda)^2\ {}_2F_1(2,2;(n/2)+2;\lambda) - 2\frac{(n-p+1)}{n}(1-\lambda)\ {}_2F_1(1,1;(n/2)+1;\lambda)+1 - 2\lambda\left\{1-\frac{(n-p+1)}{n}(1-\lambda)\ {}_2F_1(1,1;(n/2)+1;\lambda)\right\} + \lambda^2.$$

After some simplifications (see Appendix A.1),

$$(2.8)\qquad E(R-\lambda)^2 = \frac{4\lambda(1-\lambda)^2}{n} + \frac{(p+1)(p-1)(1-\lambda)^2}{n^2} - \frac{12(p+1)\lambda(1-\lambda)^2}{n^2} + \frac{56\lambda^2(1-\lambda)^2}{n^2} + o(n^{-2}).$$

Note that

$$(2.9)\qquad \mathrm{Var}(R) = \mathrm{Var}(R-\lambda) = E(R-\lambda)^2 - [E(R-\lambda)]^2 = \frac{4\lambda(1-\lambda)^2}{n} + O(n^{-2}).$$

Second order variance or risk of $\hat\lambda_{a,b}$ can be derived easily from (2.6), (2.9) and using the moment expressions in Appendix A.2. Risk of $\hat\lambda_{a,b}$ is also presented in Section 3.

2.2 *Second order admissibility of* $\hat\lambda_{a,b}$

Consider the class of invariant estimators of λ of the form

$$(2.10)\qquad \hat\lambda(c) = R + \frac{c(R)}{n}$$

where the function $c(\cdot)$ is assumed to admit a Taylor series expansion at λ. Given an estimator $\hat\lambda(c)$, we can consider another estimator $\hat\lambda(d) = R + d(R)/n$ (in the same class (2.10)) and the risk difference (RD) between the risks of these two estimators is defined as

$$\mathrm{RD} = R(\hat\lambda(d),\lambda) - R(\hat\lambda(c),\lambda).$$

Under the ordinary quadratic loss function (1.12), the risk difference can be expressed as

$$\mathrm{RD} = E\left[\left\{R+\frac{d(R)}{n}-\lambda\right\}^2 - \left\{R+\frac{c(R)}{n}-\lambda\right\}^2\right] = E\left[\left\{2(R+\frac{c(R)}{n}-\lambda)+\frac{g(R)}{n}\right\}\left\{\frac{g(R)}{n}\right\}\right],$$

where $g(R) = d(R) - c(R)$. Using Taylor series expansions of $g(R)$ and $c(R)$ at λ and after some algebraic simplifications we obtain

$$\text{RD} = \frac{1}{n^2}[g^2(\lambda) + 2g(\lambda)b(\lambda) + 2g'(\lambda)(4\lambda(1-\lambda)^2)] + o(n^{-2}),$$

where $b(\lambda)/n$ is the bias of $\{R + c(R)/n\}$ up to $o(n^{-1})$. Therefore, the estimator $\hat{\lambda}(d)$ is second order better than $\hat{\lambda}(c)$ provided

$$g^2(\lambda) + 2g(\lambda)b(\lambda) + 2g'(\lambda)(4\lambda(1-\lambda)^2) \leq 0 \quad \forall\lambda, \tag{2.11}$$

with strict inequality for at least one λ. Following the proof of Ghosh and Sinha (1981), the only solution to the inequality (2.11) is $g(\lambda) \equiv 0$ if and only if the following two conditions are satisfied for some $\lambda_o \in (0,1)$:

(2.12)

$$\text{(a)} \int_{\lambda_o}^{1} \left[\exp\left\{-\int_{\lambda_o}^{\lambda} b(u)(4u(1-u)^2)^{-1}du\right\}\right] / (4\lambda(1-\lambda)^2)d\lambda = \infty \quad \text{and}$$

$$\text{(b)} \int_{0}^{\lambda_o} \left[\exp\left\{\int_{\lambda}^{\lambda_o} b(u)(4u(1-u)^2)^{-1}du\right\}\right] / (4\lambda(1-\lambda)^2)d\lambda = \infty.$$

In the light of above two conditions we now study the second order admissibility of a simple class of estimators of the form

$$\hat{\lambda}_{a,b} = R - \frac{1}{n}(1-R)\{a + b(1-R)\}. \tag{2.13}$$

It is enough to check the above two conditions (2.12) for the estimator $\hat{\lambda}_{a,b}$. Substituting the asymptotic bias expression of $\hat{\lambda}_{a,b}$ in (2.6) we get

