A Simple Ancillarity Paradox

Author(s): Dean P. Foster and Edward I. George
Reviewed work(s):
Source: *Scandinavian Journal of Statistics*, Vol. 23, No. 2 (Jun., 1996), pp. 233-242
Published by: Blackwell Publishing on behalf of Board of the Foundation of the Scandinavian Journal of Statistics
Stable URL: http://www.jstor.org/stable/4616399
Accessed: 28/10/2011 15:35

 Vol. 23: 233–242, 1996

A Simple Ancillarity Paradox

DEAN P. FOSTER

University of Pennsylvania

EDWARD I. GEORGE

University of Texas at Austin

ABSTRACT. For the problem of estimating the mean of a univariate normal distribution with known variance, the maximum likelihood estimator (MLE) is best invariant, minimax, and admissible under squared-error loss. It is shown that if the variance is the realized value of an ancillary statistic with known distribution, the MLE can be inadmissible with respect to the unconditional risk averaged over this ancillary distribution.

Key words: admissibility, ancillary statistics, conditional inference, risk

1. The paradox

An ancillarity paradox may be succinctly defined as a phenomenon where "a procedure which is conditionally admissible for each value of an ancillary statistic can be unconditionally inadmissible" (Berger, 1990). In this paper, we show that such an ancillarity paradox can occur even in the simple and familiar setting where we wish to estimate μ from the observation of a single normal random variable $X \sim N(\mu, v)$ with v known. Under expected squared-error loss

$$R(\delta, \mu, v) \equiv E_{\mu}^{v}(\delta - \mu)^2, \tag{1}$$

the MLE $\delta^{\text{MLE}} \equiv X$ is best invariant, minimax and admissible. Now suppose that v is the realization of an ancillary statistic V with a known distribution F_V which does not depend on μ. In spite of the widely held notion that inference about μ should be conditional on $V = v$, it turns out that it can matter very much if in the long run this experiment is repeated over and over.

The main thrust of this paper is to show that for certain F_V, δ^{MLE} is unconditionally inadmissible in the following sense. For such F_V, the estimator

$$\delta^{v} \equiv \delta^{v}(X, V) = \left(1 - \frac{V}{X^2 + V}\right)X \tag{2}$$

uniformly dominates δ^{MLE} under the unconditional risk

$$R(\delta, \mu) \equiv E_{F_V} R(\delta, \mu, V) = E_{F_V} E_{\mu}^{v}(\delta - \mu)^2. \tag{3}$$

One such F_V under which this unconditional dominance occurs has continuous density

$$f_V(v) = \alpha v^{-(\alpha+1)} I_{[v>1]}, \tag{4}$$

where $1 < \alpha < \alpha_0$ with α_0 specified in theorem 3 below. The proof of dominance for this and other F_V is the subject of section 3.

A situation in which the above setup arises naturally is the estimation of the drift of Brownian motion observed only at a random stopping time. More precisely, let $Z(t)$ be a Brownian motion with drift rate μ and variance rate 1, i.e. $Z(t) \sim N(\mu t, t)$. The problem is to estimate μ based on $Z(T)$ and T, where T is a stopping time which is independent of $Z(t)$. But this is equivalent to estimating μ based on $X \equiv Z(T)/T$ and $V \equiv 1/T$, where $X \mid V \sim N(\mu, V)$. Thus, our results show that the MLE for μ may be inadmissible for an ancillary T, and provide a dominating estimator. A related result by Brown (1988) shows that in this context the MLE can be inadmissible for random T which is not ancillary.

Our ancillarity paradox persists for a large variety of loss functions. The following lemma (whose straightforward proof is omitted) shows that our results imply that δ^v dominates δ^{MLE} with respect to a wide class of unconditional risks of the form

$$R_g(\delta, \mu) \equiv E_{F_V} g(V) R(\delta, \mu, V). \tag{5}$$

For example, letting $g(v) = 1/v$, it turns out that for certain F_V (different from F_V given by (4)), δ^v dominates δ^{MLE} with respect to the unconditional normalized risk

$$R_N(\delta, \mu) \equiv E_{F_V} \frac{1}{V} R(\delta, \mu, V). \tag{6}$$

Lemma 1

Suppose δ dominates δ^{MLE} with respect to $R(\delta, \mu) \equiv E_{F_V} R(\delta, \mu, V)$ for a distribution F_V. If there exists a distribution F_V^g such that $dF_V^g(v) \propto (1/g(v))\, dF_V(v)$, for a function $g(v)$, then δ dominates δ^{MLE} with respect to $R_G(\delta, \mu) \equiv E_{F_V^g} g(V) R(\delta, \mu, V)$.

