James-Stein type estimators for ordered normal means

Somesh Kumar\textsuperscript{a}, Yogesh Mani Tripathi\textsuperscript{a} & Neeraj Misra\textsuperscript{b}

\textsuperscript{a} Department of Mathematics, Indian Institute of Technology, Kharagpur, 721 302, India
\textsuperscript{b} Department of Mathematics, Indian Institute of Technology, Kanpur, 208 016, India

Published online: 19 Aug 2006.

To cite this article: Somesh Kumar, Yogesh Mani Tripathi & Neeraj Misra (2005) James-Stein type estimators for ordered normal means, Journal of Statistical Computation and Simulation, 75:7, 501-511, DOI: 10.1080/00949650412331272877

To link to this article: http://dx.doi.org/10.1080/00949650412331272877

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
James–Stein type estimators for ordered normal means

SOMESH KUMAR*†, YOGESH MANI TRIPATHI† and NEERAJ MISRA‡

†Department of Mathematics, Indian Institute of Technology, Kharagpur 721 302, India
‡Department of Mathematics, Indian Institute of Technology, Kanpur 208 016, India

(Received 29 October 2003; in final form 16 June 2004)

Suppose independent observations $X_1, X_2, \ldots, X_k$ are available from $k \geq 2$ normal populations having means $\theta_1, \theta_2, \ldots, \theta_k$, respectively, and common variance unity. The means are known to be ordered, that is, $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Two commonly used estimators for simultaneous estimation of $\hat{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ are $\delta_{\text{MLE}}$, the order restricted maximum likelihood estimator (MLE), and $\delta_p$, the generalized Bayes estimator of $\theta$ with respect to the uniform prior on the restricted space $\Omega = \{ \hat{\theta} \in \mathbb{R}^k; \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k \}$, where $\mathbb{R}^k$ denotes the $k$-dimensional Euclidean space. Both $\delta_{\text{MLE}}$ and $\delta_p$ improve the usual unrestricted MLE $\hat{X} = (X_1, X_2, \ldots, X_k)$ and are minimax. But $\delta_{\text{MLE}}$ is inadmissible for $k \geq 2$ and $\delta_p$ is inadmissible for $k \geq 3$. However, no dominating estimators are yet known. Using Brown’s [Brown, L.D., 1979, A heuristic method for determining admissibility of estimators – with applications. *Annals of Statistics*, 7, 960–994] heuristic approach for proving admissibility or inadmissibility of estimators, we propose some classes of James–Stein type estimators and show, through a simulation study, that many of these estimators dominate $\delta_p$ and $\delta_{\text{MLE}}$.

Keywords: Ordered parameters; Maximum likelihood estimator; Pitman estimator; Generalized Bayes estimator; Mixed estimators; James–Stein estimator; Translation invariance

1. Introduction

Let $X_1, X_2, \ldots, X_k$ be independent normal random variables with means $\theta_1, \theta_2, \ldots, \theta_k$ and the common variance unity. It is assumed that $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. The problem of estimating order restricted parameters has its origins in the study of isotonic regression and has found applications in areas such as bio-assays, reliability and life testing and various agricultural and industrial experiments. For example, computer software usually contains errors called bugs. A common technique of debugging is to run the program until a bug appears and then correcting the fault and continuing the execution. If errors are corrected in stages without introducing new errors, the reliability of the software increases at each stage. If bugs are removed in $k$ stages and $\theta_i$ denotes the reliability at the $i$th stage, $i = 1, \ldots, k$, then we have $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Similarly in a dosage-mortality experiment on animals in bio-assays,
different groups of animals are administered different doses of a drug. It is known that with an increase in the dose more animals will die. Suppose in the $i$th group, quantity $t_i$ of the drug is administered to each animal and, as a result, the proportion $\theta_i$ of animals die, $i = 1, 2, \ldots, k$. If $t_1 \leq t_2 \leq \cdots \leq t_k$, then it is natural to expect $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$.

Estimation of ordered parameters has received the attention of several researchers during the past 50 years. Most of the earlier works are concerned with obtaining the maximum likelihood estimators (MLEs). Barlow et al. [1] and Robertson et al. [2] discuss in detail results on existence, uniqueness and algorithms for finding MLEs for various distributions. Lee [3] showed that for estimation of $k$ ($\geq 2$) ordered normal means, when variances are known, the restricted MLE has componentwise smaller mean squared error (MSE) than that of the usual sample mean. Kelly [4] strengthened the result of Lee [3] by establishing that the restricted MLE stochastically dominates the usual MLE componentwise. As a consequence, it follows that the restricted MLE dominates the usual MLE componentwise, with respect to any loss $L(|\delta_i - \theta_i|)$, where $L(\cdot)$ is an increasing function. Hwang and Peddada [5] extended the results of Lee [3] and Kelly [4] to location and scale probability models. Hwang and Peddada [6] and Fernandez et al. [7] have studied properties of confidence intervals based on order restricted MLEs. The dominance of the restricted MLE over the unrestricted MLE with respect to the LINEX loss function was proved by Kumar and Kumar [8] for estimating locations of two exponential populations. Fernandez et al. [9] have compared the restricted and unrestricted MLEs using the universal domination and the squared error criterion when linear functions of parameters are estimated.

