

Expression (16) suggests that it might be impossible to find a prior density that produces confidence limits having coverage error of order  $O(n^{-3/2})$ ; see DiCiccio, Keller and Martin (1992).

Many of the likelihood adjustments and distributional corrections discussed in the paper can be viewed, at least to error of order  $O_p(n^{-1})$ , in terms of the quantities  $z_0$  and  $a$  that arise in Efron's (1987)  $BC_a$  confidence limits. Efron defined  $z_0 = \Phi^{-1}\{\text{pr}(\hat{\psi} \leq \psi_0)\}$ , and  $a$  is related to the skewness of the score function; both  $z_0$  and  $a$  are of order  $O(n^{-1/2})$ . In the setting of Section 4.2, DiCiccio and Efron (1992) and Efron (1993) showed that  $E(r_p) = -z_0 + O(n^{-1})$  and that  $r_p + z_0$  has the standard normal distribution to error of order  $O(n^{-1})$ . Moreover,

$$E\{l'_p(\psi)\} = (a - z_0)\{-l''_p(\psi)\}^{1/2} + O(n^{-1})$$

## Comment

A. P. Dawid and C. Goutis

Nancy Reid has presented a clear and valuable overview of the uses of conditioning, and of associated techniques of analysis. We wish to focus on some difficulties which can arise from too uncritical an attitude to conditional inference.

It is implicit in Reid's account, as in most others, that the goal of conditional inference has been achieved when we have identified the appropriate conditional "frame of reference" (Dawid, 1991). From that point on, it is implied, we should be free to use any favourite method of inference within that new frame. However, a more thorough-going analysis casts doubt on this assumption. This doubt may be evidenced in several related ways.

First there is the problem of nonuniqueness of (maximal) ancillary statistics, and the consequent arbitrariness, in general, of the conditional frame of reference. The collected works of Basu (1988), which deal thoroughly with these topics, should be required reading for anyone contemplating conditional inference. For example, if  $(X_i, Y_i)$  have a bivariate normal distribution with known vari-

and

$$l_c(\psi) = l_p(\psi) - (a - z_0)\{-l''_p(\psi)\}^{1/2} + O(n^{-1}).$$

As many authors have noted, adjustment of the log profile likelihood function  $l_p(\psi)$  reduces the bias of the profile score. Also,  $E(r_c) = -a + O(n^{-1/2})$ , and  $r_c + a$  has the standard normal distribution to error of order  $O(n^{-1})$ . Further details are given in DiCiccio and Efron (1995).

### ACKNOWLEDGMENTS

The first author's research is supported by NSF Grant DMS-93-05547. This is paper BU-1283-M in the Biometrics Unit, Cornell University, Ithaca, New York 14853. The third author's research is supported by NIH Grant No. RO1-CA61120.

ances and unknown correlation  $\rho$ , each of  $\mathbf{X} = (X_1, \dots, X_n)$  and  $\mathbf{Y} = (Y_1, \dots, Y_n)$  is ancillary, and inference conditional on either appears equally justified. We cannot, however, condition on both, since  $(\mathbf{X}, \mathbf{Y})$  reproduces the whole sample.

Next, there is Birnbaum's (1962) celebrated demonstration that acceptance of both the sufficiency and conditionality principles demands acceptance of the likelihood principle—and is thus incompatible with any method of inference which does not respect that principle. A much weaker version of this argument and conclusion, which nevertheless implies the irrelevance of optional stopping and is hence incompatible with many common forms of inference, is given by Dawid (1986).

Then there is the "conflict between conditioning and power" mentioned in Section 6.2. A concrete example, based on Cox (1958a), is analysed in Dawid (1983, pages 99–100). In a problem with point null and alternative hypotheses, and a simple experimental ancillary, the rule "use the likelihood ratio test with size  $\alpha = 0.05$ ," if applied conditionally on the ancillary, does not agree with any unconditional likelihood ratio test and is thus less powerful than the overall 0.05-level test (which has differing conditional  $\alpha$ -levels). However, the Neyman–Pearson lemma, which simply requires use of some likelihood ratio test, can nonetheless be ap-

