Sankhyā : The Indian Journal of Statistics
Special issue in memory of D. Basu
2002, Volume 64, Series A, Pt. 3, pp 588-610

COMBINING SAMPLE INFORMATION IN ESTIMATING ORDERED NORMAL MEANS

By CONSTANCE VAN EEDEN
and
JAMES V. ZIDEK
University of British Columbia, Canada

SUMMARY. In this paper we answer a question concerned with the estimation of θ_1 when $Y_i \sim^{ind} \mathcal{N}(\theta_i, \sigma_i^2), i = 1, 2$, are observed and $\theta_1 \leq \theta_2$. In this case θ_2 contains information about θ_1 and we show how the relevance weights in the so-called weighted likelihood might be selected so that Y_2 may be used together with Y_1 for effective likelihood-based inference about θ_1. Our answer to this question uses the Akaike entropy maximization criterion to find the weights empirically. Although the problem of estimating θ_1 under these conditions has a long history, our estimator appears to be new. Unlike the MLE it is continuously differentiable. Unlike the Pitman estimator for this problem, but like the MLE, it has a simple form. The paper describes the derivation of our estimator, presents some of its properties and compares it with some obvious competitors. One of these competitors is the inadmissible maximum likelihood estimator for which we present a dominator. Finally, a number of open problems are presented.

> But, what is information? No other concept in statistics is more elusive in its meaning and less amenable to a generally agreed definition.
>
> — D. Basu (1975).

1. Introduction

Through this article, we pay tribute to the immensely important contributions made by Professor D. Basu to the understanding of the foundations of statistical inference. Referring in particular to his well-known foundational paper on information and likelihood (Basu 1975), we show how the likelihood can be extended to use "the whole of the relevant information contained in the data", a desideratum Basu attributes to Fisher. In particular, this paper shows how through that extension bias can be traded

Paper received November 2001.
AMS (1991) *subject classification.* 62F30, 62F10, 62C15, 62C20.
Keywords and phrases. Likelihood, maximum likelihood, weighted likelihood, estimation, admissibility, minimaxity, normal means, restricted parameter spaces, relevance weighting.

to gain extra information and a resulting increase in precision in statistical estimation.

The problem arises when an investigator has data from a population other than that of his or her inferential interest. Do these auxiliary data contain information of value for estimating parameters in the population of interest? If so, how can the bias in the auxiliary sample be traded off for precision in the required parameter estimators.

The specific problem we consider is that of estimating the mean θ_1 of a univariate normal population from which an observation y_1 has been drawn. We suppose an independent observation y_2 has also been drawn from another normal population with mean θ_2 when $\theta_1 \leq \theta_2$. Now the general questions we ask above can be stated more specifically by asking how y_2 can be used in conjunction with y_1 to create an estimator that improves on the estimator y_1 based only on data from the first population.

Heuristics suggest an affirmative answer. The event $y_2 < y_1$ combined with the knowledge that $\theta_1 \leq \theta_2$ suggests $\theta_1 \approx \theta_2$. That suggests a better estimator of θ_1 would be obtained by taking the BLUE that would be used if the population means were equal.

We describe a new method for operationalizing these heuristics in Section 2. That method is an extension of the maximum likelihood method and, like its predecessor, is a very generally applicable tool for the practitioner's toolbox. A primary purpose of this article is the comparison of the estimator produced by this method against its purpose-built competitors designed specifically for the problem considered in this paper. That in turn will help to determine the degree of confidence we can place in this new tool for general use.

A number of authors have found such competitors. They differ from the one we obtain with our new method for exploiting y_2 in the estimation of θ_1. Unlike the classical unbiased MLE *viz* y_1 (hereafter denoted by ULE), the alternative estimators obtained by those authors are biased like ours. However, these estimators can have substantially smaller mean-squared-errors (MSE's) than their classical counterpart over portions of the parameter space deemed to be of particular importance. At the same time, their MSE's are either smaller or not appreciably larger over the rest of the parameter space than the MSE of the ULE. Thus an effective bias-variance trade-off is indeed possible; information in the sample from the second population can help in estimating the mean of the first.

These "old" estimators, described in Section 3 to make that trade-off, depend very fundamentally on the assumption that $\theta_1 \leq \theta_2$. However, in practice θ_2 may well be slightly less than θ_1. These practical considerations point to the need for good performance in this circumstance. Therefore the "robustness" of these old estimators is of concern and will be investigated.

At the same time we describe in Section 2 a new approach to estimation that incorporates the assumption $\theta_1 \leq \theta_2$ at a "secondary" level of modelling. Its robustness against violation of that assumption is therefore expected.

That new approach uses an extension of Fisher's classical likelihood that Hu (1994) introduces and calls the "Relevance Weighted Likelihood" (REWL). It generalizes the local likelihood defined in the context of non-parametric regression by Tibshirani and Hastie (1987) that was extended as a local likelihood by Staniswalis (1989) and as a quasi-local-likelihood by Fan et al. (1995).

In contrast to the local likelihood, the REWL and its extension, the "weighted likelihood (WL)" described in Section 2, can be a global likelihood and in one of the applications developed by Hu and Zidek (2002, hereafter HZ), it is shown how the celebrated James-Stein estimator can be found as a maximum (weighted) likelihood estimator when the weights are estimated from the data.

The weights allow bias to be traded for precision in the likelihood setting, as bias is traded for variance in the nonparametric regression setting. The need for such a theory has become increasingly important as the scale of modern experimental science has grown in its space-time scales thanks to demand (*eg.* environmental science) combined with feasibility (*eg.* through information technology). On these scales, the replicated experiment seems completely unrealistic as an experimental paradigm, leading to the need for a theory that embraces bias without sacrificing the goals of efficiency and precision enshrined in Fisher's foundational works.

The theory described in Section 2 enables the bias-precision trade-off to be made without relying on the Bayesian approach (see Berger 1985). The latter permits the bias-variance trade-off to be made in a conceptually straightforward manner. Reliance on empirical Bayes methods softens the demands for realistic prior modelling in complex problems. Efron (1996) illustrates the empirical Bayes approach in such problems and uses the term "relevance" in a manner similar to that of Hu (1994).

Our theory is proposed as a simpler alternative to the empirical Bayesian approach for use in complex problems. [Fuller introductions are given by Hu and Zidek (2001) and HZ.] However, the weighted likelihood and inferential procedures deriving from it are not seen as competitors to Bayesian methods. Indeed, the latter would generally be seen as preferable when they can be used. However, the weighted likelihood can be considered a practical alternative in situations when they cannot. In those situations, a large number of uncertain quantities may obtain and even non-Bayesian methods like those obtained using an "empirical Bayesian" or "improper Bayesian" approach may be difficult to use. In this article, we demonstrate that the WL may offer a compromise that allows some prior knowledge to be used in constructing the weights, while eliminating the need to attempt what could be an unrealistic elicitation of prior knowledge about all of the model parameters.

