HILL, B. M. (1980b). Robust analysis of the random model and weighted least squares regression. In *Evaluation of Econometric Models* (J. Kmenta and J. Ramsey, eds.) 197–217. Academic, New York.

HILL, B. M. (1988). On the validity of the likelihood principle. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 119–132. Springer, Berlin.

HILL, B. M. (1990). A theory of Bayesian data analysis. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of George Barnard* (S. Geisser, J. Hodges, S. J. Press and A. Zellner, eds.) 383–395. North-Holland, Amsterdam.

HILL, B. M. and LANE, D. (1985). Conglomerability and countable additivity. *Sankhyā Ser. A* **47** 366–379.

KOLMOGOROV, A. N. (1950). *Foundations of Probability*. Chelsea, New York.

LINDLEY, D. and SMITH, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc., Ser. B* **34** 1–41.

RAMAKRISHNAN, S. and SUDDERTH, W. (1988). A sequence of coin-toss variables for which the strong law fails. *Amer. Math. Monthly* **95** 939–941.

RUSSELL, B. (1937). *The Principles of Mathematics*. Allen and Unwin, London.

SAVAGE, L. J. (1972). *The Foundations of Statistics*, 2nd rev. ed. Dover, New York.

SCHERVISH, M., SEIDENFELD, T. and KADANE, J. (1984). The extent of nonconglomerability. *Z. Wahrsch. Verw. Gebiete* **66** 205–226.

SCOZZAFAVA, R. (1984). A survey of some common misunderstandings concerning the role and meaning of finitely additive probabilities in statistical inference. *Statistica* **44** 21–45.

WHITROW, G. J. (1980). *The Natural Philosophy of Time*, 2nd ed. Oxford Univ. Press.

THE UNIVERSITY OF MICHIGAN
DEPARTMENT OF STATISTICS
ANN ARBOR, MICHIGAN 48109

IAIN M. JOHNSTONE[1]

Stanford University

In Section 5 of his thought-provoking paper, Professor Brown discusses Cox's ancillarity example and draws a distinction between point estimation and confidence procedures. He argues for the conditional validity of his proposed point estimation procedures, since in point estimation no conditionally interpretable stochastic claim is made.

It is, however, possible to make a data-dependent stochastic statement concerning a point estimate without going so far as to provide a confidence set. This may be done by estimating the (squared) error $(\tilde{\alpha} - \alpha)^2$. The issue has been considered in various point estimation settings recently by Rukhin (1988), Lu and Berger (1989) and Johnstone (1988) (the last hereafter denoted by J). Here I shall indicate briefly how some of these ideas extend to Brown's context.

In the setting and notation of Section 3, let $L = L(\{V_i\}, \{Y_i\})$ be an estimate of the squared error $(\delta - \alpha)^2$ of a point estimator $\delta = \delta(\{V_i\}, \{Y_i\})$. The quality of L may be evaluated in turn by using (for simplicity) a quadratic loss $E[L - (\delta - \alpha)^2]^2$, where the expectation is taken over the joint distribution of (V_i, Y_i).

[1]Research supported in part by NSF Grants DMS 84-51750 and 86-00235.

In Section 3, attention is restricted to estimators of the form (3.3.1), $\tilde{\delta} = \overline{Y} - \overline{V}\tilde{\beta}(\hat{\beta}, S)$, and Lemma 3.3.1 shows the close connection for these estimators of estimation of α with estimation of β under ordinary quadratic loss $\|\tilde{\beta} - \beta\|^2$. Consider loss estimators of $n(\tilde{\delta} - \alpha)^2$ which have the form $\tilde{L} = \sigma^2 + \psi(\hat{\beta}, S)$. This form is motivated by the identity (contained in the proof of Lemma 3.3.1)

$$En(\tilde{\delta} - \alpha)^2 = \sigma^2 + E\|\tilde{\beta} - \beta\|^2.$$

The following analogue of Lemma 3.3.1 says that admissibility of $\tilde{L}$ within the class of estimators of the same form is equivalent to admissibility of $\psi(\hat{\beta}, S)$ as an estimator of $\|\tilde{\beta} - \beta\|^2$ under quadratic loss.

