This can raise issues which are largely external to the ordinary statistical model and deserve more attention from the scientific side of statistics. It seems necessary in this development to use the physical context as a guide to the choice of operating model. In such contexts the issue of marginal optimality is not of interest: only the conditional calculations matter. Our statistician in Section 4 who advertises the shorter confidence intervals is guilty of professional misconduct.

Recent directions in conditional inference have deemphasized the "principle" aspect of conditioning. One motivation for this is that conditioning can provide a means to eliminate nuisance parameters and focus on the parameter of interest. Another is that conditional distributions are often much easier to calculate, which is especially useful in high-dimensional problems.

REFERENCES

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

EVANS, M., FRASER, D. A. S. and MONETTE, G. (1985). Mixtures, embedding, and ancillarity. *Canad. J. Statist.* **13** 1–6.

EVANS, M., FRASER, D. A. S. and MONETTE, G. (1986). On principles and arguments to likelihood. *Canad. J. Statist.* **14** 181–200.

FRASER, D. A. S. (1968). A black box or a comprehensive model. *Technometrics* **10** 219–229.

FRASER, D. A. S. (1979). *Inference and Linear Models*. McGraw-Hill, New York.

FRASER, D. A. S. and MCDUNNOUGH, P. (1980). Some remarks on conditional and unconditional inference for location-scale models. *Statistische Hefte* **21** 224–231.

KALBFLEISCH, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62** 251–259.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

REID, N. and FRASER, D. A. S. (1989). Comment on "The geometry of asymptotic inference," by R. E. Kass. *Statist. Sci.* **4** 231–233.

WELCH, B. L. (1939). On confidence limits and sufficiency with particular reference to parameters of location. *Ann. Math. Statist.* **10** 58–69.

DEPARTMENT OF MATHEMATICS
YORK UNIVERSITY
DOWNSVIEW
CANADA M3J 1P3

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
TORONTO
CANADA M5S 1A1

LEON JAY GLESER[1]

Purdue University and University of Pittsburgh

Put briefly, Brown's paradox is that an estimator can be conditionally admissible given each value of an ancillary statistic, but inadmissible unconditionally. Brown is to be congratulated for his insight in pointing out the conflict between frequentist criteria of good performance for point estimators and widely held notions concerning ancillary statistics. Brown supports use of unconditional frequentist measures to guard against "inconsistency" (uncon-

[1]Research supported in part by NSF Grant DMS-89-01922.

ditional inadmissibility) of conditionally based procedures. In contrast, R. A. Fisher's introduction of the concept of ancillarity and his concern with "recognizable subsets" in statistical design (which led him to modify the Latin square design) both arose from recognition that procedures which had good unconditional properties could be unacceptable conditionally.

Although the inadmissibility results which Brown presents have important practical consequences [for example, in Monte Carlo simulation; see Gleser (1987a, b)], my discussion here will concentrate upon theoretical issues.

In Section 1, I give a result (Theorem 1) concerning inadmissibility of the usual estimator of any scalar component of the mean vector of a multivariate normal distribution when the covariance matrix of this distribution is random, observable and ancillary. This is a specialization of the general context of Theorem 2.2.1 of Brown, but a different dominating estimator is given and the proof provides some insight that may be helpful. The context of Section 1 (and Theorem 1) serves a background to my discussion in Section 2 of the philosophical position taken by Brown. The main point made in Section 2 is that, except for some very specialized applications, unconditional inadmissibility is generally only of interest to a statistician seeking to do well in many similar problems, but not to users of the inference (estimate) presented by the statistician in any particular problem. Some attempts are also made to resolve the ongoing dispute between frequentists and Bayesians.

1. Estimation with a random observable ancillary covariance matrix. Let V be a random positive definite matrix and y be a random p-dimensional column vector. Conditional on V, assume that y has a multivariate normal distribution with mean vector μ and covariance matrix $\sigma^2 V$. Both y and V are observable, while μ is an unknown parameter. The positive scalar σ^2 can be known or unknown. The (marginal) distribution of V is assumed not to depend on μ (or on σ^2, if σ^2 is unknown). Thus, V is ancillary for estimation of μ.

Let

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where y_1, μ_1 and V_{11} are scalars.