We thereby gain a theory that formally links a diverse collection of statistical domains such as weighted least squares, nonparametric regression, meta-analysis and shrinkage estimation. Starting with the likelihood in these domains yields new methods and suggests new problems as we will attempt to show. At the same time, the WL comes with an (as yet incomplete) underlying general theory including extensions of Wald's theory for the maximum likelihood estimator (Hu 1997).

In Section 3 we study the bias-variance trade-off made by a number of biased estimators proposed as solutions to the problem central to this paper. Included is the estimator we propose in Section 2. Numerical assessments of their properties point to a number of conjectures and questions in that section and deeper analysis in Section 4.

In Section 4 we answer a number of questions raised in Section 3. A dominator for the inadmissible maximum likelihood estimator is also presented there. However, many of the conjectures remain unproven and questions unanswered.

In the concluding Section 5 we summarize the results of our inquiry and the possible value of the WLE-based methodology. In particular we look at the robustness of the various estimators considered in this paper against violations of the basic assumption that $\theta_1 \leq \theta_2$.

2. Weighted Likelihood Estimation

In this section we describe for completeness the weighted likelihood (WL) in the general case and then apply it to the specific problems of interest in this paper. Assume $\{Y_i\}$ are independently distributed random variables or vectors, each having an associated population distribution with probability density and cumulative distribution (PDF and CDF, respectively) f_i and F_i. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be the vector or matrix of these measurable attributes.

From each population i, $n_i \geq 0$ items are randomly and independently sampled, yielding $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, Y_{ij} representing the Y_i measured on the j-th item sampled from the i-th population $j = 1, \ldots, n_i,\ i = 1, \ldots, n$

(the null vector when $n_j = 0$). Assume the $Y_{ij},\ j = 1,\ldots,n_i$ are independent as well as identically distributed, each having its associated population distribution. Denote the realization of $\mathbf{Y}_i$ by $\mathbf{y}_i,\ i = 1,\ldots,n$.

In this paper inferential interest concerns attributes of population 1. However, in general Hu and Zidek (2001) and HZ consider other possibilities such as simultaneous inference about parameters of all the populations.

Starting from the Akaike entropy maximization principle (1973, 1977, 1978, 1982, 1983, 1985), HZ derive the WL in the nonparametric and parametric cases. To be precise they suppose (when the Y are discrete) that a predictive distribution say g of Y_1 must be chosen to maximize $\int \log g(y) dF_1(y)$ where F_1 denotes the true "conceptual" population distribution for the first population. This maximization must be done subject to knowledge that F_1 resembles each of the other $F_j,\ j \neq 1$, that is subject to $\int \log g(y) dF_j(y) > c_j,\ j \neq 1$ for specified $\{c_j\ j \neq 1\}$. A Lagrangian argument then implies that g maximizes a linear combination of the $\int \log g(y) dF_j(y),\ j = 1,\ldots,n$. However, since the $\{F_j\}$ are unknown they are estimated by $\{\hat{F}_j\}$, their empirical distribution functions. When only one observation y_j is available from population $j = 1,\ldots,n$, the empirical distribution for that population becomes a point mass at that observation.

In any event, with these heuristics the optimum g maximizes the nonparametric relevance likelihood function that, viewed as a function of g, is

$$g \to \prod_{j=1}^{n} \prod_{l=1}^{n_j} g^{\lambda_{ij}/n_j}(y_{jl}). \tag{2.1}$$

Similar heuristics apply to the case of interest in this paper, *i.e.* the parametric case, where for the likelihood we have instead

$$\theta \to \prod_{j=1}^{n} \prod_{l=1}^{n_j} f_i^{\lambda_{ij}/n_j}(y_{jl} \mid \theta_i) \tag{2.2}$$

where $\theta = (\theta_1,\ldots,\theta_n)$. In both cases we take $\lambda_{ij}/n_j = 0$ when $n_j = 0$ for all i and j.

In fact, the heuristic derivation gives $\lambda_{ij} \geq 0$ as originally required in the REWL (Hu 1994). However, we will drop the latter restriction and thereby obtain the WL rather than REWL. This formalistic extension enables our theory to embrace estimators seen below that were obtained previously by other authors. Wang (2001) studies this extension and in particular shows that it can also be obtained in a very general setting from the Akaike entropy maximization criterion.

The weights $\{\lambda_{ij}\}$ enable the investigator to trade off bias for precision in estimating the likelihood for population 1 using the data from the remaining populations. Ideally the choice of these weights (equivalently the specification of the $\{c_j\}$ above) will be context dependent. However, HZ suggest a general method for their selection based on a suggestion of Stigler (1990). That method, again based on the use of the maximization of entropy approach with follow-up estimation, is the one used in this paper. Rather than describe it in general we demonstrate it below in specific problems. Wang (2001) gives an alternative approach to estimating the weights, his being based on cross-validation.

The WLE for θ_i is found by maximizing (2.2). Hu (1997) shows that the theory of Wald for the classical MLE extends to the REWL under a suitable adaptation of Wald's assumptions. Wang et al. (to appear) extend Wald's theory in a somewhat different direction for the WL. In particular, the weights are allowed to be data-dependent.

We apply the WL to the case of two normal populations $Y_i \sim N(\theta_i, \sigma_i^2)$ for which the $\{\sigma_i^2\}$ $i = 1, 2$ are known. Now $n_1 = n_2 = 1$ for the two populations involved and for simplicity we denote the weights by $\lambda_{i1} = \lambda_i,\ i = 1, 2$ for those populations. The WLE for θ_1 or WLE for short is easily shown to be

$$\delta_{WLE}(Y_1, Y_2) = Y_1 + W\alpha$$

where $W = Y_2 - Y_1$ and $\alpha \in [0, 1]$ obtains from the weights and needs to be specified. The weight ratio defines α through

$$\frac{\lambda_2}{\lambda_1} = \frac{\sigma_2^2}{\sigma_1^2} \frac{\alpha}{(1-\alpha)}. \tag{2.3}$$

The maximization of entropy criterion above may be applied to find the weights. That criterion leads to the minimization of the MSE in this case of normal population distributions. Hence the optimal choice of α if $\Delta = \theta_2 - \theta_1$ were known would be

$$\alpha_{optimal} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2 + \Delta^2}.$$

However, since Δ is unknown it must be estimated. The appropriate estimator for the case considered in Section 3 where $\Delta \geq 0$ would be

$$\hat{\alpha}(W) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2 + W_+^2} \tag{2.4}$$

where $W_+ = \max\{0, W\}$.