Katz [10] introduced mixed estimators for simultaneous estimation of two ordered binomial parameters and showed that they are better than the unrestricted MLE. Kumar and Sharma [11] investigated properties of mixed estimators for two ordered normal means. In particular, these are shown to be minimax but inadmissible. Mixed estimators for ordered parameters of two exponential populations have been studied by Vijayshree and Singh [12, 13], Kaur and Singh [14], Kumar and Kumar [8, 15] and Misra and Singh [16]. Some new types of estimators for ordered parameters using different approaches have been proposed by van Eeden and Zidek [17], Iliopoulos [18], Lillo and Martin [19] and Misra et al. [20].

Katz [10] also considered simultaneous estimation of two ordered normal means $\theta = (\theta_1, \theta_2)$, $\theta_1 \leq \theta_2$, and proposed an analogue of the Pitman estimator $\delta_p$, that is, the generalized Bayes estimator of $\theta$ with respect to the uniform prior on the space $\{\theta \in R^2; \theta_1 \leq \theta_2\}$. Blumenthal and Cohen [21] obtained sufficient conditions for minimaxity and admissibility of $\delta_p$ in a more general set up of two location parameter densities. In particular, the conditions are satisfied for normal populations. This minimaxity result was generalized by Kumar and Sharma [11] to two location parameter densities with different scale parameters. In particular, normal, uniform and gamma densities are seen to satisfy the sufficient conditions. Kumar and Sharma [22] proved that $\delta_p$ is minimax for simultaneous estimation of $k$ ($\geq 2$) ordered normal means when variances are known and equal. When variances are known but unequal, $\delta_p$ is shown to be minimax with respect to a scale invariant loss [23] but non-minimax with respect to the squared error loss [24].

For simultaneous estimation of $k$ ordered normal means, MLEs and mixed estimators are inadmissible as they are not smooth (see for example, ref. [11]). Further, the Pitman estimator $\delta_p$ is also inadmissible for $k \geq 3$ [22]. However, so far no estimators improving upon these have been obtained. In this paper, new estimators of $\theta$ based on well known James–Stein and James–Stein positive part estimators are proposed. The risk functions of these estimators are evaluated using Monte Carlo simulations for $k = 3$ and it is shown that some of these estimators improve upon the Pitman estimator $\delta_p$ and the MLE uniformly over the restricted parameter space. In section 2, we introduce the notation and propose estimators. In section 3, numerical comparisons of risk values are made.
2. James–Stein type estimators for ordered normal means

Let $X_1, X_2, \ldots, X_k$ be independent normal random variables with means $\theta_1, \theta_2, \ldots, \theta_k$, respectively, and the common variance unity, where it is known a priori that $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. We estimate $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ when the loss function is the sum of squared errors, that is,

$$L(\theta, a) = \sum_{i=1}^{k} (\theta_i - a_i)^2,$$

where $a = (a_1, a_2, \ldots, a_k)$.

The Pitman estimator $\delta_p = (\delta_{p1}, \delta_{p2}, \ldots, \delta_{pk})$ of $\theta$ is the generalized Bayes estimator of $\theta$ with respect to the uniform prior on $\Omega_1 = \{\theta \in \mathbb{R}^k : \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k\}$ and is given by

$$\delta_{pi}(x) = \frac{\int_{\Omega} \theta_i p(x, \theta) d\theta}{\int_{\Omega} p(x, \theta) d\theta}, \quad i = 1, 2, \ldots, k,$$

where

$$p(x, \theta) = \prod_{i=1}^{k} \phi(x_i - \theta_i),$$

$x = (x_1, \ldots, x_k)$, $d\theta$ denotes $d\theta_1, \ldots, d\theta_k$ and $\int_{\Omega}$ denotes the multiple integral with respect to $\theta_1, \ldots, \theta_k$. Here $\phi$ denotes the probability density function of a standard normal random variable.

We can write

$$\delta_{pi}(x) = x_i + \gamma_i(x),$$

where

$$\gamma_i(x) = \frac{\alpha_i(x)}{D(x)}, \quad \alpha_i(x) = \int_{\Omega} (\theta_i - x_i) p(x, \theta) d\theta, \quad i = 1, \ldots, k \quad \text{and}$$

$$D(x) = \int_{\Omega} p(x, \theta) d\theta.$$

For $k = 3$, we obtain explicit expressions for $\alpha_i(x), i = 1, 2, 3$.

$$\alpha_1(x) = \int_{\Omega} (\theta_1 - x_1) p(x, \theta) d\theta$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} (\theta_1 - x_1)\phi(x_1 - \theta_1)\phi(x_2 - \theta_2)\phi(x_3 - \theta_3) d\theta_1 d\theta_2 d\theta_3.$$

Integrating with respect to $\theta_1, \theta_2$ and $\theta_3$, in that order, we get

$$\alpha_1(x) = -Q_1(x),$$

where

$$Q_1(x) = \frac{1}{\sqrt{2}} \Phi \left( \frac{x_1 - x_2}{\sqrt{2}} \right) \Phi \left( \frac{2x_3 - x_1 - x_2}{\sqrt{6}} \right)$$

and $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable.
Similarly, we get

$$\alpha_2(x) = Q_1(x) - Q_2(x) \quad \text{and} \quad \alpha_3(x) = Q_2(x),$$

where

$$Q_2(x) = \frac{1}{\sqrt{2}} \phi \left( \frac{x_2 - x_3}{\sqrt{2}} \right) \Phi \left( \frac{x_2 + x_3 - 2x_1}{\sqrt{6}} \right).$$

Thus the estimator $\delta_p$ can be expressed as

$$\delta_p(X) = \left( X_1 - \frac{Q_1(X)}{D(X)}, X_2 + \frac{Q_1(X) - Q_2(X)}{D(X)}, X_3 + \frac{Q_2(X)}{D(X)} \right). \quad (2)$$

It has been mentioned in section 1 that $\delta_p$ is minimax but inadmissible. We propose some James–Stein type estimators to improve it.

