Ann. Inst. Statist. Math.
Vol. 43, No. 1, 147–156 (1991)

A CLASS OF MULTIPLE SHRINKAGE ESTIMATORS

C. S. WITHERS

Applied Mathematics Division, Department of Scientific and Industrial Research, P. O. Box 1335, Wellington, New Zealand

(Received May 9, 1988; revised December 11, 1989)

Abstract. Based on a sample of size n, we investigate a class of estimators of the mean θ of a p-variate normal distribution with independent components having unknown covariance. This class includes the James-Stein estimator and Lindley's estimator as special cases and was proposed by Stein. The mean squares error improves on that of the sample mean for $p \geq 3$. Simple approximations for this improvement are given for large n or p. Lindley's estimator improves on that of James and Stein if either n is large, and the "coefficient of variation" of θ is less than a certain increasing function of p, or if p is large. An adaptive estimator is given which for large samples always performs at least as well as these two estimators.

Key words and phrases: Shrinkage estimates, multivariate normal, loss.

1. Introduction and summary

Here we state our main results. Proofs are given in Section 2.

Suppose we observe a random sample of size n from $N_p(\theta, vI)$ with $p \geq 3$ and $v > 0$ unknown. Let $\bar{X}$ be the sample mean. Its risk is

$$E|\bar{X} - \theta|^2 = pvn^{-1}.$$

Suppose $\hat{v} \sim v\chi_\nu^2/(\nu + 2)$ independently of $\bar{X}$.

We seek an estimate of θ with smaller risk than that of $\bar{X}$.

This problem arises in 1-way analysis of variance with equal observations per cell and more generally in regression analysis with normal residuals with ν equal to n less a constant.

Let H be any $p \times p$ idempotent matrix of rank $r_H \geq 3$. The estimator for θ that we shall consider is

$$\hat{\theta}_H = \bar{X} - n^{-1}\hat{v}(r_H - 2)H\bar{X}|H\bar{X}|^{-2}. \tag{1.1}$$

Versions of this estimate were proposed in (2.34) of Stein (1966), (4.3) of Sclove *et al.* (1972), (1.6a) of George (1986*b*) and in (1.4) of George (1986*c*).

Unlike these papers our concern is primarily on how the estimator performs as the sample size n increases. Like George, we consider an adaptive version of (1.1). However, our adaptive estimator is much simpler than his.

When $H = I_p$, it is the estimate of James and Stein (1961)

$$\hat{\theta}_{JS} = \bar{X} - n^{-1}\hat{v}(p-2)\bar{X}|\bar{X}|^{-2}.$$

When $H = I_p - \mathbf{1}\ \mathbf{1}'/p$ it is known as Lindley's estimate

$$\hat{\theta}_L = \bar{X} - n^{-1}\hat{v}(p-3)(\bar{X} - \mathbf{1}\bar{X}_.)(|\bar{X}|^2 - p\bar{X}_.^2)^{-2} \quad \text{where} \quad \bar{X}_. = \mathbf{1}'\bar{X}/p.$$

For a numerical example using $\hat{\theta}_L$, see Efron and Morris (1973*a*, 1973*b*).

We shall study the risk of $\hat{\theta}_H$ firstly for large n and then for large p. In either case we shall show that for H_1 and H_2 of the same rank, $\hat{\theta}_{H_1}$ has smaller risk than $\hat{\theta}_{H_2}$ if $|H_1\theta| < |H_2\theta|$. This is not directly helpful as θ is unknown.

However, our main result, Theorem 1.3, tells us to choose $\hat{\theta}_{H_1}$ rather than $\hat{\theta}_{H_2}$ if $|H_1\bar{X}| < |H_2\bar{X}|$.

More generally it gives us a rule to choose H efficiently from a predetermined set—and thus, in general, how to improve on both the James-Stein estimate and the Lindley estimate.