This approach yields a smooth estimator since α is "fitted" to the data only after the MSE has been computed. In particular it is a differentiable function of W in contrast to the truncated MLE of θ_1 which is not. Now the performance of the proposed estimator needs to be explored and we do this both theoretically and numerically in the next section.

However, HZ emphasize that the specification of the weights should best be done in the context of the specific inferential context. This suggestion may be followed in the restricted means problem above since a variety of estimators that exploit Y_2 in the estimation of θ_1 have already been proposed. Moreover, each may be written in the form above for the WLE with an estimated α. Thus each entails an implicit choice of the weight ratio that can be exploited through the equation above relating that ratio to α. In this paper we will explore these various choices and compare the associated estimators in the next section.

Before leaving this section, we note that linear estimators like those above arise quite generally in Bayesian settings. The empirical Bayesian will then be faced with the problem of estimating the coefficients as we are. Presumably, the estimation of those weights can sometimes be simplified by appeal to the form of the prior and the estimation of its hyperparameters. At the same time the need to limit the estimates to conform to the prior could be unduly limiting. For example, in spatially mapping the incidence of a disease, a problem Wang (2001) studies, the weights in the WL can be left quite arbitrary. In contrast, specifying a prior in a hierarchical Bayesian or empirical Bayesian analysis may force the adoption of a "convenience" model such as a Markov random field, that may be quite inappropriate, in view of the difficulty of specifying realistic neighbourhood structures with non-homogeneous spatial media.

Of course, if the weights are chosen appropriately in the linear case described in the previous paragraph, the WLE will agree with the empirical Bayes estimate. HZ give an example of such a serendipitous outcome. There the means of a collection of Poisson population of counts are simultaneously estimated and the WL is employed, the weight being an estimated Akaike optimum weight. The result turns out to be precisely the same estimator as that given by Berger (1985, p. 297) and obtained by an empirical Bayes argument.

Note that in some albeit artificial examples the WL is exactly a Bayesian integrated likelihood so that the MLE could be a posterior modal estimate when a uniform prior is selected. In one hypothetical example of a meta analysis, Zidek et al. (1999) used a generalized weighted likelihood ratio test to combine several normally and independently distributed test statistics to test the null hypothesis of no treatment effect. The p-value for this test was substantially smaller than that for any one generalized likelihood ratio test and was similar to a posterior probability of the null hypothesis for a Gaussian/Gaussian normal/conjugate normal model. However, in general the degree of agreement between the Bayes, empirical Bayes and weighted likelihood approaches remains to be more fully explored.

3. The Bias-Variance Trade-off

The bias-variance trade-off goes back at least as far as Stein's discovery that it could be made in the simultaneous estimation of independent normal population means. That celebrated discovery stimulated the study of biased estimation. The feasibility of the trade-off was demonstrated in a wide variety of contexts. One such context was that of the present paper wherein a number of biased estimators of ordered normal means were proposed.

We now examine that trade-off and the way it has been made by those estimators. Specifically we compare five estimators of θ_1 based on (Y_1, Y_2). They are: $\delta_{WLE}(Y_1, Y_2)$ the WLE as defined and discussed in Section 2; $\delta_{MLE}(Y_1, Y_2)$ the MLE, i.e. the first co-ordinate of the MLE for (θ_1, θ_2) under the restriction $\theta_1 \le \theta_2$; $\delta_{ULE}(Y_1, Y_2) = Y_1$ the unrestricted MLE of θ_1 based on Y_1; $\delta_{MIN}(Y_1, Y_2)$ the minimum of Y_1 and Y_2; and $\delta_P(Y_1, Y_2)$ the so-called Pitman estimator, i.e. the first co-ordinate of the generalized Bayes estimator of (θ_1, θ_2), that estimator being computed from the uniform prior on $\{(\theta_1, \theta_2) \mid \theta_1 \le \theta_2\}$. Apart from $\delta_{ULE}(Y_1, Y_2) = Y_1$ these estimators are given explicitly below:

$$\delta_{WLE}(Y_1, Y_2) = Y_1 + W\hat{\alpha}(W) \quad \text{where } \hat{\alpha}(W) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2 + W_+^2}; \tag{3.1}$$

$$\left.\begin{aligned} \delta_{MLE}(Y_1, Y_2) &= \min\left(Y_1, \frac{\sigma_2^2 Y_1 + \sigma_1^2 Y_2}{\sigma_1^2 + \sigma_2^2}\right) = Y_1 + \frac{1}{1+\tau} W_- \\ \text{with } \tau &= \sigma_2^2/\sigma_1^2 \quad \text{and} \quad W_- = \min(0, W); \end{aligned}\right\} \tag{3.2}$$

$$\delta_{MIN}(Y_1, Y_2) = \min(Y_1, Y_2) = Y_1 + W_-; \tag{3.3}$$

$$\delta_P(Y_1, Y_2) = Y_1 - \sqrt{\frac{\sigma_1^4}{\sigma_1^2 + \sigma_2^2}} \, \frac{\phi\left(\dfrac{W}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)}{\Phi\left(\dfrac{W}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)}. \tag{3.4}$$

Figure 1: Graphs of the mean squared errors: selected estimators.

The Pitman estimator was proposed and studied by Cohen and Sackrowitz (1970). Note, however, that our formula for $\delta_P(Y_1, Y_2)$ is not the same as the one given by them. They claim, erroneously, that one can suppose, without loss of generality, that one of the two variances equals 1, making their formula valid for that special case only.

REMARK 3.1 *Note the differences in the way the above estimators depend on σ_1^2 and σ_2^2. The estimators Y_1 and $\min(Y_1, Y_2)$ are independent of these variances, the weighted and the Pitman estimator depend on both of them, while the MLE depends on σ_1^2 and σ_2^2 only through their ratio.*

We begin by examining in Figure 1 the MSE's of the estimators plotted as functions of $\Delta = \theta_2 - \theta_1$.

We consider cases below where the population variances are unequal. For that reason we will in general divide all the MSE's by σ_1^2 to enable us to compare MSE plots. Therefore in all such plots the one for ULE has constant value 1 for all Δ whatever be σ_1. As the classical (uniform minimum variance unbiased) estimator of θ_1, the ULE provides a natural benchmark for assessing the performance of the alternatives considered in this paper.

Figure 2: Graphs of relevance weight ratios for population 2 versus 1 for selected estimators.