The motivation in using James–Stein type estimators came from Brown’s [25] heuristic approach for proving admissibility or inadmissibility of estimators. Consider an estimator $\delta_\psi(X) = \delta_p(X) + \psi(X)$, where $\psi(X) = (\psi_1(X), \psi_2(X), \psi_3(X))$ and let $\Delta(\theta) = R(\theta, \delta_p) - R(\theta, \delta_\psi)$. If there is $\psi$ for which $\Delta(\theta) \geq 0$ for all $\theta \in \Omega$, with strict inequality for some $\theta \in \Omega$, then the corresponding estimator $\delta_\psi$ will be better than $\delta_p$. However, it is difficult to guess such a $\psi$ in general. Brown suggested approximating the risk difference by means of Taylor series expansion and obtaining solutions to the approximate inequalities. We derive below the approximate expression for $\Delta(\theta)$ using his approach.

Consider a more general situation of loss function

$$\sum_{i=1}^{3} W(\theta_i - a_i),$$

a sum of individual losses in estimating $\theta_i$. Then risk difference between $\delta_p$ and $\delta_\psi$ is

$$\Delta(\theta) = \sum_{i=1}^{3} E[W(X_i + \gamma_i(X) - \theta_i) - W(X_i + \gamma_i(X) + \psi_i(X) - \theta_i)].$$

Let $W'$, $W''$ and $W'''$ denote the first-, second- and third-order derivatives of $W$, respectively. In addition, let $\psi_{ij}(\theta)$ denote the first-order partial derivatives of $\psi_i(\theta)$ with respect to $\theta_j$; $i, j = 1, 2, 3$. A Taylor expansion of $W(X_i + \gamma_i(X) + \psi_i(X) - \theta_i)$, about $X_i + \gamma_i(X) - \theta_i$, after ignoring the terms of third and higher order derivatives of $W$, $i = 1, 2, 3$, gives the approximation

$$\Delta(\theta) \approx \sum_{i=1}^{3} E \left[ -\psi_i(X)W'(X_i + \gamma_i(X) - \theta_i) - \frac{1}{2} \psi_i^2(X)W''(X_i + \gamma_i(X) - \theta_i) \right].$$

In addition,

$$W'(X_i + \gamma_i(X) - \theta_i) = W'(X_i - \theta_i) + \gamma_i(X)W''(Z_i),$$

and

$$W''(X_i + \gamma_i(X) - \theta_i) = W''(X_i - \theta_i) + \gamma_i(X)W'''(Z_i^*),$$

where $Z_i$ and $Z_i^*$ are points lying between $X_i - \theta_i$ and $X_i + \gamma_i(X) - \theta_i$. 

Ignoring the terms involving $W'''$, we get the approximation

\[
\Delta(\theta) \approx \sum_{i=1}^{3} E \left[ -\psi_i(X) W'(X_i - \theta_i) + \gamma_i(X) W''(Z_i) - \frac{1}{2} \psi_i^2(X) W''(X_i - \theta_i) \right].
\]

It can be noted that, for the squared error loss, the above expression of $\Delta(\theta)$ is exact. Further, expanding $\psi_i(X)$ and $\psi_i^2(X)$ about $\theta$ in Taylor series, we get

\[
\psi_i(X) = \psi_i(\theta) + \sum_{j=1}^{3} (X_j - \theta_j) \psi_{ij}(\theta) + \cdots
\]

and

\[
\psi_i^2(X) = \psi_i^2(\theta) + \cdots
\]

This yields the approximation

\[
\Delta(\theta) \approx \sum_{i=1}^{3} \left\{ E \left[ -\psi_i(\theta) W'(X_i - \theta_i) - \sum_{j=1}^{3} (X_j - \theta_j) \psi_{ij}(\theta) W'(X_i - \theta_i) \right.ight.
\]

\[
\left. - \psi_i(\theta) \gamma_i(X) W''(Z_i) - \sum_{j=1}^{3} (X_j - \theta_j) \psi_{ij}(\theta) \gamma_i(X) W''(Z_i) - \frac{1}{2} \psi_i^2(\theta) W''(X_i - \theta_i) + \cdots \right\}.
\]

Finally $\gamma_i(X) = \gamma_i(\theta) + \cdots$ and as an approximation, we take

\[
\Delta(\theta) \approx \sum_{i=1}^{3} \left\{ E \left[ -\psi_i(\theta) W'(X_i - \theta_i) - \sum_{j=1}^{3} (X_j - \theta_j) \psi_{ij}(\theta) W'(X_i - \theta_i) \right.ight.
\]

\[
\left. - \psi_i(\theta) \gamma_i(\theta) W''(Z_i) - \sum_{j=1}^{3} (X_j - \theta_j) \psi_{ij}(\theta) \gamma_i(\theta) W''(Z_i) - \frac{1}{2} \psi_i^2(\theta) W''(X_i - \theta_i) \right]\}

Replacing $W$ by the squared error loss and substituting the values of $\gamma_1, \gamma_2$ and $\gamma_3$, the approximation up to second-order terms takes the form

\[
\Delta(\theta) = - \left[ \frac{1}{2} \left( \psi_{11}(\theta) + \psi_{22}(\theta) + \psi_{33}(\theta) - \frac{Q_1(\theta)}{D(\theta)} \psi_1(\theta) + \frac{Q_1(\theta) - Q_2(\theta)}{D(\theta)} \psi_2(\theta) \right.ight.
\]

\[
\left. + \frac{Q_2(\theta)}{D(\theta)} \psi_3(\theta) \right) + \psi_1^2(\theta) + \psi_2^2(\theta) + \psi_3^2(\theta) \right].
\]