(2.14) LHS of condition (a)

$$= \int_{\lambda_o}^{1} \{4\lambda(1-\lambda)^2\}^{-1}$$

$$\times \exp\left\{-\int_{\lambda_o}^{\lambda}[(p-1-a-b)(1-u)]/[4u(1-u)^2]\,du - \int_{\lambda_o}^{\lambda}[u(1-u)(b-2)]/[4u(1-u)^2]\,du\right\} d\lambda$$

$$= \int_{\lambda_o}^{1} \frac{k_2^*}{\lambda^{(3+p-a-b)/4}(1-\lambda)^{(11-p+a)/4}}\,d\lambda,$$

where k_2^* is a suitable positive constant. The convergence or divergence of the last integral is not affected by the term $\lambda^{-(3+p-a-b)/4}$ for any $\lambda \in (\lambda_o, 1)$. Therefore, condition (a) holds provided

$$\int_{\lambda_o}^{1} (1-\lambda)^{-(11-p+a)/4}\,d\lambda = \infty$$

which is true provided $11 - p + a \geq 4$; i.e,

$$\text{LHS of condition (a)} = \infty \quad \Leftrightarrow \quad p \leq a + 7. \tag{2.15}$$

Similarly,

$$\text{LHS of condition (b)} = \int_0^{\lambda_o} \frac{k_1^*}{\lambda^{(3+p-a-b)/4}(1-\lambda)^{(11-p+a)/4}}\, d\lambda, \tag{2.16}$$

where k_1^* is a suitable positive constant. Again

$$\text{LHS of condition (b)} = \infty \quad \Leftrightarrow \quad p \geq 1 + a + b. \tag{2.17}$$

The following results is now obvious from the above derivations.

THEOREM 2.1. *For estimating λ, an estimator $\hat{\lambda}_{a,b}$ of the form* (2.1) *is SOA in the class* (2.10) *provided* $1 + a + b \leq p \leq a + 7$.

PROOF OF THEOREM 2.1. Clear from the inequality (2.11) and combining (2.15) and (2.17).

The following corollary now follows easily from the above theorem.

COROLLARY 2.1. (a) *The MLE $\hat{\lambda}_{o,o}$ is SOA in the class* (2.10) *provided the dimension is less than 8, i.e., $2 \leq p \leq 7$. For $p \geq 8$, the estimator $\hat{\lambda}_{o,o}$ is second order inadmissible.*

(b) *If we take $a = (p-3)$ and $b = 2$, then the estimator $\hat{\lambda}_{(p-3),2}$, a slight variation of the first order unbiased estimator $\hat{\lambda}_{u(1)}$, is SOA in the class* (2.10) *for any dimension $p \geq 2$.*

Remark 2.1. The result in Corollary 2.1 (a) came as a surprise to us. It is similar to Stein's (1956) result on a multivariate normal mean estimation where the MLE is inadmissible for $p \geq 3$. In the present problem, the critical dimension is 8. When $p \geq 8$, the MLE $\hat{\lambda}_{o,o}$ is second order inadmissible.

3. Improvements over the MLE

Following the proof of Ghosh and Sinha (1981), it is now possible to find an estimator $\hat{\lambda}(d) = R + d(R)/n$ which is second order better than $\hat{\lambda}_{o,o}$ for $p \geq 8$. Define $h_\psi(\lambda)$ as

$$h_\psi(\lambda) = -\frac{q(\lambda)}{i(\lambda)}\left[\frac{1}{2}\int_\lambda^1 \left\{\frac{i^2(u)}{q(u)} - \psi(u)\right\}du\right] \quad \text{where} \tag{3.1}$$

$$i(\lambda) = \frac{1}{4\lambda(1-\lambda)^2} \quad \text{and} \quad q(\lambda) = i(\lambda)\exp\left\{\int_{\lambda_o}^{\lambda} b_o(u)i(u)\ du\right\},$$

$b_o(\lambda)/n$ being the first order bias of $\hat{\lambda}_{o,o}$, and the function $\psi(u)$ is a negative function which is continuous and integrable over $(0, u)$ $\forall$ $0 < u < 1$. If we write $d_\psi(u) = h_\psi^{-1}(u)$ then the estimator $\hat{\lambda}(d_\psi) = R + d_\psi(R)/n$ is second order better than $\hat{\lambda}_{o,o}$ for $p \geq 8$. Ghosh and Sinha (1981) also indicated how to choose the function $\psi(\cdot)$ (in (3.1)) so that the resultant estimator $\hat{\lambda}(d_\psi)$ is SOA in the subclass (2.10). Unfortunately, in our present problem such an improved estimator $(\hat{\lambda}(d_\psi))$ has a very complicated risk expression and therefore it is impossible to study the risk improvement (second order) attained by $\hat{\lambda}(d_\psi)$.

