GEORGE CASELLA[1]

Cornell University

In his customary, penetrating way, Professor Brown has discovered and illuminated a fascinating admissibility paradox. This paradox brings together the elusive concept of ancillarity with the (still somewhat puzzling) Stein phenomenon and, through this synthesis, perhaps explains both a little better. The goal in this discussion is to understand and explain Brown's admissibility paradox in a simple intuitive way.

Using the notation of Section 3, we observe $Y_{n\times 1}$ and $V_{n\times r}$, where $EY = \alpha 1 + V\beta$, and we want to estimate α using an estimator that is a function of Y and V, say $d(Y, V)$. The loss function given by (3.1.2) is squared error loss

$$L(\alpha, d) = (\alpha - d)^2. \tag{1}$$

In a regression problem, we estimate α based on observing values $Y = y$ and $V = v$. Brown's paradox asserts that the admissibility of $\hat{\alpha}$, the least squares estimator, depends on whether V is treated as constant or as a realized value of an ancillary random variable.

An important distinction between the two problems lies in the risk functions: Although the loss function remains the same, the risk function changes depending on whether we consider the matrix V to be fixed or random. If V is fixed, then the risk of estimating α is conditional on the value $V = v$, that is,

$$R(\alpha, d|V = v) = E\Big[(\alpha - d(Y, v))^2\Big|V = v\Big]. \tag{2}$$

Here the expectation is over the distribution of Y given $V = v$ which, of course, depends on α. If V is considered a random variable, then the risk of estimating α is unconditional on the value $V = v$, that is,

$$R(\alpha, d) = \int R(\alpha, d|V = v)\, f_V(v)\, dv, \tag{3}$$

where $f_V(\cdot)$ denotes the density of V.

Keeping the risk relationship (3) in mind, we can now reexamine the admissibility/inadmissibility results of Proposition 3.1.1 and Theorem 3.2.2 (or their predecessors, Proposition 2.1.1 and Theorem 2.1.2). The admissibility results relate to the risk function $R(\alpha, d|V = v)$ of (2), while the inadmissibility results relate to the risk function $R(\alpha, d)$ of (3). Furthermore, the relationship in (3) amplifies the paradoxical nature of Brown's results. Note that from (3) we immediately get the implication that if an estimator $d(Y, v)$ is dominated for every v by $d^*(Y, v)$ using $R(\alpha, d|V = v)$ of (2), it is inadmissible under $R(\alpha, d(Y, V))$ of (3). But this does not happen for $d(Y, v) = \hat{\alpha}$. Since the least squares estimator is admissible under $R(\alpha, \hat{\alpha}|V = v)$, this implies that it cannot be dominated in risk for every v by the same estimator.

[1]Research supported by NSF Grant No. DMS-89-0039.

We therefore see that the inadmissibility of α under $R(\alpha, \hat{\alpha})$ is really tied to the distribution of V. Since the estimator $\hat{\alpha}$ cannot be dominated for each $V = v$, the domination is through the integration against $f_V(v)$. The dominating estimators (such as those in Section 3.3) must improve greatly for values of v yielding large values of $f_V(v)$, for the risk must surely be larger than that of $\hat{\alpha}$ for other values of v. The distribution of the ancillary statistic is central to the dominance result. (Consider an analogous, although different, occurrence in the classic Stein problem of estimating a multivariate normal mean. Although the usual estimator can be dominated using sum-of-squared-errors loss, it cannot be dominated componentwise. Thus, on some loss components the improved estimators must offer a large improvement, since they surely will lose on other components.)