If we take $\psi(x) = -a x / \|x\|^2$, with $0 < a \leq 2$, which is the choice for the James–Stein [26] estimator, then $\Delta(\theta) \geq 0$ for all $\theta_1 \leq \theta_2 \leq \theta_3$. The error terms in reaching the approximation are complicated. However, there is a feeling that $\hat{\theta}_p(X)$ with $\psi(x) = -a x / \|x\|^2$, improves upon $\hat{\theta}_p(X)$. This prompted us to consider estimators based on $a x / \|x\|^2$. Let $\lambda_a(x) = (1 - a / \|x\|^2)$ and $\lambda_a^+(x) = \max\{\lambda_a(x), 0\}, where $\|x\|$ represents the Euclidean norm of $x$ and $a$ is some real constant. Then $\lambda_a^+(X)X$ and $\lambda_a^+(X)X$ are the James–Stein and the James–Stein
positive part estimators, respectively, for estimating $\theta$. When there are no restrictions on the parameter space, both estimators improve $X$. Also for dimension three, $a = 1$ is the best choice for the James–Stein estimator and $a = 2.37$ is a reasonable choice for the James–Stein positive part estimator [27]. We suggest the following classes of estimators:

$$
\hat{\delta}_{1,a}(X) = \lambda_a(X)X + \gamma(X), \quad \hat{\delta}_{2,a}(X) = \lambda_a(X)\hat{\delta}_p(X),
$$

$$
\hat{\delta}_{3,a}(X) = \lambda_a^+(X)X + \gamma(X), \quad \text{and} \quad \hat{\delta}_{4,a}(X) = \lambda_a^+(X)\hat{\delta}_p(X),
$$

where

$$
\gamma(X) = (\gamma_1(x), \gamma_2(x), \gamma_3(x)).
$$

(3)

It is to be noted that among these estimators only $\hat{\delta}_{4,a}(X)$ lies completely inside the restricted space $\Omega$. However, we are able to show using simulations that for specific choices of $a$, estimators dominating $\hat{\delta}_p$ exists in all these classes. We further propose estimators $\hat{\delta}^*_{i,a}; i = 1, 2, 3$ which are obtained by restricting estimators $\hat{\delta}_{i,a}; i = 1, 2, 3$ to the parameter space $\Omega$. When the loss function is the sum of squared errors (1) the estimator $\hat{\delta}^*_{i,a}$ can be shown to improve over $\hat{\delta}_{i,a}$ for all $i = 1, 2, 3$. The estimator $\hat{\delta}^*_{i,a}$ is derived using the ‘Pool-Adjacent-Violators Algorithm’ for finding isotonic regression [1]. An estimator $\hat{\delta} = (\delta_1, \delta_2, \delta_3)$ not entirely lying in $\Omega$, is improved by the restricted estimator (we will call it isotonized version of $\hat{\delta}$) $\hat{\delta}^* = (\delta^*_1, \delta^*_2, \delta^*_3)$ defined as

$$
\hat{\delta}^* = (\delta_1, \delta_2, \delta_3), \quad \text{if} \quad \delta_1 < \delta_2 < \delta_3,
$$

$$
= \left( \frac{\delta_1 + \delta_2 + \delta_3}{3}, \frac{\delta_2 + \delta_3}{2}, \frac{\delta_1 + \delta_2 + \delta_3}{3} \right), \quad \text{if} \quad \delta_1 < \delta_2 \geq \delta_3 \text{ and } \delta_1 < \frac{\delta_2 + \delta_3}{2},
$$

$$
= \left( \frac{\delta_1 + \delta_2 + \delta_3}{3}, \frac{\delta_1 + \delta_2 + \delta_3}{2}, \frac{\delta_1 + \delta_2 + \delta_3}{3} \right), \quad \text{if} \quad \delta_1 < \delta_2 \geq \delta_3 \text{ and } \delta_1 \geq \frac{\delta_2 + \delta_3}{2},
$$

$$
= \left( \frac{\delta_1 + \delta_2 + \delta_3}{2}, \frac{\delta_1 + \delta_2 + \delta_3}{2}, \delta_3 \right), \quad \text{if} \quad \delta_1 \geq \delta_2 < \delta_3 \text{ and } \delta_1 + \frac{\delta_2}{2} < \delta_3,
$$

$$
= \left( \frac{\delta_1 + \delta_2 + \delta_3}{2}, \frac{\delta_1 + \delta_2 + \delta_3}{3}, \frac{\delta_1 + \delta_2 + \delta_3}{3} \right), \quad \text{if} \quad \delta_1 \geq \delta_2 < \delta_3 \text{ and } \delta_1 + \frac{\delta_2}{2} \geq \delta_3,
$$

$$
= \left( \frac{\delta_1 + \delta_2 + \delta_3}{3}, \frac{\delta_1 + \delta_2 + \delta_3}{3}, \frac{\delta_1 + \delta_2 + \delta_3}{3} \right), \quad \text{if} \quad \delta_1 \geq \delta_2 \geq \delta_3.
$$

In section 3, we exhibit choices of $a$ for which estimators $\hat{\delta}^*_{i,a}; i = 1, 2, 3$ are better than $\hat{\delta}_p$.

