Comment

Thomas A. Severini

The papers by Reid and by Liang and Zeger cover several important areas of statistics. I will confine my comments to three topics considered in Professor Reid's excellent review of conditional inference.

1. APPROXIMATE SUFFICIENCY IN THE PRESENCE OF A NUISANCE PARAMETER

In Section 5 Reid suggests that it may be possible to approximately eliminate a nuisance parameter by conditioning on a statistic that is approximately sufficient for fixed values of the parameter of interest. I now present some recent results regarding this issue; further details and additional results are available in Severini (1993, 1994a).

Consider a model parameterized by a scalar parameter of interest ψ and a nuisance parameter λ; for simplicity, take λ also to be a scalar although the results hold more generally. A natural approach to consider, given its optimality in exponential family models, is to take $S_1 = \hat{\psi}$, the MLE of ψ, and $S_2 = \hat{\lambda}_\psi$, the MLE of λ for fixed ψ. In exponential family models in which ψ is a linear function of the canonical parameter this leads to exact methods of inference with well-known optimality properties.

In general, under standard regularity conditions,

$$\begin{aligned}&\Pr(S_1 \geq s_1 | S_2 = s_2;\ \psi, \lambda_0 + \varepsilon n^{-1/2})\\ &\quad - \Pr(S_1 \geq s_1 | S_2 = s_2; \psi, \lambda_0)\\ &\quad = \frac{i_\theta}{2}\gamma\varepsilon^2 n^{-1/2} + O(n^{-1}),\end{aligned}$$

where i_θ is the asymptotic variance of $\sqrt{n}(\hat{\psi} - \psi)$,

$$\gamma = \frac{1}{n}E\left[\left(\frac{\partial l(\psi, \lambda_0)}{\partial \psi} - \beta\frac{\partial l(\psi, \lambda_0)}{\partial \lambda}\right)\frac{\partial^2 l(\psi, \lambda_0)}{\partial \lambda^2};\ \psi, \lambda_0\right],$$

$$\beta = \frac{E[\partial^2 l(\psi, \lambda_0)/\partial\psi\,\partial\lambda; \psi, \lambda_0]}{E[\partial^2 l(\psi, \lambda_0)/\partial\lambda^2; \psi, \lambda_0]}$$

and l is the log-likelihood function. Hence, this approach is successful in approximately eliminating the nuisance parameter, to the order considered, provided that $\gamma = O(n^{-1/2})$. In particular, if the distribution of the data for fixed ψ forms a full exponential family model, then $\gamma = 0$; in general, γ is a measure of how close the distribution is to full exponential family form.

Consider an exponential family model with log-likelihood function of the form

$$l(\psi, \lambda) = g(\psi, \lambda)t_1 + \lambda t_2 - k(\psi, \lambda),$$

where (t_1, t_2) is the sufficient statistic and is of order $O_p(n)$ and $\partial g(\psi, \lambda)/\partial\psi \neq 0$. Then $\gamma = O(n^{-1/2})$ requires that

$$\frac{\partial^2 g(\psi, \lambda_0)}{\partial \lambda^2} = 0. \quad (1)$$

Under this condition, $g(\psi, \lambda)$ is linear in λ for each fixed ψ so that there exists a one-dimensional sufficient statistic for λ for each fixed ψ and the conditional distribution of the data given $\hat{\lambda}_\psi$ is exactly free of λ. When (1) does not hold, we could attempt to eliminate λ by conditioning on $S_2 = (\hat{\lambda}_\psi, A)$ for some statistic A, but it is clear that such a statistic S_2 would have to be equivalent to the sufficient statistic in the original, unrestricted, model. Hence, in full exponential family models when the exact theory of conditional inference fails, an approximate theory fails as well, at least using the approach considered here.

This suggests that, for inference in the presence of a nuisance parameter, methods based on the marginal distribution of some quantity, such as the modified likelihood ratio statistic r^*_ψ, are likely to be more generally applicable than methods based on a theory of approximate conditional inference.

2. THE RELATIONSHIP BETWEEN BAYESIAN INFERENCE AND CONDITIONAL INFERENCE

One advantage of Bayesian inference over non-Bayesian methods of inference is in the treatment of problems involving a nuisance parameter. In Bayesian inference, any nuisance parameter can be eliminated by integrating it out, at least in principle. Given the formal appeal of Bayesian methods, as well as some well-known optimality properties, it is of interest to determine when a non-Bayesian method of eliminating a nuisance parameter, such as conditional inference, corresponds to Bayesian inference with respect to some prior distribution.

Thomas Severini is Associate Professor, Department of Statistics, Northwestern University, 2006 Sheridan Road, Evanston, Illinois 60208-4070.