THEOREM 1. *Assume that* $E[\mathrm{tr}(V)] < \infty$. *Define*

$$C = \{V_{22}\colon E[WW'|V_{22}] \text{ exists and has rank} \geq 3\},$$

where $W = V_{22}^{-1}V_{21}$. *If* $P\{V_{22} \in C\} > 0$, *then the usual estimator* y_1 *of* μ_1 *is inadmissible under squared error loss.*

PROOF. Let 1_C be the indicator function of C. Consider an estimator of μ_1 of the form

$$(1) \qquad \delta(y, V) = y_1 - W'h(y_2, V_{22})1_C,$$

where the choice of the $(p-1)$-dimensional column vector $h(y_2, V_{22})$ is yet to be determined. The difference in risks between y_1 and $\delta(y, V)$ is

$$
\begin{aligned}
\Delta &= E(y_1-\mu_1)^2 - E(\delta(y,V)-\mu_1)^2 \\
&= 2E[(y_1-\mu_1)W'h(y_2,V_{22})1_C] \\
&\quad + E[W'h(y_2,V_{22})h'(y_2,V_{22})W1_C].
\end{aligned} \tag{2}
$$

Take expected values on the right-hand side of (2) in the order $E_{y_2,V}E_{y_1|y_2,V}$ and note that

$$E[(y_1-\mu_1)|y_2,V] = W'(y_2-\mu_2).$$

Consequently, if

$$\delta^*(y_2,V_{22}) = y_2 - h(y_2,V_{22})1_C, \tag{3}$$

it is easily seen that

$$
\begin{aligned}
\Delta &= E[(y_2-\mu_2)'WW'(y_2-\mu_2)] \\
&\quad - E[(\delta^*(y_2,V_{22})-\mu_2)'WW'(\delta^*(y_2,V_{22})-\mu_2)].
\end{aligned} \tag{4}
$$

Conditional on $V_{22}\in C$, W is random and

$$\Omega = \Omega(V_{22}) = E[WW'|V_{22}]$$

has rank greater than or equal to 3. Further, conditional on V_{22}, y_2 has a multivariate normal distribution with mean vector μ_2 and covariance matrix $\Sigma = \sigma^2 V_{22}$. Calculating the expected values on the right-hand side of (4) in the order $E_{V_{22}}E_{y_2,W|V_{22}}$, we see that Δ is the expected value (over V_{22}) of the difference in the risks of $W'y_2$ and $W'\delta^*(y_2,V_{22})$ conditional on V_{22}. Theorem 2.1.2 and the remarks in Section 4.1 of Brown's paper now apply to suggest an estimator $\delta^*(y_2,V_{22})$, and thus the vector function $h(y_2,V_{22})$, such that this conditional risk difference is always positive for $V_{22}\in C$. Since $P\{V_{22}\in C\}>0$, this implies that Δ is positive, completing the proof of the theorem. □

REMARK 1. Brown indicates two distinct contexts in which his ancillarity paradox arises: (1) When it is desired to estimate a randomly chosen linear combination of the elements of the mean vector of a normal distribution (or the parameters of a linear regression model) and (2) when estimation of a mean vector (or regression parameters) is desired when the observed vector y has a random observable covariance matrix (design matrix). The former context arises naturally in prediction problems, but the latter context is more directly related to the classic concept of ancillarity, in which the ancillary statistic provides information about the accuracy with which the unknown parameter can be estimated. Context 1 is treated by Brown's Theorem 2.1.2, but Brown does not directly treat context 2, preferring to give instead a more general result (Theorem 2.2.1). Theorem 2.2.1 of Brown with

$$Q = e_1e_1', \qquad e_1 = (1,0,\dots,0)',$$

applies to the context of Theorem 1 given here. Brown's result is more general, applying when $p = 3$, whereas because W is $(p - 1)$ dimensional, Theorem 1 is applicable only when $p \geq 4$. On the other hand, Theorem 1 provides an explicit and possibly useful estimator, while the domination results in Theorem 2.2.1 suffer from the defects noted by Brown in his discussion. Note that, interestingly enough, the proof of Theorem 1 makes use of the existence of a dominating estimator in context 1 described above to provide a dominating estimator in context 2.

REMARK 2. The conditions on V in Theorem 1 are satisfied, for example, when V has a nonsingular Wishart distribution with degrees of freedom $\nu \geq p + 1$. The condition $E[\text{tr}(V)] < \infty$ is needed only to ensure that the risks of y_1 and y_2 are finite.

REMARK 3. It is not hard to see that Theorem 1 can be used to provide an alternative proof of Brown's Theorem 3.2.2.

Conditional on V, y_1 is an admissible estimator of μ_1. The dominating estimator

$$\delta(y, V) = y_1 - W'(y_2 - \delta^*(y_2, V_{22}))$$

makes use of an estimator $h(y_2, V_{22}) = y_2 - \delta^*(y_2, V_{22})$ of $y_2 - \mu_2$ to improve accuracy. Loosely speaking, this requires that y_1 and y_2 are correlated sufficiently often as V varies so that information from $y_2 - \mu_2$ is useful, and also requires that enough components of y_2 are correlated with y_1 that the bias in $y_2 - \delta^*(y_2, V_{22})$ for $y_2 - \mu_2$ can be made small by Stein-type shrinkage.

