



Conditional Confidence Statements and Confidence Estimators: Comment

Author(s): Lawrence D. Brown

Source: *Journal of the American Statistical Association*, Vol. 72, No. 360 (Dec., 1977), pp. 810-812

Published by: American Statistical Association

Stable URL: <http://www.jstor.org/stable/2286462>

Accessed: 25/03/2010 14:59

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=astata>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*.

<http://www.jstor.org>

us are prepared, and even prefer, for instance, in estimating a normal μ , to consider our statements as relating to a reference set in which the estimated precision s/\sqrt{n} is fixed and the population σ is regarded as variable. The interpretation is still in terms of relative frequency, with its objectivistic connotations. To deny that such an interpretation is a frequency interpretation is to suggest that those who, like myself, prefer it, are somehow committed thereby to an exclusively subjective view of probability, or that we do not subscribe to the law of large numbers.

One may suspect that the insistence by the $N(P)W$ † school on their special form of frequency interpretation arises partly from confusion between the planning and the inference stage of an investigation, and partly from confusion of the fully and exactly parameterized inferences at issue here with the simple test of significance and procedures related to it. In the latter case, where the likelihood ratio obviously does not exist, we may need to know the stopping rule to guard against, for example, mistaken interpretations of procedures in which observations continue until an observed mean differs from zero by a fixed multiple of its standard error.

This leads to my final point. I am (and have been—see Barnard et al. 1962, p. 334, para. 2, lines 1–5) an “advo-

† The (P) is bracketed because it does not appear to me that Pearson insists on this interpretation.

cate of the likelihood principle” “in the realm of hypothesis testing” only if this latter phrase is understood in the NPW sense, i.e., as relating to problems of discrimination between fully and exactly parameterized hypotheses with no special ordering or group structure, for example. Apart from the simple test of significance there is the case of what I have called pivotal inference, where there is some uncertainty about the form of the probability distribution so that the parameters must be defined by reference to their effect on the observations. The likelihood function here is not precisely defined, though it may become so with additional information. Without such information, the likelihood principle is inapplicable here also; and even with this information, two results giving the same likelihood function may differ in their pivotal structure.

REFERENCES

- Barnard, George A. (1949), “Statistical Inference,” *Journal of the Royal Statistical Society, Supplement*, 11, 115–139.
 ——— (1975), “Conditional Inference Is not Inefficient,” *Scandinavian Journal of Statistics*, 3, 132–134.
 ———, Jenkins, G.M., and Winsten, C.B. (1962), “Likelihood Inference and Time Series,” *Journal of the Royal Statistical Society, Ser. A*, 125, 334.
 Cox, D.R. (1975), “The Role of Significance Tests,” address to Meeting of European Statisticians, Grenoble, France.
 Pierce, David A. (1973), “On Some Difficulties in a Frequency Theory of Inference,” *Annals of Statistics*, 1, 241–250.

Comment

LAWRENCE D. BROWN*

A procedure justifiable by frequentist theory must be “anticipatory” as well as frequentist. For example, consider Wald’s sequential probability ratio test (SPRT) for testing between two simple hypotheses. It does not suffice merely to look at conditional properties of the terminal decision rule given the stopping time—even though these possess a valid (conditional) frequentist interpretation. Instead, all possible consequences of the actions must be anticipated from the outset. Thus the sequential test procedure consists of both a stopping rule as well as a terminal decision rule; and various procedures are compared according to the combined frequency properties of their stopping and terminal decision rules.

To some, it had seemed to pose an irresolvable paradox to have a theory of conditional confidence which also

possessed this anticipatory frequentist property. Kiefer has decisively demonstrated that this is not so. The key step is to anticipate from the outset the class of conditioning statements which will later be made. In this fashion the apparent paradox can thus be resolved. Furthermore, as Kiefer discusses here and demonstrates elsewhere (Kiefer 1976, 1977), a potentially satisfactory theory can be developed, at least for multidecision problems involving simple hypotheses.

