On the Bias and Mean-square Error of Order-restricted Maximum Likelihood Estimators[1]

Sanjay Chaudhuri[a] and Michael D. Perlman[b]

Department of Statistics, University of Washington, Seattle, WA 98195, USA

[a] *sanjay@stat.washington.edu*
[b] *michael@ms.washington.edu*

26 August 2003

Abstract

An order-restricted (OR) statistical model can be expressed in the general form $\{P_\theta \mid \theta \in C\}$, where C is a convex cone in $\mathbb{R}^p$. In general, no unbiased estimator exists for θ. In particular, the OR maximum likelihood estimator (ORMLE) is biased, although its aggregate mean square error is usually less than that of the unrestricted MLE (URMLE). Nonetheless, the bias and mean-square error (MSE) of a single component or single linear contrast of the ORMLE can exceed those of the corresponding component or contrast of the URMLE by amounts that approach infinity as the dimension increases. This phenomenon is examined in detail for three examples: the orthant cone, the tree-order cone, and the simple-order cone. The geometric features of the cone that determine the growth rate of the bias and MSE are studied, and bias-reducing adjustments for certain components or contrasts of the ORMLE are suggested for the orthant and tree-order models.

Key Words: Order-restricted inference, bias, mean-square error, maximum likelihood estimator, nonnegative orthant, tree order, simple order, bias-reducing adjustment.

1 The inevitability of bias in order-restricted estimation.

One of the most influential papers on order-restricted (OR) inference is Chernoff (1954), which dealt with the problem of large-sample tests for OR models. An order-restricted (OR) statistical model $\{P_\theta \mid \theta \in C\}$ can be viewed as one whose parameter space C is a closed convex cone in a real Euclidean space $\mathbb{R}^p$ – cf. Robertson *et al* (1988). We assume that p is finite, although infinite-dimensional OR models also occur. In this paper we consider the question of estimation in an OR model.

Under very mild regularity conditions, no unbiased estimator of θ can exist in an OR model $\{P_\theta \mid \theta \in C\}$. Assume that the cone[2] C has nonempty boundary

[1] This research was supported in part by the U. S. National Science Foundation

[2] In fact, this argument holds if C is any closed convex set with nonempty boundary and nonempty interior.

∂C and nonempty interior C°. If $\tilde{\theta}$ is a valid estimator of θ (i.e., $\tilde{\theta} \in C$ for all sample values) such that

$$\Pr_{\theta'}[\tilde{\theta} \in C^\circ] > 0 \quad \text{for some } \theta' \in \partial C, \tag{1}$$

(a mild condition generally satisfied by estimators in families of continuous distributions with common support), then $\tilde{\theta}$ cannot be unbiased at θ'. For, if $\mathrm{E}_{\theta'}(\tilde{\theta}) = \theta'$ then $\Pr_{\theta'}[\tilde{\theta} \in H] = 1$ for any supporting hyperplane H for C through θ', hence $\Pr_{\theta'}[\tilde{\theta} \in \partial C] = 1$, contradicting (1). Thus $\tilde{\theta}$ cannot be unbiased at θ' and therefore, under any regularity conditions that guarantee continuity of expectations, cannot be unbiased in a neighborhood of θ'.

It is not surprising, therefore, that the order-restricted maximum likelihood estimator (ORMLE) is biased in many OR estimation problems. More striking, however, is the fact that the magnitude of the bias and mean-square error (MSE) of a single component (or single normalized linear combination, or single normalized linear contrast) of an ORMLE may increase without bound as the dimension p increases, which those of the corresponding unrestricted MLE (URMLE) do not vary with p (cf. Lee (1988), Robertson *et al.* (1988), Fernandez *et al.* (1999)). On this basis Cohen and Sackrowitz (2002) have deemed the ORMLE "undesirable", while Hwang and Peddada (1994) state that it "fails disastrously".

It is well known, however, that the ORMLE has favorable MSE properties relative to the URMLE in the aggregate, that is, when the loss considered is the total MSE $\mathrm{E}_\theta \|\hat{\theta} - \theta\|^2$ (cf. Brunk (1965), Barlow *et al.* (1972, Thm. 2.1), Robertson *et al.* (1988, p.41), Hwang and Peddada (1994, p.68)). In this paper we examine this apparent contradiction in closer detail for three important examples of polyhedral convex cones: the nonnegative orthant, the tree-order cone, and the simple-order cone. Assuming independent univariate normal distributions for simplicity, in the first two cases (Sections 2 and 3) we determine the exact rate at which the bias and MSE of the central[3] linear combination or central linear contrast of the means approach infinity as $p \to \infty$. For the simple-order cone (Section 4) we give an approximate argument suggesting that in fact the bias and MSE of the central linear contrast either remain bounded or grow at a much slower rate than those for the other two cones. The geometric features of general polyhedral cones that determine the growth rates of the MSE are discussed in Section 5.

In Section 6, for the orthant and tree-order cones we discuss bias-reducing adjustments for single components or linear contrasts of the ORMLE that partially or fully alleviate the unbounded growth of their biases and MSEs. For these cones, Chaudhuri and Perlman (2003a) suggest that the problem of estimating a single parameter or single contrast for the tree-order cone can be viewed as an estimation problem with many nuisance parameters. From this viewpoint, the need to adjust an ORML-based estimator ($\equiv$ ORMLBE) in order to reduce its bias is reminiscent of the classical Neyman-Scott problem[4]

[3]Fernandez *et al.* (1999) suggest that bias and MSE are most extreme for linear combinations or contrasts determined by the central direction of the cone.

[4]Neyman and Scott (1948, Example 2)

where the MLE of a common normal variance (the target parameter) in the presence of a growing number of observations from normal populations with unknown means (the nuisance parameters) is inconsistent. In both problems, the ORMLBE for the target parameter can be adjusted to reduce its bias, and even to achieve consistency[5] if information about the target parameter accumulates sufficiently rapidly. We conclude that the method of maximum likelihood neither is "undesirable" nor "fails disastrously" for order-restricted estimation, but may require adjustment if interest is restricted to a single target parameter.

In order to focus on the role of the geometry of the cones, we restrict attention to the simple case of independent normal observations with known and equal variances. Thus the data consists of a single random vector $X \equiv (X_1, \ldots, X_p)$, where $X_1, \ldots, X_p$ are mutually independent with $X_i \sim N(\mu_i, 1)$, $i = 1, \ldots, p$. Here $\theta = \mu \equiv (\mu_1, \ldots, \mu_p)$ and the assumed order restriction is expressed as $\mu \in C$. By normality, the URMLE is just X itself and is unbiased for μ, with aggregate MSE $\mathrm{E}_\mu \|\hat{\mu} - \mu\|^2 = p$. The ORMLE $\hat{\mu}$ is the orthogonal projection of X onto C and will be described in more detail for each example.

2 The nonnegative orthant cone

First consider the case where

$$C = C^p_{\mathrm{or}} \equiv \{\mu \mid \mu_1 \geq 0, \ldots, \mu_p \geq 0\}, \tag{2}$$

the nonnegative orthant in $\mathbb{R}^p$. Here the ORMLE $\hat{\mu}$ is given by

$$\hat{\mu} \equiv (\hat{\mu}_1, \ldots, \hat{\mu}_p) = (X_1^+, \ldots, X_p^+), \tag{3}$$

where $y^+ = \max(y, 0)$. As noted above, $\hat{\mu}$ cannot be unbiased for μ. It is straightforward to show that the MSE of each component X_i^+ of the ORMLE $\hat{\mu}$ is less than that of the corresponding component X_i of the URMLE X. This fails dramatically, however, for other normalized linear combinations of $\hat{\mu}$.

In particular, let e denote the unit vector $(1/\sqrt{p}, \ldots, 1/\sqrt{p})$, which represents the central direction of the orthant C^p_{or}. The unrestricted ML-based[6] estimator (URMLBE) of the central linear combination $\mu e^t \equiv \frac{1}{\sqrt{p}}(\mu_1 + \cdots + \mu_p)$ is given by $X e^t$ and is unbiased for μe^t. The order-restricted ML-based estimator (ORMLBE) of μe^t is given by

$$\hat{\mu} e^t \equiv \frac{1}{\sqrt{p}} \sum_{i=1}^{p} X_i^+ \tag{4}$$

[5]See Chaudhuri and Perlman (2003a) for a discussion of consistency for the tree-order model.