Theorem 1 can be extended to cover estimation of any arbitrary linear combination $a'\mu$ of μ. It is enough to consider cases where $a'a = 1$. Let Γ be an orthogonal matrix with first row equal to a' and make the transformation

$$y \to \Gamma y, \qquad \mu \to \Gamma\mu, \qquad V \to \Gamma V \Gamma'$$

to reduce to the problem treated by Theorem 1. The basic condition of Theorem 1 stated in terms of the original matrix V is that

$$E\Big[\big(\Gamma_2 V \Gamma_2'\big)^{-1}\Gamma_2 V a a' V \Gamma_2'\big(\Gamma_2 V \Gamma_2'\big)^{-1}\Big|\Gamma_2 V \Gamma_2'\Big] \tag{5}$$

exists and has rank greater than or equal to 3 with positive probability under the marginal distribution of $\Gamma_2 V \Gamma_2'$, where $\Gamma' = (a, \Gamma_2')$. Again, this condition is satisfied by a nonsingular Wishart matrix V with degrees of freedom $\nu \geq p + 1$.

If V satisfies condition (5) for all a, or even only for a running over the columns of the p-dimensional identity matrix I_p, then as a corollary to Theorem 1, we have the very strong result: *There exists an estimator $\delta(y, V)$ of μ that dominates y coordinatewise in risk under squared error loss*. Recall, however, that Theorem 1 can hold only when $p \geq 4$.

Theorem 1 can be extended to the general context of Brown's Theorem 2.2.1. However, verification of this assertion is tangential to the purpose of this discussion and thus will not be given.

2. Conditional versus unconditional admissibility. Brown in his Section 5 makes a distinction between point estimators on the one hand, and hypothesis tests and confidence intervals on the other. In the latter types of inference, the correctness of an accompanying stochastic claim of accuracy (level of significance, probability of coverage) is said to be essential in determining the validity of the conclusion (inference). On the other hand, Brown claims that stochastic measures of accuracy of point estimators are less crucial for accepting the results of point estimation, and that no conditionally interpretable stochastic claim is made of accuracy when a point estimator is presented. Rather, the statistician is asked only to show that no uniformly better estimator exists (admissibility) and that the estimator is also "reasonable in the fact of whatever generally acceptable a priori evaluations can be made about the parameter."

In actuality, point estimators are frequently presented along with a stochastic measure of accuracy. Brown, himself, mentions the use of confidence intervals as such a measure of accuracy. Alternatively, the estimated standard deviation or covariance matrix may be given. Those individuals who read the statistician's report will certainly use such measures to evaluate the validity of the estimate. The above measures of accuracy do not directly reflect risk under squared-error (or alternative loss functions). Recently, Lu and Berger (1989), Johnstone (1987) and others have suggested ways to remedy that deficiency.

Further, a conditional stochastic claim is being made even if we accept Brown's position. Readers of the statistician's report will certainly interpret the statistician's claim of admissibility as being *conditional* upon the accuracy of the experiment used, and his estimates of risk as being similarly conditional. Thus, in the context of Theorem 1, the statistician's claims will be interpreted as conditional on the value of V. If the observed value of V shows small correlations between y_1 and the remaining coordinates of y, such readers may question the use of $\delta(y, V)$. Since y_1 is conditionally (given V) admissible, these readers may feel that use of y_1 is more valid than use of $\delta(y, V)$.

The statistician will respond that this objection leads to inconsistency in that y_1 is unconditionally dominated in risk by $\delta(y, V)$. However, such inconsistency is only from the statistician's point of view. Whereas the statistician is concerned with his or her own accuracy or risk over a "stream" of similar problems in which V varies randomly (and may often show large correlations between y_1 and the other coordinates of y), the readers are only concerned with the given experiment (and, if the readers are frequentists, repetitions of the experiment under the same covariance matrix V). Thus, consistency on the part of the statistician can lead to conflict with readers' notions of reasonableness.

REMARK 4. If V has a continuous distribution and $\delta(y, V)$ is unconditionally admissible, one can change $\delta(y, V)$ at any value v of V without affecting the unconditional risk (and thus unconditional admissibility) of the resulting estimator (since $P\{V = v\} = 0$). Consequently, an unconditionally admissible estimator can be very poor conditionally (when $V = v$). Brown covers this point by requiring that estimators be both conditionally and unconditionally admissible.