The underlying intuition behind our proof in section 3 that δ^v unconditionally dominates δ^{MLE} is based on observing how $R(\delta^v, \mu, v)$, the conditional risk function of δ^v, changes as v varies, and how the unconditional risk $R(\delta^v, \mu)$ can average these risks to yield smaller unconditional risk than δ^{MLE}. This general feature of ancillarity paradoxes is described by Casella (1990). To see how $R(\delta^v, \mu, v)$ changes as a function of v, Fig. 1 provides $(1/v)R(\delta^v, \mu, v)$ (obtained by numerical integration), for $v = 1, 10, 100, 1000, 10000$, where the horizontal axis has been compressed to present $(-\infty, +\infty)$. Consider first the shape of $R(\delta^v, \mu, 1)$, the risk of δ^v when $V = 1$, which corresponds to the solid line curve with maxima closest to $\mu = 0$. $R(\delta^v, \mu, 1)$ obtains its minimum at $\mu = 0$, where $R(\delta^v, 0, 1) \approx 0.46704$, increases to its maximum, $\sup_\mu R(\delta^v, \mu, 1) = 1.25$, on either side of 0, and then asymptotes to 1 as $\mu \to \pm\infty$. As v is increased, the successive risk functions have wider regions where they obtain "improved" risk. It turns out that as a result of this feature, these risk functions can be averaged to widen the region of improved risk further. Moreover, $R(\delta^{\mathrm{MLE}}, \mu, v) = v$,

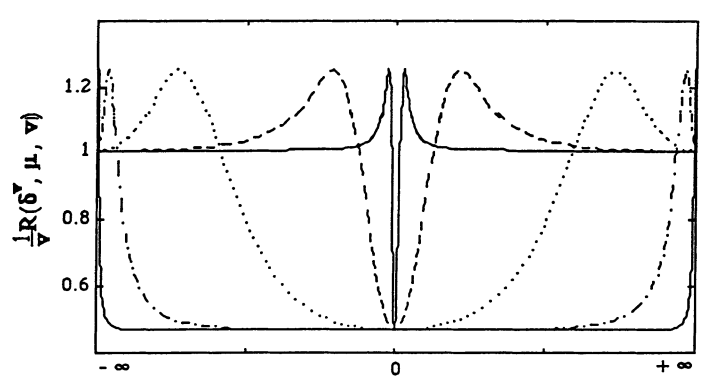

Fig. 1. Normalized risk functions $(1/v)R(\delta^v, \mu, v)$ for $v = 1, 10, 100, 1000, 10000$.

which implies that $R(\delta^{\mathrm{MLE}}, \mu) = E_{F_V} V$. Thus, unconditional dominance of δ^{MLE} by δ^v occurs under F_V for which,

$$R(\delta_i^v, \mu) = \int R(\delta^v, \mu, v)\, dF_V(v) < \int v\, dF_V(v) = R(\delta^{\mathrm{MLE}}, \mu) \quad \text{for all } \mu. \tag{7}$$

It may be of interest to note that the general results of Brown (1966) on the admissibility of the best invariant estimator of a location parameter can be applied to our setup. It follows from his th. 2.4.1 that our ancillarity paradox cannot occur when $EV^{3/2} < \infty$. Furthermore, when $EV = \infty$, δ^{MLE} has infinite unconditional risk and so can be trivially dominated by any $\delta^c \equiv c$. The distribution F_V satisfying (7) falls within these bounds. Indeed $E_{F_V} V < \infty$ whereas $E_{F_V} V^{3/2} = \infty$.

