ON THE INADMISSIBILITY OF UNBIASED ESTIMATORS *

by

James O. Berger
Purdue University

Technical Report #89-09C

Department of Statistics
Purdue University

April 1989

* Research supported by the National Science Foundation, Grant DMS–8702620

★

ON THE INADMISSIBILITY OF UNBIASED ESTIMATORS *

by

James O. Berger
Purdue University

Abstract

It is observed that unbiased estimators are always inadmissible when the parameter (or function of the parameter) being estimated has either a maximum or a minimum at a parameter value for which the probability distribution is nondegenerate. Examples of problems where this is so include variance components problems, problems with restricted parameter spaces, and estimation of the risk or variance of shrinkage estimators.

1. INTRODUCTION

It has long been recognized that unbiased estimators can be unsatisfactory. For instance, from the decision–theoretic perspective, unbiased estimators can frequently be shown to be inadmissible.

One of the most famous instances of inadmissibility arises in estimation of a multivariate normal mean (in at least three dimensions), where the usual unbiased estimator is inadmissible under quadratic loss (cf., Stein, 1981). Curiously, the study of improved estimators, the so–called shrinkage estimators, has led to a resurgence in interest in unbiased estimation. This is because shrinkage estimators have risk functions (or mean squared error functions) that are nonconstant and typically unavailable in closed form. This makes reporting the accuracy of shrinkage estimators potentially difficult. But it so happens that relatively simple closed form unbiased estimators of the risk (or mean squared error) functions can be developed (based on ideas in Stein, 1981); these have hence come under considerable scrutiny (cf., Johnstone, 1981, Lu and Berger, 1989, and Berger and Robert, 1988).

In Berger and Robert (1988), an extensive study of estimating the accuracy of shrinkage estimators was undertaken, focusing on comparison of unbiased measures of accuracy and Bayesian measures of accuracy. It was observed that unbiased estimators of risk (say) often have extreme and undesirable behavior near maxima or minima of the risk function.

* Research supported by the National Science Foundation, Grant DMS–8702620.

The reason is apparent after a little thought, and immediately generalizes to an inadmissibility result, given herein, for unbiased estimation of any quantities that have global maxima or minima at parameter values for which the probability distribution is nondegenerate. Perhaps the main insight to be gained from this note is that inadmissibility of unbiased estimators is likely to be the rule, rather than the exception.

2. INADMISSIBILITY

Suppose one observes $X \in \mathcal{X}$ having density $f(x|\theta)$ with respect to a measure ν on $\mathcal{X}$, where the unknown θ is in the parameter space Θ. It is desired to estimate a real valued function $h(\theta)$ under a loss function $L(d, h(\theta))$ which is strictly increasing in d for $d > h(\theta)$ and strictly decreasing in d for $d < h(\theta)$ (i.e., the loss increases as the estimate moves away from the actual $h(\theta)$). The key assumption follows.

Condition 1. The function $h(\theta)$ is nonconstant and has a global maximum or a global minimum at a point $\theta^* \in \Theta$ for which $f(x|\theta^*) > 0$ for almost all (w.r.t. ν) $x \in \mathcal{X}$.

Considerably more general conditions could be utilized, but Condition 1 covers virtually all the relevant examples and is comparatively simple. The following lemma contains the key idea.

Lemma 1. Suppose that Condition 1 holds, where θ^* minimizes $h(\theta)$, and that $\hat{h}(x)$ is an unbiased estimator of $h(\theta)$. Define $A_\varepsilon = \{x \in \mathcal{X} : \hat{h}(x) < h(\theta^*) - \varepsilon\}$. Then there exists $\varepsilon > 0$ such that $P_{\theta^*}(A_\varepsilon) > 0$. If θ^* maximizes $h(\theta)$, the conclusion is true with the inequality in the definition of A_ε reversed.