The MSE of another classical estimator, the MLE also appears in Figure 1. It appears to be uniformly smaller than that of the ULE but the two are in close agreement for large Δ. That agreement encourages optimism about the quality of the ULE since generally the MLE performs well. In fact, it would appear that ULE and MLE are minimax estimators. At the same time the MLE appears to dominate the ULE suggesting that the ULE is inadmissible and dominated by the MLE. In fact, we wonder if the MLE is admissible, a point we return to in the sequel along with others suggested by the figures.

Figure 1 shows the MSE for the WLE (as well as the MLE) to be much smaller than that of the ULE for small Δ-values. Moreover its MSE resembles the MLE's for such values.

How do the MLE and the WLE achieve their seeming superiority over the ULE? The immediate answer is that they exploit the information in Y_2 and they do so in a similar way. Figure 2 confirms this. That figure depicts for all estimators other than the MIN, the implied or explicit weight ratios as functions of $W = Y_2 - Y_1$. The ratios for the MLE and WLE are broadly similar. However, the WLE - ratio decreases to zero more slowly than that

Figure 3: Graphs of bias functions for selected estimators

of the MLE. Thus it makes more liberal use of that information than does MLE. (It does so at the cost of greater bias.)

To gain a better understanding of how that superior performance is achieved by the WLE and the MLE relative to the other two estimators we turn to Figure 3 and see the bias functions of the various estimators. Note the comparatively small absolute biases for both estimators when Δ is close to zero compared to those of PIT and MIN. So we see that both WLE and MLE gain their superiority over ULE by aggressively exploiting the relevant information in Y_2 to reduce their variances while controlling their biases for small Δ.

At the same time, Figure 1 shows that as Δ grows larger the MSE for the WLE increases and eventually exceeds that of the ULE. This observation coupled with our earlier conjectures suggests the WLE is not a minimax estimator. We wonder if it is admissible.

Unlike the WLE, which is (twice) differentiable, the MLE is not differentiable. The well-known necessary condition for admissibility (see Sacks (1963)) that estimators must be regular functions of the data, therefore encourages the belief that the MLE is not admissible. The remaining two estimators under consideration in this paper, PIT and MIN, also seem to successfully trade bias for variance. In fact Figure 1 suggests PIT and MIN dominate ULE. Moreover, this analysis suggests that both PIT and MIN are minimax when the population variances are identical. That figure also shows that neither estimator performs especially well when Δ is close to zero. (They effect the bias-variance trade-off in rather subtle ways.) We are led to wonder if MIN and PIT are admissible when the population variances are equal. Observe in Figure 1 that the ULE-MSE uniformly exceeds that of the Pitman estimator. Moreover the comparative advantage of the Pitman estimator obtains not at $\Delta = 0$ but rather for Δ around 2. To interpret this observation note that the Pitman prior does not put high weight on $\theta_1 = \theta_2$. In fact its uniform prior on the range of (θ_1, θ_2) forces PIT to optimize by requiring a negative weight ratio (see Figure 2). It "pushes away" the information in Y_2 when the WLE and MLE embrace it (when $\Delta = 0$) since under the prior this possibility would be remote. Instead PIT saves the trade-off for values of more realistic Δ's under the assumed prior. Nevertheless, like the other alternatives to the ULE considered here other than the WLE, PIT proves to be negatively biased; it tends to underestimate θ_1 (see Figure 3).

MIN succeeds in making the bias-variance tradeoff (see Figure 1) but the mechanism by which it does this proves elusive. The weight ratio for the MIN cannot even be plotted in Figure 2, being infinite when $W < 0$ since in that case the estimator puts all the weight on Y_2 and none on Y_1. On the other hand, when $W \geq 0$ that ratio becomes zero. How does MIN so successfully exploit Y_2? The answer seems to be that since $\Delta \geq 0$, $Y_2 \leq Y_1$ suggests Y_1 is an overestimate of θ_1. We can then profitably shrink it down to Y_2. To test this explanation we consider its implication when $\sigma_2 < \sigma_1$, *i.e.* when Y_2 is a measurement of higher quality than Y_1 (even if biased as an estimator of θ_1). In this case Y_2 would indicate quite reliably when Y_1 overestimates θ_1.

Figure 4 validates this heuristic reasoning. The relative gain in MIN's performance over that of ULE exceeds its gain when the population variances are unequal.

On the other hand the explanation also suggests that when Y_2 is of low quality it will not help much to show when Y_1 overestimates θ_1. Again the implication is validated, this time by Figure 5. MIN now performs poorly against the other estimators as measured by its MSE.

These numerical assessments thus tend to support our explanation of how MIN works and when it would perform well. It also points to the desirability of making MIN depend on the population variances. This leads us to wonder if a minimax estimator resembling MIN can be found for estimating θ_1 when $\sigma_2 > \sigma_1$.

Figure 4: Graphs of mean squared error functions for selected estimators when population 1 (variance = 3) is overdispersed relative to 2 (variance = 1).

4. Performance of the estimators

In this section we answer some of the questions raised by the analysis of the previous section. Those answers are stated as theorems whose proofs can be found in the Appendix. We begin by stating in the next theorem the mean-squared-errors of the estimators considered in Section 3. There $\Delta = \theta_2 - \theta_1$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$, $\theta = (\theta_1, \theta_2)$ and $\tau = \sigma_2^2/\sigma_1^2$.

Theorem 4.1 *For the MSEs we have :*

(a) The MSE of δ_{WLE} *is given by*

$$\begin{aligned} &\mathcal{E}_\theta\left(\delta_{WLE}(Y_1,Y_2) - \theta_1\right)^2 = \\ &\sigma_1^2 + \frac{2}{1+\tau}\mathcal{E}_\theta W\hat{\alpha}(W)(\Delta - W) + \mathcal{E}_\theta\hat{\alpha}^2(W)W^2 = \\ &\sigma_1^2 + \sigma_1^4\,\mathcal{E}_\theta\frac{W^2\left(2I(W>0)+1\right) - 2(\sigma_1^2+\sigma_2^2)}{\left(\sigma_1^2+\sigma_2^2+W_+^2\right)^2}; \end{aligned} \tag{4.1}$$

Figure 5: Graphs of mean squared error functions for selected estimators when population 2 (variance = 3) is overdispersed relative to 1 (variance = 1).

