Conditional Inference

George Casella

Cornell University

BU-1001-M November, 1988

This paper is written in honor of Professor D. Basu

on the occasion of his 65^{th} birthday

Abstract

Ideas of conditional inference have grown out of many different schools of statistical thought. The development of these ideas is traced, starting with some original ideas of Fisher. The influence of other researchers, such as Basu and Buehler, is also discussed. The development is traced to the present, through the work of Pierce and Robinson, to current work in conditional inference.

1. Introduction

The development of conditional inference has followed many paths. There are now a number of inferential methods that use this name. For example, the likelihood based methods of Hinckley (1983), or Cox and Reid (1986), are conditional inference methods. The attempt of Kiefer (1977), to merge conditional ideas with frequentist theory is also conditional inference.

The one common factor in the different conditional inferences is the requirement for reasonable (coherent) post-data inference. That is, inferential statements made after the data have been seen should have some logical consistency. Another approach to conditional inference, one that gained structure through the work of Buehler (1959) and Robinson (1979a,b), provides an objective framework for assessing post-data validity. It is this version of conditional inference on which we will concentrate.

The different versions of conditional inference have a common origin in ideas of Fisher. These ideas of Fisher are somewhat intuitive, and leave some gaps in development (but not to Fisher!). The origins in Fisher were later refined by Basu, who relied on ideas of Bayesian inference to close any gaps. This is where our review begins.

1.1. The Seeds of Conditional Inference

Many influential ideas in statistical inference can be attributed to Sir Ronald Fisher. One of the most elusive, perhaps, is that of conditional inference. In Fisher (1959) we find the ideas of a *reference set*:

> "In attempting to identify a test of significance $\cdots$ with a test for acceptance, one of the deepest dissimilarities lies in the population, or reference set, available for making statements of probability."

Interpreting Fisher, we find that he is concerned with the range of the inferences, that is, with the set in the population to which the inference should apply. In this sense, he is concerned with conditional inference, inference conditional on some subset of the sample space. The exact nature of his concern is not, at first, clear. It does emerge in some later statements, again from Fisher (1959). In talking of inference from Student's t distribution, he says

> "The reference set for which this probability statement holds is that of the values of μ, $\overline{x}$ and s corresponding to the same sample $\cdots$ there is no possibility of *recognizing* any subset of cases $\cdots$ for which any different value of the probability should hold." (my italics)

In this statement we see one of the keystones of conditional inference. There should not be a subset of the sample space (a *recognizable* subset) on which the inference from a procedure can be substantially altered. If such subsets exist, then inference from the procedure is suspect.

If such a recognizable subset existed, then Fisher would no doubt find it, however, there does not seem to be any general methodology used. Although ideas of estimating and eliminating nuisance parameters are used, and also ideas of ancillarity are used, no general scheme is defined.

One famous example is Fisher's criticism of Welch's solution to the Behrens-Fisher problem. If $\overline{x}_i$, s_i^2, i=1,2, are the sample mean and variance from a sample of size n from independent normal population with unknown parameters μ_i and σ_i^2, Fisher (1956) derived the following fact. Under the hypothesis H_O: $\mu_1=\mu_2$, for any value t,

$$P\left(\left|\frac{\sqrt{n}\left(\overline{X}_1-\overline{X}_2\right)}{\sqrt{S_1^2+S_2^2}}\right| > t \,\middle|\, S_1^2=S_2^2\right) = P\left(\left|T_{2(n-1)}\right| > \tau t\right), \tag{1.1}$$

where $T_{2(n-1)}$ has Student's t distribution with 2(n-1) degrees of freedom, and τ is an unknown parameter satisfying $0 \le \tau \le 1$. In other words, conditional on $S_1^2=S_2^2$, the random variable $\sqrt{n}\left|\overline{X}_1-\overline{X}_2\right|/\sqrt{S_1^2+S_2^2}$ is stochastically greater than $\left|T_{2(n-1)}\right|$. Fisher used this fact to show that Welch's solution suffered from the property that the probability of rejecting a true H_O, given that $s_1^2=s_2^2$, was bounded below by the nominal level. Thus, on the recognizable subset $\left\{(s_1^2,s_2^2)\colon s_1^2=s_2^2\right\}$, Welch's solution has an actual error rate greater than the nominal level.