Proof. Supposing θ^* minimizes $h(\theta)$ (the maximizing case is done similarly), we will proceed by contradiction. Thus suppose that, for all $\varepsilon > 0$, $\hat{h}(X) > h(\theta^*) - \varepsilon$ with P_{θ^*} – probability one. It follows that

$$P_{\theta^*}(\hat{h}(X) - h(\theta^*) \geq 0) = 1. \tag{1}$$

Unbiasedness of $\hat{h}$ implies that

$$E_{\theta^*}[\hat{h}(X) - h(\theta^*)] = 0. \tag{2}$$

But (1) and (2) can hold at the same time only if

$$P_{\theta^*}(\hat{h}(X) = h(\theta^*)) = 1.$$

Because of Condition 1, this can hold only if

$$\hat{h}(x) = h(\theta^*) \text{ a.e. } (\nu). \tag{3}$$

But, if (3) holds, it is clear that $E_\theta[\hat{h}(X)] = h(\theta^*)$ for all $\theta \in \Theta$, which contradicts the nonconstancy of $h(\theta)$ in Condition 1. □

Lemma 1 establishes that, with positive probability, $\hat{h}(X)$ must be less than the known minimum of $h(\theta)$ (and/or greater than the known maximum). This is clearly unreasonable; one should never use an estimate of $h(\theta)$ that lies outside the known range of $h(\theta)$. The following theorem states this rigorously.

Theorem 1. Under the assumptions of Lemma 1, if $\hat{h}$ has finite risk function

$$R(\hat{h}, h(\theta)) = E_\theta L(\hat{h}(X), h(\theta)),$$

then $\hat{h}$ is an inadmissible estimator of $h(\theta)$. A better estimator is obtained by truncating $\hat{h}(x)$ at the minimum (and/or maximum) $h(\theta^*)$.

Proof. For the case where $h(\theta^*)$ is a minimum (the other case is handled similarly), the truncated estimator is

$$\tilde{h}(x) = \begin{cases} \hat{h}(x) & \text{if } \hat{h}(x) \geq h(\theta^*) \\ h(\theta^*) & \text{if } \hat{h}(x) < h(\theta^*). \end{cases}$$

The difference of the risk functions of $\hat{h}$ and $\tilde{h}$ is

$$\begin{aligned} R(\hat{h}, h(\theta)) - R(\tilde{h}, h(\theta)) &= E_\theta[L(\hat{h}(X), h(\theta)) - L(\tilde{h}(X), h(\theta))] \\ &= \int\limits_{\{x:\hat{h}(x)<h(\theta^*)\}} [L(\hat{h}(x), h(\theta)) - L(h(\theta^*), h(\theta))] f(x|\theta) d\nu(x) \\ &\geq \int\limits_{A_\varepsilon} [L(h(\theta^*) - \varepsilon, h(\theta)) - L(h(\theta^*), h(\theta))] f(x|\theta) d\nu(x) \\ &= P_\theta(A_\varepsilon)[L(h(\theta^*) - \varepsilon, h(\theta)) - L(h(\theta^*), h(\theta))]. \end{aligned}$$

By the assumption on L, the term in brackets is positive for all ε, so that we immediately have

$$R(\tilde{h}, h(\theta)) \leq R(\hat{h}, h(\theta)). \tag{4}$$

Also, $P_{\theta^*}(A_\varepsilon) > 0$ by Lemma 1 (for some $\varepsilon > 0$), so that the inequality in (4) is strict for $\theta = \theta^*$. Hence $\tilde{h}$ is better than $\hat{h}$. □