[6]We reserve the phrase "maximum likelihood estimator" to indicate the MLE of the *entire parameter vector*, rather than of a single parameter, because the underlying rationale for ML inference is to estimate the entire distribution (actually, probability density function) under which the observed data is "most likely" – see Bahadur (1971, §9). We therefore use the phrase "ML-based estimator" (MLBE) for a single component of the MLE (or for any non 1-1 functional of the MLE).

and is biased, with bias given by

$$\begin{aligned} \mathrm{E}_\mu(\hat{\mu}e^t) - \mu e^t &= \frac{1}{\sqrt{p}} \sum_{i=1}^p [\mathrm{E}_\mu(X_i^+) - \mu_i] \\ &= \frac{1}{\sqrt{p}} \sum_{i=1}^p \left(\varphi(\mu_i) - \mu_i[\bar{\Phi}(\mu_i)]\right), \end{aligned} \tag{5}$$

where φ and Φ denote the standard normal pdf and cdf, respectively, and $\bar{\Phi} = 1 - \Phi$. Since $\varphi(u) - u\bar{\Phi}(u)$ is nonnegative and decreasing for $0 \le u < \infty$, the bias (5) is greatest when $\mu = 0$, where it attains the value $\sqrt{p/2\pi}$, hence in this case the squared bias $p/2\pi$ approaches ∞ at rate $\mathrm{O}(p)$ as $p \to \infty$. By independence, the variance of $\hat{\mu}e^t$ when $\mu = 0$ is given by

$$\begin{aligned} \mathrm{Var}_0(\hat{\mu}e^t) &= \frac{1}{p} \sum_{i=1}^p \mathrm{Var}_0(X_i^+) \\ &= \mathrm{Var}_0(X_1^+) \\ &= \frac{1}{2}\left(1 - \frac{1}{\pi}\right), \end{aligned} \tag{6}$$

which does not vary with p, so when $\mu = 0$ the dominant contribution to the MSE of the ORMLBE $\hat{\mu}e^t$ comes from its bias rather than its variance. Precisely, the MSEs of the URMLBE Xe^t and the ORMLBE $\hat{\mu}e^t$ at $\mu = 0$ are, respectively,

$$\mathrm{E}_0[(Xe^t)^2] = 1, \tag{7}$$

$$\mathrm{E}_0[(\hat{\mu}e^t)^2] = \frac{p - 1 + \pi}{2\pi} = .159p + .341. \tag{8}$$

Finally, the total MSEs of the URMLE and the ORMLE at $\mu = 0$ are given by

$$\mathrm{E}_0\|X\|^2 \equiv \mathrm{E}_0(X_1^2 + \cdots + X_p^2) = p, \tag{9}$$

$$\mathrm{E}_0\|\hat{\mu}\|^2 \equiv \mathrm{E}_0((X_1^+)^2 + \cdots + (X_p^+)^2) = \frac{p}{2}. \tag{10}$$

From these calculations we see that in the least favorable case that $\mu = 0$, the squared bias and MSE of the ORMLBE $\hat{\mu}e^t$ both approach ∞ at a linear rate[7] as $p \to \infty$ and account for considerable fractions of the total bias and MSE of the ORMLE, whereas this does not hold for the URMLBE Xe^t. However, the total MSE of the ORMLE is substantially smaller than that of the URMLE.

3 The tree-order cone

Another example of this apparent contradiction is provided by the case where C is the *tree-order cone* C_{to}^{s+1} in $\mathbb{R}^p \equiv \mathbb{R}^{s+1}$, defined by

$$C_{\mathrm{to}}^{s+1} = \{\mu \equiv (\mu_0, \mu_1, \ldots, \mu_s) \in \mathbb{R}^{s+1} \mid \mu_0 \le \mu_i,\ i = 1, \ldots, s\}, \tag{11}$$

[7] A similar linear growth rate was noted by Fernandez *et al.* for the case where C is a proper circular cone – see Section 5 below.

a non-pointed obtuse cone with spine[8] given by the line

$$L^{s+1} = \{\mu \mid \mu_0 = \mu_1 = \cdots = \mu_s\} \equiv \text{span}\{e\}, \tag{12}$$

where now $e = (1, 1, \ldots, 1) \in \mathbb{R}^{s+1}$. Let the data vector $X \equiv (X_0, X_1, \ldots, X_s)$ consist of independent normal observations with $X_i \sim N(\mu_i, 1)$, $i = 0, 1, \ldots, s$. Here $p = s+1$, $\theta = \mu$, and the order restrictions are given by $\mu_0 \leq \mu_1, \ldots, \mu_0 \leq \mu_s$. The tree-order restrictions arise naturally when considering the problem of comparing several treatments to a control. Estimation for the tree-order model has been investigated by Lee (1988), Robertson *et al.* (1998), and Fernandez *et al.* (1999).

The ORMLE $\hat{\mu} \equiv (\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_s)$ is the orthogonal projection of X onto C_{to}^{s+1}, hence satisfies[9] $\hat{\mu} e^t = X e^t$, i.e.,

$$\hat{\mu}_0 + \hat{\mu}_1 + \cdots + \hat{\mu}_s = X_0 + X_1 + \cdots + X_s. \tag{13}$$

Furthermore, $\hat{\mu}$ can be expressed explicitly as (cf. Lee (1988) eqn. (1.6))

$$\begin{aligned}
\hat{\mu}_0 &= \min_{\sigma \subseteq \{1,2,\ldots,s\}} \left\{ \frac{X_0 + \sum \left(X_i \mid i \in \sigma \right)}{1 + |\sigma|} \right\}, && (14) \\
&= \min_{k=0,\ldots,s} \left\{ \frac{X_0 + \sum \left(X_{(i)} \mid i = 1, \ldots, k \right)}{1 + k} \right\}, && (15) \\
&= \min_{k=0,\ldots,r^*-1} \left\{ \frac{X_0 + \sum \left(X_{(i)} \mid i = 1, \ldots, k \right)}{1 + k} \right\}, && (16) \\
\hat{\mu}_i &= \max\,(\hat{\mu}_0, X_i), \qquad i = 1, \ldots, s, && (17)
\end{aligned}$$

where $X_{(1)}, \ldots, X_{(s)}$ are the order statistics for $X_1, \ldots, X_s$ and r^* is the rank of X_0 among $X_0, X_1, \ldots, X_s$. Lee (1988, Thms. 4.1 and 2.1) showed that for each $i = 1, \ldots, s$, the MSE of the ORMLBE $\hat{\mu}_i$ of the treatment mean μ_i is less than that of the URMLBE X_i, but that the opposite is true for the ORMLBE of the control mean μ_0. Here we study the growth rate of the MSE of $\hat{\mu}_0$ as s, the number of treatments, increases.

From (14),

$$\begin{aligned}
\hat{\mu}_0 &\sim \mu_0 + \min_{\sigma \subseteq \{1,2,\ldots,s\}} \left\{ \frac{N(0,1) + \sum \left(N(\mu_i - \mu_0, 1) \mid i \in \sigma \right)}{1 + |\sigma|} \right\} && (18) \\
&<_{\text{stoch}} N(\mu_0, 1), && (19)
\end{aligned}$$

so the bias $\hat{b}_0(\mu)$ is strictly negative, i.e.,

$$\hat{b}_0(\mu) \equiv \mathrm{E}_\mu(\hat{\mu}_0) - \mu_0 < 0, \tag{20}$$

[8] A convex cone $C \in \mathbb{R}^p$ is non-pointed if it contains a nontrivial linear subspace. Its spine L is the maximal such subspace, and C can be uniquely represented as the product $L \times (C \cap L^\perp)$, where $C \cap L^\perp$ is a pointed cone in $\mathbb{R}^{p - \dim L}$ that is also obtained by projecting C onto $L^\perp$. We refer to $C \cap L^\perp$ as the *reduced cone.*

[9] Since $C_{\text{to}}^{s+1} = L^{s+1} \times ((C_{\text{to}}^{s+1}) \cap (L^{s+1})^\perp)$.

and is greatest in magnitude when $\mu_0 = \mu_1 = \cdots = \mu_s$, i.e., when[10] $\mu \in L^{s+1}$. From (15),

$$\min(X_0', X_{(1)}') \leq \hat{\mu}_0 - \mu_0 \leq \frac{X_0' + X_{(1)}'}{2}, \tag{21}$$

where $X_i' = X_i - \mu_0 \sim N(\mu_i - \mu_0, 1)$ and $X_{(1)}' = \min(X_1', \ldots, X_s')$. It follows from extreme value theory (Pickands (1968), Galambos (1978, §2.3.2), Johnson and Kotz (1970, Ch.21)) that when $\mu \in L^{s+1}$ (so $\mu_i = \mu_0$ for $i = 1, \ldots, s$),

$$\mathrm{E}_\mu(X_{(1)}') \sim -\sqrt{2\log s} \tag{22}$$

$$\mathrm{E}_\mu(X_{(1)}')^2 \sim 2\log s \tag{23}$$

as $s \to \infty$, hence from (21),

$$|\mathrm{E}_\mu(\hat{\mu}_0) - \mu_0| = \mathrm{O}(\sqrt{2\log s}), \tag{24}$$

$$\mathrm{E}_\mu(\hat{\mu}_0 - \mu_0)^2 = \mathrm{O}(2\log s). \tag{25}$$

Thus in the least favorable case where $\mu \in L^{s+1}$, the bias and MSE of the ORMLBE $\hat{\mu}_0$ again become arbitrarily large as $s \to \infty$, although at a slower rate than for the orthant cone, whereas the bias and MSE of the URMLBE X_0 are 0 and 1, respectively, for every s.