2. Understanding the paradox

The value of statistical paradoxes lies in what we learn from them. The key to understanding our ancillarity paradox is to note that the main issue has to do with the choice of a long run sequence for risk evaluation. More precisely, suppose we are interested in the long run consequence of using an estimator δ in our setup. Two different kinds of long run may be considered. One long run sequence of interest would consist of repetitions of the unconditional experiment where $X \sim N(\mu, V)$ after $V \sim F_V$, namely

$$X_1^{V_1}, X_2^{V_2}, \ldots. \tag{8}$$

The long run consequence of using δ on (8) is measured by the unconditional risk $R(\delta, \mu) = E_{F_V} E_\mu^v (\delta - \mu)^2$ in (3), which is equal to the limit of the average loss over (almost) any realization of this sequence. The other kind of long run of interest would consist only of repetitions of the conditional experiment $X \sim N(\mu, v)$ for a fixed value of v, namely

$$X_1^v, X_2^v, \ldots. \tag{9}$$

The long run consequence of using δ on (9) is measured by the conditional risk $R(\delta, \mu\ v) \equiv E_\mu^v (\delta - \mu)^2$ in (1), which is equal to the limit of the average loss over (almost) any realization of this sequence.

From this long run sequence perspective (which we call the "sequence man perspective", see also Berger (1984) and Neyman (1977)), conditioning on $V = v$ in (8) corresponds to restricting evaluation to a sequence of the form (9). (Strictly speaking, this is only true if F_V is a discrete distribution. For continuous F_V, this interpretation is only obtained as the limit of suitable discrete approximations to F_V.) In contrast, unconditional evaluation via (8) corresponds to evaluation over a mixture of subsequences of the form (9) according to the distribution of V, namely F_V. Seen in this way, our ancillarity paradox is simply an example where admissibility over conditional sequences of the form (9) does not guarantee admissibility over unconditional sequences of the form (8). That this can happen is not so surprising when it is understood that risk values are limits of average loss over these sequences. Indeed, in our example the effect on the limit of conditional subsequences where δ^v fares worse than δ^{MLE} can be offset in the unconditional sequence by mixing in many other conditional subsequences where δ^v fares much better than δ^{MLE}.

At first glance, it appears that the only paradox here is that the relative merits of δ^{MLE} and δ^v depend on whether or not we condition on the ancillary statistic V, which has nothing to do with μ. However, the real issue is whether it suffices to base risk evaluation on the conditional sequence (9), which is the effect of conditioning on the ancillary statistic. This is a serious problem because we do not know in advance which sequence we are facing and because, as our paradox shows, good conditional performance may not guarantee good unconditional performance. It seems to us that the only way to avoid such potential problems is to carry out evaluations from both the conditional and the unconditional points of view.

Our ancillarity paradox has several precedents in the literature. To begin with, Perng (1970) and Fox (1981), building on the work of Brown (1966), provide explicit examples where the best invariant estimator X of a location parameter μ is inadmissible. Although these examples do not explicitly refer to an ancillarity paradox, the distributions under which X is inadmissible are obtained by mixing a conditional density of the form $f(x-\mu \mid v)$ over a distribution on v which does not depend on μ. Thus, an ancillarity paradox such as ours is implicit in their results. From a purely technical point of view, the only contribution of our example is that it is simpler and more clearly exposes the nature of the paradox.

The main precedent of our ancillarity paradox is the discussion paper by Brown (1990) which demonstrates an ancillarity paradox in multiple regression, namely that the least squares intercept estimate which, conditionally on the design matrix, is best invariant, minimax and admissible, can become unconditionally inadmissible under certain distributions on the design matrix, which is ancillary in this problem. Using our above "sequence man perspective" with X and V representing the intercept estimate and the design, this paradox also occurs because risk evaluation over an unconditional sequence (8) can differ markedly from risk evaluation over a conditional sequence (9). In this case, the offending distribution on V combines the conditional subsequences in such a way that a multivariate Stein phenomenon occurs in the unconditional sequence.

Although there are other ancillarity paradoxes in the literature (see e.g. Cox, 1958), these are fundamentally different in nature from ours. Berger (1990) observes that Brown's paradox can be distinguished from these other paradoxes by the fact that it is "impossible to determine its inadequacy using conditional reasoning". He points out that other ancillarity paradoxes can be resolved by noting that the conditional procedures are Bayes or generalized Bayes with respect to different priors. Our paradox shares this distinguishing feature of Brown's paradox since it also cannot be resolved by such conditional reasoning.