However, it is my feeling that the theory is still incomplete. True, the conditioning is anticipated. However, there is no provision allowing for an accurate anticipation of the goals for such confidence statements.

Consider the problem discussed in Section 6, Example 1(b.2). For simplicity set $n = 2$. Then X_1, X_2 are a sample from the uniform distribution on $[\omega - \frac{1}{2}, \omega + \frac{1}{2}]$, $T = \bar{X}$, $W = X_2 - X_1$ and the (unconditional) confidence interval with coefficient $\gamma = 1 - (1 - d)^2$ is $(T - h/2, T + h/2)$, where $h = \min(d, 1 - W)$. (While

* Lawrence D. Brown is Professor, Department of Statistics, Rutgers University, New Brunswick, NJ 08903. Research was supported in part by National Science Foundation Grant MCS 75-23343 A01.

this procedure is not the “optimum” procedure of Example 1(a), it too has an optimality property which may be suited to certain applications: It has the maximum probability of coverage among unbiased procedures having confidence intervals of length $\leq d$.) The conditional procedure which uses this interval and the conditioning $Z = W$ is “of more interest here,” as Kiefer himself observed. Indeed, I suggest that most careful practitioners would prefer to condition on Z (which is an ancillary statistic here) rather than quote the unconditional confidence level. However, both procedures are admissible by Kiefer’s optimality criteria. In particular for $d = \frac{1}{2}$ the unconditional confidence is $\gamma = \frac{3}{4}$; for the unconditional procedure

$$G(t) = 0 \quad t \geq \frac{3}{4}$$

$$= \frac{3}{4} \quad t < \frac{3}{4} ,$$

and for the procedure conditioning on Z

$$\tilde{G}(t) = 0 \quad t \geq 1$$

$$= \frac{1}{2}t - \frac{1}{4} \quad \frac{1}{2} \leq t < 1$$

$$= \frac{3}{4} \quad t < \frac{1}{2} .$$

Hence in particular $\tilde{G}(4/5) = 3/8 > 0 = G(4/5)$ while $\tilde{G}(3/5) = 7/12 < 3/4 = G(3/5)$, so that neither procedure dominates the other according to Kiefer’s criteria.

Why then should the conditional procedure be preferred in practice? I think it is because intuitively in such a problem one compares procedures on the basis of the stated conditional confidence level, which is

$$\Gamma^Z = \Pr\{\omega \in (T - h/2, T + h/2) | Z\}$$

here, rather than on the basis of

$$G(t) = P_\omega\{\Gamma_\omega > t; \omega \in (T - h/2, T + h/2)\} ;$$

and that the utility, $U(\Gamma)$, of such a statement is increasing and convex in Γ . To be precise one wishes to condition so that $(\Gamma_\omega)^Z$ will be constant in ω , and the expectation $E(U(\Gamma)) = \int U(\gamma) d \Pr(\Gamma < \gamma)$ will be large. It can be shown that for this problem the conditioning $Z = W$ is optimum in this sense—for any increasing convex U it maximizes $E(U(\Gamma^Z))$ among all possible partitions. (In the terminology of Brown (1977), the partition $Z = W$ is the “totally admissible (second sense)” partition for the confidence intervals $(T - h/2, T + h/2)$.)

(However, if U is not convex then the above conditioning will generally not be admissible relative to the given U . This shows the importance of making some determination of this utility function before deciding what conditional procedure is to be used. Statisticians (and their clients) are not yet used to considering such an issue, but Kiefer’s anticipatory theory plus the admissibility criteria outlined above indicate that some effort should be spent on such considerations.)