A competing estimator to $\hat{\delta}_p$ is the restricted maximum likelihood estimator $\hat{\delta}_{\text{MLE}}$. Although it is inadmissible, it performs better than the Pitman estimator $\hat{\delta}_p$ when means are close to each other. Kumar and Sharma [11] compared the risk functions of $\hat{\delta}_{\text{MLE}}$ and $\hat{\delta}_p$ for estimating two ordered normal means and concluded that none of them dominates the other. When the two means are close, the MLE is better and when their difference is large, the Pitman estimator has superior risk performance. Similar observations have been made for estimating three ordered normal means by Kumar et al. [24].

The MLE for $\theta = (\theta_1, \theta_2, \theta_3)$, when $\theta_1 \leq \theta_2 \leq \theta_3$, is simply the isotonized version (call it $\hat{\delta}_{\text{MLE}}$) of the unrestricted MLE $\hat{X} = (X_1, X_2, X_3)$ and is thus defined as $\hat{\delta}$ earlier with $\delta_i = X_i$ for $i = 1, 2, 3$. We propose below two classes of James–Stein type estimators for improving $\hat{\delta}_{\text{MLE}}$:

$$
\hat{\delta}_{5,a}(X) = \lambda_a(X)\hat{\delta}_{\text{MLE}} \quad \text{and} \quad \hat{\delta}_{6,a}(X) = \lambda_a^+(X)\hat{\delta}_{\text{MLE}}.
$$

Once again, of these two estimators, only $\hat{\delta}_{6,a}(X)$ lies completely in $\Omega$. However, we are able to show using simulations that there are choices of $a$ for which both of these improve
over $\delta_{\text{MLE}}$. We can further improve upon $\delta_{2,a}(X)$ by $\delta_{5,a}^+(X)$ which is obtained by restricting the former estimator to $\Omega$. The restricted estimator is derived as $\delta^*$ is derived from $\delta$ earlier in this section.

Next, we observe that the estimation problem is invariant under the location transformation $(X_1, X_2, \ldots, X_k) \rightarrow (X_1 + c, X_2 + c, \ldots, X_k + c)$. Under this transformation, $(\theta_1, \theta_2, \ldots, \theta_k) \rightarrow (\theta_1 + c, \theta_2 + c, \ldots, \theta_k + c)$ and so the inequality restrictions $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$ are preserved. When the loss function is the sum of squared errors, the form of a translation equivariant estimator $\hat{\delta}(X) = (\delta_1, \delta_2, \delta_3)$ is given by $\delta_i = X_i + \phi(U_1, U_2); i = 1, 2, 3,$ where $U_1 = X_2 - X_1$ and $U_2 = X_3 - X_2$. The Pitman estimator $\delta_p$ and the MLE $\delta_{\text{MLE}}$ are equivariant. However, our proposed estimators are not so. This is because they rely on James–Stein shrinkage toward a single point, chosen $(0, 0, \ldots, 0)$ here. However, Lindley in his discussion of Stein [28] and Efron and Morris [27] suggested shrinking toward the one-dimensional line $\theta_1 = \theta_2 = \cdots = \theta_k$. Let $\bar{X} = (1/k)(X_1 + \cdots + X_k)$, $S = \sum_{i=1}^{k}(X_i - \bar{X})^2$ and $1 = (1, \ldots, 1)$ be a $k \times 1$ vector. For estimating $\theta = (\theta_1, \ldots, \theta_k)$ when there are no restrictions on the parameter space, Lindley–Efron–Morris estimator is, then, derived as

$$\hat{\delta}_L = \bar{X}1 + \left(1 - \frac{c(k-3)}{S}\right)(X - \bar{X}1).$$

The positive part version $\delta_L^+$ of $\hat{\delta}_L$ dominates it and is given by

$$\hat{\delta}_L^+ = \bar{X}1 + \left(1 - \frac{c(k-3)}{S}\right)^+(X - \bar{X}1).$$

Both $\hat{\delta}_L$ and $\hat{\delta}_L^+$ are shown to improve $X$ for $k \geq 4$ and the best choice of $c$ is $c = 1$.

For the problem of estimating $\theta = (\theta_1, \theta_2, \theta_3)$, when $\theta_1 \leq \theta_2 \leq \theta_3$, we propose estimators based on $\delta_L^+$ in order to improve $\hat{\delta}_p$ and $\delta_{\text{MLE}}$. These are defined as

$$\delta_{L1,a}(X) = \bar{X}1 + \left(1 - \frac{a}{S}\right)(X - \bar{X}1) + \gamma(X), \quad \delta_{L2,a}(X) = \bar{X}1 + \left(1 - \frac{a}{S}\right)^+(X - \bar{X}1) + \gamma(X),$$

$$\delta_{L3,a}(X) = \bar{X}1 + \left(1 - \frac{a}{S}\right)^+(\hat{\delta}_p - \bar{X}1), \quad \text{and} \quad \delta_{L4,a}(X) = \bar{X}1 + \left(1 - \frac{a}{S}\right)^+(\delta_{\text{MLE}} - \bar{X}1).$$

Here $\gamma(X)$ is as defined in equation (3). We note that estimators $\delta_{L1,a}$ and $\delta_{L2,a}$ do not always lie in the restricted space $\Omega$. By isotonizing these we can improve upon them by estimators $\delta_{L1,a}^+$ and $\delta_{L2,a}^+$, respectively. On the other hand, $\delta_{L3,a}$ and $\delta_{L4,a}$ lie completely in $\Omega$. However, it is interesting to observe that while we get values of $a$ for which $\delta_{L1,a}^*$ and $\delta_{L2,a}^*$ improve on $\hat{\delta}_p$, there is no choice of $a$ for which $\delta_{L3,a}$ and $\delta_{L4,a}$ improve upon $\delta_p$ and $\delta_{\text{MLE}}$, respectively.