3. The proof

In this section we prove that for certain F_V, δ^v uniformly dominates δ^{MLE} with respect to the unconditional risk (3). However, rather than prove this directly, we first show unconditional dominance with respect to variance normalized risk

$$R_N(\delta, \mu, v) \equiv \frac{1}{v} R(\delta, \mu, v). \tag{10}$$

That is, we show that for particular F_V, δ^v-dominates δ^{MLE} with respect to

$$R_N(\delta, \mu) \equiv E_{F_V} R_N(\delta, \mu, V), \tag{11}$$

the unconditional normalized risk in (6). The dominance result with respect to $R(\delta, \mu)$ in (3) will then follow as a result of lemma 1.

The advantage in working with $R_N(\delta, \mu)$ is that it is easier to see what is driving the main result. Indeed, the normalized risk functions in Fig. 1 are just special cases of $R_N(\delta^v, \mu, v) = (1/v)R(\delta, \mu, v)$. As v is increased, the successive $R_N(\delta^v, \mu, v)$ are obtained by "stretching" $R_N(\delta^v, \mu, 1)$, a consequence of the fact that

$$R_N(\delta^v, \mu, v) = R_N(\delta^v, \mu/\sqrt{v}, 1). \tag{12}$$

Moreover, $R_N(\delta^{\text{MLE}}, \mu, v) \equiv 1$, which implies that $R(\delta^{\text{MLE}}, \mu) \equiv 1$. Thus, it suffices to show that for some F_V

$$R_N(\delta^v, \mu) = \int R_N(\delta^v, \mu, v)\, dF_V(v) < R(\delta^{\text{MLE}}, \mu) \equiv 1 \quad \text{for all } \mu. \tag{13}$$

We now proceed to construct F_V which satisfy (13). Our construction is facilitated by considering first, an F_V with discrete support. Based on this discrete F_V, we then obtain continuous F_V which also satisfy (13). For some $0 < r < 1$ and $s > 1$ (to be specified), a discrete F_V satisfying (13) has a discrete density of the form

$$P_V(s^i) = (1-r)r^i, \quad \text{for } i = 0, 1, \ldots. \tag{14}$$

where $[s^0, s^1, s^2, \ldots\}$ is the support of F_V. In order to satisfy (13), r and s must be chosen so that F_V puts enough weight on successive $R_N(\delta^v, \mu, v)$.

For example, it turns out that such an F_V is obtained by $r = 0.6$ and $s = 10$. For this choice, Fig. 2 displays successive conditional risks $E_k(\mu) \equiv E_{F_V}[R_N(\delta^v, \mu, V) \mid V \leqslant 10^k]$ for $k = 0, \ldots, 4$. Note that each $E_k < 1$ (the dotted line) in a region surrounding $\mu = 0$. This region becomes wider as k increases because in going from E_{k-1} to E_k, the "hump" of E_{k-1} is canceled out only to be replaced by a new hump of E_k which is smaller and further away from zero. Continuing in this manner yields $R_N(\delta^v, \mu) = \lim_{k\to\infty} E_k(\mu) < 1$ for all μ.

The choice of F_V in (14) is based on the following construction of an upper bound function for $[R_N(\delta^v, \mu, 1) - 1]$. This function, denoted $G(\mu)$, is defined by four positive constants A, B, a, b to be