This conditional behavior would be even more disturbing if the set $\left\{(s_1^2,s_2^2)\colon s_1^2=s_2^2\right\}$ is taken as a reference set, i.e., a set on which the conditional inference should be applied. Fisher's argument for conditioning on this set, or more generally on the ratio s_1^2/s_2^2, is elusive. The fact that Fisher considers this a reasonable reference set appears again in Fisher (1959), where he discusses his solution to the Behrens-Fisher problem.

The fact remains, however, that the mechanism of choice of a reference set is elusive. Although concepts of ancillarity and elimination of nuisance parameters are considered, a general mechanism for choosing a conditional reference set is not known.

1.2. Basu's Refinement

In doing conditional, or post-data, inference the evidential meaning of the inference becomes increasingly important. Fisher's idea of a reference set has some meaning, i.e., it defines a part of the sample space on which inference it is meaningful to restrict inference. On the other hand, the connotation of a recognizable set does not carry this distinction.

A recognizable set is only a set that is in the sample space, and may give no meaningful inference base. Poor conditional (post-data) performance of a procedure on a recognizable set is taken as criticism, but if this recognizable set is not a meaningful reference set, then the criticism may be vacuous.

Fisher had the intuition to choose recognizable subsets that were also meaningful reference sets. Thus, when he leveled criticism (or praise) of the conditional performance of a procedure using a particular recognizable set, this set was also a meaningful reference set. One of the major clues left to us by Fisher, on how to chose these reference sets, is that they should use ancillary information.

Alas, many of us are not possessed with Fisher's intuition in choosing reference sets. When Basu started to think about this, he realized that basing conditioning sets on ancillary information was not, in itself, a reasonable technique in general. In Basu (1964), he says

> "The ancillary argument of Fisher cannot be extended $\cdots$. We end this discourse with an example where $\cdots$ the ancillary argument leads us to a rather curious and totally unacceptable 'reference set'."

Basu then gives an example to illustrate his point. The point that we should be concerned with is that the choice of the reference set is not automatic. Of course, Basu does not give us a recipe for choosing a reference set, but rather argues that the only reasonable procedures are free of conditional defects.

1.3. Conditional and Unconditional Inference

Inference made conditional on the data must, necessarily, connect a statement about the unknown parameters to the data actually observed. This fact separates conditional inference from unconditional, or pre-data, inference. This latter inference, that of the frequentist (Neyman-Pearson) school, need not apply in any way, to the data at hand. A frequentist inference merely states how the procedure will perform in repeated trials, even if such a statement is ludicrous in the face of the observed data.

This dichotomy, between conditional and unconditional inference, most often results in a statistician choosing one stand and rejecting the other. Fisher rejected unconditional inference in favor of conditional. Basu, although starting out in the Newman-Pearson camp, ultimately rejected unconditional inference in favor of Bayesian conditional inference. Indeed, perhaps Basu stated his belief most elegantly in Basu (1981)

> "With E_X as the (Neyman-Pearson) confidence set corresponding to the observed sample x, can any evidential meaning be attached to the assertion $\theta \in E_X$? Suppose on the basis of sample X one can construct a 95% confidence interval estimator for the parameter θ, then does it mean that (the random variable) X has information on θ in some sense?"