### 3. Numerical comparisons

The Pitman estimator $\delta_p$ and the restricted MLE $\delta_{\text{MLE}}$, both are known to improve the unrestricted MLE $\bar{X}$. However, the proofs do not involve explicit computations of risk functions ([22] for result on $\hat{\delta}_p$ and to Brunk (1965) for result on $\delta_{\text{MLE}}$). In fact, risk functions of both estimators are quite complicated and it is difficult to compare them theoretically with those of our proposed estimators. Further, estimator $\delta_{L1,a}^*$ improves $\delta_{j,a}$; for $j = 1, 2, 3, 5$ and estimators $\delta_{L1,a}^*$ and $\delta_{L2,a}^*$ improve upon $\delta_{L1,a}$ and $\delta_{L2,a}$, respectively. Hence for comparison with $\hat{\delta}_p$ and $\delta_{\text{MLE}}$, we consider risk functions of dominating estimators only.

In this section, we have numerically evaluated the risk functions of $\hat{\delta}_p$, $\delta_{\text{MLE}}$, $\delta_{L1,a}^*$, for $j = 1, 2, 3, 5$; $\delta_{j,a}$ for $j = 4, 6$; $\delta_{L1,a}^*$, for $i = 1, 2$ and $\delta_{L2,a}^*$, for $i = 3, 4$ (for selected values
of \( a \) using simulations based on 5000 generations of values of \((X_1, X_2, X_3)\). The comparisons are made for various values of \( \theta = (\theta_1, \theta_2, \theta_3) \). For estimators \( \delta_{j,a} \), for \( j = 1, 2, 5; a = 1 \) is natural choice, being optimal for the traditional James–Stein estimator. Similarly for estimators \( \delta_{j,a} \), for \( j = 3, 4, 6; a = 2.37 \) is a natural choice, being optimal for the James–Stein positive part estimator [27]. However, as we observe, for estimators \( \delta_{1,a} \) and \( \delta_{3,a} \), a small value of \( a \) such as \( a = 0.1 \) gives improvement over \( \delta_{p} \), whereas \( a = 0.5, 1 \) etc. leads to both \( \delta_{1,a}^* \) and \( \delta_{3,a}^* \) being worse than \( X \) itself when \( \theta_i \)s are all equal and large. In fact the risk values of both estimators decrease rapidly (as \( a \) increases) near the origin and when differences between \( \theta_i \)s are large; but on the line \( \theta_1 = \theta_2 = \theta_3 \) they become worse, when \( \theta_i \)s increase. Therefore, risk values of \( \delta_{1,a}^* \) and \( \delta_{3,a}^* \) are tabulated for \( a = 0.1 \) only. For \( \delta_{j,a}^* \), \( j = 2, 5 \); and \( \delta_{j,a} \), \( j = 4, 6 \), we take choices \( a = 0.5, 1 \) and 1.5. In table 1, the risk values of estimators \( \delta_{\text{MLE}}, \delta_{p}, \delta_{1,a}^* \) and \( \delta_{2,a}^* \) with \( a = 0.1 \) are presented. It can be seen that \( \delta_{1,a}^* \) and \( \delta_{3,a}^* \) are better than \( \delta_{p} \) for all configurations of values of \( \theta = (\theta_1, \theta_2, \theta_3) \); \( \theta_1 \leq \theta_2 \leq \theta_3 \). In table 2(a)–(c), risk values of estimators \( \delta_{j,a}^* \), for \( j = 2, 5 \) and \( \delta_{j,a} \), for \( j = 4, 6 \) for \( a = 0.5, a = 1 \) and \( a = 1.5 \), respectively, are tabulated. In table 3, risk values of estimators \( \delta_{1,1.05}^*, \delta_{1,0.1}^*, \delta_{1,3.1}^*, \delta_{1,4.05}^* \) are presented.

The following conclusions are drawn from our simulation studies and tables 1–3:

(i) \( \delta_{2,a}^* \) and \( \delta_{4,a}^* \) are better than \( \delta_{p} \) for all choices of \( a \) in \((0, 2)\). The amount of improvement increases substantially as \( a \) increases when \( \theta_i \)s are close. When differences among \( \theta_i \)s are large, \( a = 1 \) gives marginally better risk performance than \( \delta_{p} \) as compared to choices \( a = 0.1, 0.5 \) and 1.5. Also for the choice \( a = 2.37 \), the optimal choice for the James–Stein positive part estimator, \( \delta_{4,a} \) is worse than \( \delta_{p} \). Thus, the simulation study suggests that \( \delta_{2,a}^* \) and \( \delta_{4,a}^* \) are better than \( \delta_{p} \) for all \( 0 < a < 2 \). However, \( \delta_{4,a} \) seems to be uniformly better than \( \delta_{2,a}^* \) for all values of \( a \) in \((0, 2)\). The amount of improvement is large when differences among \( \theta_i \)s are small but marginal when they are large.

(ii) For estimators \( \delta_{1,a}^* \) and \( \delta_{3,a}^* \), only a small value of \( a \) such as \( a = 0.1 \) seems to give uniform improvements over \( \delta_{p} \). As \( a \) increases, the risk values of \( \delta_{1,a}^* \) and \( \delta_{3,a}^* \) decrease rapidly near the origin and also when differences among \( \theta_i \)s are moderate/large and thus they offer substantial improvement over \( \delta_{p} \) (and in some cases over \( \delta_{2,a}^* \) and \( \delta_{4,a} \) too). But their performance is not good when values of \( \theta_i \)s are very close to each other and large. In this case, these estimators become worse than \( X \) itself.