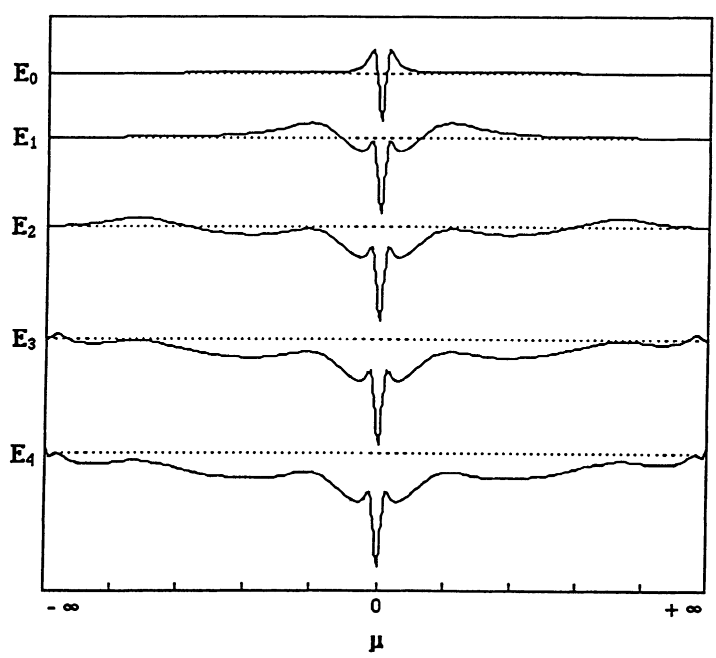

Fig. 2. Conditional risks $E_k(\mu) \equiv E_{F_V}[R_N(\delta^v, \mu, V) \mid V \leqslant 10^k]$ with $r = 0.6$ and $s = 10$ for $k = 0, \ldots, 4$. Each dotted line is 1.

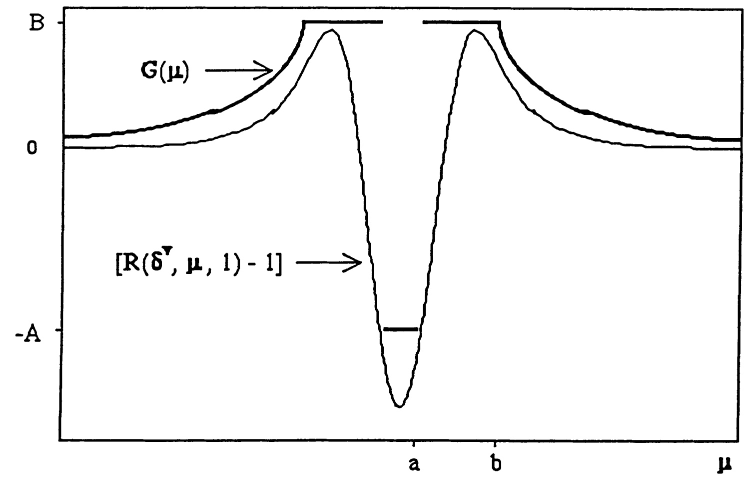

Fig. 3. $G(\mu)$ and $[R_N(\delta^v, \mu, 1) - 1]$.

$$G(\mu) = \begin{cases} -A & \text{for } |\mu| \leqslant a \\ B & \text{for } a < |\mu| \leqslant b \\ Bb^2/\mu^2 & \text{for } b < |\mu| < \infty \end{cases} \tag{15}$$

where these constants are chosen such that

$$[1 - R_N(\delta^v, 0, 1)] > A > B > \left[\sup_\mu R_N(\delta^v, \mu, 1) - 1\right] \tag{16}$$

and

$$G(\mu) > [R_N(\delta^v, \mu, 1) - 1] \quad \text{for all } \mu. \tag{17}$$

An example of such a choice is displayed in Fig. 3. Although it can be rigorously proved that such a $G(\mu)$ can be chosen using limits of approximations, this is tedious and we leave it to the interested reader. For practical purposes, these constants can be obtained using numerical methods. Analysis of $R_N(\delta^v, \mu, v)$ has also been carried out by Thompson (1968).

We need the following four constants r, ε, c and s specified in order. Choose r to satisfy

$$\frac{B}{A} < r < 1. \tag{18}$$

Now choose $\varepsilon > 0$ such that

$$r = \frac{B + 2\varepsilon}{A}. \tag{19}$$

Now choose $c > a$ to satisfy

$$G(\mu) = Bb^2/\mu^2 < \varepsilon r \quad \text{for all } \mu > c. \tag{20}$$

Finally, for the support set $\{s^0, s^1, s^2, \ldots\}$ of F_V, let s satisfy

$$s > c^2/a^2 \quad \text{and} \quad rs > 2. \tag{21}$$

Now define a sequence of functions,

$$G_i(\mu) \equiv G(\mu/s^{i/2}), \quad i = 0, 1, 2, \ldots \tag{22}$$

so that by (12) and (17),

$$G_i(\mu) > [R_N(\delta^v, \mu, s^i) - 1] \quad \text{for all } \mu, \tag{23}$$

and

$$G_i(\mu) = \begin{cases} -A & \text{for } |\mu| \leqslant as^{i/2} \\ B & \text{for } as^{i/2} < |\mu| \leqslant bs^{i/2} \\ Bb^2 s^i/\mu^2 & \text{for } bs^{i/2} < |\mu| < \infty \end{cases} \tag{24}$$

