

Carl Morris and I worried about this a lot in our 1971 and 1972 papers, and also in the specific examples of 1975. Our hard-working 18 baseball players were offered as a simplified test case for thinking about the trade-offs between d_0 and d_1 ; see also Section 8 of Efron (1982).

REFERENCES

- EFRON, B. (1982). Maximum likelihood and decision theory. *Ann. Statist.* **10** 340–356.
 EFRON, B. and MORRIS, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators, I: The Bayes case. *J. Amer. Statist. Assoc.* **66** 801–815.
 EFRON, B. and MORRIS, C. (1972). Limiting the risk of Bayes and Empirical Bayes estimators, II: The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.
 EFRON, B. and MORRIS, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70** 311–319.

DEPARTMENT OF STATISTICS
 STANFORD UNIVERSITY
 STANFORD, CALIFORNIA 94305

D. A. S. FRASER AND N. REID

York University and University of Toronto

1. Introduction. Professor Brown has presented a comprehensive discussion of multiple regression in relation to admissibility and the ancillarity principle. He concludes that there is a paradox: That the results with multiple regression contradict “the widely held notion that statistical inference in the presence of ancillary statistics should be independent of the distribution of these ancillary statistics.” The reader thus receives the impression that there is something wrong or inappropriate with conditional inference. The basic assumption of conditional inference is that only the conditional model is examined and that information from the marginal model is ignored. This is not a “notion” that inference “should” be independent of the marginal model as interpreted by Professor Brown, but that inference should not use or make reference to that model.

The technical point then is that there is a conflict between conditional methods and classical optimality criteria. We feel that this should be no surprise, let alone paradox. In Section 5 we present a simple example that also illustrates the conflict.

Our broader viewpoint is that the familiar optimality criteria of statistics are in fact in conflict with scientific principles and that this provides the explanation for the issues raised in the paper; see Section 2.

In a concluding Section 6, we argue that conditional methods are close to the core of the scientific method, and note that conditional inference from both a theoretical and pragmatic orientation is a recently active area of research and presents exciting possibilities for research development.

Standard statistical theory uses a range of optimality criteria, such as maximum power for a test at a given size α , minimum length for a confidence

interval with a given coverage β , minimum variance for an unbiased estimate, minimax risk and admissibility. In any problem an optimum procedure may or may not exist. If not, then from a pragmatic view, the criteria provide some directions in which to search.

The theory of conditional inference has evolved from Fisher's recommendation to condition on an ancillary, often taken to be a variable with a fixed distribution, although Fisher clearly had more in mind. We prefer to emphasize that the variable should describe some objective characteristics of the physical problem, and we feel that this captures a main part of the intent in Fisher.

We thus have the classical *optimality criteria* and the *conditional inference prescription*. We feel the conflict between these is well documented in the literature; Professor Brown adds admissibility to the documented criteria. We argue in Section 3 that the conflict should be no surprise and in Section 4 that the resolution censures the broad-base use of the criteria and favours current directions of conditional inference.

2. Ancillaries qualified. An ancillary as a variable with a fixed distribution needs to be qualified in various ways, even to express the intentions in Fisher's early papers on conditioning. A few examples close to his discussions are:

1. Random choice of measuring instrument.
2. Random choice of sample size.
3. Location, location scale and transformation models.
4. Random choice of design matrix for the regression linear model;

for a survey of these and others, see Cox and Hinkley (1974) and Fraser (1979).

In these examples the conditioning variable represents some physical aspect that has been singled out in the investigation at hand. In some cases it may have a fixed distribution but that is typically secondary. For example, the error distribution underlying the third model may depend on additional parameters; the conditioning then is on the observed characteristics of that error; for examples, see Fraser [(1979) Sections 2.1.1, 3.2, 6.1.2 and 7.1.3]. This is consistent with Fisher but requires scientific aspects of an investigation that are ordinarily not included in the standard model. Some discussion of aspects omitted by the standard model may be found in Fraser (1968; 1979, Chapter 1), Evans, Fraser and Monette (1986) and Kalbfleisch (1975). These aspects bear on the scientific choice of conditioning variable and we view this issue as being very central to the development of theoretical statistics. For some comment see Reid and Fraser (1989).

The conditionality principle, as it is usually formulated, requires conditioning on any variable that has a distribution free of the parameter. One difficulty with defining a conditioning variable as just a variable with a fixed distribution is clearly illustrated by easily constructed 2×2 tables where both the row

totals and the column totals are ancillary, but the row and column totals combined do not form an ancillary. This example, first discussed by Fisher, motivated the definition of the cross-embedded model [Evans, Fraser and Monette (1985)] and led to the discussion in Evans, Fraser and Monette (1986) showing that the conditionality principle implies the likelihood principle. The likelihood principle, of course, undercuts all the optimality criteria mentioned in Section 1: One just records a likelihood interval without any reference to asymptotic or distribution properties, such as variance, power, etc.

The conflict mentioned in Section 1 that Professor Brown discusses is one aspect of something much bigger: the complete rejection of the optimality criteria based on *conventional* conditioning.