(iii) Estimators \( \delta_{5,a}^* \) and \( \delta_{6,a}^* \) are uniformly better than \( \delta_{\text{MLE}} \) for \( 0 < a \leq 1 \). The improvements, are substantial when \( \theta_i \)s are close to each other and marginal when their differences are large. As \( a \) increases (\( \leq 2 \)), risk values of both \( \delta_{5,a}^* \) and \( \delta_{6,a}^* \) decrease rapidly when \( \theta_i \)s are close to each other. However, for some configurations of values of \( (\theta_1, \theta_2, \theta_3) \) away from the line \( \theta_1 = \theta_2 = \theta_3 \), their values become marginally higher.

<table>
<thead>
<tr>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \theta_3 )</th>
<th>( \delta_{\text{MLE}} )</th>
<th>( \delta_{p} )</th>
<th>( \delta_{1,0.1}^* )</th>
<th>( \delta_{2,0.1}^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.826405</td>
<td>3.000000</td>
<td>2.890227</td>
<td>2.899278</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
<td>1.2</td>
<td>1.937766</td>
<td>2.430911</td>
<td>2.364992</td>
<td>2.364548</td>
</tr>
<tr>
<td>-0.6</td>
<td>-0.4</td>
<td>-0.2</td>
<td>1.858256</td>
<td>2.672793</td>
<td>2.573238</td>
<td>2.572137</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.826405</td>
<td>3.000000</td>
<td>2.965047</td>
<td>2.964538</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>2.274496</td>
<td>2.066286</td>
<td>1.962943</td>
<td>1.962380</td>
</tr>
<tr>
<td>-3.0</td>
<td>0.0</td>
<td>3.0</td>
<td>2.927100</td>
<td>2.479722</td>
<td>2.465288</td>
<td>2.465289</td>
</tr>
<tr>
<td>0.0</td>
<td>2.0</td>
<td>3.0</td>
<td>2.509887</td>
<td>2.103406</td>
<td>2.084609</td>
<td>2.084609</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>2.5</td>
<td>2.409047</td>
<td>2.051670</td>
<td>2.015415</td>
<td>2.015331</td>
</tr>
<tr>
<td>-3.0</td>
<td>1.0</td>
<td>2.5</td>
<td>2.773419</td>
<td>2.442769</td>
<td>2.427422</td>
<td>2.427422</td>
</tr>
<tr>
<td>-3.0</td>
<td>-1.0</td>
<td>3.0</td>
<td>2.839865</td>
<td>2.488576</td>
<td>2.475432</td>
<td>2.475432</td>
</tr>
</tbody>
</table>
Table 2. MSEs of $\delta^*_2,a$, $\delta^*_4,a$, $\delta^*_5,a$ and $\delta^*_6,a$.

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\delta^*_2,a$</th>
<th>$\delta^*_4,a$</th>
<th>$\delta^*_5,a$</th>
<th>$\delta^*_6,a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.936638</td>
<td>1.912452</td>
<td>1.367426</td>
<td>1.326531</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
<td>1.2</td>
<td>2.016581</td>
<td>2.000145</td>
<td>1.742656</td>
<td>1.730283</td>
</tr>
<tr>
<td>−0.6</td>
<td>−0.4</td>
<td>−0.2</td>
<td>1.890431</td>
<td>1.876201</td>
<td>1.523623</td>
<td>1.439646</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.547458</td>
<td>2.536042</td>
<td>2.070674</td>
<td>2.097547</td>
</tr>
<tr>
<td>−1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1.769005</td>
<td>1.759193</td>
<td>2.005566</td>
<td>1.996185</td>
</tr>
<tr>
<td>−3.0</td>
<td>0.0</td>
<td>3.0</td>
<td>2.439892</td>
<td>2.439888</td>
<td>2.888414</td>
<td>2.882915</td>
</tr>
<tr>
<td>0.0</td>
<td>2.0</td>
<td>3.0</td>
<td>2.047286</td>
<td>2.047029</td>
<td>2.476619</td>
<td>2.476619</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>2.5</td>
<td>1.949412</td>
<td>1.947140</td>
<td>2.376427</td>
<td>2.376427</td>
</tr>
<tr>
<td>−3.0</td>
<td>1.0</td>
<td>2.5</td>
<td>2.395149</td>
<td>2.395151</td>
<td>2.736289</td>
<td>2.736289</td>
</tr>
<tr>
<td>−3.0</td>
<td>−1.0</td>
<td>3.0</td>
<td>2.447210</td>
<td>2.486151</td>
<td>2.558530</td>
<td>2.558530</td>
</tr>
</tbody>
</table>

Table 3. MSEs of $\delta^*_L,0.05$, $\delta^*_L,0.01$, $\delta^*_L,1.0$ and $\delta^*_L,4.05$.