Examination of another situation may further clarify the importance of constructing an anticipatory frequentist framework which includes an evaluation of the

utility of the eventual confidence statement. Kiefer mentions briefly (in Section 6) the work of Buehler (1959) and others concerning the conditional properties of Student’s t intervals for an unknown normal mean. This work branches from the discussion of relevant subsets in Fisher (1956). The pertinent technical fact is that the conditional coverage of the usual Γ confidence interval is $\geq \Gamma + \epsilon > \Gamma$ given $(\bar{X}, s) \in C$. (It is shown in Brown (1967) that one may choose $C = \{(\bar{X}, s) : |\bar{X}/s| \leq k\}$ for a wide range of constants k .) This observation is initially disturbing, for it suggests that if $(\bar{X}, s) \in C$, one may quote a higher value of Γ or alternatively use a shorter confidence interval. However, to do merely this destroys the usual (unconditional) frequentist interpretation for the confidence procedure, since it does not anticipate the possibility that $(\bar{X}, s) \notin C$. A formally proper resolution in this case would be to also quote a lower conditional confidence value when $(\bar{X}, s) \notin C$ or to lengthen the interval. Such a procedure would be consistent with the principles Kiefer lays down.

An interesting recent paper by Robinson (1976) contains the evaluation that the existence of “positively biased relevant” conditioning sets such as C , in the preceding paragraph, does not constitute a serious deficiency of the usual Student’s t procedure. (See also Bondar 1977.) His point of view is the conservative one that it is a benefit that the conditional coverage is $\geq \Gamma + \epsilon$ given $(\bar{X}, s) \in C$. Basically, I agree with this part of his point of view. After all, the usual definition of a Γ confidence set requires only that the probability of covering the true value be at least Γ ; not that it be exactly Γ (or even that its infimum be Γ). See, e.g., Lehmann (1959, p. 193).

But again, the proper frequentist interpretation must anticipate the possibility that $(\bar{X}, s) \notin C$. Robinson plays down the existence of this “negatively biased semi-relevant” subset, \bar{C} , for which the conditional probability of coverage is always $< \Gamma$ but the supremum is Γ . This is not justified. By the same conservative logic and definition of Γ confidence, one is led to the necessity of quoting the infimum of the conditional coverage given $(\bar{X}, s) \in \bar{C}$. Take $n = 2$ for simplicity of computation. Then, for the usual 90 percent interval, $\bar{X} \pm 6.31s$, the conditioning $|\bar{X}/s| < 12.7$ as suggested in Brown (1967, Section 3) yields $\inf \Gamma_{\mu, \sigma^C} = .947$ but, unfortunately, $\inf \Gamma_{\mu, \sigma^{\bar{C}}} = 0$. A more reasonable two-subset conditioning rule would be, for example, based on the conditioning $|\bar{X}/s| < 1$. Then $\inf \Gamma_{\mu, \sigma^C} = .933$ and $\inf \Gamma_{\mu, \sigma^{\bar{C}}} = .8$. It can be seen that $E_{\mu, \sigma}(\inf_{\mu', \sigma'} \Gamma_{\mu', \sigma'})$ ranges from a high value of .867 (at $\mu = 0$) to a low value of .8 (as $\mu/\sigma \rightarrow \pm \infty$). This illustrates the general phenomenon: If one intends to make conditional Γ confidence statements with the usual (conservative) interpretation for such a statement, then it does not pay to condition as long as the utility of such a statement is linear in Γ . If the utility is not linear in Γ , then conditioning may be advantageous.

The conclusions of the preceding paragraph and those

in the preceding example are entirely within the anticipatory frequentist framework which Kiefer sets forth. However, the admissibility criteria implied by the consideration of $\inf_{\mu, \sigma} \Gamma_{\mu, \sigma}$ are different from any raised by Kiefer. I have defined and have discussed some features of such criteria in Brown (1977). I argued at more length there that for *some* multiple decision problems involving conditional confidence statements the consideration of $\inf_{\omega} \Gamma_{\omega}$ is unambiguously dictated by the structure of the problem. By extension, there are *many* such problems for which such admissibility criteria are appropriate.

