it is unconditionally advantageous to use them. This is an example of Brown's phenomenon at the level of loss estimators.

For more general point estimators $\tilde{\delta}$ of the form (3.3.1), the Lemma indicates how one might apply existing work to construct reasonable loss estimators for $(\tilde{\delta} - \alpha)^2$. If one works conditionally on S, as in (3.3.3), then it is plausible that an improvement on the unbiased estimate of loss of $(\tilde{\delta} - \alpha)^2$ will follow as in Section 5 of J and an improvement on the upper bound $\sigma^2 + \sigma^2 \operatorname{tr} S^{-1}$ as in Lu and Berger (1989). Construction of loss estimates corresponding to (3.3.4) and (3.3.5) is less clear, but an interesting problem perhaps deserving further study.

REFERENCES

JOHNSTONE, I. (1988). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 361–379. Springer, New York.

LU, K. L. and BERGER, J. O. (1989). Estimated of normal means: Frequentist estimation of loss. *Ann. Statist.* **17** 890–906.

RUKHIN, A. (1988). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 409–420. Springer, New York.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

KUN-LIANG LU

University of Nebraska—Lincoln

The fundamental ancillarity paradox introduced by Brown can be observed in many other settings. As an example, we herein extend the results of Brown to the confidence set scenario:

Let X be a p-dimensional normal random variable with mean $\mu \in R^p$ and covariance matrix Σ. Consider the confidence procedure

$$C_\delta(X) = \{\mu: (\delta(X) - \mu)'\Sigma^{-1}(\delta(X) - \mu) \le c^2\},$$

where Σ^{-1} is an inverse or generalized inverse of Σ. The coverage probability of C_δ, $P_\mu(C_\delta(X)$ contains $\mu)$, is the usual criterion used for evaluating procedures of a fixed size (determined by c). It is convenient to rephrase this as a decision problem, with $\delta(X)$ being thought of as an estimator and $1 - P_\mu(C(X)$ contains $\mu)$ being the risk function corresponding to the loss function.

$$L_c(\mu, d) = \begin{cases} 1, & \text{if } (d - \mu)'\Sigma^{-1}(d - \mu) \ge c^2, \\ 0, & \text{otherwise.} \end{cases}$$

Brown (1966) and Joshi (1969) independently showed that $\delta_0(X) = X$ is admissible if $p = 1, 2$ and inadmissible if $p \ge 3$. Hwang and Casella (1982, 1984) proved that the positive part James–Stein estimator is an improved estimator under the above loss L_c.

Suppose, following Brown, that $w \in R^p$ is a fixed vector and define

$$\theta = \sum_{i=1}^{p} w_i \mu_i = w'\mu.$$

It is obvious that the confidence set for θ given by

$$C_{\delta_0}(X) = \{\theta = w'\mu : |w'X - w'\mu| \le c\}$$

cannot be improved, since $w'X \sim N_1(\theta, w'\Sigma w)$, a one-dimensional problem.

Now assume that the values of $(w_1, \dots, w_p)$ are observed coordinate values of a random variable, $W \in R^p$ and define

$$\Omega = E(WW').$$

The customary confidence set for θ in this problem is still

$$C_0(x, w) = \{\theta = w'x : |w'x - w'\mu| \le c\},$$

but as in Brown's Section 2.1 we can establish:

THEOREM 1. *Let X, W be independent. Suppose $W \sim N_p(0, \Omega)$ and Ω is nonsingular and $p \ge 3$. Then the confidence set $C_0(x, w)$ can be improved in terms of coverage probability* [*i.e.*, *$\delta_0(X, W)$ is inadmissible as an estimator under loss L_c*].

PROOF. Let $\delta(X)$ be any estimator of μ:

$$\begin{aligned}
\Pr(|W'\delta(X) - W'\mu| \ge c) &= E_\mu^X \Pr(|W'\delta(X) - W'\mu| \ge c \mid X) \\
&= E_\mu^X \Pr(|(\delta(X) - \mu)'W| \ge c \mid X) \\
&= E_\mu^X \Pr\left(\frac{|(\delta(X) - \mu)'W|}{[(\delta(X) - \mu)'\Omega(\delta(X) - \mu)]^{1/2}} \right. \\
&\qquad \left. \ge \frac{c}{[(\delta(X) - \mu)'\Omega(\delta(X) - \mu)]^{1/2}} \,\middle|\, X \right) \\
&= E_\mu^X 2\left[1 - \Phi\left(\frac{c}{[(\delta(X) - \mu)'\Omega(\delta(X) - \mu)]^{1/2}}\right)\right] \\
&= E_\mu^X L_1(\|\delta(X) - \mu\|_\Omega),
\end{aligned}$$

where $\Phi(x)$ is the normal distribution function,

$$L_1(t) = 2\left[1 - \Phi\left(\frac{c}{t}\right)\right]$$

and

$$\|x\|_\Omega = (x'\Omega x)^{1/2}, \qquad x \in R^p.$$

Brown (1966) found sufficient conditions on the loss such that there exists an estimator $\delta(X)$, which is better than $\delta_0(X) = X$ (Theorem 3.3.1 in Brown's paper). Since the loss function $W(d - \mu) = L_1(\|d - \mu\|_\Omega)$ is bounded and sufficiently smooth, the conditions of Theorem 3.3.1 are satisfied. The theorem implies that there are constants $a > 0$ and a constant matrix B such that

$$\delta^*(X) = \big(I + B/(a + \|X\|^2)\big)X$$

is a better estimator. Therefore, the confidence set

$$C^*(X) = \{\theta = W'\mu : |W'\delta^*(X) - W'\mu| \le c\}$$

has larger coverage probability than C_0. □

Next, consider the usual normal multiple linear regression as in Section 3 of Brown. Consider the confidence set

$$C_0(Y, V) = \{\alpha : |\delta_0 - \alpha| \le c\}$$

and the alternate confidence set

$$C_\delta(Y, V) = \{\alpha : |\delta(Y, V) - \alpha| \le c\},$$

where

$$\delta(Y, V) = \bar{Y} - \bar{V}\tilde{\beta}(\hat{\beta}, S) = \alpha + \bar{V}\big(\hat{\beta} - \tilde{\beta}(\hat{\beta}, S)\big).$$