An important contribution of this paper is that we have found very simple estimators of λ which are uniformly second order better than $\hat{\lambda}_{o,o}$ (for $p > 8$) and some of them are also SOA in the class (2.10). This is shown in the following.

The risk expression of $\hat{\lambda}_{a,b}$ is

$$\begin{aligned}
R(\hat{\lambda}_{a,b}, \lambda) &= E(R-\lambda)^2 + \frac{2a}{n}E(1-R)^2 - \frac{2a}{n}(1-\lambda)E(1-R) \\
&\quad + \frac{2b}{n}E(1-R)^3 - \frac{2b}{n}(1-\lambda)E(1-R)^2 + \frac{a^2}{n^2}E(1-R)^2 \\
&\quad + \frac{2ab}{n^2}E(1-R)^3 + \frac{b^2}{n^2}E(1-R)^4 \\
&= \frac{1}{n}4\lambda(1-\lambda)^2 + \frac{1}{n^2}(p^2-1)(1-\lambda)^2 - \frac{1}{n^2}12(p+1)\lambda(1-\lambda)^2 \\
&\quad + \frac{1}{n^2}56\lambda^2(1-\lambda)^2 - \frac{1}{n^2}4a(p-1)(1-\lambda)^2 + \frac{1}{n^2}16a\lambda(1-\lambda)^2 \\
&\quad + \frac{1}{n^2}2a(p-1)(1-\lambda)^2 - \frac{1}{n^2}4a\lambda(1-\lambda)^2 - \frac{1}{n^2}6b(p-1)(1-\lambda)^3 \\
&\quad + \frac{1}{n^2}36b\lambda(1-\lambda)^3 + \frac{1}{n^2}4b(p-1)(1-\lambda)^3 - \frac{1}{n^2}16b\lambda(1-\lambda)^3 \\
&\quad + \frac{1}{n^2}a^2(1-\lambda)^2 + \frac{1}{n^2}2ab(1-\lambda)^3 + \frac{1}{n^2}b^2(1-\lambda)^4 + o(n^{-2}).
\end{aligned}$$

After much simplifications we get

$$\begin{aligned}
(3.2)\quad & R(\hat{\lambda}_{a,b}, \lambda) \\
&= \frac{(1-\lambda)^2}{n^2}\{4n\lambda + (p^2-1) - 12(p+1)\lambda + 56\lambda^2 \\
&\qquad -2a(p-1) + 12a\lambda + a^2 - 2b(p-1)(1-\lambda) \\
&\qquad + 20b\lambda(1-\lambda) + 2ab(1-\lambda) + b^2(1-\lambda)^2\} \\
&\quad + o(n^{-2}).
\end{aligned}$$

We will look for improved estimators of λ in the following two subclasses

$$(3.3)\qquad \mathcal{C}_b = \{\hat{\lambda}_{0,b} = \hat{\lambda}_{a,b} \quad \text{with} \quad a = 0 \quad \text{and} \quad b \in \mathcal{R}\},$$

$$(3.4)\qquad \mathcal{C}_a = \{\hat{\lambda}_{a,0} = \hat{\lambda}_{a,b} \quad \text{with} \quad b = 0 \quad \text{and} \quad a \in \mathcal{R}\}.$$

The second order risk of an estimator $\hat{\lambda}_{0,b}$ in $\mathcal{C}_b$ is

$$\begin{aligned}
(3.5)\quad R_2(\hat{\lambda}_{0,b}, \lambda) = \frac{(1-\lambda)^2}{n^2}\{&4n\lambda + (p^2-1) - 12(p+1)\lambda + 56\lambda^2 \\
&-2b(p-1)(1-\lambda) + 20b\lambda(1-\lambda) + b^2(1-\lambda)^2\}.
\end{aligned}$$

Note that

$$\frac{\partial}{\partial b} R_2(\hat{\lambda}_{0,b}, \lambda) = \frac{2(1-\lambda)^3}{n^2}\{b(1-\lambda) + 10\lambda - (p-1)\}$$
$$\geq \quad \text{or} \quad \leq 0 \Leftrightarrow b \quad \geq \quad \text{or} \quad \leq b_*(\lambda)$$

where

$$b_*(\lambda) = \frac{(p-1)-10\lambda}{(1-\lambda)}.$$

Also, the second order risk $R_2(\hat{\lambda}_{0,b}, \lambda)$ is convex in b. It is observed that $b_*^{'}(\lambda) = (p-11)/(1-\lambda)^2$. We now consider the following two cases.

Case (i) $p \leq 10$. Then $b_*(\lambda)$ is strictly decreasing and $\sup_\lambda b_*(\lambda) = (p-1)$. Therefore, for $p \leq 10$, if $b \geq (p-1)$ then the second order risk of $\hat{\lambda}_{0,b}$ is increasing in b and one should prefer $b = (p-1)$ over $b > (p-1)$.