To prove our main result, we define

$$H_k(\mu) = (1-r) \sum_{i=0}^{k} r^i G_i(\mu). \tag{25}$$

Obviously, $\lim_{k\to\infty} H_k(\mu) \geqslant R_N(\delta^v, \mu) - 1$. Our final result, $R_N(\delta^v, \mu) < 1$ $(\equiv R_N(\delta^{\mathrm{MLE}}, \mu))$, is then obtained by showing that for all μ, $\lim_{k\to\infty} H_k(\mu) < 0$.

Lemma 2
For $|\mu| \leqslant as^{k/2}$, (i) $H_k(\mu) < H_{k-1}(\mu)$ and (ii) $H_k(\mu) < 0$.

Proof. (i) follows from the decomposition $H_k(\mu) = H_{k-1}(\mu) + (1-r)r^k G_k(\mu)$ and the fact that $G_k(\mu) < 0$ for $|\mu| \leqslant as^{k/2}$.

(ii) is shown by induction. Since $G_0(\mu) \equiv G(\mu)$, the assertion is true for $k = 0$. Now assume it is true for $k-1$. It then follows from (i) that $H_k(\mu) < 0$ for $|\mu| \leqslant as^{(k-1)/2}$. Thus it suffices to show

$$H_k(\mu) < 0 \quad \text{for} \quad as^{(k-1)/2} < |\mu| \leqslant as^{k/2}. \tag{26}$$

Consider the decomposition

$$H_k(\mu) = H_{k-2}(\mu) + (1-r)r^{k-1}[G_{k-1}(\mu) + rG_k(\mu)] \tag{27}$$

For $as^{(k-1)/2} < |\mu| \leqslant as^{k/2}$,

$$\begin{aligned} H_{k-2}(\mu) &= (1-r) \sum_{i=0}^{k-2} r^i G_i(\mu) < \varepsilon(1-r)r \sum_{i=0}^{k-2} (rs)^i / s^{(k-2)} \\ &= \varepsilon(1-r)r \frac{((rs)^{k-1} - 1)}{s^{k-2}(rs-1)} < \varepsilon(1-r)r^{k-1} \frac{rs}{rs-1} < 2\varepsilon(1-r)r^{k-1} \end{aligned} \tag{28}$$

where the first inequality above follows because for $\mu > as^{(k-1)/2} > cs^{(k-2)/2}$,

$$G_i(\mu) < \varepsilon rs^{i-(k-2)}, \quad \text{for } i = 0, \ldots, k-2 \tag{29}$$

which in turn follows from (20) and (24). Since $G_{k-1}(\mu) \leqslant B$, and $G_k(\mu) = -A$ for $|\mu| \leqslant as^{k/2}$, it follows using (19) that,

$$(1-r)r^{k-1}[G_{k-1}(\mu) + rG_k(\mu)] \leqslant -2\varepsilon(1-r)r^{k-1} \quad \text{for } |\mu| \leqslant as^{k/2} \tag{30}$$

Combining (27), (28) and (30) yields (26). □

Theorem 1
For r and s obtained by construction (15)–(21), *under discrete F_V with density*

$$p_V(s^i) = (1-r)r^i, \quad \text{for } i = 0, 1, \ldots, \tag{31}$$

δ^v dominates δ^{MLE} with respect to the unconditional normalized risk $R_N(\delta, \mu) = E_{F_V} R_N(\delta, \mu, V)$ in (6).

Proof. Because $s > 1$, it follows from lemma 2 that for all μ, the sequence $H_k(\mu)$ is eventually monotone decreasing. Further, once it starts decreasing, it starts out less than 0. Thus for all μ, $\lim_{k\to\infty} H_k(\mu) < 0$ which implies $R(\delta^v, \mu) < 1(\equiv R(\delta^{\text{MLE}}, \mu))$. □

Theorem 2

For r and s obtained by the construction (15)–(21), *under continuous F_V with density*

$$f_V(v) = (\alpha - 1)v^{-\alpha} I_{[v>1]}, \tag{32}$$

where $1 < \alpha < 1 - (\ln(1-r)/\ln s)$, δ^v *dominates* δ^{MLE} *with respect to the unconditional normalized risk* $R_N(\delta, \mu) = E_{F_V} R_N(\delta, \mu, V)$ *in* (6).