Of course, Basu gave examples of 95% Neyman-Pearson confidence intervals with no information at all about θ. For example, if $\theta \in [0,1]$, and $X \sim U(0,1)$ (X is θ-free), then for any fixed set $B \subset (0,1)$, the set

$$E_X = \begin{cases} B \text{ if } 0 < X \leq .05 \\ (0,1) \text{ if } .05 < X < .95 \\ B^c \text{ if } .95 \leq X < 1 \end{cases}$$

is a 95% unconditional confidence set for θ. But, of course, we cannot attach any evidential meaning to the statement "$\theta \in E_X$." (We note, in passing, that the conditional behavior of this set is wretched. For example, $P\left(\theta \in E_X | 0 < X \leq .05\right) = P(\theta \in B)$ and $P\left(\theta \in E_X | .95 \leq X < 1\right) = P\left(\theta \in B^c\right)$. One of these two probabilities must be smaller than .95. Further, $P\left(\theta \in E_X | .05 < X < .95\right) = 1$, showing that the post-data inference can be moved all over.)

As we trace the development of conditional inference, we will see that Basu's teachings are there. Many papers take the approach of verifying good conditional properties by verifying Bayesianity. However, this might be a case where some good can come out of greed. Why should we be satisfied with only good post-data behavior or good pre-data behavior? Why can't we try for both? The answer is that we can not only try for both, we can sometimes attain it. The procedures that do can be acclaimed by both camps – conditional and unconditional.

2. Formalizing Conditional Inference

The work of Buehler (1959) was a landmark attempt in examining post-data validity of Neyman-Pearson procedures. Buehler's work is pioneering for two reasons. One, he examined post-data behavior of frequency based rules (not necessarily Bayes rules) and two, he developed criteria for carrying out this evaluation in an objective manner. Buehler's work was based on other seminal ideas of Tukey (1958) and Stein (1961), and was ultimately generalized and formalized by Robinson (1979a,b). We briefly describe Robinson's set-up.

The random variable X has density $f(x|\theta)$ and, based on observing X=x, a confidence procedure $<C(x),\gamma(x)>$ is constructed. A confidence procedure consists of a set C(x) and a probability assertion $\gamma(x)$. The validity of $\gamma(x)$ as a confidence assertion is measured by the ability of $<(C(x),\gamma(x)>$ to maintain its confidence even when evaluated conditionally. To be specific, we consider $\gamma(x)$ to be an evaluation of the coverage properties of C(x) in the sense that

$$E_\theta \gamma(X) \approx P_\theta\Big(\theta \in C(X)\Big) . \tag{2.1}$$

Suppose now that a recognizable subset, $\mathcal{A}$, of the sample space, and an $\epsilon > 0$ exists such that

$$E_\theta\Big(\gamma(X)|X\epsilon\mathcal{A}\Big) - P_\theta\Big(\theta\epsilon C(X)|X\epsilon\mathcal{A}\Big) \geq \epsilon . \quad \forall\theta \tag{2.2}$$

Then, we have qualitatively changed the confidence behavior. On the set $\mathcal{A}$, our conditional assertion is suspect: The asserted probability, $\gamma(x)$, is, on the average, uniformly greater than the actual conditional coverage.

In Robinson's terminology, (2.2) is a special case of a relevant betting function, defined as follows:

Definition 2.1: A function k(x), $-1 \leq k(x) \leq 1$ is *relevant* for $<C(x),\gamma(x)>$ if

$$E_\theta\Big\{\Big(I(\theta\epsilon C(X) - \gamma(X)\Big)k(X)\Big\} \geq \epsilon E_\theta|k(X)| \tag{2.3}$$

for all θ and some $\epsilon > 0$. If $\epsilon=0$ k(x) is *semirelevant.*

For statistical purposes, the most interesting forms of functions k(x) are indicator functions. Such functions reduce (2.3) to forms like (2.2), and allow interpretations in terms of conditional coverage probabilities. If $k(x) < 0$ is relevant, it is called *negatively biased.* If $k(x) = -I(X \epsilon \mathcal{A})$ then (2.3) would reduce to (2.2). *Positively-biased* sets can similarly be defined. In the previously mentioned criticism by Fisher of Welch's solution to the Behrens-Fisher problem, Fisher identified a *negatively-biased relevant* subset.

