Comment

Bruce G. Lindsay and Bing Li

The two papers before us consider the same basic problem: statistical inference for a finite dimensional parameter, possibly in the presence of nuisance parameters. The strikingly different results arise from the differing approaches to making modelling assumptions. Whereas Professors Liang and Zeger would have us make the minimal assumptions necessary to achieve the inference, Professor Reid shows that a completely believed parametric model assumption can be turned into a gold mine of more precise asymptotic approximations. We wish to discuss here some aspects of the middle ground between these two extremes and how it relates to conditional inference.

Perhaps it is useful to make a distinction between the goals that we attempt to achieve by employing conditional inference and the natural consequences to which conditional inference leads. These goals are, for example, (i) to make the assessment of the precision of a statistical method as true to the experiment that actually occurred as possible and (ii) to make the inference about the interest parameter as accurate as possible by minimizing the effect of the estimation of the nuisance parameter. If an appropriate parametric model is applicable, as it is in many important examples, then conditional inference is a powerful means to achieve these purposes. However, we want a statistical procedure to possess these desirable properties whether or not we have suitable ancillary statistics to condition on, and whether or not we have a fully prescribed model under which we can talk about conditional probability in accurate terms. Although the principle of conditioning on ancillary statistics or on sufficient statistics for the nuisance parameter is very clear when we have rigidly prescribed a parametric model, we ask what its statistical meaning might be outside those contexts. We offer here some illustrations of how the idea of projection, as used by Liang and Zeger, can be useful in achieving these goals under such circumstances.

Bruce G. Lindsay is Professor and Bing Li is Assistant Professor, Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802.

We start with a basic tool, the Bhattacharyya scores. Let $X = (X_1, \dots, X_n)$ be independent random observations with density $f(x; \theta)$. The Bhattacharyya scores B_i, $i = 0, 1, 2, \dots$, are defined by

$$(1)\quad B_0 = 1, \quad B_k = \frac{\partial^i f(x;\theta)/\partial\theta^i}{f(x;\theta)}, \quad i = 1, 2, \dots.$$

For example, if we write the log-likelihood function as $l(\theta; X)$, then B_1 is the score function $\dot{l}$ and B_2 is $\ddot{l} + \dot{l}^2$. We consider $\{B_i\colon\ i = 0, 1, \dots, k\}$ as vectors in the Hilbert space of square-integrable functions $h(x, \theta)$, with inner product $E\{h_1(X, \theta)h_2(X, \theta); \theta\}$. We use $\mathscr{B}$ to denote the subspace spanned some or all Bhattacharyya scores; $P_{\mathscr{B}}$, the orthogonal projection onto $\mathscr{B}$; and $I - P_{\mathscr{B}}$, the orthogonal projection onto the orthogonal complement of $\mathscr{B}$. We wish to discuss how one can use these scores to evaluate or improve conditional-type properties of inference.

Our first illustration is an optimality property of the observed Fisher information. Let $\hat{\theta}$ be the maximum likelihood estimate. Conditional inference suggests that the assessment of the precision of $\hat{\theta}$ be conditioned on any ancillary statistics. The results of Efron and Hinkley (1978) indicate that, for translation families and numerous other cases, the variance of $\hat{\theta}$, conditioned upon an ancillary, is approximated by the inverse of the observed Fisher information. Here, the goal is to make the precision assessment more relevant to the realized experiment, and it is achieved by drawing inference conditioning on ancillary statistics.

It is possible, however, to achieve this goal without specifying an ancillary statistic, either exact or approximate. Consider the following (unconditional) minimization problem: choose a statistic $T(X)$ that minimizes the mean squared error

$$E_{\theta_0}\{(\hat{\theta} - \theta_0)^2 - T(X)\}^2.$$

Lindsay and Li (1995) demonstrated that, among a wide class of statistics, the asymptotically optimal choice of $T(X)$ is once again the inverse of the observed Fisher information. The result relies not on the specification of ancillaries or approximate ancillaries, which may be difficult to obtain under some circumstances, but rather an asymptotic Cramér–Rao–type argument based on the projection of $(\hat{\theta} - \theta_0)^2 - T(X)$ onto $\mathscr{B} = \text{span}\{B_i\colon\ i = 0, 1, 2\}$.