Next, consider the problem of estimating the central contrast

$$\mu f^t \equiv \frac{1}{\sqrt{s(s+1)}} \sum_{i=1}^{s} (\mu_i - \mu_0) \tag{26}$$

$$\equiv \sqrt{\frac{s}{s+1}}\, \bar{\Delta}, \tag{27}$$

where $\Delta_i = \mu_i - \mu_0$ and

$$f \equiv \frac{1}{\sqrt{s(s+1)}}(-s, 1, \ldots, 1) \tag{28}$$

is the unit vector that indicates the central direction[11] of the cone C_{to}^{s+1}, cf. Robertson *et al.* (1988, pp.180-1). (Note that $f \in (C_{\mathrm{to}}^{s+1}) \cap (L^{s+1})^{\perp}$.) By (13),

$$\mathrm{E}_\mu(\hat{\mu}_0 + \hat{\mu}_1 + \cdots + \hat{\mu}_s) = \mu_0 + \mu_1 + \cdots + \mu_s, \tag{29}$$

so by symmetry,

$$\mathrm{E}_\mu(\hat{\mu}_0 - \mu_0) = -s\mathrm{E}_\mu(\hat{\mu}_1 - \mu_1) \tag{30}$$

[10] Because $\hat{\mu}$ is location-equivariant, the bias and MSE of both $\hat{\mu}_0$ and $\hat{\mu} f^t$ are constant for $\mu \in L^{s+1}$.

[11] See Footnote 3

when $\mu \in L^{s+1}$. Thus, when $\mu \in L^{s+1}$, the bias of the ORMLBE

$$\hat{\mu} f^t \equiv \frac{1}{\sqrt{s(s+1)}} \sum_{i=1}^{s} (X_i - \hat{\mu}_0)^+ \tag{31}$$

is

$$\mathrm{E}_\mu(\hat{\mu} f^t) - \mu f^t = -\sqrt{\frac{s+1}{s}}[\mathrm{E}_\mu(\hat{\mu}_0) - \mu_0], \tag{32}$$

which is again $\mathrm{O}(\sqrt{2 \log s})$ by (24). The MSE of $\hat{\mu} f^t$ when $\mu \in L^{s+1}$ is given by

$$\begin{aligned}
\mathrm{E}_\mu(\hat{\mu} f^t - \mu f^t)^2 &= \frac{1}{s(s+1)} \mathrm{E}_\mu[(\hat{\mu}_1 - \mu_1) + \cdots + (\hat{\mu}_s - \mu_s) - s(\hat{\mu}_0 - \mu_0)]^2 \\
&= \frac{1}{s(s+1)} \mathrm{E}_\mu(X'_0 + X'_1 + \cdots + X'_s - (s+1)(\hat{\mu}_0 - \mu))^2 \qquad (33) \\
&\equiv \frac{s+1}{s} \mathrm{E}_\mu(\bar{X}'_{s+1} - (\hat{\mu}_0 - \mu))^2 \qquad (34) \\
&= \frac{s+1}{s} \left[\mathrm{E}_\mu(\hat{\mu}_0 - \mu_0)^2 + \mathrm{o}(1)\right] \qquad (35) \\
&= \mathrm{O}(2 \log s) \qquad (36)
\end{aligned}$$

by (25), where (33) follows from (13) and (35) follows since $\bar{X}'_{s+1} = \mathrm{O}(s^{-1/2})$. Thus, both the squared bias and MSE of the ORMLBE $\hat{\mu} f^t$ again approach infinity at a logarithmic rate in the least favorable case $\mu \in L^{s+1}$.

It is this behavior of the ORMLBE, as well as that demonstrated in Section 2, that has led others to conclude that the ORMLE "fails disastrously", in apparent contradiction to the fact that the aggregate MSE of the ORMLE is smaller than that of the URMLE, often by a substantial amount. Specifically, for the orthant cone the aggregate MSEs of the URMLE and ORMLE in the least favorable case $\mu = 0$ are given by (9) and (10), respectively, while for the tree-order cone these aggregate MSEs in the least favorable case $\mu \in L^{s+1}$ are, respectively,

$$\begin{aligned}
\mathrm{E}_\mu \|X - \mu\|^2 &= 1 + s, \qquad (37) \\
\mathrm{E}_\mu \|\hat{\mu} - \mu\|^2 &\equiv \mathrm{MSE}_\mu(\hat{\mu}_0) + s\mathrm{MSE}_\mu(\hat{\mu}_1) \\
&= \mathrm{O}(2 \log s) + s\mathrm{MSE}_\mu(\hat{\mu}_1) \qquad (38) \\
&\ll 1 + s
\end{aligned}$$

by symmetry, (25), and the fact that $\mathrm{MSE}_\mu(\hat{\mu}_1) < 1$ (Lee (1988), Theorem 4.1).

4 The simple-order cone

Finally, consider the case where $C = C^p_{\mathrm{so}}$, the well-known *simple order cone* in $\mathbb{R}^p$ (cf. Robertson*et al.* (1988)) given by

$$C^p_{\mathrm{so}} = \{\mu \equiv (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p \mid \mu_1 \leq \cdots \leq \mu_p\}, \tag{39}$$

which is a non-pointed acute polyhedral cone with spine given by the line

$$L^p = \text{span}\{e_p\} \equiv \{\mu \mid \mu_1 = \cdots = \mu_p\}, \tag{40}$$

where $e_p = (1, \ldots, 1) \in \mathbb{R}^p$. As in Section 2, estimation of μ is to be based on $X \equiv (X_1, \ldots, X_p)$, where the X_i are independent and $X_i \sim N(\mu_i, 1)$, with order restrictions given by $\mu_1 \leq \cdots \leq \mu_p$.

Lee (1981) showed that for each $i = 1, \ldots, p$, the MSE of the individual ORMLBE $\hat{\mu}_i$ is less than that of the URMLBE X_i. On the other hand, Fernandez *et al.* (1999) showed that when $\mu_1 = \cdots = \mu_p$, i.e., when $\mu \in L^p$, the MSE of the ORMLBE $\hat{\mu}c^t$ for the contrast μc^t determined by the central direction c of C^p_{so}. is greater than that of the URMLBE Xc^t, but they did not specify the magnitude of the exceedance.[12]

Here we address the following question: when $\mu \in L^p$, do the bias and MSE of the ORMLBE for this central contrast approach ∞ as the dimension $p \to \infty$, as for the tree-order and orthant cones, or do they remain bounded? Such questions involving the simple-order cone are more complex to analyze, since this cone possesses fewer symmetries than the orthant and tree-order cones. We give an approximate argument that suggests that when $\mu \in L^p$ (which is assumed for the rest of this section), the squared bias and MSE of the central contrast[13] either remain bounded as the dimension p increases or else grow at an even slower rate than in the tree-order case, namely, rate $\log \log p$.

The unit edge vectors $u_1, \ldots, u_{p-1}$ of the reduced cone $C^p_{\text{so}} \cap (L^p)^\perp \equiv C^*$ are given by

$$u_i = \frac{1}{\sqrt{i(p-i)p}}(\overbrace{-(p-i), \ldots, -(p-i)}^{i}, \overbrace{i, \ldots, i}^{p-i}). \tag{41}$$

For $1 \leq i \leq j \leq p-1$, the cosine of the angle between u_i and u_j is given by

$$c_{ij} \equiv \sqrt{\frac{i(p-j)}{j(p-i)}}. \tag{42}$$

The unit vector c that defines the central direction of C^p_{so} has been determined by Abelson and Tukey (1963) and also appears in Robertson *et al.* (1988, eqn. (4.2.5)). It is defined to be the vector in C^* that maximizes the minimum angle between it and the edge vectors $u_1, \ldots, u_{p-1}$, and is equiangular with these edges. Fernandez *et al* (1999) denote the cosine of this maximin angle by $r \equiv r_{\min}$, so that

$$r = u_i\, c^t, \qquad i = 1, \ldots, p-1. \tag{43}$$

The exact form of c is somewhat complex, but Abelson and Tukey showed that

$$r^2 \approx \frac{2}{2 + \log(p-1)} \quad \text{for large } p. \tag{44}$$

[12] As in the cases of the orthant and tree-order cones, it is expected that the bias and MSE of the ORMLBE $\hat{\mu}c^t$ are greatest when $\mu \in L^p$.

[13] Because $\hat{\mu}$ is also location-equivariant, its bias and MSE are constant for $\mu \in L^p$.

Let $\hat{\mu}^{\perp}$ denote the orthogonal projection of the MLE $\hat{\mu}$ onto $(L^p)^{\perp}$. (This was denoted by $X_{\perp}^{*}$ by Fernandez *et al.* (1999).) Note that $\hat{\mu}^{\perp} \in C_{\text{so}}^{p} \cap (L^p)^{\perp}$ and $\hat{\mu}c^t = \hat{\mu}^{\perp}c^t$. Thus, the MSE of $\hat{\mu}c^t$ when $\mu \in L^p$ is given by

$$\begin{aligned} \mathrm{E}_{\mu}(\hat{\mu}^{\perp}c^t)^2 &= \mathrm{E}_{\mu}[\mathrm{E}_0[(\hat{\mu}^{\perp}c^t)^2 \mid F]] \\ &= \mathrm{E}_{\mu}[\mathrm{E}_0[\|\hat{\mu}^{\perp}\|^2 \cos^2(\alpha(\hat{\mu}^{\perp}, c)) \mid F]] \qquad (45) \\ &= \mathrm{E}_{\mu}[K_F \cdot \mathrm{E}_{\mu}[\cos^2(\alpha(\hat{\mu}^{\perp}, c)) \mid F]], \qquad (46) \end{aligned}$$

where $F \equiv F(\hat{\mu}^{\perp})$ denotes the face of C^* in which $\hat{\mu}^{\perp}$ lies, $\alpha(\hat{\mu}^{\perp}, c)$ is the angle between $\hat{\mu}^{\perp}$ and c, and K_F is the dimension of F $(0 \leq K_F \leq p-1)$. The equality (46) follows from Lemma B of Robertson *et al.* (1988, p.70).