Buehler and Fedderson (1963) identify, in a special case, a positively-biased relevant subset for the one-sample t interval (they also attribute a similar results to Stein (1961) – Oh! to have a copy!). Later, Brown (1967) generalized this result to any one-sample t interval. For a random sample $X_1,\cdots,X_n$ from $n(\mu,\sigma^2)$, Brown identified constants k and ϵ so that

$$\text{(2.4)} \qquad P\left(\mu \epsilon \bar{X} \pm tS \middle| |\bar{X}|/S > K\right) \geq 1\text{-}\alpha+\epsilon \qquad \forall \mu,\sigma^2 ,$$

where t is the cutoff yielding a nominal $1\text{-}\alpha$ interval. This can be interpreted as saying that the conditional coverage of the t interval, after accepting $H_O:\mu=0$, is uniformly greater than the nominal level.

Identification of semirelevant subsets is less interesting than identification of relevant subsets, as most procedures with a frequentist guarantee will allow them. For example, from (2.4) we can deduce

$$\text{(2.5)} \qquad P\left(\mu \epsilon \bar{X} \pm ts \middle| |\bar{X}|/S > K\right) \leq 1\text{-}\alpha \qquad \forall \mu,\sigma^2 ,$$

identifying a negatively-biased semirelevant set for the t interval. However, Robinson (1976) showed that the t interval allows no negatively biased relevant sets. This led him to conclude that elimination of negatively-biased semirelevant sets was too strong a conditional criterion, but elimination of negatively-biased relevant sets was about right. (The elimination of positively biased sets is not of major concern, as this corresponds to being conservative.)

An interesting set of papers are those by Olshen (1973), and Sheffé (1977) with a rejoinder by Olshen (1977). In the 1973 paper, Olshen established a result like (2.5) for the Sheffé multiple comparisons procedure. Specifically, Olshen showed that the conditional coverage of the Sheffé procedure, given that the ANOVA F test rejects H_O, is less than or equal to the nominal level. Thus, Olshen generalized Brown (1967) in one direction, identifying a negatively biased semirelevant set for the Sheffé intervals. Sheffé took exception to this criticism, and answers Olshen in the 1977 article.

The connection between Bayes sets and conditional performance is very strong, as shown by Pierce (1973) and Robinson (1979a). If $\pi(\theta)$ is a proper prior, and we define $<C^{\pi}(x),\gamma^{\pi}(x)>$ by

$$\gamma^{\pi}(x) = \int_{C^{\pi}(x)} \pi(\theta|x)d\theta \,, \tag{2.6}$$

where $\pi(\theta|x) = f(x|\theta)\pi(\theta)/\int f(x|\theta)\pi(\theta)d\theta$, then no semirelevant functions exist for $<C^{\pi}(x),\gamma^{\pi}(x)>$. Thus, proper Bayes procedures have the strongest possible conditional properties.

Although the connection between Bayesianity and conditional performance is very strong, the exact link has not yet been established. That is, necessary and sufficient conditions for elimination of relevant, or semirelevant, functions have not yet been established. Although the work of Pierce and Robinson, and also Bondar (1977), establishes links between (possibly improper) Bayes procedures and nonexistence of relevant sets, the ultimate theorem, giving a necessary and sufficient condition, is still not known. It seems that the answer, although still unproven due to mathematical technicalities, is that elimination of relevant functions will occur if, and only if, the procedure is a limit of Bayes rules. This, however, remains an open question in the conditional inference literature.

3. Frequentist Conditional Inference

Although proper Bayes rules have strong conditional properties they do not, in general, have good frequentist properties. However, limits of Bayes rules, or generalized Bayes rules, can have good frequentist properties. Furthermore, such procedures may also have acceptable conditional properties. It is within this class that we can find procedures that have acceptable frequentist (or pre-data) properties and acceptable conditional (or post-data) properties.