Also, $\inf_\lambda b_*(\lambda) = -\infty$ and as a result we can't find a lower bound of b.

Case (ii) $p \geq 11$. In this case $b_*(\lambda)$ is nondecreasing and $\inf_\lambda b_*(\lambda) = (p-1)$. Hence for $p \geq 11$, $R(\hat{\lambda}_{0,b}, \lambda)$ is decreasing in b and as a result $b = (p-1)$ is preferable over $b < (p-1)$.

The following result is immediate.

THEOREM 3.1. *The estimator* $\hat{\lambda}_{0,(p-1)}$ *is optimal in the subclass* (a) $\{\hat{\lambda}_{0,b} \mid b \leq (p-1)\}$ *for* $p \geq 11$; *and* (b) $\{\hat{\lambda}_{0,b} \mid b \geq (p-1)\}$ *for* $p \leq 10$.

Remark 3.1. From the above theorem (part(a)), it is clear that $\hat{\lambda}_{0,(p-1)}$ is second order better than the MLE ($\hat{\lambda}_{0,b}$ with $b = 0$) for $p \geq 11$. The above result doesn't produce any improved estimator for $8 \leq p \leq 10$ (for $2 \leq p \leq 7$ the MLE is SOA in the class (2.10)).

If we look at the subclass $\mathcal{C}_a$ (in (3.4)) then it is possible to dominate the MLE for any $p \geq 8$. The second order risk of an estimator $\hat{\lambda}_{a,0}$ in $\mathcal{C}_a$ is

$$R_2(\hat{\lambda}_{a,0}, \lambda) = \frac{(1-\lambda)^2}{n^2}\{4n\lambda + (p^2-1) - 12(p+1)\lambda + 56\lambda^2 - 2a(p-1) + 12a\lambda + a^2\}. \tag{3.6}$$

Note that

$$\frac{\partial}{\partial a} R_2(\hat{\lambda}_{a,0}, \lambda) = \frac{2(1-\lambda)^2}{n^2}\left\{a - (p-1) + 6\lambda\right\}$$
$$\geq \text{ or } \leq 0 \quad \Leftrightarrow \quad a \geq \text{ or } \leq a_*(\lambda)$$

where $a_*(\lambda) = (p-1) - 6\lambda$. Second derivative of $R_2(\hat{\lambda}_{a,0}, \lambda)$ with respect to a shows that it is a convex function. Also, $\inf_\lambda a_*(\lambda) = (p-7)$ and $\sup_\lambda a_*(\lambda) = (p-1)$. The following theorem now holds easily.

THEOREM 3.2. *Let $p \geq 8$. The estimator* (a) $\hat{\lambda}_{(p-7),0}$ *is optimal in the subclass* $\{\hat{\lambda}_{a,0} | a \leq (p-7)\}$; *and* (b) $\hat{\lambda}_{(p-1),0}$ *is optimal in the subclass* $\{\hat{\lambda}_{a,0} | a \geq (p-1)\}$.

Remark 3.2. Obviously, the estimator $\hat{\lambda}_{(p-7),0}$ is uniformly second order better than the MLE for any $p \geq 8$.

3.1 *Second order risk improvements due to* $\hat{\lambda}_{(p-7),0}$ *and* $\hat{\lambda}_{0,(p-1)}$ *over the MLE*

From the expression (3.2) it is easy to get the second order risk of the MLE as

$$R_2(\hat{\lambda}_{o,o}, \lambda) = \frac{(1-\lambda)^2}{n^2}\{4n\lambda + (p^2-1) - 12(p+1)\lambda + 56\lambda^2\}. \tag{3.7}$$

Using (3.5)–(3.6) we get the relative risk improvements (RRIs) of $\hat{\lambda}_{0,(p-1)}$ and $\hat{\lambda}_{(p-7),0}$ over MLE. RRI of an estimator $\hat{\lambda}$ over $\hat{\lambda}_{o,o}$ is:
$\text{RRI}(\hat{\lambda}) = [\{R_2(\hat{\lambda}_{o,o}, \lambda) - R_2(\hat{\lambda}, \lambda)\}/R_2(\hat{\lambda}_{o,o}, \lambda)] \times 100\%$.

$$\text{RRI}\,(\hat{\lambda}_{0,(p-1)}) = \frac{(1-\lambda)(p-1)\{(p-1)(1+\lambda) - 20\lambda\}}{\{4n\lambda + (p^2-1) - 12(p+1)\lambda + 56\lambda^2\}} \times 100\% \tag{3.8}$$
$$\to \begin{cases} 0 & \text{as} \quad \lambda \to 1 \\ \dfrac{(p-1)}{(p+1)} \times 100\% & \text{as} \quad \lambda \to 0. \end{cases}$$

Therefore RRI of $\hat{\lambda}_{0,(p-1)}$ can be almost 100 at $\lambda = 0$ for large p. The following table (Table 1) gives the RRI of $\hat{\lambda}_{0,(p-1)}$ for various values of p ($p \geq 11$).