(b) The MSE of δ_{MLE} is given by

$$\begin{aligned}
&\mathcal{E}_\theta(\delta_{MLE}(Y_1, Y_2) - \theta_1)^2 = \\
&\sigma_1^2 + \frac{1}{(1+\tau)^2}\Big(2\Delta\mathcal{E}_\theta W I(W<0) - \mathcal{E}_\theta W^2 I(W<0)\Big) = \quad (4.2)\\
&\sigma_1^2 + \frac{1}{(1+\tau)^2}\left\{(\Delta^2 - \sigma^2)\left(1 - \Phi\left(\frac{\Delta}{\sigma}\right)\right) - \Delta\sigma\phi\left(\frac{\Delta}{\sigma}\right)\right\};
\end{aligned}$$

(c) The MSE of δ_{MIN} is given by

$$\begin{aligned}
&\mathcal{E}_\theta(\delta_{MIN}(Y_1, Y_2) - \theta_1)^2 = \\
&\sigma_1^2 + \frac{1}{1+\tau}\mathcal{E}_\theta W I(W<0)(2\Delta - W(1-\tau)) = \quad (4.3)\\
&\sigma_1^2 + \left(\Delta^2 - \sigma^2\frac{1-\tau}{1+\tau}\right)\left(1 - \Phi\left(\frac{\Delta}{\sigma}\right)\right) - \Delta\sigma\phi\left(\frac{\Delta}{\sigma}\right);
\end{aligned}$$

(d) The MSE of the Pitman estimator is given by

$$\mathcal{E}_\theta(\delta_P(Y_1,Y_2)-\theta_1)^2=\sigma_1^2-\frac{\sigma_1^4}{\sigma^3}\,\Delta\,\mathcal{E}_\theta\frac{\phi\left(\dfrac{W}{\sigma}\right)}{\Phi\left(\dfrac{W}{\sigma}\right)}\,. \tag{4.4}$$

In the next theorem some of the MSEs are compared. Like all comparisons between MSEs in this paper, they are made on the restricted parameter space $\Theta=\{\theta\mid\theta_1\le\theta_2\}$. Thus, an estimator δ is inadmissible for estimating θ_1 if there exists an estimator δ^* dominating it on Θ and δ is minimax if it minimizes, among estimators δ^*, $\sup_{\theta\in\Theta}R(\delta^*,\theta)$, where $R(\delta^*,\theta)$ is the MSE of δ^* at the parameter point θ.

We note here that the first result stated in Theorem 4.2 below can be obtained from Kubokawa's (1994) integral-expression-of-risk-difference method. This method provides sufficient conditions for $Y_1+\varphi(W)$ to dominate Y_1. For our case of two ordered normal means, Kubokawa and Saleh (1994) use Kubokawa's method to show that δ_P and δ_{MLE} dominate δ_{ULE}.

THEOREM 4.2 *Each of the estimators δ_{MLE} and δ_P dominates δ_{ULE}. The estimator δ_{MIN} dominates δ_{ULE} when $\tau\le 1$. The MSE's of δ_{MIN} (for $\tau<1$) and δ_{MLE} are strictly smaller than σ_1^2 for all $\Delta\ge 0$ with their limits, as $\Delta\to\infty$, equal to σ_1^2. For δ_{MIN} with $\tau=1$, the MSE equals σ_1^2 for $\Delta=0$. For δ_P equality holds for $\Delta=0$ as well as for $\Delta\to\infty$.*

The previous theorem proves the conjectures made earlier (for MIN only when $\tau\le 1$). The next theorem shows that the WLE, the MLE as well as MIN are inadmissible and a class of dominators of the MLE is given. (Such dominators can be found in Shao and Strawderman (1996)). Thus the next result answers questions we raised in the previous section, the latter for MIN.

THEOREM 4.3 *The estimators δ_{WLE}, δ_{MLE} and δ_{MIN} are inadmissible. Further, δ_{MLE} is dominated by*

$$\frac{\tau Y_1+Y_2}{1+\tau}-\delta_2^*\left(\frac{Y_2-Y_1}{1+\tau}\right)$$

where δ_2^* is a dominator of the maximum likelihood estimator of a non-negative normal mean based on a single observation with unit variance.

In the following theorem we state a result of Cohen and Sackrowitz (1970) concerning the admissibility and minimaxity of PIT. In the Appendix we give a simpler proof of the admissibility. A simpler proof of the minimaxity of PIT can be found in Kumar and Sharma (1988, Theorem 2.3). That theorem thus proves our earlier conjecture for PIT and it answers affirmatively a question we raised in Section 3 concerning that estimator.

THEOREM 4.4 *The Pitman estimator is admissible and minimax. The minimax value for our problem equals* σ_1^2.

The following theorem contains more minimaxity results and proves the first conjecture for the MLE made in Section 3 about it, as well as the conjecture there for the MIN when $\tau \leq 1$.

THEOREM 4.5 *The estimators* δ_{MLE} *and* δ_{ULE} *are minimax and so is* δ_{MIN} *when* $\tau \leq 1$. *Further,* δ_{MIN} *is not minimax when* $\tau > 1$.

5. Discussion

In this article we have tried to show how the intuitively natural idea of the weighted likelihood can be used in parametric estimation to trade bias for precision and thereby reduce the MSE in fortuitous circumstances. The resulting estimators use all the relevant information and not just the direct sample information from the population of interest. By comparing those estimators with others that were obtained earlier for the same purpose we find the weighted likelihood to be promising.

Although we demonstrate the value of our method in a specific normal means estimation context the method itself has wide applicability. Methods of the type described here seem likely to assume increasingly greater importance as the space-time scales of modern experiments continue to expand thanks to need and technological feasibility. Indeed the classical repeated sampling paradigm on which Fisher bases his theory of the MLE will become increasingly untenable as that scale grows. Reliance on biased but relevant sample data will become increasingly imperative.

Brewster and Zidek (1974) show that both the MLE and PIT can be obtained by the method presented in their paper. In fact they consider the case of $p \geq 2$ populations and show how relevant information in samples from populations $2, \ldots, p$ may be used in estimating θ_1, the mean of the first population when $\theta_1 \leq \theta_i$, $i = 2, \ldots, p$. Kubokawa and Saleh (1994) consider this same problem for distributions with monotone likelihood ratio. Further consideration of the extension of our results to more than two means will be left for future work.

We assessed numerically the performance of the various estimators of this paper when $-2 \leq \theta_2 - \theta_1 < 0$. We found that the risk functions of MLE and WLE are similar to the left of $\Delta = 0$. Both appear to dominate ULE at least in the range $-1 \leq \Delta \leq 0$, although that of the WLE seems uniformly lower than that of the MLE.

In contrast, both MIN and PIT do worse than ULE when $\Delta < 0$. In fact the risk of MIN is a rapidly decreasing function in that range so that at $\Delta = -1$ for example the risk of MIN is about 2.25 while that of PIT, ULE, MLE and WLE is respectively about 0.5, 1.0, 0.8, 0.7. This provides evidence of the anticipated robustness of the WLE under departures from the parameter restrictions while at the same time offering some support in favour of the MLE in these circumstances.