It is convenient to describe the (open) face F of the $(p-1)$-dimensional reduced cone C^* as the positive span of a unique subset

$$\{u_i \mid i \in I_F\} \subseteq \{u_1, \ldots, u_{p-1}\}, \qquad (47)$$

where $|I_F| = K_F$. (Set $I_{\{0\}} = \emptyset$ and $K_{\{0\}} = 0$.) Thus, when $\hat{\mu}^{\perp} \in F$,

$$\hat{\mu}^{\perp} = \sum_{i \in I_F} \lambda_i \cdot u_i \ , \qquad (48)$$

where each $\lambda_i \equiv \lambda_i(\hat{\mu}^{\perp}) > 0$. Thus if $K_F > 0$, it follows from (43) that

$$\begin{aligned} \cos^2(\alpha(\hat{\mu}^{\perp}, c)) &= \frac{(\hat{\mu}^{\perp}c^t)^2}{\|\hat{\mu}^{\perp}\|^2} \qquad (49) \\ &= \frac{r^2 \cdot \left(\sum_{i \in I_F} \lambda_i\right)^2}{\|\sum_{i \in I_F} \lambda_i \cdot u_i\|^2} \qquad (50) \\ &= \frac{r^2}{\|\bar{u}_{F,w_F}\|^2}, \qquad (51) \end{aligned}$$

where

$$\begin{aligned} w_F &\equiv w_F(\hat{\mu}^{\perp}) = (w_i \mid i \in I_F), \\ w_i &\equiv w_i(\hat{\mu}^{\perp}) = \frac{\lambda_i}{\sum_{j \in I_F} \lambda_j} \\ \bar{u}_{F,w_F} &\equiv \bar{u}_{F,w_F}(\hat{\mu}^{\perp}) = \sum_{i \in I_F} w_i \cdot u_i \ . \end{aligned}$$

(Note that $\sum_{i \in I_F} w_i = 1$). Thus from (46),

$$\begin{aligned} \mathrm{E}_{\mu}(\hat{\mu}^{\perp}c^t)^2 &\leq r^2 \cdot \mathrm{E}_{\mu}\left[\frac{K_F}{m_F^+}\right] \qquad (52) \\ &= r^2 \cdot \mathrm{E}_{\mu}\left[\frac{K_F}{m_{K_F}^+}\right], \qquad (53) \end{aligned}$$

where

$$m_F^+ = \min_{w \in \Omega_F^+} \Big\| \sum_{i \in I_F} w_i u_i \Big\|^2, \tag{54}$$

$$w_F^+ = \Big\{ w \equiv (w_i \mid i \in I_F) \Bigm| \text{all } w_i > 0,\ we_F^t = 1 \Big\}$$

$$e_F = (1, \ldots, 1): \quad \{1\} \times I_F,$$

$$\frac{1}{m_k^+} = \mathrm{E}_\mu \left[\frac{1}{m_F^+} \middle| K_F = k \right], \tag{55}$$

and where we set $\frac{0}{0} = 0$. Here, m_F^+ is the squared distance between the origin and the simplex S_F determined by the unit edge vectors $\{u_i \mid i \in I_F\}$ of the face F.

An upper bound possibly slightly less sharp than (53) but which can be given in a more explicit form is obtained by replacing m_k^+ by $m_k \leq m_k^+$, where

$$\frac{1}{m_k} = \mathrm{E}_\mu \left[\frac{1}{m_F} \middle| K_F = k \right], \tag{56}$$

$$m_F \equiv \min_{w \in \Omega_F} \Big\| \sum_{i \in I_F} w_i u_i \Big\|^2 \tag{57}$$

$$\equiv \min_{w \in \Omega_F} \|wU_F\|^2 \tag{58}$$

$$= (e_F A_F^{-1} e_F^t)^{-1} \tag{59}$$

$$= \left[1 - \frac{\det(\tilde{A}_F)}{\det(A_F)} \right]^{-1}. \tag{60}$$

Here m_F is the minimum squared distance between the origin and the $(K_F - 1)$-dimensional hyperplane $H_F \equiv \{wU_F \mid w \in \Omega_F\}$ affinely spanned by $(u_i \mid i \in I_F)$,

$$\Omega_F = \Big\{ w \equiv (w_i \mid i \in I_F) \Bigm| we_F^t = 1 \Big\},$$

$$A_F = U_F U_F^t$$

$$\equiv (a_{ij} \mid (i,j) \in I_F \times I_F): \quad I_F \times I_F,$$

$$\tilde{A}_F = \begin{pmatrix} 1 & e_F \\ e_F^t & A_F \end{pmatrix},$$

U_F is the $I_F \times \{1, \ldots, p\}$ matrix whose rows are $(u_i \mid i \in I_F)$, and

$$a_{ij} = c_{i \wedge j, i \vee j}, \tag{61}$$

(recall (42)). The relation (58) = (59) is a standard result from generalized least squares theory (e.g., see (1f.1.3) in Rao (1973)), and this minimum occurs for

$$w \equiv \omega_F = (e_F A_F^{-1} e_F^t)^{-1} e_F A_F^{-1}. \tag{62}$$

Therefore the MSE of $\hat{\mu}c^t$ when $\mu = 0$ satisfies

$$\begin{aligned} \mathrm{E}_\mu(\hat{\mu}^\perp c^t)^2 &\leq r^2 \cdot \mathrm{E}_\mu \left[\frac{K_F}{m_{K_F}} \right] &(63)\\ &\equiv \mathrm{E}_\mu \left[K_F \cdot c_{K_F} \right], &(64) \end{aligned}$$

where

$$c_k = \mathrm{E}_\mu \left[\cos^2 \alpha_F \;\middle|\; K_F = k \right] \quad (65)$$

and α_F is the angle between the face F and the central direction c.

By (62), the minimum in (57) is attained at the point $h_F \in H_F$ given by

$$h_F = w_F U_F \equiv (e_F A_F^{-1} e_F^t)^{-1} e_F A_F^{-1} U_F. \quad (66)$$

Therefore

$$h_F U_F^t = (e_F A_F^{-1} e_F^t)^{-1} e_F \propto e_F, \quad (67)$$

so h_F is equiangular with each u_i, $i \in I_F$. In particular, when $F \equiv F^*$ is the maximal face[14] of the cone C^* (so $K_{F^*} = p - 1$), h_{F^*} is equiangular with $u_1, \ldots, u_{p-1}$, hence h_{F^*} must lie along the central direction c, i.e., $h_{F^*} = \beta c$ for some real β. Thus by (43) and (44),

$$\begin{aligned} m_{F^*} &\equiv \|h\|^2 \\ &= \beta^2 \\ &= (\beta c c^t)^2 \\ &= \left[w_{F^*} U_{F^*} c^t \right]^2 \\ &= \left[r \cdot \omega_{F^*} e_{F^*}^t \right]^2 \\ &= r^2 &(68)\\ &\approx \frac{2}{2 + \log(p-1)} &(69) \end{aligned}$$

for large p. Because $m_F > m_{F^*}$ when $K_F \leq p - 2$, it follows from (63), (68), and (3.2.2) of Robertson *et al.* (1988, p.121) that the MSE of $\hat{\mu}c^t$ when $\mu = 0$ satisfies

$$\begin{aligned} \mathrm{E}_\mu(\hat{\mu}^\perp c^t)^2 &\leq \mathrm{E}_\mu[K_F \mid K_F > 0] \mathrm{P}_\mu[K_F > 0] \\ &= \mathrm{E}_\mu[K_F] \\ &= \sum_{k=2}^{p} \frac{1}{k} \\ &\leq \log p, &(70) \end{aligned}$$

which bound is of the same order of magnitude as the bound (36) found in the tree-order case.

[14] That is, F^* is the interior of the cone C^*

However, the lower bound $m_F > m_{F^*}$ is not sharp when $1 \leq K_F \leq p-2$; for example, $m_F = 1$ when $K_F = 1$. Thus it seems likely that the upper bound $\log p$ in (70) can be improved substantially. Recall that (59) and (60) provide explicit expressions for m_F, but these seem difficult to evaluate in closed form. Instead we present the following approximate argument.