Even the general notion that a conditioning variable should represent some clear physical aspect of an investigation needs some qualifying. For finite population sampling and for experimental design, a fixed objective distribution is imposed to obtain the randomness needed to make statistical statements. The randomness was imposed to obtain a statistical model; conditioning on the observed randomness would eliminate the model. The use of such randomization is thus a scientific issue.

3. Conflict: Optimality versus conditioning. The conflict between the *optimality criteria* and the conventional *conditional inference* prescription can be documented from quite general considerations. Compare the two possibilities:

1. Examine all statistical procedures of a given type as discussed in Section 1.
2. Examine all statistical procedures of a given type, conditionally for each value of the ancillary.

For the second a procedure must satisfy more conditions. It follows that the second class is a subset of the first class and thus that an optimum procedure in the second class can be equalled or exceeded by one in the first class. In other words, you can do better, certainly as well, by using a marginal procedure over a conditional procedure.

Nevertheless, examples do have merit; they can bring into focus the mechanism by which the "conflict" arises. An early example (Welch, 1939) involves a sample y_1, y_2 from the uniform $(\theta \pm \frac{1}{2})$ distribution. Conditional tests are less powerful and confidence intervals longer, when defined conditionally given the configuration statistic $y_2 - y_1$; Welch favoured the unconditional. A general discussion of conditional and unconditional confidence intervals with examples may be found in Fraser and McDunnough (1980). Examples are also given in Fraser (1979, Chapters 3 and 4).

4. The mechanism. The mechanism by which conditioning frustrates the standard optimality criteria is a nonlinearity implicit in the process of choosing an optimal procedure.

The mechanism can perhaps be seen more clearly by examining it in a larger context. Consider a statistician who advertises that his 95% confidence inter-

vals are shorter on the average than all his competitors; in fact, he works on an overall 95% rate rather than a conditional 95% rate for each client contact. He achieves his special performance by giving longer intervals (above 95%) to his clients with precise measurements, and shorter intervals (below 95%) to his clients with imprecise measurements, still maintaining the overall 95% rate. His intervals will be shorter on the average [for details, see Fraser and McDunnough (1980)].

From a betting viewpoint he buys his short intervals where they are statistically cheap for him—from the imprecise clients. Mean length and coverage are not linearly related. They can be traded off. Should one conclude that this is better than conditioning on client contact or on the investigation itself? This is where the scientific aspects enter; the unconditional model, certainly if made expansive enough, is inappropriate.

5. A simple example. Consider the estimation of θ for the 2×2 table with probabilities $(\theta/3, (1 - \theta)/3; (1 - \theta)/3, (1 + \theta)/3)$. Let the response be the generalized indicator $(x_{ij}; i, j = 1, 2)$; for simplicity we record results for the $n = 1$ case.

Recall that for an observation x from the Bernoulli (θ) distribution, x is the uniformly minimum variance unbiased (UMVU) estimate of θ and that $(x + 1/2)/(1 + 1) = x/2 + 1/4$ is an admissible estimate of θ and is minimax with constant risk $1/16$ with respect to squared error loss [Lehmann (1983), Section 4.2, 4.3].

Conditional on the first column total ($x_{11} + x_{21} = 1$), the estimate x_{11} is UMVU; it has mean θ and variance $\theta(1 - \theta)$. The minimax estimate is $(3/4)x_{11} + (1/4)x_{21}$; it is admissible and has risk $1/16$. Conditional on the second column total ($x_{12} + x_{22} = 1$), the estimate $-x_{12} + x_{22}$ is UMVU, with mean θ and variance $1 - \theta^2$. The admissible minimax estimate is $-(1/2)x_{12} + (1/2)x_{22}$, with risk $1/4$. Because either x_1 or x_2 is 1, the unbiased and admissible estimates can be written, respectively, in the form $t_u = x_{11} - x_{12} + x_{22}$ and

$$t_a = (3/4)x_{11} + (1/4)x_{21} - (1/2)x_{12} + (1/2)x_{22}.$$

The unconditional variance of t_u is $(1/3)(1 - \theta)(2 + 3\theta)$, which has the value $11/16$ at $\theta = 1/4$. The unconditional risk of t_a is $3/16$.

Improved estimates can be obtained by averaging cells x_{21} and x_{12} , that is, by using the minimal sufficient statistic. Let $t_u^* = x_{11} - (1/2)(x_{12} + x_{21}) + x_{22}$ and $t_a^* = (3/4)x_{11} - (1/8)(x_{12} + x_{21}) + (1/2)x_{22}$. The unconditional variance of t_u^* is $(1/2)(1 - \theta)(1 + 2\theta)$, which takes the value $9/16$ at $\theta = 1/4$. The unconditional risk of t_a^* is $(3/32)(1 + \theta)$, which is equal to 0.1172 at $\theta = 1/4$.

Thus the conditional UMVU estimator and the conditionally admissible minimax estimator are dominated in the marginal setting.

6. Conditioning: Now and directions. What sort of conclusions should we draw from the present discussion?

We feel that the notion of conditioning on variables that define the physical context in which the interest parameter is being examined is very compelling.

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 GLESER¹

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-

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