Since so much of the literature on Stein estimation uses baseball averages as examples, I cannot resist using the following example. A manager must choose a pinch hitter in a crucial moment of a ball game. Of the available players, player A has the highest batting average (.300, say) against all pitchers, but hits only .260 against left-handed pitchers. Player B has the highest available average (.310) against left-handed pitchers, but an overall average of only .275. In the long run, against a random stream of right-handed and left-handed pitchers, player A would be the manager's uniformly best choice. But if the opposing pitcher is left-handed, the manager would face considerable criticism from fans if he used player A instead of player B. Here, identify the type of opposing pitcher with V, the choice of player A with an unconditionally admissible (indeed, best) inference procedure, choice of player B with a conditionally reasonable procedure, the manager with the statistician and the fans with readers of the statistician's report. Although this example is not one of estimation (or even inference), it does illustrate the conflict between the conditional perspective of a statistician's audience and the unconditional perspective of the statistician.

For statistical inference, a resolution of this conflict (from Brown's perspective) is possible. As Brown remarks, proper Bayes procedures are (except for pathological cases) both conditionally and unconditionally admissible. Since Brown's stated criteria for an estimator are that it be conditionally and unconditionally admissible and reasonable in the light of generally accepted prior information about the parameter, I recommend that Brown adopt a proper Bayes approach to estimation. Unfortunately, a generally acceptable prior opinion about the parameter may not exist, so that the statistician's prior distribution may be criticized. One could advocate a robust Bayes approach, but (again as Brown notes) estimators constructed by such an approach may fail to be unconditionally admissible. One advantage of a Bayesian approach is that the statistician's prior opinion is available for inspection, not hidden (as it so often is).

The difference between the conditional and unconditional perspectives on inference is basic to the dispute between frequentists and Bayesians. Frequentist measures of performance implicitly assume a stream of similar experiments and inference problems, in which no single experiment is singled out for particular attention. The Bayesian (and also likelihood) perspective concentrates on what can be learned from a single experiment taken from this stream. Both perspectives have their own validity and applicability. Although the frequentist perspective is clearly appropriate in automatic repetitions of a certain type of inference problem (quality control, laboratory assay), the Bayesian approach appears better suited to general scientific and practical inquiry. In any case, these approaches address quite different issues. For this reason, any apparent conflict between them is illusory.

What both frequentists and Bayesians tend to overlook is the fact that data (and data summaries) are put to many uses and viewed from many perspectives (prior opinions) by a statistician's audience. Concentration upon which single procedure to use in analyzing data tends to neglect the diverse interests of this audience. More attention needs to be paid to the design of the experiment, The founders (Neyman, Pearson, Wald) of the modern frequentist approach to inference were aware of this point, and discussed designs based on minimax procedures as a way to satisfy all users of the data. Where prior opinion is not highly variable (as assumed by the minimax approach), designs constructed from a robust Bayesian perspective may be more efficient in satisfying the needs and interests of a statistician's audience. A step in this direction has been made in hypothesis testing by my student Burt (1989), and for estimation by DasGupta and Studden (1989).

REFERENCES

BURT, J. (1989). Towards Agreement: Bayesian Experimental Design. Ph.D. dissertation, Dept. of Statist., Purdue Univ.

DASGUPTA, A. and STUDDEN, W. J. (1989). Towards a theory of compromise designs: Frequentist, Bayes and robust Bayes. Mimeo Series No. 89-20, Dept. of Statist., Purdue Univ.

GLESER, L. J. (1987a). Improved estimators of mean response in simulation when control variates are used. Mimeo Series No. 87-32, Dept. of Statist., Purdue Univ.

GLESER, L. J. (1987b). Improved estimators of mean response in simulation when control variates with unknown covariance are used. Mimeo Series No. 87-34, Dept. of Statist., Purdue Univ.

JOHNSTONE, I. (1987). 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). Estimation of normal means: Frequentist estimation of loss. *Ann. Statist.* **17** 890–906.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA 15260

B. M. HILL

The University of Michigan

I am delighted for the opportunity to discuss this interesting paper by Professor Brown. I have long admired his work both for its technical virtuosity and for his thoughtful, philosophical discussions of the statistical approach he advocates. I have often expressed the hope that frequentists would offer Bayesians some interesting challenges, which Professor Brown does here. He shows that the frequentist admissibility paradigm (FA) is not compatible with the principle of ancillarity (AP) and suggests that the latter should be abandoned. AP is an intuitively compelling idea, a version of which is implied by the