To conclude we summarize the results of our investigation in Section 4 of the conjectures and questions suggested by the numerical work in Section 3. We have answered negatively these questions on the admissibility of the MLE and WLE respectively (in Theorem 4.3). Theorem 4.3 gives a negative answer on the admissibility of MIN when $\sigma_1 = \sigma_2$. However, Theorem 4.4 answers positively the same question for PIT.

The question on the form of the MIN when the population variances are unequal remains open.

Theorem 4.5 proves the earlier claim that the ULE and MLE are minimax. Theorem 4.2 proves the claim here that the ULE is inadmissible and dominated by the MLE. We have been unable to prove or disprove the claim that the WLE is minimax in spite of the very strong numerical evidence against it.

Theorem 4.2 proves the MIN dominates the ULE when $\tau \leq 1$ and the PIT in any case. Theorem 4.5 proves an earlier conjecture that the MIN is minimax when $\sigma_2 \leq \sigma_1$. At the same time, Theorem 4.4 shows that PIT is minimax whatever be the σ's.

Appendix

This Appendix contains the proofs of the results presented in Section 4. In these proofs the following four results (stated in the form of lemmas and a corollary) are used.

The first lemma contains the well-known Stein identity.

LEMMA $A.1$ *For a $\mathcal{N}(\nu, \gamma^2)$ random variable Z and a function g which is almost everywhere (with respect to Lebesgue measure) differentiable, $\mathcal{E}(Z - \nu)g(Z) = \gamma^2 \mathcal{E} g'(Z)$.*

The following corollary gives expressions for the mean-squared-error of the estimator $Y_1 + \varphi(W)$ of θ_1. These expressions follow immediately from Lemma A.1 and the fact that the distribution of Y_1, conditional on W, is

$$\mathcal{N}\left(\theta_1 + \frac{\Delta - W}{1+\tau}, \frac{\sigma_2^2}{1+\tau}\right) = \mathcal{N}\left(\frac{\sigma_2^2\theta_1 + \sigma_1^2(\theta_2 - W)}{\sigma^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma^2}\right). \tag{A.1}$$

COROLLARY *A.1 The mean-squared-error of the estimator* $Y_1 + \varphi(W)$ *of* θ_1 *is given by*

$$\mathcal{E}_\theta(Y_1 - \theta_1 + \varphi(W))^2 = \sigma_1^2 + 2\mathcal{E}_\theta(Y_1 - \theta_1)\varphi(W) + \mathcal{E}_\theta\varphi^2(W),$$

where

$$\mathcal{E}_\theta(Y_1 - \theta_1)\varphi(W) = \frac{1}{1+\tau}\,\mathcal{E}_\theta(\Delta - W)\varphi(W). \tag{A.2}$$

Further, if $\varphi(W)$ *is differentiable almost everywhere,*

$$\mathcal{E}_\theta(\Delta - W)\varphi(W) = -\sigma^2\,\mathcal{E}_\theta\varphi'(W). \tag{A.3}$$

In our next lemma, a rotation technique used by Blumenthal and Cohen (1968a) (see also Cohen and Sackrowitz (1970)) is applied.

LEMMA *A.2 Let*

$$X_1 = \frac{\tau Y_1 + Y_2}{1+\tau} \quad X_2 = \frac{-Y_1 + Y_2}{1+\tau} \tag{A.4}$$

and let, for $i = 1, 2$, $\mu_i = \mathcal{E}_\theta X_i$. *Then* $Y_1 + \varphi(W)$ *is inadmissible for estimating* θ_1 *based on* (Y_1, Y_2) *under the condition* $\theta_1 \le \theta_2$ *if* $\delta_2(X_2) = X_2 - \varphi((1+\tau)X_2)$ *is inadmissible for estimating* μ_2 *based on* X_2 *under the condition* $\mu_2 \ge 0$. *Further, if* $\delta_2^*(X_2)$ *dominates* $\delta_2(X_2)$ *for estimating* μ_2 *under the condition* $\mu_2 \ge 0$ *based on* X_2, *then* $X_1 - \delta_2^*(X_2)$ *dominates* $Y_1 + \varphi(W)$ *for estimating* θ_1 *under the condition* $\theta_1 \le \theta_2$ *based on* (Y_1, Y_2).

PROOF. Note that, under $\theta_1 \le \theta_2$, μ_1 is unrestricted while $\mu_2 \ge 0$. Further $Y_1 + \varphi(W) = X_1 - \delta_2(X_2)$. The result then follows from the fact that X_1 is unbiased and that X_1 and X_2 are independent. □

The following result is used several times in our proofs. Its proof is straightforward.

LEMMA *A.3*

$$\mathcal{E}_\theta W I(W < 0) = \Delta\left(1 - \Phi\left(\frac{\Delta}{\sigma}\right)\right) - \sigma\phi\left(\frac{\Delta}{\sigma}\right),$$

$$\mathcal{E}_\theta W^2 I(W < 0) = (\Delta^2 + \sigma^2)\left(1 - \Phi\left(\frac{\Delta}{\sigma}\right)\right) - \Delta\sigma\phi\left(-\frac{\Delta}{\sigma}\right).$$

We are now ready to give the proofs of the results in Section 4.

Proof of the formula for the Pitman estimator δ_P (see (3.4)). The proof of this formula is similar to the one given by van Eeden and Zidek (2001) for the Pitman estimator when the difference between the normal means is bounded. □

Proof of theorem 4.1. The results (4.1),(4.2) and (4.3) follow from Corollary A.1 and Lemma A.3. The expression (4.4) for the MSE of δ_P can be obtained by generalizing a proof of Al-Saleh (1997) and by one given by Kumar and Sharma (1993). These authors assume $\sigma_1 = \sigma_2$. □

Proof of theorem 4.2. That δ_{MLE} dominates δ_{ULE} follows from a result of Lee (1981). He shows that for independent $Y_i \sim \mathcal{N}(\theta_i, 1)$, $i = 1, \ldots, k$, with $\theta_1 \leq \ldots \leq \theta_k$, the i-th component of the order-restricted MLE dominates Y_i, $i = 1, \ldots, k$. For our particular case, where $k = 2$, the result can more easily be proved by using the second line of (4.2) and the following inequalities

$$\Delta \mathcal{E}_\theta W I(W < 0) \leq 0 \text{ for all } \Delta \geq 0 \tag{A.5}$$

$$\mathcal{E}_\theta W^2 I(W < 0) > 0 \text{ for all } \Delta \geq 0. \tag{A.6}$$

From (4.4) it is immediately clear that δ_P dominates δ_{ULE}.