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\delta^*_L,0.05$</th>
<th>$\delta^*_L,0.01$</th>
<th>$\delta^*_L,1.0$</th>
<th>$\delta^*_L,4.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.981797</td>
<td>2.942540</td>
<td>1.569029</td>
<td>1.536455</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
<td>1.2</td>
<td>2.384612</td>
<td>2.337783</td>
<td>1.651914</td>
<td>1.726126</td>
</tr>
<tr>
<td>−0.6</td>
<td>−0.4</td>
<td>−0.2</td>
<td>2.645213</td>
<td>2.595927</td>
<td>1.560827</td>
<td>1.595940</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.981940</td>
<td>2.942540</td>
<td>1.569029</td>
<td>1.536455</td>
</tr>
<tr>
<td>−1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>2.031236</td>
<td>1.996846</td>
<td>2.100416</td>
<td>2.249125</td>
</tr>
<tr>
<td>−3.0</td>
<td>0.0</td>
<td>3.0</td>
<td>2.497311</td>
<td>2.476957</td>
<td>2.558596</td>
<td>2.942896</td>
</tr>
<tr>
<td>0.0</td>
<td>2.0</td>
<td>3.0</td>
<td>2.088515</td>
<td>2.074339</td>
<td>2.340738</td>
<td>2.563918</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>2.5</td>
<td>2.075240</td>
<td>2.003647</td>
<td>2.240890</td>
<td>2.435469</td>
</tr>
<tr>
<td>−3.0</td>
<td>1.0</td>
<td>2.5</td>
<td>2.441453</td>
<td>2.440496</td>
<td>2.525851</td>
<td>2.809094</td>
</tr>
<tr>
<td>−3.0</td>
<td>−1.0</td>
<td>3.0</td>
<td>2.487210</td>
<td>2.486151</td>
<td>2.558530</td>
<td>2.558530</td>
</tr>
</tbody>
</table>

than that of $\delta^*_L,0.05$ for $a > 1$. Also both $\delta^*_L,0.05$ and $\delta^*_L,4.05$ dominate $\tilde{X}$ for $0 < a \leq 2$. Further, $\delta^*_L,4.05$ always dominates $\delta^*_L,1.0$.

(iv) The risk comparison amongst estimators $\{\delta^*_L,i,a,i=1,2,3,\delta^*_4,a\}$ and $\{\delta^*_L,5,a,\delta^*_L,6,a\}$ is similar to that between $\delta^*_p$ and $\delta^*_MLE$. That is, estimators $\delta^*_L,5,a$ and $\delta^*_L,6,a$ perform better than estimators $\delta^*_L,i,a,i=1,2,3$ and $\delta^*_L,4,a$ when $\theta_i$s are close to each other and worse when
differences among them are large. Thus, if we have some prior information about values of \( \theta_i \)'s, a clear recommendation about use of improving estimators can be made.

(v) Among estimators \( \hat{\theta}_{4,a} \), \( a = 0.1, 0.5, 1.0, 1.5 \), the estimator \( \hat{\theta}_{4,1} \) seems to be the best choice. Similarly, among estimators \( \hat{\theta}_{6,a} \), \( a = 0.1, 0.5, 1.0, 1.5 \), the estimators \( \hat{\theta}_{6,1} \) seems to be the best choice.

(vi) In the class of Lindley–Efron–Morris type estimators, \( \hat{\theta}_{L1,a} \) and \( \hat{\theta}_{L2,a} \) seem to improve over \( \hat{\theta}_p \) for very small values of \( a \) only (we take \( a = 0.05 \) and 0.1, respectively, in table 3). Clearly, as estimators do not differ too much from \( \hat{\theta}_p \), the margins of improvement are also small. Surprisingly, \( \hat{\theta}_{L3,a} \) and \( \hat{\theta}_{L4,1,a} \) do not seem to provide uniform improvement over \( \hat{\theta}_p \) and \( \hat{\theta}_{MLE} \), respectively for any value of \( a \), though \( \hat{\theta}_{L3,1} \) and \( \hat{\theta}_{L4,0.5} \) improve upon \( X \). The reason for not so good performance of Lindley–Efron–Morris type estimators in our problem could be due to the fact that their estimators \( \hat{\theta}_{L} \) and \( \hat{\theta}_{L}^+ \) provided improvements over \( X \) for dimension \( k \geq 3 \) [27]. In addition, the choice of \( \psi(X) \) corresponding to the estimator \( \hat{\theta}_{L} \) does not yield \( \Delta(\hat{\theta}) \geq 0 \) for all \( \theta_1 \leq \theta_2 \leq \theta_3 \) for any choice of \( a \), where \( \Delta(\hat{\theta}) \) is the approximate risk difference obtained in Brown’s heuristic approach (as derived in section 2). It would be interesting to explore the case of \( k \geq 4 \) (though simulations become very tedious).

(vii) Estimators \( \hat{\theta}_{6,1} \) performs better than \( \hat{\theta}_{4,1} \) when \( \theta_i \)'s are close to each other and worse than \( \hat{\theta}_{4,1} \) when differences among \( \theta_i \)'s are large.

(viii) Since the estimator \( \hat{\theta}_{4,1} \) and \( \hat{\theta}_{6,1} \) lie inside \( \Omega \) and they perform better than \( \hat{\theta}_p \) and \( \hat{\theta}_{MLE} \), respectively, we recommend their use in practical applications. A clear choice between \( \hat{\theta}_{4,1} \) and \( \hat{\theta}_{6,1} \) can be made if one has additional information on the differences between \( \theta_i \)'s (see conclusion (vi)).

(ix) Finally, we observe that our proposed estimators are easy to use and in many cases they offer substantial improvements over \( \hat{\theta}_p \) and \( \hat{\theta}_{MLE} \), respectively. The amount of improvement is seen to be as large as 55% in some cases.

(x) Numerical comparisons in all cases were carried out for various other configurations of \( (\theta_1, \theta_2, \theta_3) \) (not presented in tables). However, we observed trends similar to those mentioned in comments (i)–(ix).

Acknowledgement

The authors are thankful to a referee for some very useful suggestions which have substantially improved the paper.

References