Table 1. RRI of $\hat{\lambda}_{0,(p-1)}$ at $\lambda = 0$.

p	11	15	20	25	50	∞
RRI	83.5%	87.5%	90.4%	92.3%	96.1%	100%

Similarly,

$$\text{RRI}(\hat{\lambda}_{(p-7),0}) = \frac{(p-7)(p+5-12\lambda)}{\{4n\lambda + (p^2-1) - 12(p+1)\lambda + 56\lambda^2\}} \times 100\% \tag{3.9}$$
$$\to \begin{cases} \dfrac{(p-7)(p+5)}{(p^2-1)} \times 100\% & \text{as} \quad \lambda \to 0 \\ \dfrac{(p-7)^2}{\{4n + (p^2-1) - 12(p+1) + 56\}} \times 100\% & \text{as} \quad \lambda \to 1. \end{cases}$$

Near $\lambda = 1$, $\hat{\lambda}_{(p-7),0}$ seems to perform better than $\hat{\lambda}_{0,(p-1)}$ since $\hat{\lambda}_{(p-7),0}$ has nonzero RRI. Again, RRI of $\hat{\lambda}_{(p-7),0}$ approaches 100% at $\lambda = 0$ for large p.

Table 2. RRI of $\hat{\lambda}_{(p-7),0}$ at $\lambda = 0$.

p	8	11	15	20	50	∞
RRI	20.6%	53.3%	71.4%	81.4%	94.6%	100%

3.2 *Are* $\hat{\lambda}_{0,(p-1)}$ *and* $\hat{\lambda}_{(p-7),0}$ *SOA ?*

Since our proposed estimators $\hat{\lambda}_{(p-7),0}$ and $\hat{\lambda}_{0,(p-1)}$ are better than the MLE (for $p \geq 8$ and $p \geq 11$ respectively) and competes with $\hat{\lambda}_{u(1)}$ one might be interested to know whether these two estimators are SOA or not. We can use our earlier Theorem 2.1 to answer this question. An estimator $\hat{\lambda}_{a,b}$ is SOA provided $1 + a + b \leq p \leq a + 7$. The estimator $\hat{\lambda}_{(p-7),0}$ is SOA since ($a = p - 7, b = 0$) satisfies the above condition whereas $\hat{\lambda}_{0,(p-1)}$ is not SOA since ($a = 0, b = (p-1)$) does not satisfy the condition.

Remark 3.3. We can evaluate the exact risk functions of the estimators discussed above (in terms of hypergeometric functions). From numerical calculations it has been observed that the second order risk approximates the exact risk fairly well for $k = n/p \geq 5$.

In Figs. 1–3, we have plotted the exact and second order risk curves of the estimators $\hat{\lambda}_{o,o}$, $\hat{\lambda}_{u(1)}$, $\hat{\lambda}_{(p-7),0}$ and $\hat{\lambda}_{0,(p-1)}$. The values of n and p are selected