The influence of the ancillary statistic V, or of any conditioning statistic, can also be seen through the following decomposition. For any estimator $d(Y, V)$, the risk given by (3), $R(\alpha, d) = E[\alpha - d(Y, V)]^2$, can be decomposed into

$$(4) \quad E[\alpha - d(Y,V)]^2 = E[\alpha - E(d(Y,V)|V)]^2 + E[\mathrm{Var}(d(Y,V)|V)].$$

Although this decomposition works for any conditioning statistic, risk improvement will obtain only if the statistic behaves in a certain way. In order to improve on the unconditional risk [the left-hand side of (4)], knowing that the conditional risk (the first term on the right-hand side) cannot always be improved, the variance of the estimator, with respect to the conditioning statistic, must behave correctly. For the estimator d to dominate $\hat{\alpha}$ we must have

$$(5) \quad \begin{aligned} 0 &\le R(\alpha, \hat{\alpha}) - R(\alpha, d) \\ &= E\left[(\alpha - E(\hat{\alpha}|V))^2 - (\alpha - E(d|V))^2\right] + E[\mathrm{Var}(\hat{\alpha}|V) - \mathrm{Var}(d|V)], \end{aligned}$$

so dominance requires that the statistic V imparts the correct influence on the variance (precision) of the estimator d.

Another remark of Brown's becomes clearer when seen in the light of (3) together with the admissibility of $\hat{\alpha}$ using $R(\alpha, d|V = v)$. In Remark 4.3.2 it is conjectured that $\hat{\alpha}$ remains admissible if the mean of V is unknown, but is inadmissible if this mean can be independently estimated. From (3) we know that $\hat{\alpha}$ is only inadmissible when the risk function $R(\alpha, \hat{\alpha}|V = v)$ is integrated against $f_V(v)$. Hence, this distribution (or some independent estimate of it) must be known for the inadmissibility result to hold. In the case considered here $f_V(v)$ is a member of a parametric family, so only the parameters need to be known.

Another explanation of Brown's ancillarity paradox, one that is more closely tied to Stein estimation, emerged from discussions with David Lansky, a Cornell graduate student. If we look more closely at Brown's Lemma 3.3.1, we can write [using the definition of $\tilde{\delta}$ and $\tilde{\beta}$ in (3.3.1)]

$$(6) \quad \begin{aligned} &E(\alpha - \hat{\alpha})^2 - E(\alpha - \tilde{\delta})^2 \\ &\quad = E\left[(\hat{\beta} - \beta)'\bar{V}'\bar{V}(\hat{\beta} - \beta)\right] - E\left[(\tilde{\beta} - \beta)'\bar{V}'\bar{V}(\tilde{\beta} - \beta)\right]. \end{aligned}$$

Notice that (6) is valid whether $\bar{V}$ is fixed or random, and it links the dominance of $\hat{\alpha}$ by $\tilde{\delta}$ under squared error loss to the dominance of $\hat{\beta}$ by $\tilde{\beta}$ under weighted squared error loss.

If $\bar{V}$ is random, we can further simplify the expectation on the right-hand side of (6) and obtain Lemma 3.3.1. If $\bar{V}$ is fixed, however, the term $\bar{V}'\bar{V}$ cannot be removed from the right-hand side of (6). Furthermore, the matrix $\bar{V}'\bar{V}$ is of rank 1, making the right-hand side of (6) equivalent to a one-dimensional risk difference. In such a case it is well known that $\hat{\beta}$ is admissible. Thus, in the case of fixed $\bar{V}$ the estimation of α is a one-dimensional problem, but if $\bar{V}$ is random the assumption that $E(\bar{V}'\bar{V}) = I/n$ turns the estimation of α into a multivariate problem (and leads to the inadmissibility results). Although this argument is restricted to the class of estimators given in (3.3.1), it clearly shows the influence of V on the admissibility of α.

Perhaps one of the most important effects of Brown's work will be the rethinking of the notion of ancillarity. Fisher, in his wonderfully vague way, left us with this idea of ancillarity, but not with any unequivocal definition. [The article by Buehler (1982) ably demonstrates this.] The main problem seems to be that ancillarity is defined by a marginal distribution, $f_V(v)$, when the more useful definition would be in terms of a conditional distribution, $f_{Y|V}(y|v)$. This idea may be closer to how Fisher thought of ancillaries, for he was most interested in the case where the pair (Y, V) is a sufficient statistic (but Y alone is not), and the marginal distribution of V does not depend on the parameter of interest. In such cases Fisher recommended using $f_{Y|V}(y|v)$ for inference. The major reason for this, which bears on the reasoning behind (4) and (5), is that the ancillary statistic contains information about precision. Fisher (1936) states "Ancillary information never modifies the value of our estimate; it determines its precision."