The considerations previously outlined imply the use of procedures for which Γ_{ω}^b is constant in ω for each value of $Z = d$; see Brown (1967, Section 6). While this symmetry is appropriate for many situations, it should be emphasized that there are other problems for which it is not. Suppose, for example X_1, X_2, \dots, X_n are iid $N(\mu, 8)$ random variables with $\mu = \pm 1$. It may be desired to condition with $B = (\omega, s)$ (as in Kiefer's Example 2(b)) so that $\Gamma_1^s \geq .9$ and $\Gamma_{-1}^s \geq .7$. If $n = 16$ this implies setting (approximately)

$$C_1^s = \{\bar{X} \in (1.6, \infty)\} \quad C_{-1}^s = \{\bar{X} \in (-\infty, -2.4)\} .$$

However, in such a situation one must be careful to accurately anticipate the use of the procedure in order to construct an entirely proper frequentist interpretation. Note that the decision partition of C^w is not determined by the conditions so far stated for the problem. It really may be the case that an observation in C^w will be followed by a continuation of sampling, rather than by the terminal decision that $\mu = +1$ or that $\mu = -1$. In other words, the actual procedure has the following structure: Observe X_1, \dots, X_{16} . If a strongly believed terminal decision (i.e., with $\Gamma_1^s = .9$ and $\Gamma_{-1}^s = .7$) can be made, then make it. If not, then continue sampling X_{17}, \dots, X_{n_2} , etc.

A natural modification of this scenario is a fully sequential problem in which X_1, X_2, \dots are observed sequentially. At each stage one may either stop and make a strongly believed terminal decision (conditional on the sample size of the stopped sample) or continue sampling. It appears to us to be difficult to calculate what should be the optimum procedure in such a situation, even if one neglects overshoot in the computations. The usual SPRT does not have the desired conditional properties;

though it is a plausible conjecture that one would wish to use some sort of generalized SPRT.

Kiefer also raises the possibility that the statistician may wish to use procedures which estimate the conditional confidence, rather than state its exact value (or its approximate value as one might do in sequential problems with calculations neglecting overshoot). The suggestion is an interesting one and should be a challenge for future research. Kiefer's development of this idea is as yet unsatisfactory when compared to the presentation of conditional confidence procedures in the body of his paper. The formulation is not yet anticipatory in a way which leads to a suitable frequentist comparison of various procedures. There are various ways in which this may be corrected. For example, one might include as an additional component of the loss function a cost due to making an incorrect estimate of the conditional confidence (depending on the difference between the true and estimated values). In theory, various procedures could then be compared on frequentist grounds. The computations required appear to us to be extremely difficult. Furthermore, it is not clear to us that a loss function like that described previously really fits the practical situations which call for estimated conditional confidence. Clearly, considerably more thought concerning the formulation and more mathematical analysis will be required to begin to develop this interesting and challenging suggestion.

REFERENCES

- Bondar, James V. (1977), "On a Conditional Confidence Principle," to appear in *Annals of Statistics*.
- Brown, Lawrence D. (1967), "The Conditional Level of Student's *t*-Test," *Annals of Mathematical Statistics*, 38, 1068-1071.
- (1977), "An Extension of Kiefer's Theory of Conditional Confidence Procedures," to appear in *Annals of Statistics*.
- Buehler, R.J. (1959), "Some Validity Criteria for Statistical Inference," *Annals of Mathematical Statistics*, 30, 845-863.
- Fisher, R.A. (1956), *Statistical Methods and Scientific Inference*, Edinburgh: Oliver and Boyd.
- Kiefer, Jack (1976), "Admissibility of Conditional Confidence Procedures," *Annals of Statistics*, 4, 836-865.
- (1977), "Conditional Confidence Approach in Multidecision Problems," to appear in *Proceedings of the Fourth International Symposium on Multivariate Analysis*.
- Lehmann, Erich L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley & Sons.
- Robinson, G.K. (1976), "Properties of Student's *t* and of the Behrens-Fisher Solution to the Two Means Problem," *Annals of Statistics*, 5, 963-971.