Proof. Let W be a random variable with density proportional to $w^{-\alpha} I_{[1 \leqslant w \leqslant s]}$. Pick r such that $\alpha < 1 - (\ln(1-r)/\ln s)$. Let $V_1 \sim p_V$ in (14) with parameters r and s. Let $V_2 = W \cdot V_1$ which then has density f_V in (32). Then under F_{V_2}

$$\begin{aligned} R_N(\delta, \mu) &\equiv E_{F_{V_2}} R_N(\delta, \mu, V_2) = E_{F_W} E_{F_{V_1}}[R_N(\delta, \mu, W \cdot V_1) \mid W] \\ &= E_{F_W} E_{F_{V_i}}[R_N(\delta, \mu/\sqrt{W}, V_1) \mid W] < 1 \end{aligned} \tag{33}$$

because $E_{F_{V_1}} R_N(\delta, \mu', V_1) < 1$ for all μ' by theorem 1. □

Theorem 3

For r and s obtained by the construction (15)–(21), *under discrete F_V with density*

$$p_V(s^i) = \left(1 - \frac{r}{s}\right)\left(\frac{r}{s}\right)^i, \quad \text{for } i = 0, 1, \ldots, \tag{34}$$

and under continuous F_V with density

$$f_V(v) = \alpha v^{-(\alpha+1)} I_{[v>1]}, \tag{35}$$

where $1 < \alpha < 1 - (\ln(1-r)/\ln s)$, δ^v *dominates* δ^{MLE} *with respect to the unconditional risk* $R(\delta, \mu) = E_{F_V} R(\delta, \mu, V)$ *in* (3)

Proof. Apply lemma 1 to theorems 1 and 2. □

More generally, it follows immediately from lemma 1 that δ^v may dominate δ^{MLE} with respect to risks of the form $R_g(\delta, \mu) \equiv E_{F_V} g(V) R(\delta, \mu, V)$ (for g satisfying certain regularity conditions). Note that such cases include the estimation of μ under bounded loss functions such as truncated squared-error loss.

Finally, we conclude with a graph which displays how the normalized risk $R_N(\delta^v, \mu)$ is obtained from the conditional risk $R_N(\delta^v, \mu, v)$ under the continuous distribution F_V in (32). This graph is based on the re-expression of the unconditional risk under this F_V as

$$R_N(\delta^v, \mu) = \int R_N(\delta^v, \mu, v)\, dF_V(v) = \int_0^1 R_N(\delta^v, \mu, w^{1/(1-\alpha)})\, dw \tag{36}$$

which follows from the fact that for $W \sim$ Uniform [0, 1], $W^{1/(1-\alpha)} \sim f_V(v) = (\alpha - 1)v^{-\alpha} I_{[v>1]}$ in (32). Figure 4 displays the surface and contour plot of $R_N(\delta^v, \mu, w^{1/(1-\alpha)}) = R_N(\delta^v, \mu w^{1/2(\alpha-1)}, 1)$ when $\alpha = 1.39$ ($< 1 - (\ln(1-r)/\ln s)$ for $r = 0.6$ and $s = 10$) over the range $-20 < \mu < 20$ and $0 < w < 1$. This graph shows how quickly the region of improvement is moved out towards $(-\infty, \infty)$. This region would move out even more rapidly for heavier tailed f_V obtained by α closer to 1.