Connecting ancillarity with precision is a subtly different notion from that expressed by Brown in the first paragraph of his introduction. The quote from Savage, in Section 5, is closer to this sentiment, since the values of the independent variable have a direct impact on precision. In Brown's examination of the influence of ancillary statistics, he has substituted "admissibility" for "precision", bringing new understanding to this elusive concept. [Some problems with the influence of ancillarity on precision are discussed by Basu (1964, 1981).]

Rather than classify a statistic as ancillary or otherwise, a more useful categorization may be whether the statistic can influence an inference. This notion can also be traced back to Fisher and is made mathematically precise by Robinson (1979). Although Robinson is mainly concerned with conditional inference from confidence statements, conditional evaluations can also be tied into admissibility of point estimators. A decomposition such as (4) shows that the usual definition of ancillarity is not strong enough to be of use in decision theory. We need to take into account the conditional distribution of Y given V, and the effect that V has on the precision of our estimator. This consideration is, perhaps, equivalent to what Brown means in the last sentence of his article.

REFERENCES

BASU, D. (1964). Recovery of ancillary information. In *Contributions to Statistics* 7–29. Pergammon Press, Oxford. Republished in *Sankhyā* **26** 3–16 (1964). Also in *Statistical Information and Likelihood, A Collection of Critical Essays by Dr. D. Basu* (J. K. Ghosh, ed.). Springer, New York (1988).

BASU, D. (1981). On ancillary statistics, pivotal quantities, and confidence statements. In *Topics in Applied Statistics* (Y. P. Chaubey and T. D. Disivedi, eds.) 1–20. Concordia University, Montreal. Also in *Statistical Information and Likelihood, A Collection of Critical Essays by Dr. D. Basu* (J. K. Ghosh, ed.). Springer, New York (1988).

BUEHLER, R. (1982). Some ancillary statistics and their properties (with discussion). *J. Amer. Statist Assoc.* **77** 581–593.

FISHER, R. A. (1936). Uncertain inference. *Proc. Amer. Acad. Arts Sci.* **71** 245–258.

ROBINSON, G. K. (1979). Condition properties of statistical procedures. *Ann. Statist.* **7** 742–755.

BIOMETRICS UNIT
337 WARREN HALL
CORNELL UNIVERSITY
ITHACA, NY 14853

J. B. COPAS

University of Birmingham

The paper gives a counterexample against two common assumptions about shrinkage estimation: First, that shrinkage applies only to vector parameters with a single loss function allowing exchangeability of errors between component estimates; second, that shrinkage estimates are necessarily biased. In this important paper Dr. Brown provides a shrinkage estimate which is both scalar and unbiased. Particularly interesting and surprising is the finding that improved estimation of α is possible only when the mean of the V_i's is known. Are there intuitive grounds for expecting this to be so?

In Section 3 of the paper, suppose that the values of the V_i's are temporarily lost. Can anything then be said about α? Usually the answer is no, but if we have the additional information that the V_i's are *known* to have mean zero, then $\hat{\alpha}_0 = \bar{Y}$ becomes a (globally) unbiased estimate. With the full data $\hat{\alpha}$ is available as a second unbiased estimate, suggesting that we could do even better with the combined unbiased estimate

$$\hat{\alpha}(\lambda_1) = (1 - \lambda_1)\hat{\alpha}_0 + \lambda_1\hat{\alpha}.$$

Evaluating global moments (averaging over Y as well as V) we obtain the variances of $\hat{\alpha}_0$ and $\hat{\alpha}$ as $n^{-1}(\sigma^2 + \beta'\beta)$ and $n^{-1}\sigma^2(1 + r/(n - r - 2))$, respectively, with covariance $n^{-1}\sigma^2$. The variance of $\hat{\alpha}(\lambda_1)$ is therefore minimized when

$$\lambda_1 = \frac{\beta'\beta(n - r - 2)}{\beta'\beta(n - r - 2) + r\sigma^2}.$$