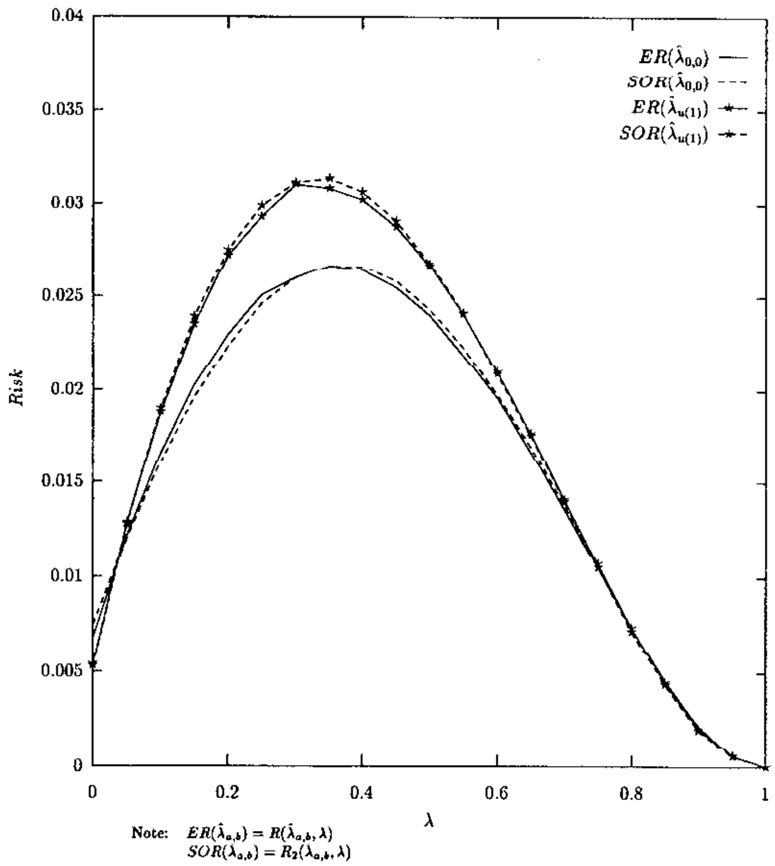

Fig. 1. Exact risks (ERs) and second order risks (SORs) of $\hat{\lambda}_{0,0}$ and $\hat{\lambda}_{u(1)}$ for $p = 2$ and $n = 20$.

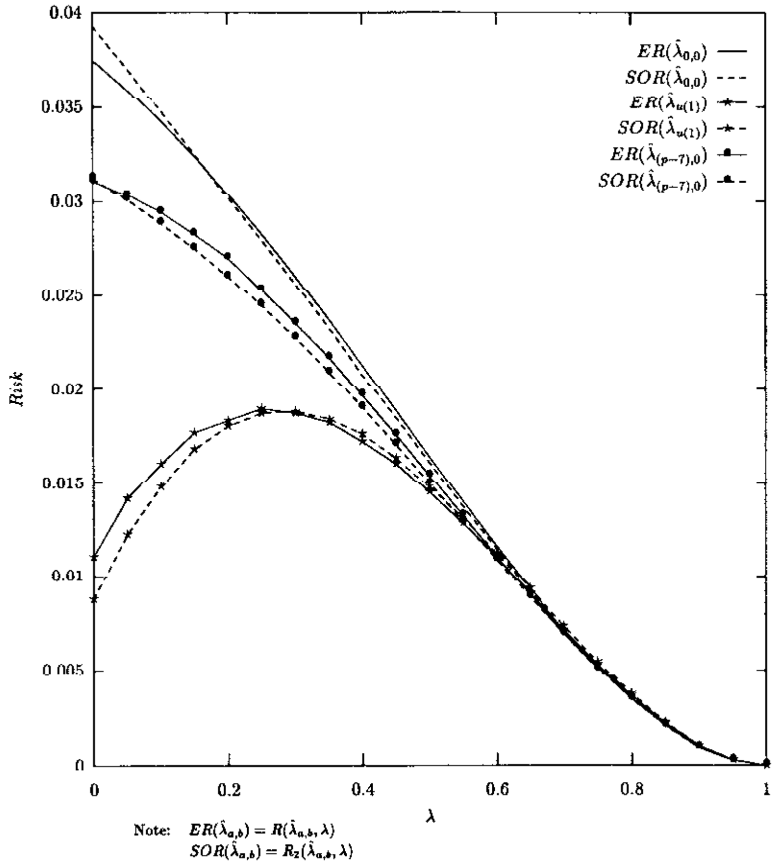

Fig. 2. Exact risks (ERs) and second order risks (SORs) of $\hat{\lambda}_{0,0}$, $\hat{\lambda}_{u(1)}$ and $\hat{\lambda}_{(p-7)}$ for $p = 8$ and $n = 40$.

(for Figs. 1–3) such that $k = n/p \geq 5$. Note that the second order risk curves of $\hat{\lambda}_{(p-7),0}$ and $\hat{\lambda}_{0,(p-1)}$ are available only for $p \geq 8$ and $p \geq 11$ respectively.

Concluding Remark. We have seen that when $2 \leq p \leq 7$, $\hat{\lambda}_{o,o}$ is SOA and $\hat{\lambda}_{u(1)}$ is SOA always (both in the class (2.10)). Numerically we have found that $\hat{\lambda}_{o,o}$ performs quite well compared to $\hat{\lambda}_{u(1)}$ for $2 \leq p \leq 7$. $\hat{\lambda}_{u(1)}$ over takes the MLE only when λ is close to either 0 or 1. For $p \geq 8$, we know that $\hat{\lambda}_{o,o}$ is second order inadmissible and interestingly it is the $\hat{\lambda}_{u(1)}$ which seems to outperform the other estimators, i.e., for large dimensions ($p \geq 8$), a shrinkage estimator ($\hat{\lambda}_{u(1)}$) performs well than the traditional estimator. Note that $\hat{\lambda}_{u(1)}$ can be improved further by $\hat{\lambda}^+_{u(1)} = \max(0, \hat{\lambda}_{u(1)})$. In terms of exact risk $\hat{\lambda}_{u(1)}$ is inadmissible even though it is SOA in the class (2.10). Therefore, the risk difference between $\hat{\lambda}_{u(1)}$ and $\hat{\lambda}^+_{u(1)}$ must be of order $o(n^{-2})$.