Motivated by the relation (69), we conjecture that for all $1 \leq k \leq p-2$,

$$\frac{1}{m_k} \approx 1 + \frac{1}{2}\log k, \tag{71}$$

which is valid for $k = p-1$ and can be verified explicitly for small values of k and numerically for some larger values of k. If so, then it would follow from (63) that the MSE of $\hat{\mu}c^t$ when $\mu = 0$ approximately satisfies the inequality

$$\begin{aligned} \mathrm{E}_\mu(\hat{\mu}^\perp c^t)^2 &\leq r^2 \cdot \mathrm{E}_\mu\left[K_F\left(1 + \frac{1}{2}\log K_F\right)\right] \\ &= r^2 \cdot \mathrm{E}_\mu(K_F)\left[1 + \frac{1}{2}\mathrm{E}^*(\log K_F)\right] \\ &\leq r^2 \cdot \mathrm{E}_\mu(K_F)\left[1 + \frac{1}{2}\log(\mathrm{E}^*(K_F))\right] \qquad (72) \\ &= r^2 \cdot \mathrm{E}_\mu(K_F)\left[1 + \frac{1}{2}\log\left(\frac{\mathrm{E}_\mu(K_F^2)}{\mathrm{E}_\mu(K_F)}\right)\right]. \end{aligned}$$

Here, (72) follows from Jensen's inequality and E^* denotes expectation with respect to the distribution of K_F given by the probabilities $\pi_k^* \propto k\pi_k$, $k = 1, \ldots, p-1$, where $(\pi_0, \pi_1, \ldots, \pi_{p-1})$ is the distribution of K_F when $\mu = 0$. By (3.2.2) of Robertson *et al.* (1988, p.121), however,

$$\mathrm{E}_\mu(K_F) = \sum_{k=2}^{p}\frac{1}{k} \approx \log p, \tag{73}$$

$$\mathrm{E}_\mu(K_F^2) \leq \left(\sum_{k=2}^{p}\frac{1}{k}\right)^2 + \sum_{k=2}^{p}\frac{1}{k} \approx (\log p)^2 + \log p, \tag{74}$$

so by (44), if (71) is valid then the MSE of the ORMLBE $\hat{\mu}c^t$ when $\mu = 0$ will approximately satisfy

$$\begin{aligned} \mathrm{E}_\mu(\hat{\mu}^\perp c^t)^2 &\leq \frac{2\log p}{2 + \log(p-1)}\left[1 + \frac{1}{2}\log(\log p + 1)\right] \qquad (75) \\ &\sim \log\log p \qquad (76) \end{aligned}$$

as $p \to \infty$. This would be a much slower growth rate than (8) and (36) found for the orthant and tree-order cones, respectively.

Of course, because (76) is an approximate *upper* bound, this would not eliminate the possibility that this MSE remains bounded as $p \to \infty$.

5 MSE growth rate and the cone geometry

The bounds (53) and (63) extend directly to other polyhedral cones C such as the orthant cone and the tree-order cone. The former shows that the growth rate of the MSE of the ORMLBE for the central contrast in the least favorable case where μ lies in the spine L of C is determined by three aspects of the geometry of C, specifically, by

(i) r^2, the squared cosine of the angle between the central direction and the edge vectors of the reduced cone;

(ii) m_k^+, the average of the reciprocals of the squared distances between the origin and the simplices S_F such that $K_F = k$;

(iii) the distribution of K_F (the so-called *level probabilities*, cf. Robertson *et al.* (1988)).

By (64), the information in (ii) and (iii) is equivalent to that given by

(iv) c_k, the average over all k-dimensional faces of the squared cosine of the angle between the central direction and the face.

For the orthant cone C_{or}^p, $r^2 = 1/p$, $K_F \sim \text{Binomial}(p, \frac{1}{2})$, and $m_k^+ = 1/k$,so the upper bound (53) is

$$\frac{1}{p} \cdot \mathrm{E}_\mu[K_F^2] = \frac{p+1}{4}, \tag{77}$$

which is of the same order of magnitude as the exact value given by (8). For the tree-order cone C_{to}^{s+1}, $r^2 = 1/s^2$ and $m_k^+ = \frac{1}{k} - \frac{1}{s} + \frac{1}{ks}$, so the upper bound (53) becomes

$$\frac{1}{s} \cdot \mathrm{E}_\mu \left[\frac{K_F^2}{s - K_F + 1} \right]. \tag{78}$$

The distribution of K_F and its moments for the tree-order cone currently are unavailable in the literature, but it would be of interest to determine whether (78) is of the same order of magnitude $(2 \log p)$ as (25).

For the simple-order cone, (44) and the conjecture (71) suggest that r^2 and m_k approach 0 more slowly than for the other two cones as the dimension increases. Thus, for these examples, the slower the approaches to 0, the slower the growth of the MSE. Furthermore, it can be seen that larger values of r^2 and m_k^+ occur for narrower cones. For example, if the edge vectors $\{u_i\}$ are approximately colinear then $r^2 \approx 1$ and $m_k^+ \approx 1$. Indeed, the simple-order cone is narrower than the orthant and tree-order cones, for the minimum pairwise angle between the edges of the acute reduced cone $C_{\text{so}}^p \cap (L^p)^\perp$ approaches 0 as $p \to \infty$ (see (41) and (42)), whereas the minimum pairwise edge angle[15] for the

[15]The unit edge vectors for the reduced cone $C_{\text{to}}^{s+1} \cap (L^{s+1})^\perp$ are proportional to $(-1, s, -1, \ldots, -1), \ldots, (-1, \ldots, -1, s)$.

obtuse reduced cone $C_{\rm to}^{s+1} \cap (L^{s+1})^{\perp}$ approaches $\pi/2$. Obviously, each pairwise edge angle for the orthant cone is exactly $\pi/2$.

The narrowness of the cone is not the sole feature that determines the growth rate of the MSE, however, for although the tree-order cone is somewhat wider[16] than the orthant, its MSE grows at a slower rate. Furthermore, when the cone C either degenerates to a single ray or expands to an entire halfspace, estimation of the central contrast becomes a one-dimensional problem and the MSE does not depend on the dimension of the space at all. Thus, it is the interplay among (i), (ii), and (iii) that determines the MSE growth rate.

Finally, as in Fernandez *et al.* (1999, §3) consider the case of a circular cone C_ω^p in $\mathbb{R}^p$ with central half-angle ω. They obtained the following lower bound for the MSE of the ORMLBE $\hat{\mu}^{\perp} c^t$ of the central linear combination in the least favorable case $\mu = 0$:

$$\mathrm{E}_0(\hat{\mu}^{\perp} c^t)^2 \geq (\cos^2 \omega)[(p-1)(\sin^2 \omega) + \frac{1}{2}]. \tag{79}$$

Because C_ω^p can be viewed approximately as a polyhedral cone with faces of dimensions 0, $p-1$, and p only, our upper bound (64) becomes

$$\begin{aligned} \mathrm{E}_0(\hat{\mu}^{\perp} c^t)^2 &\leq (p-1)\cos^2\omega(1 - \mathrm{P}_0[K_F = 0] - \mathrm{P}_0[K_F = p]) \\ &\quad + p\mathrm{P}_0[K_F = p] \\ &\equiv (p-1)(\cos^2\omega)(1 - b_0 - b_p) + pb_p, \end{aligned} \tag{80}$$

where (Fernandez *et al* (1999, p.590))

$$b_0 \equiv \mathrm{P}[K_F = 0] = \frac{1}{2}\mathrm{Beta}_{\frac{p-1}{2},\frac{1}{2}}(\cos^2\omega), \tag{81}$$

$$b_p \equiv \mathrm{P}[K_F = p] = \frac{1}{2}\mathrm{Beta}_{\frac{p-1}{2},\frac{1}{2}}(\sin^2\omega), \tag{82}$$

and $\mathrm{Beta}_{\frac{p-1}{2},\frac{1}{2}}(\cdot)$ is the cdf of the beta distribution. Together, (79) and (80) confirm that this MSE is $\mathrm{O}(p)$, a linear growth rate as in the orthant case.

6 Bias adjustments

The discussions in Sections 2, 3, and 4 show that in those cases where the MSE of an ORMLBE approaches infinity as the dimension increases, it is its squared bias rather than its variance that is unbounded. Thus, in order to reduce the MSE of the ORMLBE, bias-reducing adjustments might be sought.

For the orthant cone model (Section 2), a bias-reducing adjustment for the ORMLBE $\hat{\mu}e^t$ is suggested by (5), leading to the bias-adjusted ORMLE ($\equiv$ BAORMLE) $\check{\mu}$ with components given by

$$\check{\mu}_i = \left[X_i^+ - \left(\varphi(X_i^+) - X_i^+[\bar{\Phi}(X_i^+)]\right)\right]^+ \leq \hat{\mu}_i, \qquad i = 1, \ldots, p. \tag{83}$$

[16] The maximum edge angle for the tree-order cone $C_{\rm to}^{s+1}$ is $\cos^{-1}(-1/s) > \pi/2$.

This in turn yields the BAOHMLBE for μe^t given by

$$\check{\mu}e^t \equiv \frac{1}{\sqrt{p}} \sum_{i=1}^{p} \check{\mu}_i. \tag{84}$$

When $\mu = 0$,

$$\begin{aligned} \mathrm{E}_0(\check{\mu}e^t) &= \sqrt{p}\mathrm{E}_0(\check{\mu}_1) &&\approx .338\sqrt{p}, &(85)\\ \mathrm{Var}_0(\check{\mu}e^t) &= \frac{1}{p}\sum_{i=1}^{p} \mathrm{Var}_0(\check{\mu}_i) = \mathrm{Var}_0(\check{\mu}_1) &&\approx .332, &(86)\end{aligned}$$

where the first equality in (86) again follows by independence and where the two approximations were obtained by simulation. Thus the MSE of the BAORMLBE $\check{\mu}e^t$ at $\mu = 0$ is approximated by

$$\mathrm{MSE}_0(\check{\mu}e^t) \approx .114p + .332, \tag{87}$$

which is only about 71% of the MSE of the unadjusted ORMLBE $\hat{\mu}e^t$ (cf. (8)). Nonetheless, this MSE remains $\mathrm{O}(p)$, unlike that of the URMLBE Xe^t, which is $\mathrm{O}(1)$ (cf. (7)).