Our first result is well known, although it does not appear to be specifically stated anywhere. It shows that $\hat{\theta}_H$ has smaller risk than $\bar{X}$.

$$\text{(1.2)} \qquad \text{Set} \quad h(\lambda, r) = r^2 E(r+2K)^{-1} \quad \text{for} \quad r \geq 0$$

and K Poisson with mean $\lambda/2 > 0$

$$\lambda_H = nv^{-1}|H\theta|^2 \quad \text{and}$$
$$\Delta(H) = h(\lambda_H, r_H - 2).$$

THEOREM 1.1.

$$\text{(1.3)} \qquad E|\hat{\theta}_H - \theta|^2 = n^{-1}v\{p - b\Delta(H)\} \quad \textit{where} \quad b = \nu/(\nu+2).$$

Note 1.1. This was proved for $H = I$ by James and Stein (1961) and for $H_L = I - \mathbf{1}\ \mathbf{1}'/p$ by Lindley—see the discussion to Stein (1962). We denote their corresponding values of

$$\tau_H = |H\theta|^2/p \quad \text{by} \quad \tau_{JS} = |\theta|^2/p \quad \text{and} \quad \tau_L = p^{-1}\sum_1^p(\theta_i - \bar{\theta})^2$$

where $\bar{\theta} = p^{-1}\sum_1^p \theta_i$. Also $\tau_H \leq \tau_{JS}$. Thus τ_H is bounded as p increases if say $\{\theta_i\}$ are bounded.

Note 1.2. By (1.3), $E|\hat{\theta}_H - \theta|^2/E|\bar{X} - \theta|^2$ lies between $1 - b(r_H - 2)/p$ at $|H\theta| = 0$ and 1 at $|H\theta| = \infty$.

COROLLARY 1.1. *If the direction of θ, $i = \theta/|\theta|$, is* ***known*** *and $H = I - ii'$ then $E|\hat{\theta}_H - \theta|^2 = n^{-1}v(3\nu + 2p)/(\nu+2)$ for all $|\theta|$, so that the risk of $\hat{\theta}_H$ relative to that of $\bar{X}$ is about $3/p$ for ν/p large.*

This is an exceptional situation and will not be referred to again.

The following expressions for $h(\lambda, r)$ are due to Stein (1966) and Efron and Morris (1973*b*).

$$h(\lambda, r) = r^2 e^{-\lambda/2} \sum_{h=0}^{\infty} (r+2h)^{-1}(-\lambda/2)^h/h!, \tag{1.4}$$

$$\begin{aligned} h(\lambda, r) &= r \sum_{h=0}^{\infty} (-\lambda/2)^h \Gamma(r/2+1)/\Gamma(r/2+1+h) \\ &= r\{1 - \lambda(r+2)^{-1} + \lambda^2(r+2)^{-1}(r+4)^{-1} - \cdots\} \\ &= r\,{}_1F_1(1;\ r/2+1;\ -\lambda/2), \end{aligned} \tag{1.5}$$

where ${}_pF_q$ is the hypergeometric function; for $r \geq 0$ an even integer, (1.5) can be written

$$h(\lambda, r) = (r/2)!(-\lambda/2)^{-r/2} \left\{ e^{-\lambda/2} - \sum_{h=0}^{r/2} (-\lambda/2)^h/h! \right\}. \tag{1.6}$$

These expressions and others easily follow by noting that $h(\lambda, r)$ is equivalent to the incomplete gamma function

$$\gamma(a, x) = \int_0^x t^{a-1}e^{-t}dt, \qquad a > 0.$$

THEOREM 1.2.

$$h(\lambda, r) = r^2 e^{-\lambda/2}(-\lambda/2)^{-r/2}\gamma(r/2, -\lambda/2)/2. \tag{1.7}$$

For an approximation to the risk for large n, by (1.3) we need to approximate $h(\lambda, r)$ for large λ. From (6.5.32) of Abramowitz and Stegun (1964) it follows that for $I \geq 1$,

$$\begin{aligned} h(\lambda, r)/r^2 &= -\sum_{i=0}^{I-1} (-\lambda/2)^{-i-1}(r/2-1)_i/2 + R_I \\ &= \lambda^{-1} - (r-2)\lambda^{-2} + (r-2)(r-4)\lambda^{-3} - \cdots + R_I \end{aligned} \tag{1.8}$$

where

$$|R_I| \leq (\lambda/2)^{-I-1}|(r/2-1)_I|/2 \quad \text{and}$$
$$(a)_i = a!/(a-i)! = a(a-1)\cdots(a-i+1).$$

If $I < r/2$,

$$R_I = (-1)^I (r/2-1)_I (\lambda/2)^{-r/2} e^{-\lambda/2} \int_0^{\lambda/2} x^{r/2-I-1} e^x \, dx$$

(this follows from integration by parts).

For $r \geq 0$ even this implies (1.6), while for r odd and $I = (r-1)/2$ it reduces to a result of Egerton and Laycock (1982).