3. EXAMPLES

Example 1. (Shrinkage Estimation). Suppose $X = (X_1, \ldots, X_p) \sim \mathcal{N}_p(\theta, I)$, where $\theta = (\theta_1, \ldots, \theta_p)$, and it is desired to estimate θ under the quadratic loss $(d-\theta)^t Q(d-\theta)$, where Q is a positive definite matrix. Shrinkage estimators of θ are of the form $\delta(x) = (I - H(x))x$, where $H(x)$ is the shrinkage factor (possibly matrix–valued). Shrinkage estimators always have risk functions, $h(\theta, \delta) = E_\theta[(\theta - \delta(X))^t Q(\theta - \delta(X))]$, that have a global minimum at some point θ^* (which need not be unique), so that Condition 1 is clearly satisfied. Furthermore, using integration by parts (cf. Stein (1981), Berger and Robert (1988)), one can find an unbiased estimator, $\hat{h}_\delta(x)$, of $h(\theta, \delta)$. Under any reasonable loss $L(\hat{h}, h)$ (e.g., any bounded loss or squared error loss), one can establish that $E_\theta L(\hat{h}(X), h(\theta))$ is finite, so that Theorem 1 applies; it is thus better to truncate $\hat{h}_\delta(x)$ at $h(\theta^*, \delta)$.

For many shrinkage estimators, $h(\theta, \delta)$ also has a maximum at some θ^*. (Not all shrinkage estimators have this property; e.g., for minimax shrinkage estimators the maximum of $h(\theta, \delta)$ is typically at $|\theta| = \infty$.) Then Lemma 1 and Theorem 1 also apply to the maximum.

The amount by which $\hat{h}_\delta(x)$ is excessively small or large can be very substantial, as illustrated in Berger and Robert (1989). Similar conclusions also apply to unbiased estimation of variances of shrinkage estimators.

Example 2. (Variance Components). Although Theorem 1 will apply to general variance components problems, the main ideas can be seen by looking at the simple one–way model

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \ldots, I \text{ and } j = 1, \ldots, J,$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and α_i are $\mathcal{N}(0, \tau^2)$, all independently, and μ, σ^2, and τ^2 are unknown.

Consider, first, estimation of $h(\mu, \tau^2, \sigma^2) = \tau^2$. This function is clearly minimized at $\tau^2 = 0$ (with any values of μ and σ^2). Furthermore, even at $\tau^2 = 0$ the support of the probability distribution is the entire sample space. Thus Condition 1 is satisfied and Theorem 1 applies: for any loss function for which the risks exist, any unbiased estimator of τ^2 is inadmissible because it must, with positive probability, assume values that are negative (less than zero, the minimum of τ^2). It is, of course, well known that the best unbiased estimator of τ^2 can be negative, and that this is inadmissible.

Next, consider $h(\mu, \tau^2, \sigma^2) = \sigma^2$. This function likewise is minimized at $\sigma^2 = 0$ (and any values of μ and τ^2), but now Condition 1 is not satisfied. (The probability distribution is degenerate when $\sigma^2 = 0$, with $X_{i1} = X_{i2} = \ldots = X_{iJ}$ for all i.) Indeed, the unbiased estimator for σ^2 does here only assume values in the range, $[0, \infty)$, of σ^2.

Example 3. (Restricted Parameter Spaces). For completeness, we note that Condition 1 and Theorem 1 typically will apply to situations in which unbiased estimators exist for a general parameter space Θ^*, but the actual problem has the parameter restricted to a subset, Θ, of Θ^*. An estimator that is unbiased with the general Θ^* is also clearly unbiased over the smaller Θ. And, over the smaller Θ, Condition 1 will typically be satisfied. The inadmissibility of unbiased estimators here is, of course, not a surprise, since again they virtually always assume values outside the actual parameter space Θ.

REFERENCES

[1] Berger, J. and Robert, C. (1988). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. Technical Report #88–24C, Department of Statistics, Purdue University.

[2] Johnstone, I. (1988). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics IV* (S.S. Gupta and J. Berger, eds.). Springer–Verlag, New York.

[3] Lu, K.L. and Berger, J. (1989). Estimation of normal means: frequentist estimation of loss. *Ann. Statist.* **17**.

[4] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135–1151.