---

*Philip Dawid is Professor of Statistics and Costas Goutis is Lecturer in Statistics, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom.*

plied in ways which do not conflict with conditioning. For example, the rule “use the likelihood ratio test which minimises  $c\alpha + \beta$ ” (for fixed  $c$ ;  $c = 1$  might appear attractive) yields the same conclusions whether applied unconditionally or conditionally: this is related to the fact that it may be re-expressed as requiring the acceptance of the alternative hypothesis when the likelihood ratio in its favour exceeds  $c$ , and is thus compatible with the likelihood principle. Here we have an indication that, if we wish to make conditional inference, some prescriptions for inference may be more acceptable than others. In particular, pre-fixing the size  $\alpha$  of a likelihood ratio test (a rule for which we are not aware of any theoretical justification), rather than the cutoff value  $c$ , is problematic.

Ideally, we should use forms of inference (such as likelihood or Bayesian methods) which deliver the *same* solution, whether applied unconditionally or conditionally on any relevant ancillary (this interpretation of the “conditionality principle” is discussed in Dawid, 1991, Rejoinder). There can then be no conflict between conditioning and other inferential desiderata. In particular, in the asymptotic context which Reid emphasises, this suggests the use of those formulae which are insensitive to the specific choice of approximate ancillary. This typically can hold only up to  $O(n^{-1})$ , and this approach thus suggests that analysis of further terms in asymptotic expansions of conditional distributions is fundamentally pointless.

Another variation on the theme of conflicting frames of reference is explored by Dawid (1975), building on Durbin (1969). In a simple example, there is a choice between two irreconcilable frames of reference for inference about a parameter  $\theta$ : one based on marginalisation to the minimal sufficient statistic, the other based on conditioning on a natural experimental ancillary (the example was phrased in terms of  $S$ -sufficiency and  $S$ -ancillarity, but this is not crucial). Many forms of inference give necessarily different results in these two frames (see, e.g., Dawid, 1977, Example 6). However, once again the conflict can be resolved by confining attention to those forms of inference (e.g., Bayesian inference) which deliver the same conclusion (exactly or asymptotically) in both frames.

When we come to deal with problems with nuisance parameters, we should be careful not to use terminology or notation which adds further to the already considerable difficulties. Thus, after (3.2), Reid says that “the nuisance parameter  $\lambda$  has been eliminated in the conditional distribution”  $f(s_1|s_2; \psi)$  and calls  $s_2$  “sufficient for  $\lambda$ .” But what is  $\lambda$ ? If, for example,  $\theta = (\mu, \sigma^2)$  and  $\psi = \sigma^2$ , as in the special case of Example 3.6, we can define  $\lambda = \mu$  or  $\lambda = \mu/\sigma$ , or many other choices. The “parameter of interest” is always clearly defined; “the” nuisance parameter is not. That is one reason why it is not appropriate, in Example 3.6, to regard  $\bar{y}$  as “sufficient for  $\mu$ ”; a more accurate description would be “sufficient given  $\sigma^2$ .”

Similarly, it is inappropriate to consider  $s^2$  as “sufficient for  $\sigma^2$ ,” although this is even more tempting since the distribution of  $s^2$  depends on  $\sigma^2$  alone. However, even within the framework of standard frequentist decision theory one can improve upon inference based on  $s^2$  alone, for example, using the improved estimator of variance given by Stein (1964), which uses information contained in  $\bar{y}/s$ . A similar phenomenon appears in inference about  $\mu$ : the usual  $t$ -pivot is not necessarily the quantity one should base inference on. There is again information hidden in  $\bar{y}/s$ , which can be suitably exploited for specific purposes (see, e.g., Goutis and Casella, 1992). Perhaps ironically, in both cases one is able to improve upon standard inference using conditional arguments, using the distributions of  $s^2$  or  $\bar{y}$ , conditionally on  $\bar{y}/s$ .

Reid mentions in passing some aspects of a Bayesian approach to sufficiency and ancillarity in the presence of nuisance parameters. A nonasymptotic study of these Bayesian concepts, and their relation to their classical counterparts, was undertaken by Dawid (1980). It is encouraging to see recent work on modified likelihood functions delivering solutions with an approximate Bayesian justification. However, unless such likelihood functions are used appropriately for inference, there will still be possibilities for internal inconsistency. However much they might wish it otherwise, those who would take seriously the use of conditional frames of reference cannot ignore the constraints on their inferential freedom that this imposes.