Note 1.3. (1.8) with $I = \infty$ yields the expansion for ${}_2F_0(1;\ 1 - r/2;\ 2/\lambda)$; this is divergent if r is not even.

Since $R_I = O(r^I \lambda^{-I-1})$ as $\lambda/r \to \infty$, (1.8) gives a useful expansion for λ/r large and I fixed.

Copas (1983) gives expressions equivalent to the following approximations ((3.12) and p. 349);

(C1) $\quad h(\lambda,\, r) \doteq r^2/(\lambda + r - 2),$

(C2) $\quad h(\lambda,\, r) \doteq r^2(\lambda + r + 2)/(\lambda^2 + 2r\lambda + r^2 + 2r).$

Table 1 gives $h(\lambda,\, r)/r$ and the deviations from it for (C1), (C2) and (1.8) with $I = 2$ and 3—referred to as (I2) and (I3). For λ/r small (I2) and (I3) perform poorly, as expected.

Table 1. A comparison of 4 approximations to $h(\lambda,\, r)$.

λ	r	$h(\lambda,\, r)/r$	C1	C2	I2	I3
100	3	.030	.000	.000	.000	.000
	8	.075	.000	.000	.000	.000
	18	.155	.000	.000	−.004	.000
20	3	.142	−.001	−.002	.000	.000
	8	.302	−.006	−.002	−.022	.002
	18	.487	−.013	−.001	−.307	.197
10	3	.265	−.008	−.008	.005	.002
	8	.474	−.026	−.003	.154	.038
	18	.659	−.033	.000	−1.74	2.29
5	3	.446	−.054	−.017	−.034	.010
	8	.652	−.075	−.003	−.972	.564
	18	.797	−.060	−.001	−8.72	23.5

COROLLARY 1.2. *Define τ_H, τ_{JS} and τ_L as in Note* 1.1.

$$\begin{aligned}(1.9)\qquad \Delta(H) &= (r_H-2)^2\lambda_H^{-1}\{1+O(r_H/\lambda_H)\}\\ &= v(r_H-2)^2(np\tau_H)^{-1}\{1-v(r_H-2)(np\tau_H)^{-1}+O(n^{-2})\}\\ &= \begin{cases} O(n^{-1}p) & if \quad \tau_H>0\\ r_H-2 & if \quad \tau_H=0.\end{cases}\end{aligned}$$

Hence for b of (1.3) *and $\tau_L>0$*

$$(1.10)\qquad \begin{aligned} &E|\hat\theta_{JS}-\theta|^2 - E|\hat\theta_L-\theta|^2 = pn^{-2}bv^2B_p\{1+O(n^{-1})\} \qquad and\\ &E|\hat\theta_{JS}-\theta|^2/E|\hat\theta_L-\theta|^2 = 1+n^{-1}bvB_p+O(n^{-2})\end{aligned}$$

where $B_p=\tau_L^{-1}(1-3/p)^2-\tau_{JS}^{-1}(1-2/p)^2$; thus $\hat\theta_L$ improves on $\hat\theta_{JS}$ for large n (that is, $B_p>0$) $\Longleftrightarrow$

$$(1.11)\qquad |\theta|/|\bar\theta| < (p-2)/(2-5/p)^{1/2} \Longleftrightarrow$$

$$(1.12)\qquad CV(\theta) < (p-3)(2p-5)^{-1/2}$$

where $CV(\theta)=\tau_L^{1/2}/|\bar\theta|$, the "coefficient of variation of θ"; if $<$ in (1.12) *is replaced by $>$, then $\hat\theta_{JS}$ improves on $\hat\theta_L$ for large n. Similarly, if $\tau_{JS}>0$ and $\tau_H>0$ then $\hat\theta_H$ improves on $\hat\theta_{JS}$ for large $n \Longleftrightarrow$*

$$\begin{gathered}|H\theta|/|\theta| < (r_H-2)/(p-2) \Longleftrightarrow\\ |H\theta|/|(I-H)\theta| < (r_H-2)(p-r_H)^{-1/2}(p+r_H-4)^{-1/2}.\end{gathered}$$

Note 1.4. RHS of (1.11) and (1.12) both increase with p.