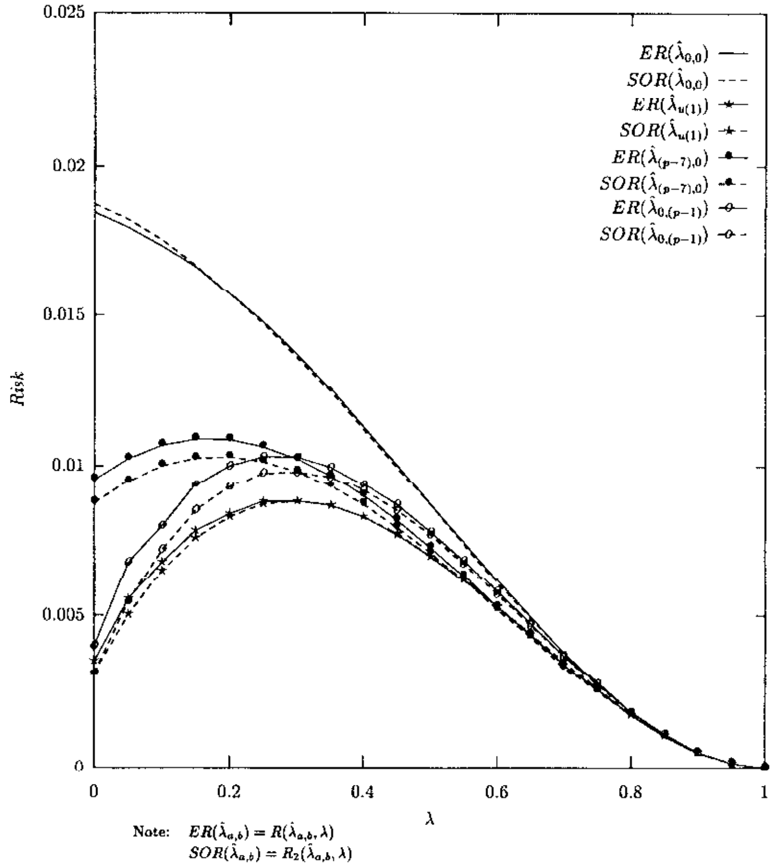

Fig. 3. Exact risks (ERs) and second order risks (SORs) of $\hat{\lambda}_{0,0}, \hat{\lambda}_{u(1)}, \hat{\lambda}_{(p-7)}$ and $\hat{\lambda}_{0,(p-1)}$ for $p = 11$ and $n = 60$.

Based on our work on λ estimation, we propose $\hat{\lambda}_{o,o}$ (MLE) if $2 \leq p \leq 7$ and $\hat{\lambda}_{u(1)}$ (UMVUE) if $p \geq 8$ for practical applications.

Acknowledgements

The authors sincerely thank an anonymous Associate Editor as well as two anonymous referees for valuable comments and suggestions on an earlier version of this article.

Appendix

A.1 *Derivation of* (2.8) *from* (2.7)

$$
\begin{aligned}
E(R-\lambda)^2 &= \frac{(n-p+1)(n-p+3)}{n(n+2)}(1-\lambda)^2{}_2F_1(2,2;(n/2)+2;\lambda)\\
&\quad - 2\frac{(n-p+1)}{n}(1-\lambda){}_2F_1(1,1;(n/2)+1;\lambda)+1\\
&\quad - 2\lambda\left\{1-\frac{(n-p+1)}{n}(1-\lambda){}_2F_1(1,1;(n/2)+1;\lambda)\right\}+\lambda^2.
\end{aligned}
$$