LEMMA. *Let $\tilde{\delta}$ and $\tilde{\beta}$ be related as in* (3.3.1). *If $\psi_i = \psi_i(\hat{\beta}, S)$, then*

$$\begin{aligned} E\left[\sigma^2 + \psi_1 - n(\tilde{\delta} - \alpha)^2\right]^2 - E\left[\sigma^2 + \psi_2 - n(\tilde{\delta} - \alpha)^2\right]^2 \\ = E\{\psi_1 - \|\tilde{\beta} - \beta\|^2\}^2 - E\{\psi_2 - \|\tilde{\beta} - \beta\|^2\}^2. \end{aligned} \tag{1}$$

The verification of (1), which uses the same methods as the proof of Lemma 3.3.1, is omitted.

Consider first $\tilde{\delta} = \hat{\alpha}$ and estimation of $L_0 = n(\hat{\alpha} - \alpha)^2$. Conditionally on S, $\hat{\beta} \sim N(\beta, \sigma^2 S^{-1})$ and $E[\|\hat{\beta} - \beta\|^2 | S] = \sigma^2 \operatorname{tr} S^{-1}$. An unbiased estimate for the risk of $\sigma^2(\operatorname{tr} S^{-1} - \tilde{\gamma}(\hat{\beta}))$ as an estimator of $\|\hat{\beta} - \beta\|^2$ may be obtained exactly as in Proposition 2.2 of J. Let $X = \sigma^{-1} S^{1/2} \hat{\beta} \sim N(\theta, I)$, where $\theta = \sigma^{-1} S^{1/2} \beta$ and set $\gamma(X) = \tilde{\gamma}(\hat{\beta})$. Denote the Hessian matrix $(D_{ij}\gamma)$ by $H(\gamma)$. Then

$$\begin{aligned} E\{\sigma^2[\operatorname{tr} S^{-1} - \tilde{\gamma}(\hat{\beta})] - \|\hat{\beta} - \beta\|^2\}^2 \\ = 2\sigma^4 \operatorname{tr} S^{-2} + \sigma^2 E[2 \operatorname{tr} S^{-1} H(\gamma)(X) + \gamma^2(X)]. \end{aligned} \tag{2}$$

A "completely unconditional" estimator of L_0 is given by its expectation $EL_0 = \sigma^2 + \sigma^2 E \operatorname{tr} S^{-1} = \sigma^2 + \sigma^2/(n - r - 2)$. However, this is already dominated by a conditionally (given S) unbiased estimator $\hat{L}_0 = \sigma^2 + \sigma^2 \operatorname{tr} S^{-1}$. This may be seen directly or via (1) and (2) from which $E(EL_0 - L_0)^2 - E(\hat{L}_0 - L_0)^2 = \sigma^2 E[\operatorname{tr} S^{-1} - (n - r - 2)^{-1}]^2$.

More significantly, an improvement of $\hat{L}_0$ may be obtained by solving the differential inequality $2 \operatorname{tr} S^{-1} H(\gamma) + \gamma^2 \le 0$. Again following J, an ad hoc, but convenient, choice is $\gamma(x) = c/(x^t S x)$. For the corresponding estimator $\tilde{L}_0 = \hat{L}_0 - \sigma^2 \tilde{\gamma}(\hat{\beta})$, this leads to

$$E(\tilde{L}_0 - L_0)^2 - E(\hat{L}_0 - L_0)^2 = -2c(p - 4)\sigma^2 E \frac{\sigma^4}{(\hat{\beta}^t S^2 \hat{\beta})^2},$$

which is minimised by the choice $c = 2(p - 4)$. The estimator $\tilde{L}$ [or better, $\max(\tilde{L}_0, 0)$], while not yet completely satisfactory for practice, at least demonstrates that data-dependent measures of the error of $\hat{\alpha}$ are available and that 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.