Table 2. Maximum values of $|\theta|/|\bar\theta|$ and coefficient of variation for Lindley's estimate, $\hat\theta_L$ to improve on the James-Stein estimate, $\hat\theta_{JS}$ for large n.

p	3	4	5	6	7	8	9	10	20	30	40	50	100	∞
RHS (1.11)	1.73	2.31	3	3.70	4.41	5.12	5.82	6.53	13.6	20.7	27.8	34.8	70.2	∞
RHS (1.12)	0	.577	.894	1.13	1.33	1.51	1.66	1.81	2.87	3.64	4.27	4.82	6.95	∞

Note 1.5. (1.10) is a particular case of $H_i=I-ii'$ where i is any given unit vector. Thus $\hat\theta_{H_i}$ improves on $\hat\theta_{JS}$ for large $n \Longleftrightarrow$

$$\begin{aligned}(p-3)^2|H_i\theta|^{-2} > (p-2)^2|\theta|^{-2} &\Longleftrightarrow |H_i\theta|/|i'\theta| < \text{RHS } (1.12)\\ &\Longleftrightarrow |\sin\alpha| < 1-(p-2)^{-1}\end{aligned}$$

where α is the angle between i and θ. This will be true for p sufficiently large if α is bounded away from $\pi/2$. Also $|H_i\theta|^2=|\theta-ii'\theta|^2=|\theta|^2-(i'\theta)^2$.

According to (1.9) for large n we would like to choose H to minimise $|H\theta| /(r_H - 2)$. This motivates the following adaptive estimate.

THEOREM 1.3. *Let $\boldsymbol{H}$ be a finite set of $p \times p$ idempotent H with rank $r_H \le 3$. Choose $\tilde{H}$ from $\boldsymbol{H}$ to minimise $|H\bar{X}|/(r_H - 2)$. Then for fixed p, as $n \to \infty$,*

$$E|\hat{\theta}_{\tilde{H}} - \theta|^2 = n^{-1}v\{p - b\Delta(\boldsymbol{H})\} + O(e^{-\lambda n}) \tag{1.13}$$

$$= \begin{cases} n^{-1}v\{p - b\Delta_0(\boldsymbol{H})\} + O(n^{-3}) & \text{if} \quad \tau(\boldsymbol{H}) > 0 \\ n^{-1}v\{p - b\Delta_1(\boldsymbol{H})\} + O(e^{-\lambda n}) & \text{if} \quad \tau(\boldsymbol{H}) = 0 \end{cases} \tag{1.14}$$

where

$$\Delta(\boldsymbol{H}) = \max\{\Delta(H)\colon H \in \boldsymbol{H}\},$$
$$\tau(\boldsymbol{H}) = \min\{\tau(H)\colon H \in \boldsymbol{H}\},$$
$$\Delta_0(\boldsymbol{H}) = n^{-1}v/\{\min[|H\theta|/(r_H - 2)\colon H \in \boldsymbol{H}]\}^2,$$
$$\Delta_1(H) = \max\{r_H - 2\colon H\theta = 0\}$$

and $\lambda > 0$.

This theorem shows that for large n, $\hat{\theta}_{\tilde{H}}$ performs as well as the best estimator from $\{\hat{\theta}_H\colon H \in \boldsymbol{H}\}$.

In practice one might choose $\boldsymbol{H} = \{I\} \cup \boldsymbol{H}_0$ where $\boldsymbol{H}_0 = \{I - \boldsymbol{i}\boldsymbol{i}',\ i \in \mathcal{I}\}$ for some finite collection of unit p-vectors $\mathcal{I} = \{\boldsymbol{i}\}$; according to the rule in Theorem 1.3 we can replace $\boldsymbol{H}_0$ by $\{I - \boldsymbol{i}_X\boldsymbol{i}'_X\}$ where $\boldsymbol{i}$ maximises $|\boldsymbol{i}'\bar{X}\ |$ in $\mathcal{I}$; thus $\tilde{H} = I$ or $I - \boldsymbol{i}_X\boldsymbol{i}'_X$ accordingly to whether

$$|\bar{X}|/(p-2) < \quad \text{or} \quad > |\bar{X} - \boldsymbol{i}_X\boldsymbol{i}'_X\bar{X}|/(p-3).$$