For the case of the tree-order cone (Section 3), bias-reducing adjustments for the ORMLBEs $\hat{\mu}_0$ and $\hat{\mu}f^t$ are obtained as follows. From (18) – (20), the bias of $\hat{\mu}_0$ is given by

$$\hat{b}_0(\mu) = \mathrm{E}_\mu \left[\min_{\sigma \subseteq \{1,2,\ldots,s\}} \left\{ \frac{N(0,1) + \sum \left(N(\Delta_i, 1) \bigm| i \in \sigma \right)}{1 + |\sigma|} \right\} \right] < 0, \tag{88}$$

where $\Delta_i = \mu_i - \mu_0$. This suggests a BAORMLE $\check{\mu}$ with components

$$\begin{aligned} \check{\mu}_0 &= \hat{\mu}_0 - \widehat{\hat{b}_0(\mu)}, &(89)\\ \check{\mu}_i &= \max(\check{\mu}_0, \hat{\mu}_i), \qquad i = 1, \ldots, s, &(90)\end{aligned}$$

where $\widehat{\hat{b}_0(\mu)}$ is an estimate of $\hat{b}_0(\mu)$ obtained as follows. First substitute estimates $\hat{\Delta}_i \equiv \hat{\Delta}_i(X)$ for Δ_i in (88), then approximate the expectation by repeatedly sampling from the $1 + s$ independent normal distributions represented there. The minimum of the indicated averages is computed[17] for each sample, then the expectation is approximated by the mean of these sample minima.

We consider two estimates of Δ_i, $i = 1, \ldots, s$:

$$\begin{aligned} \hat{\Delta}_i^{(1)} &= (X_i - X_0)^+, &(91)\\ \hat{\Delta}_i^{(2)} &= \hat{\mu}_i - \hat{\mu}_0. &(92)\end{aligned}$$

[17] For computational efficiency, the alternative formula (16) is useful.

The corresponding BAORMLEs are denoted by $\breve{\mu}^{(1)}$ and $\breve{\mu}^{(2)}$. Thus for $j = 1, 2$,

$$\breve{\mu}_0^{(j)} = \hat{\mu}_0 - \widehat{\text{mean}} \left[\min_{\sigma \subseteq 1,2,\ldots,k} \left\{ \frac{N(0,1) + \sum \left(N\left(\hat{\Delta}_i^{(j)}, 1\right) \mid i \in \sigma \right)}{1 + |\sigma|} \right\} \right], \quad (93)$$

$$\breve{\mu}_i^{(1)} = \max\left(\breve{\mu}_0^{(1)}, X_i\right), \qquad i = 1, \ldots, s \quad (94)$$

$$\breve{\mu}_i^{(2)} = \max\left(\breve{\mu}_0^{(2)}, \hat{\mu}_i\right), \qquad i = 1, \ldots, s. \quad (95)$$

Then μ_0 and μf^t are estimated by $\breve{\mu}_0^{(j)}$ and $\breve{\mu}^{(j)} f^t$, respectively, for $j = 1, 2$. Note that to the accuracy of the sampling approximation used in the computation of $\widehat{b_0(\mu)}$, $\hat{\mu}_0 \le \breve{\mu}_0^{(j)}$ for $j = 1, 2$.

Cohen and Sackrowitz [CS] (2002) proposed an alternative estimator $\tilde{\mu}$, which we motivate as follows. First consider the problem of estimating μ_0 based on the two observations X_0, X_i only. The bias[18] and MSE of the estimator $M_i \equiv \min(X_0, X_i)$ $(i = 1, \ldots, s)$ are, respectively, (Van Eeden and Zidek (2002, Theorem 4.1); Chaudhuri and Drton (2003))

$$\begin{aligned} E_\mu(M_i) - \mu_0 &= -\sqrt{2}\left[\varphi(\gamma_i) - \gamma_i \bar{\Phi}(\gamma_i)\right] \equiv \tilde{b}(\gamma_i) < 0, && (96) \\ E_\mu(M_i - \mu_0)^2 &= 1 - 2\gamma_i\,\varphi(\gamma_i) + 2\gamma_i^2\left[\bar{\Phi}(\gamma_i)\right] < 1, && (97) \end{aligned}$$

both following from the inequality $\bar{\Phi}(\gamma) < \varphi(\gamma)/\gamma$, where $\gamma_i = \frac{\Delta_i}{\sqrt{2}}$. Thus, each M_i dominates[19] the URMLBE X_0 as an estimator of μ_0. This motivates the CS estimator (CSE) $\tilde{\mu} \equiv (\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_s)$:

$$\tilde{\mu}_0 = \frac{1}{1+s}\left[X_0 + \sum_{i=1}^{s} M_i\right], \quad (98)$$

$$\tilde{\mu}_i = \tilde{\mu}_0 + (X_i - X_0)^+, \qquad i = 1, \ldots, s. \quad (99)$$

From (96), the magnitude of the bias of $\tilde{\mu}_0$ is again greatest when $\mu \in L^{s+1}$. CS (2002) apply (97) and the Cauchy-Schwartz inequality to conclude that for any μ,

$$\text{MSE}_\mu(\tilde{\mu}_0) < 1 \quad \forall s, \quad (100)$$

hence this MSE remains bounded as $s \to \infty$, unlike that of the ORMLBE[20] $\hat{\mu}_0$. However, because $X_0, M_1, \ldots, M_s$ are positively correlated and $M_1, \ldots, M_s$ are identically distributed when $\mu_0 \le \mu_1 = \cdots = \mu_s$, in this case (which includes the least favorable case $\mu \in L^{s+1}$) the MSE of $\tilde{\mu}_0$ decreases very slowly with s

[18] The bias (96) also follows from (5) via the identity $X_i = M_i + (X_i - X_0)^+$.

[19] CS (2002) assert that this result follows from a theorem of Lee (1981, p.687), but Lee actually considers an ORMLBE rather that the sample minimum. The result does follow from Theorem 5.1 of Cohen and Sackrowitz (1970). The proof of dominance of the minimum given here is due to Chaudhuri and Drton (2003).

[20] Of course, the MSE of the URMLBE X_0 also remains bounded as $s \to \infty$.

(see Tables 2, 4). Thus the accumulation of additional observations X_i does not markedly enhance the performance[21] of $\tilde{\mu}_0$ as $s \to \infty$.

The CSE $\tilde{\mu}$ gives the following estimator of the central contrast μf^t:

$$\tilde{\mu} f^t = \frac{1}{\sqrt{s(s+1)}} \sum_{i=1}^{s} (X_i - X_0)^+. \tag{101}$$

(Compare to (31)). Its performance is qualitatively similar to that of $\tilde{\mu}_0$ (see Tables 5 – 8).

The bias-adjusted CS estimator (BACSE) $\breve{\mu}$ is suggested by (96):

$$\breve{\mu}_0 = \frac{1}{1+s} \left\{ X_0 + \sum_{i=1}^{s} \left[M_i - \tilde{b}_i(\hat{\gamma}_i) \right] \right\}, \tag{102}$$

$$\breve{\mu}_i = \breve{\mu}_0 + (X_i - X_0)^+, \qquad i = 1, \ldots, s, \tag{103}$$

where $\hat{\gamma}_i = \frac{\hat{\Delta}_i^{(1)}}{\sqrt{2}}$. (Note that $\tilde{\mu}_0 \le \breve{\mu}_0$.)

The CSE $\tilde{\mu}_0$ in (98) was motivated by the fact that each $M_i \equiv \min(X_0, X_i)$ has lower MSE for μ_0 than X_0 itself. An alternative pairwise estimator with this property is the ORMLBE $\hat{\mu}_0(X_0, X_i)$ based on (X_0, X_i), $i = 1, \ldots, s$, specifically,

$$\hat{\mu}_0(X_0, X_i) = X_0 I_{\{X_0 \le X_i\}} + \left(\frac{X_0 + X_i}{2} \right) I_{\{X_0 > X_i\}}, \tag{104}$$

so $M_i \le \hat{\mu}_0(X_0, X_i) \le X_0$. The bias and MSE of $\hat{\mu}_0(X_0, X_i)$ are, respectively (Van Eeden and Zidek (2002, Theorem 4.1); Chaudhuri and Drton (2003)):

$$\mathrm{E}_\mu(\hat{\mu}_0(X_0, X_i)) - \mu_0 = \frac{1}{2}\tilde{b}_0(\gamma_i) < 0, \tag{105}$$

$$\mathrm{E}_\mu(\hat{\mu}_0(X_0, X_i) - \mu_0)^2 = 1 - \frac{\gamma_i}{2}\varphi(\gamma_i) + \left(\frac{\gamma_i^2 - 1}{2} \right) \bar{\Phi}(\gamma_i) < 1, \tag{106}$$

again using the inequality $\bar{\Phi}(\gamma) < \varphi(\gamma)/\gamma$, so each $\hat{\mu}_0(X_0, X_i)$ also dominates the URMLBE X_0 as an estimator of μ_0. In fact, the MSE of $\hat{\mu}_0(X_0, X_i)$ is substantially smaller than that of M_i when $\Delta_i = 0$, namely .75 vs. 1. The MSE of $\hat{\mu}_0(X_0, X_i)$ increases from .75 until it becomes essentially 1 for $\Delta_i \ge 3$. The MSE of M_i descends from 1 to .80 at $\Delta_i \approx .6$, where it achieves a slight advantage over $\hat{\mu}_0(X_0, X_i)$ and maintains this slight advantage as it too now increases and becomes essentially 1 for $\Delta_i \ge 3$. Thus, the pairwise ORMLBEs are preferable to the pairwise minima in the least favorable case that $\mu \in L^{s+1}$.