To see that δ_{MIN} dominates δ_{ULE} when $\tau \leq 1$, note that (see the second line of (4.3))

$$\begin{aligned}
\Delta \mathcal{E}_\theta W I(W < 0) &\leq 0 \text{ for all } \Delta \geq 0 && \text{(A.7)} \\
(1-\tau)\, \mathcal{E}_\theta W^2 I(W < 0) &= 0 \text{ for all } \Delta \geq 0 \text{ when } \tau = 1 && \text{(A.8)} \\
(1-\tau)\, \mathcal{E}_\theta W^2 I(W < 0) &> 0 \text{ for all } \Delta \geq 0 \text{ when } \tau < 1. && \text{(A.9)}
\end{aligned}$$

From the second line of (4.3) and the inequalities (A.7) and (A.9) it follows that the MSE of δ_{MIN} with $\tau < 1$ is strictly smaller than σ_1^2 for all $\Delta \geq 0$. For δ_{MLE}, the second line of (4.2) and the inequalities (A.5) and (A.6) imply that its MSE is strictly smaller than σ_1^2 for all $\Delta \geq 0$. As for the limits, as $\Delta \to \infty$, of the MSEs of δ_{MIN} and δ_{MLE}, use the last line of (4.3) and of (4.2) and note that $\lim_{\Delta\to\infty} \Delta^2 (1 - \Phi(\Delta/\sigma)) = 0$ and $\lim_{\Delta\to\infty} \Delta\phi(\Delta/\sigma) = 0$. That, for $\tau = 1$, the MSE of δ_{MIN} equals σ_1^2 for $\Delta = 0$ follows immediately from the last line of (4.3).

Finally (using (4.4)), the MSE of δ_P clearly equals σ_1^2 when $\Delta = 0$. That its MSE converges to σ_1^2 when $\Delta \to \infty$ can be seen from (4.4) by noting that $\Delta\mathcal{E}_\theta h(W') = \Delta\mathcal{E}h(Z+\Delta/\sigma)$, where $Z \sim \mathcal{N}(0,1)$, $W' = W/\sigma$ and $h(z) = \phi(z)/\Phi(z)$. The result then follows from the fact that, for each fixed z, $\Delta h(z+\Delta/\sigma)$ is bounded in z for $\Delta \geq 0$ and converges to zero as $\Delta \to \infty$. □

Proof of theorem 4.3. The inadmissibilities follow from Lemma A.2 as follows. For δ_{WLE} use the fact that $\delta_2(X_2)$ does not satisfy Sack's (1963) necessary condition for admissibility for estimating $\mu_2 \geq 0$ based on X_2. For δ_{MLE}, $\delta_2(X_2)$ is the MLE of $\mu_2 \geq 0$ based on X_2 which is well-known to be inadmissible. The dominator follows from Lemma A.2. Finally, for δ_{MIN} the estimator δ_2 is inadmissible because it is not monotone. □

Proof of theorem 4.4. As already noted above, Kumar and Sharma (1988, Theorem 2.3) give a proof of the minimaxity of δ_P. Their proof is very much simpler than the one given by Cohen and Sackrowitz (1970). The Kumar-Sharma proof is based on an extension of a result of Blumenthal and Cohen (1968b, Theorem 3.0).

For an alternate and simpler proof of the admissibility of δ_P, use the transformation (A.4). Then (see the proof of Lemma A.2) $\delta_P(Y_1,Y_2) = X_1 - \delta_2(X_2)$, where $\delta_2(X_2)$ can be written in the form

$$\delta_2(X_2) = X_2 + \sigma(X_2)h\left(\frac{X_2}{\sigma(X_2)}\right)$$

with $\sigma^2(X_2)$ the variance of X_2. Further, $\theta_1 = \mu_1 - \mu_2$ and $\theta_1 \leq \theta_2 \iff \mu_1 \in (-\infty,\infty), \mu_2 \geq 0$.

So, it is now sufficient to show that $\delta(X) = X_1 - \delta_2(X_2)$ is admissible for estimating $\mu_1 - \mu_2$ based on $X = (X_1, X_2)$ when $\mu_2 \geq 0$. We will show this by using Blyth's (1951) method.

Suppose that there exists an estimator $\delta'(X)$ which dominates $\delta(X)$ on $\Omega = \{\mu \mid \mu_1 \in (-\infty,\infty), \mu_2 \geq 0\}$. Then, because the risk function $R(\delta_o,\mu)$ of every estimator $\delta_o(X)$ is continuous in μ for $\mu \in \Omega$, there exists an $\varepsilon > 0$ and a rectangle $S = (\mu_{1,1},\mu_{1,2}) \times (\mu_{2,1},\mu_{2,2}) \subset \Omega$ such that

$$R(\delta,\mu) - R(\delta',\mu) > \varepsilon \quad \text{on } S. \tag{A.10}$$

Now take a sequence of priors $\lambda_n, n = 1,2,\ldots$ for $\mu \in \Omega$ where, for each n, μ_1 and μ_2 are independent, μ_1 with the improper uniform prior on $(-\infty,\infty)$ and μ_2 with density $e^{-\mu_2/n}/n$ on $\mu_2 \geq 0$. Then (A.10) implies that

$$r_n(\delta) - r_n(\delta') > \varepsilon(\mu_{1,2} - \mu_{1,1})\left(e^{-\frac{\mu_{2,1}}{n}} - e^{-\frac{\mu_{2,2}}{n}}\right) = O(\frac{1}{n}), \tag{A.11}$$

where the r_n's are Bayes risks with respect to λ_n.

Now let, for $i = 1, 2$, $\delta_{n,i}(X_i)$ be the Bayes estimator of μ_i based on X_i with respect to the (marginal) prior of μ_i. Then, by the prior independence of μ_1 and μ_2 and the conditional independence of X_1 and X_2 given μ_1 and μ_2, the Bayes estimator of $\mu_1 - \mu_2$ for the prior λ_n based on X, is given by $\delta_n(X) = \delta_{n,1}(X_1) - \delta_{n,2}(X_2)$, where (see Katz (1961))

$$\delta_{n,2}(X_2) = X_2 - \frac{\sigma^2(X_2)}{n} + \sigma(X_2) h\left(\frac{X_2 - \frac{\sigma^2(X_2)}{n}}{\sigma(X_2)}\right).$$

Further (see Katz, (1961))

$$r_n(\delta) - r_n(\delta_n) = r_{n,2}(\delta_2) - r_{n,2}(\delta_{n,2}) = o(\frac{1}{n}), \tag{A.12}$$

where $r_{n,2}$ is the Bayes risk of an estimator based on X_2 with respect to the (marginal) prior of μ_2.