Note 1.6. If we change the origin by $\delta \in R^p$ we obtain the "translated" estimator

$$\hat{\theta} = \hat{\theta}(H, \delta) = \bar{X} - ZY$$

where

$$Y = H(\bar{X} - \delta) \quad \text{and} \quad Z = n^{-1}v(r_H - 2)|Y|^{-2}.$$

Its risk is that of $\hat{\theta}_H$ with θ replaced by $\theta - \delta$. Thus its risk is that of $\hat{\theta}_H$ if δ lies in the null space of H; (this has dimension $p - r_H$). It's "positive-part" version is

$$\hat{\theta}^+ = \hat{\theta}^+(H, \delta) = \begin{cases} \hat{\theta}(H, \delta) & \text{if} \quad Z < 1 \\ \delta + (I - H)(\bar{X} - \delta) & \text{if} \quad Z \ge 1. \end{cases}$$

Also $|E\hat{\theta}^+ - \theta|^2 < E|\hat{\theta} - \theta|^2$. However one can show the difference is only $O(n^{-3})$ for fixed p.

We now extend Theorem 1.3 to include such translated estimators.

THEOREM 1.4. *Let $\boldsymbol{H}$ be a finite set of $p \times p$ idempotent H with rank $r_H \ge 3$, and $\boldsymbol{L}$ a finite subset of R^p. Choose $(\tilde{H}, \tilde{\delta})$ from $\boldsymbol{H} \times \boldsymbol{L}$ to minimise $|H(\bar{X} - \delta)|/(r_H - 2)$. Then for fixed p, as $n \to \infty$*

$$E|\hat{\theta}(\tilde{H}, \tilde{\delta}) - \theta|^2 = \begin{cases} n^{-1}v\{p - b\Delta_0(\boldsymbol{H}, \boldsymbol{L})\} + O(n^{-3}) & \text{if} \quad \tau(\boldsymbol{H}, \boldsymbol{L}) > 0 \\ n^{-1}v\{p - \Delta_1(\boldsymbol{H}, \boldsymbol{L})\} + O(e^{-\lambda n}) & \text{if} \quad \tau(\boldsymbol{H}, \boldsymbol{L}) = 0 \end{cases}$$

where

$$\Delta_0(\boldsymbol{H}, \boldsymbol{L}) = n^{-1}v/\{\min|H(\theta-\delta)|/(r_H-2)\colon H \in \boldsymbol{H},\ L \in \boldsymbol{L}\}^2,$$
$$\Delta_1(\boldsymbol{H}, \boldsymbol{L}) = \max\{r_H - 2\colon H(\theta-\delta) = 0\},$$
$$\tau(\boldsymbol{H}, \boldsymbol{L}) = \min\{|H(\theta-\delta)|\colon H \in \boldsymbol{H},\ L \in \boldsymbol{L}\} \quad \text{and} \quad \lambda > 0.$$

Note 1.7. George (1986*a*, 1986*b* and 1986*c*) considers quite a different sort of adaptive estimate—an adaptively weighted linear combination of $\{\hat{\theta}(H, \delta),\ H \in \boldsymbol{H},\ \delta \in \boldsymbol{L}\}$. Our estimator $\hat{\theta}(\tilde{\boldsymbol{H}}, \tilde{\delta})$ is much simpler.

So far we have looked at how the risk behaves as the sample size n increases with the dimension p fixed. We now consider what happens when n is fixed but $p \to \infty$.

LEMMA 1.1. *For* $r > 2$,

$$(1.15) \qquad (1 + r/\lambda)^{-1} \le \lambda h(\lambda, r) r^{-2} < 1$$

and

$$(1.16) \qquad h(\lambda, r) = r^2\lambda^{-1}\{1 + O(r/\lambda)\} \quad \text{as} \quad r/\lambda \to 0.$$

Also

$$(1.17) \qquad h(rc, r)/r \to (1+c)^{-1} \quad \text{as} \quad r \to \infty \text{ for fixed } c.$$

Note 1.8. The first inequality in (1.15) is by Casella and Hwang (1982). Inequalities are also given by Sathe and Shenoy (1986). Others are obtainable using

(i) $h(\lambda, r)/r^2$ is decreasing in (λ, r);
(ii) $h(\lambda, r)/r$ is increasing in r;
(iii) $h(\lambda, 2) = 4\lambda^{-1}(1 - e^{-\lambda/2})$.

THEOREM 1.5. *Suppose n is fixed and*

$$|H\theta|^2/r_H \to \tau^* \quad \text{and} \quad r_H/p \to \kappa \quad \text{as} \quad p \to \infty.$$

Then

$$(1.18) \qquad \Delta(H)/p \to \kappa/(1 + nv^{-1}\tau^*) \quad \text{as} \quad p \to \infty.$$

This result immediately implies.