Expanding the hypergeometric functions we get,

$$
\begin{aligned}
E(R-\lambda)^2 &= \frac{(n-p+1)(n-p+3)}{n(n+2)}(1-\lambda)^2\left\{1+\frac{8\lambda}{(n+4)}\right.\\
&\qquad\qquad \left. + \frac{72\lambda^2}{(n+4)(n+6)}+o(n^{-2})\right\}\\
&\quad - \frac{2(n-p+1)}{n}(1-\lambda)^2\left\{1+\frac{2\lambda}{(n+2)}+\frac{8\lambda^2}{(n+2)(n+4)}+o(n^{-2})\right\}\\
&\quad + (1-\lambda)^2\\
&= \frac{(n-p+1)(n-p+3)}{n(n+2)}(1-\lambda)^2+\frac{8(n-p+1)(n-p+3)}{n(n+2)(n+4)}\lambda(1-\lambda)^2\\
&\quad + \frac{72(n-p+1)(n-p+3)}{n(n+2)(n+4)(n+6)}\lambda^2(1-\lambda)^2-\frac{2(n-p+1)}{n}(1-\lambda)^2\\
&\quad - \frac{4(n-p+1)}{n(n+2)}\lambda(1-\lambda)^2-\frac{16(n-p+1)}{n(n+2)(n+4)}\lambda^2(1-\lambda)^2\\
&\quad + (1-\lambda)^2+o(n^{-2})\\
&= \frac{4}{n}\lambda(1-\lambda)^2+\frac{(p+1)(p-1)}{n^2}(1-\lambda)^2-\frac{12(p+1)}{n^2}\lambda(1-\lambda)^2\\
&\quad + \frac{56}{n^2}\lambda^2(1-\lambda)^2+o(n^{-2}).
\end{aligned}
$$

A.2 *Moment expressions for* $(1-R)$.

The general moment expressions for $(1-R)$ is given by (Muirhead (1982))

$$
E(1-R)^h = \frac{[(n-p+1)/2]_h}{(n/2)_h}(1-\lambda)^h{}_2F_1(h,h;(n/2)+h;\lambda),
$$

where h is a positive integer and $(a)_h=(a)(a+1)\cdots(a+h-1)$. For $h=1$,

$$
\begin{aligned}
E(1-R) &= \frac{(n-p+1)}{n}(1-\lambda){}_2F_1(1,1;(n/2)+1;\lambda)\\
&= (1-\lambda)-\frac{(1-\lambda)(p-1-2\lambda)}{n}-\frac{2(p+1)\lambda(1-\lambda)}{n^2}\\
&\quad + \frac{8\lambda^2(1-\lambda)}{n^2}+o(n^{-2}).
\end{aligned}
$$

The second moment is

$$\begin{aligned} E(1-R)^2 &= \frac{(n-p+1)(n-p+3)}{n(n+2)}(1-\lambda)^2{}_2F_1(2,2,(n/2)+2,\lambda) \\ &= (1-\lambda)^2 - \frac{2(p-1)(1-\lambda)^2}{n} + \frac{8\lambda(1-\lambda)^2}{n} + \frac{(p+1)(p-1)(1-\lambda)^2}{n^2} \\ &\quad + \frac{16(p+1)\lambda(1-\lambda)^2}{n^2} + \frac{72\lambda^2(1-\lambda)^2}{n^2} + o(n^{-2}). \end{aligned}$$

The third moment of $(1-R)$ is

$$\begin{aligned} E(1-R)^3 &= \frac{(n-p+1)(n-p+3)(n-p+5)}{n(n+2)(n+4)}(1-\lambda)^3{}_2F_1(3,3;(n/2)+3;\lambda) \\ &= (1-\lambda)^3 - \frac{3(1-\lambda)^3(p-1-6\lambda)}{n} - \frac{3(p-1)(p-3)(1-\lambda)^3}{n^2} \\ &\quad - \frac{108(p+1)\lambda(1-\lambda)^3}{n^2} + \frac{288\lambda^2(1-\lambda)^3}{n^2} + o(n^{-2}). \end{aligned}$$

The fourth moment is (for the risk of $\hat{\lambda}_{a,b}$ we only need $O(1)$ term)

$$\begin{aligned} E(1-R)^4 &= \frac{(n-p+1)(n-p+3)(n-p+5)(n-p+7)}{n(n+2)(n+4)(n+6)} \\ &\quad \cdot (1-\lambda)^4{}_2F_1(4,4;(n/2)+4;\lambda) \\ &= (1-\lambda)^4 + O(n^{-1}). \end{aligned}$$

References

Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility, *Ann. Math. Statist.*, **22**, 22–42.

Ghosh, J. K. and Sinha, B. K. (1981). A necessary and sufficient conditions for second order admissibility with applications to Berkson's bioassay problem, *Ann. Statist.*, **9**, 1334–1338.

Gurland, J. (1968). A relatively simple form of the distribution of the multiple correlation coefficient, *J. Roy. Statist. Soc., Ser. B*, **30**, 276–283.

Karlin, S. (1958). Admissibility for estimation with quadratic loss, *Ann. Math. Statist.*, **29**, 406–436.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Olkin, I. and Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients, *Ann. Math. Statist.*, **29**, 201–211.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution, *Proc. Third Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 197–206, University of California Press, Berkeley.