[21] This is again reminiscent of the Neyman-Scott nuisance-parameter example mentioned in Section 1 but for the case where only a single observation is taken from each normal population beyond the first one. In the tree-order model, $\mu_1, \ldots, \mu_s$ can be regarded as nuisance parameters and μ_0 the single target parameter. From this viewpoint, additional observations X_i do not directly provide information about μ_0, so no estimator of μ_0 should be expected to be consistent. Also see Chaudhuri and Perlman (2003a).

This suggests the following modified CSE (MCSE) $\dot\mu \equiv (\dot\mu_0, \dot\mu_1, \ldots, \dot\mu_s)$:

$$\dot\mu_0 = \frac{1}{1+s}\left[X_0 + \sum_{i=1}^{s} \hat\mu_0(X_0, X_i)\right], \tag{107}$$

$$\dot\mu_i = \dot\mu_0 + (X_i - X_0)^+, \qquad i = 1, \ldots, s. \tag{108}$$

As in the case of the CSE, the MCSE $\dot\mu_0$ satisfies

$$\mathrm{MSE}_\mu(\dot\mu_0) < 1 \quad \forall s, \tag{109}$$

and again the MSE of $\dot\mu_0$ decreases very slowly with s when $\mu_0 \le \mu_1 = \cdots = \mu_s$, (Tables 2, 4), so again the accumulation of additional observations X_i does not markedly enhance the performance of $\dot\mu_0$ (or of $\dot\mu f^t$) as $s \to \infty$.

We note that $\hat\mu_0 \le \tilde\mu_0 \le \dot\mu_0 \le X_0$. Also,

$$\tilde\mu f^t = \breve\mu f^t = \dot\mu f^t = \ddot\mu f^t, \tag{110}$$

i.e., each CS-type estimator $\tilde\mu$, $\breve\mu$, $\dot\mu$, $\ddot\mu$ yields the same estimator (101) of the central contrast.

From (105) we obtain the bias-adjusted MCSE (BAMCSE) $\ddot\mu$:

$$\ddot\mu_0 = \frac{1}{1+s}\left\{X_0 + \sum_{i=1}^{s}\left[\hat\mu_0(X_0, X_i) - \frac{1}{2}\tilde b_i(\hat\gamma_i)\right]\right\}, \tag{111}$$

$$\ddot\mu_i = \ddot\mu_0 + (X_i - X_0)^+, \qquad i = 1, \ldots, s. \tag{112}$$

(Note that $\dot\mu_0 \le \ddot\mu_0$.)

Besides the ORMLE $\hat\mu$, six alternative estimators of μ for the tree-order model have been discussed in this section: the BAORMLEs $\check\mu^{(1)}$ and $\check\mu^{(2)}$, the CSE $\tilde\mu$, the BACSE $\breve\mu$, the MCSE $\dot\mu$, and the BAMCSE $\ddot\mu$. The biases and MSEs of the corresponding estimators for μ_0 and μf^t are given in Tables 1 – 4 and 5 – 8, respectively. For $s = 1, 2, 3, 5, 10, 20, 50, 100$, two configurations of $\mu \equiv (\mu_0, \mu_1, \ldots, \mu_s)$ are considered: $(0, 0, \ldots, 0)$ (the least favorable configuration for $\hat\mu_0$ and $\tilde\mu_0$) and $(0, 1, \ldots, 1)$. The tabulated biases and MSEs are the results of 100,000 simulated repetitions each, except for $s = 50, 100$ where 10,000 repetitions were taken. For the BAORMLEs $\check\mu^{(1)}$ and $\check\mu^{(2)}$, the sampled estimate $\widehat{b_0(\mu)}$ is based on 1000 samples within each repetition for $s = 1, 2, 3, 5, 10, 20$ and on 100 samples within each repetition for $s = 50, 100$.

The tables indicate that, with the possible exception of the second BAORMLE $\check\mu^{(2)}$ in Table 2 and 6, none of the six alternative estimators share the unbounded growth as $s \to \infty$ of the MSEs of the ORMLBEs $\hat\mu_0$ and $\hat\mu f^t$. Instead, these MSEs appear to stabilize or decrease with s. For the least favorable case $\mu = (0, 0, \ldots, 0)$ in Tables 2 and 6 (actually, for any $\mu \in L^{s+1}$), the first BAORMLBE $\check\mu^{(1)}$ provides the smallest MSEs, with its relative improvement increasing with s through $s = 100$. When μ lies in the interior of C^{s+1}_{to} (Tables 4 and 8), the ORMLBEs themselves are preferable for small values of s, while

s	$b(\hat{\mu}_0)$	$b(\breve{\mu}_0^{(1)})$	$b(\breve{\mu}_0^{(2)})$	$b(\tilde{\mu}_0)$	$b(\breve{\mu}_0)$	$b(\dot{\mu}_0)$	$b(\ddot{\mu}_0)$
1	- 0.283	- 0.084	- 0.084	- 0.283	- 0.084	- 0.149	- 0.043
2	- 0.443	- 0.117	- 0.136	- 0.374	- 0.109	- 0.187	- 0.054
3	- 0.559	- 0.138	- 0.178	- 0.423	- 0.123	- 0.210	- 0.060
5	- 0.721	- 0.165	- 0.243	- 0.472	- 0.140	- 0.238	- 0.072
10	- 0.948	- 0.185	- 0.334	- 0.509	- 0.146	- 0.253	- 0.071
20	- 1.190	- 0.204	- 0.443	- 0.538	- 0.158	- 0.269	- 0.079
50	- 1.503	- 0.213	- 0.593	- 0.550	- 0.159	- 0.274	- 0.079
100	- 1.736	- 0.217	- 0.713	- 0.555	- 0.159	- 0.275	- 0.077

Table 1: Bias of the estimators of μ_0 when $\mu = 0$.

s	$\mathrm{M}(\hat{\mu}_0)$	$\mathrm{M}(\breve{\mu}_0^{(1)})$	$\mathrm{M}(\breve{\mu}_0^{(2)})$	$\mathrm{M}(\tilde{\mu}_0)$	$\mathrm{M}(\breve{\mu}_0)$	$\mathrm{M}(\dot{\mu}_0)$	$\mathrm{M}(\ddot{\mu}_0)$
1	0.755	0.771	0.771	0.755	0.770	0.818	0.853
2	0.730	0.670	0.667	0.691	0.676	0.758	0.796
3	0.770	0.615	0.612	0.669	0.630	0.734	0.772
5	0.883	0.544	0.551	0.647	0.579	0.704	0.739
10	1.159	0.455	0.496	0.619	0.524	0.671	0.705
20	1.607	0.396	0.503	0.619	0.504	0.667	0.698
50	2.379	0.317	0.570	0.606	0.479	0.654	0.683
100	3.105	0.281	0.675	0.603	0.473	0.646	0.676

Table 2: MSE of the estimators of μ_0 when $\mu = 0$.

both BAORMLBEs offer improvements for larger values of s, with the second BAORMLBE somewhat more desirable. The CSE $\tilde{\mu}$, while never optimal, also improves upon the ORMLBEs in many cases.

From the discussion preceding (107), it might be expected that the MSEs of the MCSE $\dot{\mu}_0$ and BAMCSE $\ddot{\mu}_0$ are less than that of the CSE $\tilde{\mu}_0$ when $\mu \in L^{s+1}$, but in fact the opposite is true, as seen from Table 2. This can be explained by the fact although the MSE (=.75) of each pairwise MLE $\hat{\mu}_0(X_0, X_i)$ is less than that of the MSE (=1) of each pairwise minimum M_i, the correlations among the pairwise MLEs are greater than those among the pairwise minima, hence the latter are more nearly antithetic than the former. Furthermore, although the BACSE $\breve{\mu}_0$ has smaller MSE than the CSE for the least favorable case $\mu \in L^{s+1}$ (Table 2), the opposite is true when μ lies in the interior of C_{to}^{s+1} (Table 4).