THEOREM 1.6. *Suppose that $r_{H_1}/r_{H_2} \to 1$ as $p \to \infty$ and r_{H_1}/p is bounded away from 0. Then $\hat{\theta}_{H_1}$ has smaller mean square error than $\hat{\theta}_{H_2}$ for large p and fixed $n \iff |H_1\theta| < |H_2\theta|$, or more precisely,*

$$\left(\liminf_{p\to\infty}\right)(|H_2\theta|^2 - |H_1\theta|^2)/p > 0.$$

COROLLARY 1.3. *Under the conditions of Theorem* 1.5

$$\left(\limsup_{p\to\infty}\right) E|\hat{\theta}_H - \theta|^2/E|\bar{X} - \theta|^2 = 1 - b\kappa/(1 + nv^{-1}\tau^*) \leq 1$$

with equality $\Longleftrightarrow$ *either* $r_H/p \to 0$ *or* $|H\theta|^2/r_H \to \infty$ *as* $p \to \infty$.

Thus the risk of $\hat{\theta}_H$ relative to that of $\bar{X}$ is generally bounded below 1 for fixed n and increasing p. But for fixed p and increasing n it tends to 1 if $|H\theta| \neq 0$.

2. Proofs

PROOF OF THEOREM 1.1. Set $r = r_H$, $H = U'\Lambda_r U$ where $U'U = I_p$ and $\Lambda_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$. Set $Y = U\bar{X} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ with $Y_1 \in R^r$. Then $Y \sim N_p(U\theta, vn^{-1}I)$ and $H\bar{X} = U'\begin{pmatrix} Y_1 \\ 0 \end{pmatrix}$. So $|\hat{\theta}_H - \theta|^2 = \left|Y - U\theta - (r-2)\hat{v}n^{-1}\begin{pmatrix} Y_1 \\ 0 \end{pmatrix}|Y_1|^{-2}\right|^2$ has mean

$$pvn^{-1} - 2bvn^{-1}EA + b(r-2)^2(v/n)^2EB$$

where

$$A = (Y - U\theta)'\begin{pmatrix} Y_1 \\ 0 \end{pmatrix}|Y_1|^{-2} = (Y_1 - U\theta)'Y_1|Y_1|^{-2}, \qquad \text{and}$$

$$B = |Y_1|^{-2}.$$

By James and Stein ((1961), pp. 364–365),

$$EA = (r-2)E(r-2+2K)^{-1} \quad \text{and} \quad EB = nv^{-1}E(r-2+2K)^{-1}$$

where $K \sim Po(\delta/2)$ and $\delta = nv^{-1}|(U\theta)_1|^2 = \lambda_H$. □

Corollary 1.1 follows from $H\theta = \mathbf{0}$.

PROOF OF THEOREM 1.2.

$$E(b+K)^{-1} = E\int_0^1 t^{K+b-1}dt = \int_0^1 t^{b-1}\exp\mu(t-1)dt$$
$$= e^{-\mu}\mu^{-b}\int_0^{\mu} s^{b-1}e^s ds = e^{-\mu}(-\mu)^{-b}\gamma(b, -\mu). \qquad \square$$

Set $|X|_s = (E|X|^s)^{1/s}$.

LEMMA 2.1. $|\hat{\theta}_H - \theta|_s = O(n^{-1/2})$ *for* $r_H > s \geq 2$.

PROOF. $|\hat{\theta}_H - \theta|_s \leq C_s + n^{-1}v(r-2)C_s$ where $r = r_H$,

$$C_s = |\bar{X} - \theta|_s = O(n^{-1/2})$$

and

$$D_s = ||H\bar{X}|^{-1}|_s.$$

Now by p.3 of James and Stein (1961),

$$|H\bar{X}|^2 = n^{-1}v\chi_r^2(\lambda_H) = n^{-1}v\chi_{r+2K}^2 \quad \text{for} \quad K \sim Po(\lambda_H/2).$$

So

$$D_s^s = E|H\bar{X}|^{-s}(n/v)^{s/2}E\Gamma\left(\frac{r-s}{2}+K\right)\Big/\Gamma\left(\frac{r}{2}+K\right).$$