But (A.11) and (A.12) imply that, for sufficiently large n,

$$\frac{r_n(\delta) - r_n(\delta')}{r_n(\delta) - r_n(\delta_n)} > 1$$

which contradicts the fact that, for each n, $\delta_n(X)$ is the Bayes estimator of $\mu_1 - \mu_2$ with respect to λ_n based on X. □

REMARK A.1 *Katz's minimaxity and admissibility proofs are incorrect for the general case of the exponential family he considers (see van Eeden (1995)), but the above quoted results of his for the normal mean are correct.*

PROOF OF THEOREM 4.5. The minimaxity results follow immediately from Theorem 4.2 and the fact that, by Theorem 4.4, the minimax value for our problem is equal to the mean square error σ_1^2 of $\delta_{ULE} = Y_1$. That δ_{MIN} is not minimax when $\tau > 1$ can be seen from the second line of (4.3) by noting that the MSE of δ_{MIN} is larger than σ_1^2 when $\Delta = 0$. □

References

AKAIKE, H. (1973). Information theory and an extension of entropy maximization principle. In *Second International Symposium on Information Theory*, B.N. Petrov and F. Csáki, eds., Akadémiai Kiadó, Budapest, 276-281.

AKAIKE, H. (1977). On entropy maximization principle. In *Applications of Statistics*, P.R. Krishnaiah, ed., North-Holland, Amsterdam, 27-41.

AKAIKE, H. (1978). A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, **30**, 9-14.

AKAIKE, H. (1982). On the fallacy of the likelihood principle, *Statist. Probab. Lett.*, **1**, 75-78.

AKAIKE, H. (1983). Information measures and model selection. In *Proc. 44th ISI Session.* **1**, 277-291.

AKAIKE, H. (1985). Prediction and entropy. In *A Celebration of Statistics, The ISI Centenary Volume*, Spring-Verlag, Berlin, 1-24.

AL-SALEH, M.F. (1997). Estimating the mean of a normal population utilizing some available information: a Bayesian approach, *J. Inform. Optim. Sci.*, **18**, 1-7.

BASU, D. (1975). Statistical information and likelihood, *Sanhkyā, Ser. A*, **37**, 1-71.

BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Second edition, Springer-Verlag, New York.

BLUMENTHAL, S. and COHEN, A. (1968a). Estimation of the larger translation parameter, *Ann. Math. Statist.*, **39**, 502-516.

BLUMENTHAL, S. and COHEN, A. (1968b). Estimation of two ordered translation parameters, *Ann. Math. Statist.*, **39**, 517-530.

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

BREWSTER, J.F. and ZIDEK, J.V. (1974). Improving on equivariant estimators, *Ann. Statist.*, **2**, 21-38.

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

EFRON, B. (1996). Empirical Bayes methods for combining likelihoods (with discussion), *J. Amer. Statist. Assoc.*, **91**, 538-565.

FAN, J., HECKMAN, N.E. and WAND, W.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions, *J. Amer. Statist. Assoc.*, **90**, 141-150.

HU, F. (1994). Relevance weighted smoothing and a new bootstrap method. Ph.D. Thesis, Department of Statistics, University of British Columbia.

HU, F. (1997). Asymptotic properties of relevance weighted likelihood estimations, *Canad. J. Statist.*, **25**, 45-60.

HU, F. and ZIDEK, J.V. (2001). The relevance weighted likelihood with applications. In *Empirical Bayes and Likelihood Inference*, S.E. Ahmed and N. Reid, eds., Springer, New York, 211-235.

HU, F. and ZIDEK, J.V. (2002). The weighted likelihood, *Canad. J. Statist.*, to appear.

KATZ, M.W. (1961). Admissible and minimax estimates of parameters in truncated spaces, *Ann. Math. Statist.*, **32**, 136-142.

KUBOKAWA, T. (1994). A unified approach to improving equivariant estimators, *Ann. Statist.*, **22**, 290-299.

KUBOKAWA, T. and SALEH, A.K.MD.E. (1994). Estimation of location and scale parameters under order restrictions, *J. Statist. Res.*, **28**, 41-51.

KUMAR, S. and SHARMA, D. (1988). Simultaneous estimation of ordered parameters, *Comm. Statist. Theory Methods*, **17**, 4315-4336.

KUMAR, S. and SHARMA, D. (1993). Minimaxity of the Pitman estimator of ordered normal means when the variances are unequal, *J. Indian Soc. Agricultural Statist.*, **45**, 230-234.

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

SACKS, J. (1963). Generalized Bayes solutions in estimation problems, *Ann. Math. Statist.*, **34**, 751-768.

SHAO, P.Y-S. and STRAWDERMAN, W.E. (1996). Improving on the MLE of a positive normal mean, *Statist. Sinica*, **6**, 259-274.

STANISWALIS, J.G. (1989). The kernel estimate of a regression function in likelihood - based models, *J. Amer. Statist. Assoc.*, **84**, 276-283.

STIGLER, S.M. (1990). The 1988 Neyman Memorial Lecture: A Galtonian perspective on shrinkage estimators, *Statist. Sci.*, **5**, 147-155.

TIBSHIRANI, R. and HASTIE, T. (1987). Local likelihood estimation, *J. Amer. Statist. Assoc.* **82**, 559-567.

VAN EEDEN, C. (1995). Minimax estimation of a lower-bounded scale parameter of a gamma distribution for scale-invariant squared-error loss, *Canad. J. Statist.*, **23**, 245-256.

VAN EEDEN, C. and ZIDEK, J.V. (2001). Estimating one of two normal means when their difference is bounded, *Statist. Probab. Lett.*, **51**, 277-284.

WANG, S. (2001). The maximum weighted likelihood estimator. Ph.D Thesis, Department of Statistics, University of British Columbia.

WANG, X., VAN EEDEN, C. and ZIDEK, J.V. (to appear). Asymptotic properties of maximum weighted likelihood estimators, *J. Statist. Plann. Inference.*

ZIDEK, J.V., WHITE, R. and LE, N.D. (1999). Using spatial data in assessing the association between air pollution episodes and respiratory morbidity. In *Statistics for the Environment* **4**, *Statistical Aspects of Health and the Environment*, V. Barnett, A. Stein and K.F. Turkman, eds., Wiley, New York, 117-135.

CONSTANCE VAN EEDEN AND JAMES V. ZIDEK
DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
6356 AGRICULTURE ROAD
VANCOUVER, BC
CANADA V6T 1Z2
E-mail: jim@stat.ubc.ca
vaneeden@stat.ubc.ca, cve@xs4all.nl