Thus it is seen that for the tree-order model of several treatments and one control with equal variances and sample sizes, the two BAORMLBEs are desirable alternatives[22] to the ORMLBE for estimating the control parameter μ_0

[22] Admittedly at the cost of additional computation, comparable to bootstrapping.

s	$b(\hat{\mu}_0)$	$b(\check{\mu}_0^{(1)})$	$b(\check{\mu}_0^{(2)})$	$b(\tilde{\mu}_0)$	$b(\breve{\mu}_0)$	$b(\dot{\mu}_0)$	$b(\ddot{\mu}_0)$
1	- 0.102	0.029	0.029	- 0.102	0.029	- 0.052	- 0.013
2	- 0.168	0.052	0.044	- 0.132	0.042	- 0.066	0.021
3	- 0.221	0.071	0.052	- 0.146	0.050	- 0.071	0.027
5	- 0.306	0.090	0.050	- 0.167	0.051	- 0.084	0.025
10	- 0.439	0.128	0.044	- 0.179	0.059	- 0.087	0.031
20	- 0.595	0.167	0.019	- 0.188	0.062	- 0.092	0.032
50	- 0.821	0.215	- 0.039	- 0.196	0.600	- 0.099	0.029
100	- 1.000	0.251	- 0.096	- 0.199	0.059	- 0.100	0.030

Table 3: Bias of the estimators of μ_0 when $\mu_0 = 0$ and $\mu_i = 1$ for $i \geq 1$.

s	$\mathrm{M}(\hat{\mu}_0)$	$\mathrm{M}(\check{\mu}_0^{(1)})$	$\mathrm{M}(\check{\mu}_0^{(2)})$	$\mathrm{M}(\tilde{\mu}_0)$	$\mathrm{M}(\breve{\mu}_0)$	$\mathrm{M}(\dot{\mu}_0)$	$\mathrm{M}(\ddot{\mu}_0)$
1	0.825	0.938	0.938	0.825	0.937	0.891	0.956
2	0.741	0.890	0.886	0.763	0.907	0.854	0.937
3	0.694	0.873	0.846	0.733	0.894	0.841	0.934
5	0.654	0.836	0.786	0.709	0.882	0.828	0.930
10	0.638	0.779	0.685	0.682	0.870	0.817	0.929
20	0.691	0.707	0.565	0.659	0.851	0.797	0.912
50	0.899	0.618	0.426	0.637	0.829	0.780	0.896
100	1.166	0.569	0.345	0.640	0.835	0.786	0.905

Table 4: MSE of the estimators of μ_0 when $\mu_0 = 0$ and $\mu_i = 1$ for all $i \geq 1$.

and the central contrast μf^t, especially in the least favorable case and when s is large. The CSE also improves upon the ORMLBEs in the least favorable case, but to a somewhat lesser extent. We conclude that the method of maximum likelihood does not "fail disastrously" for order-restricted estimation, but may require adjustment depending on the specific parameter to be estimated.

Remark. The introduction of the CSE $\tilde{\mu}$ as an alternative to the ORMLE $\hat{\mu}$ was motivated by the fact that in the least favorable case $\mu \in L^{s+1}$, the bias of the (unadjusted) ORMLBE $\hat{\mu}_0$ is unbounded as $s \to \infty$, whereas that of the CSE $\tilde{\mu}_0$ remains bounded. What can be said about the biases of the other components $\hat{\mu}_i$ and $\tilde{\mu}_i$ for $i \geq 1$? The following results can be found in Chaudhuri and Perlman (2003b).

Consider the vector of means $\mu \equiv (\mu_0, \mu_1, \ldots, \mu_s)$ in the configuration

$$\mu_j(\gamma) \equiv (0, \overbrace{0, \ldots, 0}^{j}, \overbrace{\gamma, \ldots, \gamma}^{s-j}), \qquad j = 0, \ldots, s, \tag{113}$$

for $\gamma \geq 0$. Then if $s, \gamma \to \infty$, $\frac{j}{s+1} \to \beta \in [0,1]$, and $\frac{i}{s+1} \to \alpha \in [0,1]$, the bias of

s	$b(\hat{\mu} f^t)$	$b(\check{\mu}^{(1)} f^t)$	$b(\check{\mu}^{(2)} f^t)$	$b(\tilde{\mu} f^t)$
1	0.396	0.361	0.361	0.396
2	0.546	0.438	0.444	0.462
3	0.644	0.470	0.485	0.487
5	0.788	0.507	0.535	0.516
10	0.997	0.536	0.612	0.536
20	1.219	0.551	0.684	0.550
50	1.519	0.555	0.786	0.558
100	1.744	0.554	0.871	0.557

Table 5: Bias of the estimators of μf^t when $\mu = 0$ (recall (110)).

s	$\mathrm{M}(\hat{\mu} f^t)$	$\mathrm{M}(\check{\mu}^{(1)} f^t)$	$\mathrm{M}(\check{\mu}^{(2)} f^t)$	$\mathrm{M}(\tilde{\mu} f^t)$
1	0.494	0.464	0.464	0.494
2	0.599	0.492	0.495	0.462
3	0.685	0.494	0.506	0.552
5	0.858	0.503	0.535	0.575
10	1.182	0.489	0.575	0.585
20	1.636	0.475	0.638	0.600
50	2.411	0.435	0.748	0.599
100	3.124	0.418	0.873	0.599

Table 6: MSE of the estimators of μf^t when $\mu = 0$ (recall (110)).

$\tilde{\mu}_i$ (when $\mu = \mu_j(\gamma)$) approaches

$$\begin{cases} \sqrt{2}(1-\beta)\phi(0) = 0.564(1-\beta) & \text{if } \alpha < \beta, \\ -\sqrt{2}\beta\phi(0) \quad\;\; = -0.564\beta & \text{if } \alpha > \beta. \end{cases} \tag{114}$$

By contrast, in every such case the bias of the ORMLBE $\hat{\mu}_i$ approaches 0. Thus, for example, if $\beta = \frac{1}{2}$ then *every* component of the CSE except $\tilde{\mu}_0$ has approximate bias ± 0.282 for large s and γ, whereas *every* component of the ORMLE except $\hat{\mu}_0$ has approximate bias 0.

Acknowledgement: We thank Mathias Drton, J. T. Gene Hwang, David Ragozin, Jon Wellner, Michael Woodroofe, and an anonymous referee for helpful comments and suggestions.

s	$b(\hat{\mu} f^t)$	$b(\check{\mu}^{(1)} f^t)$	$b(\check{\mu}^{(2)} f^t)$	$b(\tilde{\mu} f^t)$
1	0.140	0.101	0.101	0.140
2	0.205	0.101	0.104	0.160
3	0.256	0.097	0.105	0.171
5	0.336	0.086	0.106	0.183
10	0.459	0.055	0.105	0.187
20	0.609	0.022	0.117	0.192
50	0.828	- 0.023	0.153	0.197
100	1.005	- 0.057	0.194	0.200

Table 7: Bias of the estimators of μf^t when $\mu_0 = 0$ and $\mu_i = 1$ for $i \geq 1$.

s	$\mathrm{M}(\hat{\mu} f^t)$	$\mathrm{M}(\check{\mu}^{(1)} f^t)$	$\mathrm{M}(\check{\mu}^{(2)} f^t)$	$\mathrm{M}(\tilde{\mu} f^t)$
1	0.659	0.673	0.673	0.659
2	0.615	0.640	0.637	0.650
3	0.596	0.618	0.612	0.649
5	0.583	0.584	0.572	0.649
10	0.600	0.535	0.511	0.650
20	0.675	0.471	0.437	0.642
50	0.896	0.392	0.352	0.630
100	1.168	0.345	0.303	0.637

Table 8: MSE of the estimators of μf^t when $\mu_0 = 0$ and $\mu_i = 1$ for $i \geq 1$.

References

Abelson, R. P. and Tukey, J. W. (1963). Efficient utilization of non-numerical inofrmation in quantitative analysis: general theory and the case of simple order. *Ann. Math. Statist.* **34**, 1347-1369.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference under OrderRestrictions.* New York: Wiley.

Bahadur, R. R. (1971). *Some Limit Theorems in Statistics.* Regional Conference Series in Applied Mathematics, Vol. 4. Philadelphia: SIAM.

Chaudhuri, S. and Drton, M. (2003). On the bias and mean-square error of the sample minimum and the maximum likelihood estimator for two ordered normal means. Technical Report No. 432, Department of Statistics, University of Washington, Seattle.
`http://www.stat.washington.edu/www/research/reports/2003/tr432`

Chaudhuri, S. and Perlman, M. D. (2003a). Estimation of the minimum mean of normal populations under the tree-order restriction. In preparation.

Chaudhuri, S. and Perlman, M. D. (2003b). On the bias of the Cohen-Sackrowitz estimator for the tree-order model. Technical Report No. 437 Department of Statistics, University of Washington, Seattle.
`http://www.stat.washington.edu/www/research/reports/2003/tr437`

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573-578.

Cohen, A. and Sackrowitz, H. B. (1970). Estimating the last mean of a monotone sequence. *Ann. Math. Statist.* **41** 2021-2034.

Cohen, A. and Sackrowitz, H. B. (2002). Inference for the model of several treatments and a control. *J. Statist. Planning Inference* **107** 89-101.

Fernandez, M. A., Rueda, C. and Salvador, B.. (1999). The loss of efficiency estimating linear functions under restrictions. *Scand. J. Statist.* **26**, 579-592.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York.

Hwang, J. T. G. and Peddada, S. D. (1994). Confidence interval estimation subject to order restrictions. *Ann. Statist.* **22**, 67-93.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions – I.* Wiley, New York.

Lee, C.-I. C. (1981). The quadratic loss of isotonic regression under normality. *Ann. Statist.* **9**, 686-688.

Lee, C.-I. C. (1988). Quadratic loss of order restricted estimators fortreatment means with a control. *Ann. Statist.* **16**, 751-758.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1-32.

Pickands, J. (1968). Moment convergence of sample extremes. *Ann. Math. Statist.* **39**, 881-889.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order-Restricted Statistical Inference.* New York: Wiley.

Van Eeden, C and Zidek, J. V. (2002) Combining sample information in estimating ordered normal means. *Sankhya, Series A* **64**, 588-610.