Therefore for $r > s \geq 2$,

$$\begin{aligned}\Gamma(s/2)D_s^s(n/v)^{-s/2} &= EB\left(\frac{r-s}{2}+K,\, s/2\right)\\ &\leq EB\left(\frac{r-s}{2}+K,\, 1\right) = E\left(\frac{r-s}{2}+K\right)^{-1} < \infty\end{aligned}$$

since $B(a,\, b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ decreases with b. Therefore $D_s = O(n^{1/2})$. □

PROOF OF THEOREM 1.3. We give this for the case $\boldsymbol{H} = \{H_1,\, H_2\}$, $r_i = r_{H_i}$ and $\boldsymbol{A}(\theta)$: $|H_1\theta|/(r_1-2) < |H_2\theta|/(r_2-2)$. (Proof for the general case is similar.) By large deviation theory $\exists\lambda > 0$ such that

$$P(\tilde{H} = H_1) = P(\boldsymbol{A}(\bar{X})) = 1 - O(e^{-\lambda n}).$$

Therefore

$$\begin{aligned}E|\hat{\theta}_{\tilde{H}} - \theta|^2 &= E|\hat{\theta}_H - \theta|^2 I(\boldsymbol{A}(\bar{X})) + E|\hat{\theta}_{H_2} - \theta|^2\\ &= E|\hat{\theta}_{H_1} - \theta|^2 + \Delta\end{aligned}$$

where $\Delta = E(A_2 - A_1)B$, $B = 1 - I(\boldsymbol{A}(\bar{X}))$ and $A_i = |\hat{\theta}_{H_i} - \theta|^2$. For

$$r^{-1} + s^{-1} = 1 \quad \text{and} \quad 1 < r < \infty,$$

$$|E(A_2 - A_1)B| \leq \sum_1^2 |A_i|_r |B|_s.$$

Also $|B|_s = O(e^{-\lambda n/s})$. Take $1.5 < r \leq 2$, so $2 \leq s < 3$. By Lemma 2.1, $|A_i|_r = O(n^{-1})$, so $\Delta = O(e^{-\lambda n/s})$. Hence (1.3) implies (1.13). (1.14) follows by (1.9). □

Theorem 1.4 is proved similarly.

PROOF OF LEMMA 1.1. The 2nd inequality comes from

$$E(b+K)^{-1} < E(1+K)^{-1} = \mu(1-e^{-\mu}) \quad \text{for} \quad b > 1 \quad \text{and} \quad \mu = \lambda/2.$$

(1.15)⇒(1.16). □

Theorem 1.5 follows from (1.17). Corollary 1.3 follows from (1.3), (1.18).

Acknowledgements

I wish to thank the referees for their comments and suggestions including making a comparison with the approximations of Copas (1983).

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D. C.

Casella, G. and Hwang, J. T. (1982). Limit expressions for the risk of James-Stein estimators, *Canad. J. Statist.*, **10**, 305–309.

Copas, J. B. (1983). Regression, prediction and shrinkage, *J. Roy. Statist. Soc. Ser B*, **45**, 311–354.

Efron, B. and Morris, C. (1973*a*). Combining possibly related estimation problems, *J. Roy. Statist. Soc. Ser. B*, **35**, 379–421.

Efron, B. and Morris, C. (1973*b*). Stein's estimation rule and its competitors—an empirical Bayes approach, *J. Amer. Statist. Assoc.*, **68**, 117–130.

Egerton, M. F. and Laycock, P. J. (1982). An explicit formula for the risk of James-Stein estimators, *Canad. J. Statist.*, **10**, 199–205.

George, E. (1986*a*). Minimax multiple shrinkage estimation, *Ann. Statist.*, **14**, 188–205.

George, E. (1986*b*). A formal Bayes multiple shrinkage estimator, *Comm. Statist. A—Theory Methods*, **15**, 2099–2114.

George, E. (1986*c*). Combining minimax shrinkage estimators, *J. Amer. Statist. Assoc.*, **81**, 437–445.

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 361–380, Univ. California Press, Bekeley.

Sathe, Y. S. and Shenoy, R. G. (1986). On sharper bounds for the risk of James-Stein estimators, *Comm. Statist. A—Theory Methods*, **15**, 613–617.

Sclove, S. L., Morris, C. and Radhakrishnan, R. (1972). Nonoptimality of preliminary-test estimators for the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **43**, 1481–1490.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution, *J. Roy. Statist. Soc. Ser. B*, **24**, 265–296.

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs, *Festschrift for J. Neyman* (ed. F. N. David), 351–366, Wiley, New York.