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

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

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.

because these risks depend also on the unknown parameter $\boldsymbol{\beta}$, even though we have the weaker result that $R_1 = E[R_1(\mathbf{V})] < R_2 = E[R_2(\mathbf{V})]$ for all $\boldsymbol{\beta}$. The inequality would be more convincing if expectations were over the distribution of likely $\boldsymbol{\beta}$ values rather than the observed $\mathbf{V}$, but one might be hard pressed to find such information. Thus, it seems appropriate to be as conditional as possible, in this case using the distribution of $\mathbf{Y}$, given $\mathbf{V}$, but not $\mathbf{V}$ to make the inference.

In Section 3.4, the unconditional risk of the standard unbiased estimator δ_0 of α exceeds that of any of the three "improvements" by a fixed amount at $\boldsymbol{\beta} = \mathbf{0}$. However, when $\overline{\mathbf{V}} = \mathbf{0}$ is observed, all three estimators equal $\overline{Y}$ exactly. Furthermore, $\delta_0 = \overline{Y}$ has conditional risk σ^2/n in this case, which is less than the conditional or unconditional risks of the other three estimators. Which measure of risk is appropriate in this case $\overline{\mathbf{V}} = \mathbf{0}$ and which estimator should be preferred? Unconditionally we are saying δ_0 has largest risk. Conditionally, $\delta_0 = \overline{Y}$ has the smallest risk. Actually, all estimators are numerically equal to those in Section 3.4. I find this disturbing.

Use of the unconditional risk when $\overline{\mathbf{V}} = \mathbf{0}$ requires knowing which rule the statistician would have used if $\overline{\mathbf{V}}$ had been other than $\mathbf{0}$. We may not know this. (Of course, $\overline{\mathbf{V}} = \mathbf{0}$ occurs with probability 0, in theory at least, but it can happen in practice, or nearly so. I think this consideration does not weaken the case being made.)

2. Simplified special case of the paradox. The generality of Brown's paper yields the broad conclusions he seeks, but I find the paradox itself to be more easily stated in the simpler setting of Section 3.3 with $r \geq 3$, and we can simplify much further, to increase understanding. We may suppose that $\mathbf{S}$ is predetermined, letting only $\overline{\mathbf{V}}$ be random, imagining that somehow $\mathbf{V}$ is only partially random. This happens, for example, if the column vectors of $\mathbf{V}$ can be controlled by the statistician, except for an error term (the ith component of $\overline{\mathbf{V}}$) which is additive to the column. Many technicalities of Section 3 are avoided in this fashion without altering the concept of the paradox. Let us further assume $\mathbf{S} = c\mathbf{I}$ is diagonal, $c > 0$ known. Then the mean squared error of $\alpha^* = \overline{Y} - \overline{\mathbf{V}}\boldsymbol{\beta}^*$ is

$$E(\alpha^* - \alpha)^2 = E\big[(\overline{Y} - \alpha - \overline{\mathbf{V}}\boldsymbol{\beta}) - \overline{\mathbf{V}}(\boldsymbol{\beta}^* - \boldsymbol{\beta})\big]^2 = \frac{\sigma^2}{n} + E\big[\overline{\mathbf{V}}(\boldsymbol{\beta}^* - \boldsymbol{\beta})\big]^2, \tag{1}$$

with $\boldsymbol{\beta}^*$ depending only on

$$\hat{\boldsymbol{\beta}} \sim N_r\left(\boldsymbol{\beta}, \frac{\sigma^2}{c}\mathbf{I}\right). \tag{2}$$

Brown has $\overline{\mathbf{V}} \sim N_r(\mathbf{0}, (1/n)\mathbf{I})$, independently of $\hat{\boldsymbol{\beta}}$. Then the risk (1) is

$$\frac{\sigma^2}{n} + \frac{1}{n}E\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|^2. \tag{3}$$

This (3) is $(1 + r/c)\,\sigma^2/n$ when $\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}$. The squared error loss (3) and distribution (2) for $\hat{\boldsymbol{\beta}}$ suggest choosing $\boldsymbol{\beta}^*$ to be Stein's estimator and leads to risk dominance and Brown's paradox.

With fixed $\overline{\mathbf{V}}$, and if $\boldsymbol{\beta}^*$ is Stein's estimator, the latter term in (1) is obtained from the component risk for Stein's rule. The risk (1) may be computed from component risk formulae of Baranchik (1964) and Efron and Morris (1972). It is

$$E(\alpha^* - \alpha)^2 = \frac{\sigma^2}{n} + \|\overline{\mathbf{V}}\|^2 \frac{\sigma^2}{n} R \tag{4}$$

and R is the component risk for Stein's rule

$$R = R_s + 2(r^2 - 4)\left(p - \frac{1}{r}\right) E \frac{2J}{(r + 2J)(r - 2 + 2J)}, \tag{5}$$

where J is Poisson with mean

$$\lambda = \frac{c\|\boldsymbol{\beta}\|^2}{2\sigma^2}, \qquad p = \frac{(\overline{\mathbf{V}}\boldsymbol{\beta})^2}{\|\overline{\mathbf{V}}\|^2\|\boldsymbol{\beta}\|^2}$$

and R_s is the average risk per component of Stein's estimator, having maximum of 1. The maximum of R in (4) and (5) is about $(r + 2)/4$, occurring when $\overline{\mathbf{V}}$ and $\boldsymbol{\beta}$ are collinear (when $p = 1$), and when 2λ is near r. Numerical values are in Efron and Morris (1972, Section 5).

If $p = 1/r$, (4) is given by Stein's risk. But for $p = 1$, the risk (5) starts above σ^2/n when $\boldsymbol{\beta} = \mathbf{0}$ and increases to a value exceeding (substantially, for large r) the risk of the MLE $\hat{\alpha}$ and then diminishes as $\boldsymbol{\beta} \to \infty$.

Perhaps this risk formula (5) will help us to understand the paradox. We should advocate these estimators for practical use only if we are sure they are appropriate. I do not think we know that yet.

REFERENCE

EFRON, B. and C. MORRIS (1972). Limiting the risk of Bayes and empirical Bayes estimators II: The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

CENTER FOR STATISTICAL SCIENCES
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712

CHARLES STEIN

Stanford University

Both the mathematical results and the conceptual aspects of this paper are very interesting to me. Most of my remarks, which are numbered for convenient reference, are related to my rejection of the principle of conditionality, which is stronger than Brown's rejection of that principle.