is a point in R^2 whose expectation is $x(\lambda_j)$. Consequently,

$$h(R_\phi^{-1}(y_j - \omega)/\rho) - 1$$

has zero mean for each $j = 1, \ldots, n$, and nonzero mean off the template. These n elementary estimating functions can be combined linearly in an optimal manner to obtain estimates of the parameters of interest without involving the nuisance parameters.

It would be of considerable interest to know whether the preceding method can be extended in useful ways, possibly to nonlinear distortions of templates.

Comment: Alternative Aspects of Conditional Inference

George Casella, Thomas J. DiCiccio and Martin T. Wells

The roles of conditioning in inference are almost too varied to be summarized in one paper. Professor Reid has done a wonderful job of explaining and illustrating some of these roles. We expand on a number of her points, with particular attention to the practical uses and implementation of the methods. We also discuss some overall goals of conditional inference and alternative ways of achieving them.

1. INTRODUCTION

The techniques of conditional inference are a collection of extremely powerful tools. They allow for the construction of procedures with extraordinarily good properties, especially in terms of frequentist asymptotic behavior. In fact, in many cases these procedures are so good that one begins to wonder why they are not more widely used; that is, although statistics methodology journals often contain articles on conditional inference, such techniques have not really found their way into the arsenal of the applied statistician and thus into the subject matter journals. There are, we feel, two reasons for this. One is that, unfortunately, the procedures are fairly complex in their derivation and, hence, in their implementation, and for that reason alone they may not have received thorough consideration. The second reason is somewhat more subtle, but perhaps more important. If an experimenter uses conditional inference techniques, the goal of the analysis (and the exact type of ultimate inference to be made) is not at all clear. In Section 1, Reid recounts four roles of conditional inference that are identified by Cox (1988). However, to a prospective user of these techniques, these goals are vague, and the effort needed to actually implement these solutions can be prohibitive. For example, consider Example 3.3, used to illustrate conditional inference techniques in the estimation of the gamma shape parameter when the scale parameter is unknown. The density given by (3.3) and (3.5), which contain components that are "difficult to calculate," is offered as a conditional inference solution to the problem. This density can be used to test an hypothesis or, with some difficulty, to calculate a confidence interval, but the details of carrying out these procedures are quite complex. Moreover, if one is interested in a point estimate and evaluation of the performance of the estimate, this density will not suffice. Rather, one might use a saddlepoint approximation (Reid, 1988) for the density of the maximum likelihood estimate, yielding a density proportional to

$$\Gamma^n(\hat{\psi})\Gamma^n(\psi)\{\hat{\psi}\Delta'(\hat{\psi}) - 1\}^{1/2} \cdot \exp[n\{(\hat{\psi} - \psi)\Delta(\hat{\psi}) + \hat{\psi} - \psi \ln \hat{\psi}\}],$$

where $\Delta(\cdot)$ is the digamma function. Although the approximation is remarkably accurate, computation of the normalizing constant (which involves integrating this function with respect to $\hat{\psi}$) is quite demanding, limiting the use of the formula. Thus, the "naive" user is shortchanged. Rather than the accurate approximations and, hence, more precise inference, the user gets only halfway there and can be faced with calculations of prohibitive complexity.

George Casella is Professor, Biometrics Unit, and Thomas J. DiCiccio and Martin T. Wells are Associate Professors, Department of Social Statistics, Cornell University, Ithaca, New York 14853.