A confidence set, C(x), is a $1\text{-}\alpha$ frequentist confidence procedure for a parameter θ if

$$P_\theta\Big(\theta \in C(X)\Big) \geq 1\text{-}\alpha \qquad \text{for all } \theta \;, \tag{3.1}$$

that is, the unconditional coverage probability of C(x) is at least $1\text{-}\alpha$. Of course, this pre-data guarantee says nothing of the conditional performance of the procedure $<C(x),1\text{-}\alpha>$. Robinson was able to establish conditional properties for a number of frequentist procedures by using the fact that they are limits of Bayes rules. In particular, his results for the t-interval (Robinson, 1979b) rely on this fact. Other results (Robinson, 1979b) for frequentist intervals for location or scale families also use arguments based on limiting Bayesianity. Most conditional properties of limits of Bayes rules deal with relevant, rather than semirelevant, functions as to the existence of $\epsilon > 0$ becomes important in the limit. However, for certain procedures from location families, Robinson (1979b) established the nonexistence of semirelevant functions. In particular, if $X \sim f(x\text{-}\theta)$, then the procedure

$$<[x\text{-}c, x+c],\ 1\text{-}\alpha> \;,$$

$$1\text{-}\alpha = \int_{-c}^{c} f(t)dt \;, \tag{3.2}$$

is a $1\text{-}\alpha$ frequentist confidence procedure that allows no semirelevant functions. Using different arguments based on invariance, Bondar (1977) established conditional properties of invariant frequentist sets.

The issue that is at the heart of the frequentist/conditional dichotomy is the assignment of a confidence function to a set C(x). For example, for any set C(x), where X $\sim$ f(x|θ), if we define $\gamma(x)$ by

$$\gamma(x) = \frac{\int_{C(x)} f(x|\theta)\pi(\theta)d\theta}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}, \tag{3.3}$$

where $\pi(\theta)$ is a proper prior, then the procedure $<C(x),\gamma(x)>$ is free of semirelevant sets. However, if C(x) is also a 1-α frequentist confidence procedure, this argument does not imply any conditional properties of $<C(x),1\text{-}\alpha>$. Thus, this type of consideration leads to two questions:

(3.4)

i) Is $<C(x),\gamma(x)>$ a reasonable frequentist procedure?

ii) Is $<C(x),1\text{-}\alpha>$ a reasonable conditional procedure?

Since the work of Robinson, and the others, in the 1970s there has been some progress made on the questions in (3.4). In Casella (1987) it was argued that, with some regularity conditions, a sufficient condition for the frequentist procedure $<C(x),1\text{-}\alpha>$ to be conditionally acceptable is the existence of a (possibly improper) prior $\pi(\theta)$ such that

$$\gamma(x) = \frac{\int_{C(x)} f(x|\theta)\pi(\theta)d\theta}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \geq 1\text{-}\alpha \qquad \text{for all } x . \tag{3.5}$$

If (3.5) is satisfied, then the procedure $<C(x),1\text{-}\alpha>$ allows no negatively biased relevant sets, which is acceptable conditional performance. Furthermore, it was demonstrated that such a property held for the multivariate normal confidence set centered at the positive-part James-Stein estimator. Specifically, if X $\sim$ N(θ,I), a p-variate normal random variable ($p \geq 3$), then the confidence procedure $<C_\delta(x),1\text{-}\alpha>$ allows no negatively-biased relevant sets, where

$$C_\delta(x) = \left\{\theta\colon |\theta-\delta(x)| \le c\right\}, \qquad \delta(x) = \left(1 - \frac{p-2}{|x|^2}\right)^+ x\ ,$$