THEOREM 2. *Suppose $V_1, \dots, V_n$ are i.i.d. with common distribution $N_r(0, I)$ and for given $V, Y \sim N_n(1\alpha + V\beta, \sigma^2 I)$, σ^2 is known. Then the confidence set C_0 can be improved, in terms of coverage probability.*

PROOF. Let $\theta = \alpha + \bar{V}\beta = E(\bar{Y})$. Note that $\bar{Y}$ is conditionally independent of $\hat{\beta}$, S, given $\bar{V}$, and $\bar{V}$ is independent of $\hat{\beta}$, S. Hence

$$\begin{aligned}
\Pr(\alpha \in C_\delta(Y, V)) &= \Pr\big(|\delta(Y, V) - \alpha| \le c\big) \\
&= E_{\alpha,\beta} \Pr\big(\big|\bar{Y} - \theta - \bar{V}(\tilde{\beta} - \beta)\big| \le c \big| \hat{\beta}, S, \bar{V}\big) \\
&= E_{\alpha,\beta}\big(\Phi\big(\sqrt{n}\,\big[\bar{V}(\tilde{\beta} - \beta) + c\big]\big) - \Phi\big(\sqrt{n}\,\big[\bar{V}(\tilde{\beta} - \beta) - c\big]\big)\big) \\
&= E_{\alpha,\beta} E\big[\Phi\big(\sqrt{n}\,\big[\bar{V}(\tilde{\beta} - \beta) + c\big]\big) \\
&\qquad - \Phi\big(\sqrt{n}\,\big[\bar{V}(\tilde{\beta} - \beta) - c\big]\big) \big| \hat{\beta}, S\big] \\
&= E_{\alpha,\beta} \int_{-\infty}^{\infty} \big(\Phi(\|\tilde{\beta} - \beta\| u + \sqrt{n}\, c) \\
&\qquad - \Phi(\|\tilde{\beta} - \beta\| u - \sqrt{n}\, c)\big)\phi(u)\, du \\
&= E_{\alpha,\beta} \int_{-\infty}^{\infty} \big(2\Phi(\|\tilde{\beta} - \beta\| u + \sqrt{n}\, c) - 1\big)\phi(u)\, du \\
&= E_{\alpha,\beta}\big(1 - L_2(\|\tilde{\beta} - \beta\|)\big),
\end{aligned}$$

where

$$L_2(x) = \int_{-\infty}^{\infty} 2\big(1 - \Phi(xu + \sqrt{n}\,c)\big)\phi(u)\,du.$$

Consider $\hat{\beta} \sim N_r(\beta, S^{-1})$, given S. Then we want to find an estimator $\tilde{\beta} = \tilde{\beta}(\hat{\beta}, S)$, given S, such that

$$E_\beta\big[L_2(\|\tilde{\beta} - \beta\|)|S\big] < E_\beta\big[L_2(\|\hat{\beta} - \beta\|)|S\big].$$

This again follows from Theorem 3.3.1 of Brown (1966). □

REFERENCES

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

JOSHI, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40** 1042–1067.

HWANG, J. and CASELLA, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10** 868–881.

HWANG, J. and CASELLA, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Decisions* **1** 3–16.

DEPARTMENT OF MATHEMATICS AND STATISTICS
925 OLDFATHER HALL
UNIVERSITY OF NEBRASKA—LINCOLN
LINCOLN, NEBRASKA 68588-0323

C. N. MORRIS[1]

University of Texas

1. Conditionality. A paradox is a self-contradictory statement, and a paradox in science demands resolution. The discovery of each new paradox creates an opportunity for a new growth and deeper understanding as we seek explanation.

Professor Brown's paradox is that conditionality is at odds with unconditional admissibility. While his concluding remarks do not resolve the paradox, he seems to take sides by insisting that we account for "the unconditional frequentist structure of the situation." I see it differently, and argue for being as conditional as possible in making statistical inferences.

It can happen, and did in Brown's example, that decision rules 1 and 2 with risks R_1 and R_2 obey $R_1 < R_2$ uniformly in the parameters when we average over an ancillary $\mathbf{V}$, but that the conditional risks, given $\mathbf{V}$, satisfy $R_1(\mathbf{V}) > R_2(\mathbf{V})$ for some parameters and some values of $\mathbf{V}$. If $\mathbf{V}$ occurs and is observed and it happens to be a value for which $R_1(\mathbf{V}) > R_2(\mathbf{V})$, then rule 2 is better for that $\mathbf{V}$. It matters not at all that for most $\mathbf{V}$, $R_1(\mathbf{V}) < R_2(\mathbf{V})$. Brown's example is less clear. We do not know for the observed $\mathbf{V}$ which $R(\mathbf{V})$ is smaller,

[1]Research supported by NSF Grant DMS-89-11562.