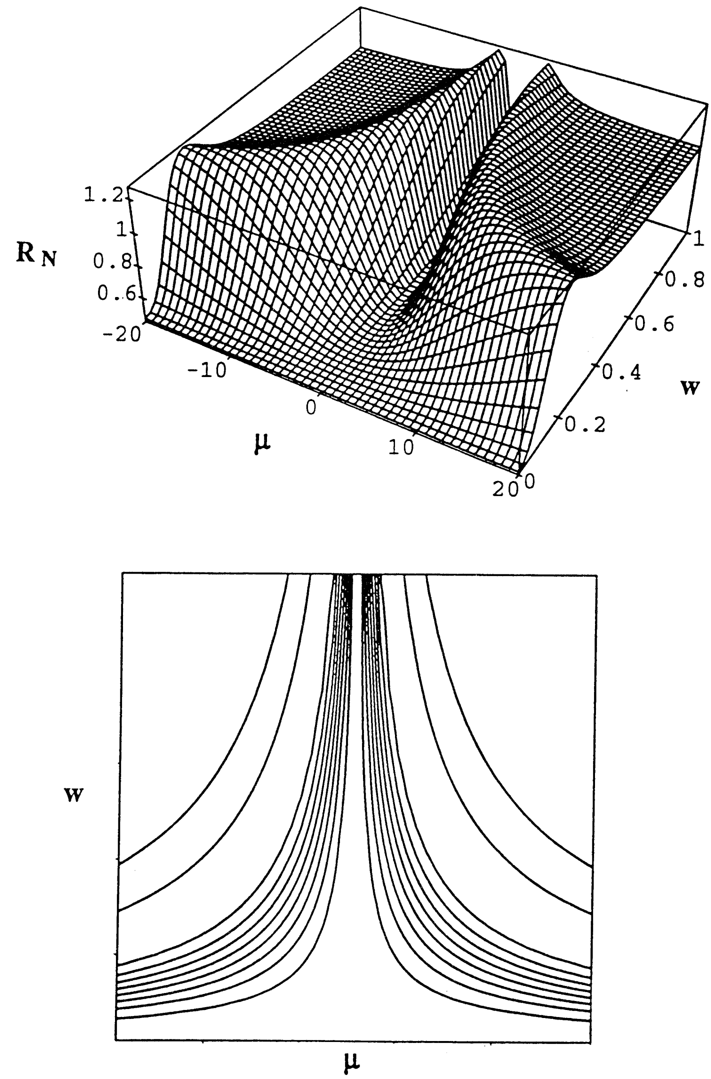

Fig. 4. Surface and contour plot of $R_N(\delta^v, \mu, w^{1/(1-\alpha)})$ when $\alpha = 1.39$.

Acknowledgement

The authors would like to thank Jim Berger, Larry Brown, George Casella, Christian Robert and Herman Rubin for helpful suggestions.

References

Berger, J. O. (1984). The frequentist viewpoint and conditioning. In *Proceedings of the Berkeley conference in honor of Kiefer and Neyman* (eds. L. LeCam & R. Olshen) Wadsworth, Belmont.

Berger, J. O. (1990). Discussion of "An ancillarity paradox which appears in multiple linear regression by L. D. Brown", *Ann. Statist.* **18,** 493–496.

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37,** 1087–1136.

Brown, L. D. (1988). The differential inequality of a statistical estimation problem. *Proceedings of the fourth Purdue symposium on statistical decision theory and related topics* (eds. S. S. Gupta & J. Berger), Springer-Verlag, New York.

Brown, L. D. (1990). An ancillarity paradox which appears in multiple linear regression (with discussion). *Ann. Statist.* **18,** 471–536.

Casella, G. C. (1990). Discussion of "An ancillarity paradox which appears in multiple linear regression by L. D. Brown". *Ann. Statist.* **18,** 497–499.

Cox, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29,** 357–372.

Fox, M. (1981). An inadmissible best invariant estimator: the i.i.d. case. *Ann. Statist.* **9,** 1127–1129.

Neyman, J. (1977). Frequentist probability and frequentist statistics. *Synthese* **36,** 97–131.

Perng, S. K. (1970). Inadmissibility of various "good" statistical procedures which are translation invariant. *Ann. Math. Statist.* **41,** 1131–1321.

Thompson, J. R. (1968). Some shrinkage techniques for estimating the mean. *J. Amer. Statist. Assoc.* **63,** 113–122.

Received October 1992, in final form March 1995

D. P. Foster, Statistics Department, University of Pennsylvania, 3009 SHDH, Wharton, Philadelphia, PA 19104-6302, USA.

E. I. George, Department MSIS, CBA 5.202, University of Texas, Austin, TX 78712-1175, USA.