(3.5)

$$P\left(\chi_p^2 \le c\right) = 1-\alpha\ .$$

Such a conditional inference strategy was also promoted in Casella (1988), and a number of other procedures were also examined. In discussing this paper, a number of alternate strategies were put forth. For example, Berger (1988) advocates an "estimated confidence" approach, where the procedure $<C(x),\gamma(x)>$ would be considered *frequency valid* if

$$E_\theta\gamma(x) \le P_\theta\left(\theta \in C(x)\right), \qquad \text{for all } \theta\ , \tag{3.6}$$

i.e., on the average, the confidence assertion is conservative. Berger and Lu (1987) have applied these ideas to Stein-type problems. Most recently, Brown and Hwang (1989) have shown that for the confidence set $[x-c,x+c]$, where $X=x$ is an observation from $f(x-\theta)$, the confidence procedure $<[x-c,x+c],1-\alpha>$ is admissible, where $1-\alpha = \int_{-c}^{c} f(t)dt$. The admissibility is with respect to the class of confidence procedures $<[x-c,x+c],\gamma(x)>$ (fixed c), where $\gamma(x)$ satisfies $E_\theta\gamma(x) \le 1-\alpha$ (frequentist validity) and the loss function is $L_C\left(\theta,\gamma(x)\right) = \left(\gamma(x) - I(\theta\epsilon[x-c,x+c])\right)^2$.

Another alternate strategy was described by Lindsay (1988), who suggested attaching both a frequentist and conditional confidence to a given set $C(x)$. Although this is a sensible approach, it is probably the case that practitioners are more comfortable with one number for a confidence assertion. Thus, this reasonable solution might not find acceptability in practice.

Returning to the questions posed in (3.4), we might now ask what is the reasonable requirement for the confidence assertion to be attached to $C(x)$. Taking into account the theories of relevant sets, and the manner in which confidence sets are used by practitioners, the following strategy seems most reasonable. For a set $C(x)$, assert confidence $\gamma(x)$ where $\gamma(x)$ satisfies (3.3) for some (possibly improper) prior $\pi(\theta)$. This strategy assures us that $<C(x),\gamma(x)>$ is conditionally acceptable. But, moreover, we require that $\gamma(x)$ be valid as a measure of frequentist confidence. Ideally, we would require that $\gamma(x)$ satisfy (3.5), which not only renders $<C(x),\gamma(x)>$ frequency valid, but also yields the conditional acceptability of $<C(x),1-\alpha>$. However, condition (3.5) may not always be attainable and, in such a case, we would settle for $\gamma(x)$ satisfying something like (3.6). This would give some frequentist acceptability to the procedure $<C(x),\gamma(x)>$.

These ideas have been investigated, in different forms, by Maatta and Casella (1987), Casella and Maatta (1988), Goutis (1989) for estimating a normal variance, and Hwang and Casella (1988) for estimation of a normal mean.

4. Discussion

The ideas behind conditional inference are deep, and here we have superficially sketched one line of work stemming from the developments of Fisher and Basu. There are many ideas in their work, both implicit and explicit, that haven't been mentioned. (For example, Basu is an advocate of the Likelihood Principle, and recent work by Casella and Robert (1988) suggest that violation of this principle immediately leads to the existence of relevant sets.) However, it is clear that the ideas of conditional inference play an important role in statistics.

Although it might be argued that searching for relevant sets is an occupation only for the theoretical statistician, we must remember that practitioners are going to make conditional (post-data) inferences. Thus, we must be able to assure the user that any inference made, either pre-data or post-data, possesses some definite measure of validity.

References

1. Basu, D. (1964). Recovery of Ancillary Information. Contribution to *Statistics*, Pergammon Press, Oxford, 7-20. Republished in *Sankhyā* 26 (1964), 3-16. Also in *Statistical Information and Likelihood. A Collection of Critical Essays by Dr. D. Basu*, 1988, J.K. Ghosh (ed.). Lecture Notes in Statistics. New York: Springer-Verlag.

2. Basu, D. (1981). On Ancillary Statistics, Pivotal Quantities, and Confidence Statements. In *Topics in Applied Statistics*, Y.P. Chaubey and T.D. Dwivedi (eds.). Concordia University, Montreal, 1-29. Also in *Statistical Information and Likelihood. A Collection of Critical Essays by Dr. D. Basu*, 1988, J.K. Ghosh (ed.). Lecture Notes in Statistics. New York: Springer-Verlag.

3. Berger, J. O. (1988). An Alternative: The Estimated Confidence Approach, in *Statistical Decision Theory and Related Topics IV*, Volume I, S.S. Gupta and J.O. Berger (eds.)., 85-90.

4. Berger, J. O. and Lu, K. L. (1987). Estimation of Normal Means: The Estimated Loss Frequentist Approach. Technical Report, Department of Statistics, Purdue University.

5. Bondar, J. V. (1977). A Conditional Confidence Principle. *Ann. Statist.* 5, 881-891.

6. Brown, L. D. (1967). The Conditional Level of Student's t-test. *Ann. Math. Statist.* 38, 1068-1071.

7. Brown, L. D. and Hwang, J. T. (1989). Admissibility of the Usual Confidence Set Using Frequentist Validity. Technical Report, Statistics Center, Cornell University.

8. Buehler, R. J. (1959). Some Validity Criteria for Statistical Inference. *Ann. Math. Statist.* 30, 845-863.

9. Buehler, R. J. and Feddersen, A. P. (1963). Note on a Conditional Property of Student's t. *Ann. Math. Statist.* 34, 1098-1100.

10. Casella, G. (1987). Conditionally Acceptable Recentered Set Estimators. *Ann. Statist.* 15, 1363-1371.

11. Casella, G. (1988). *Conditionally Acceptable Frequentist Solutions.* In *Statistical Decision Theory and Related Topics IV*, Volume I (S.S. Gupta and J.O. Berger, eds.), 73-117.

12. Casella, G. and Maatta, J. M. (1987). Conditional Properties of Generalized Bayes Variance Intervals. Biometrics Unit Technical Report, Cornell University.

13. Casella, G. and Robert, C. (1988). Nonoptimality of Randomized Confidence Sets. Biometrics Unit Technical Report, Cornell University.

14. Fisher, R. A. (1956). On a Test of Significance in Pearson's Biometrika Tables (No. 11). *J. Roy. Statist. Soc. Ser. B* 18, 56-60.

15. Fisher, R. A. (1959). *Statistical Methods and Scientific Inference*, second edition. New York: Hafner.

16. Goutis, C. (1989). Improved Interval Estimators of a Normal Variance and Generalizations. Ph.D. Thesis, Cornell University.

17. Hwang, J. T. and Casella, G. (1988). Frequentist Priors. Biometrics Unit Technical Report, Cornell University.

18. Lindsay, B. G. (1988). Discussion in *Statistical Decision Theory and Related Topics IV*, Volume I (S.S. Gupta and J.O. Berger, eds.), 94-98.

19. Maatta, J. M. and Casella, G. (1987). Conditional Properties of Interval Estimators of a Normal Variance. *Ann. Statist.* 15, 1372-1388.

20. Olshen, R. A. (1973). The Conditional Level of the F-Test. *J. Amer. Statist. Assoc.* 68, 692-698.

21. Olshen, R. S. (1977). Comment. *J. Amer. Statist. Assoc.* 72, 144-146.

22. Pierce, D. A. (1973). On Some Difficulties with a Frequency Theory of Inference. *Ann. Statist.* 1, 241-250.

23. Robinson, G. K. (1976). Properties of Student's t and the Behrens-Fisher solution to the two means problem. *Ann. Statist.* 4, 963-971.

24. Robinson, G. K. (1979a). Conditional Properties of Statistical Procedures. *Ann. Statist.* 7, 742-755.

25. Robinson, G. K. (1979b). Conditional Properties of Statistical Procedures for Location and Scale Parameters. *Ann. Statist.* 7, 756-771.

26. Scheffé, H. (1977). A Note on a Reformulation of the S-Method of Multiple Comparisons. *J. Ann. Statist. Assoc.* 72, 143-146.

27. Stein, C. (1961). Estimation of Many Parameters. Inst. Math. Statist. Wald Lectures. Unpublished.

28. Tukey, J. W. (1958). Fiducial Inference. Inst. Math. Statist. Wald Lectures. Unpublished.