On the accuracy of Approximate Studentization

Fraser, D.A.S., Wong, A.C.M.

With a parametric model, a measure of departure for an interest parameter is often easily constructed but frequently depends in distribution on nuisance parameters; the elimination of such nuisance parameter effects is a central problem of statistical inference. Fraser & Wong (1993) proposed a nuisance-averaging or approximate Studentization method for eliminating the nuisance parameter effects. They showed that, for many standard problems where an exact answer is available, the averaging method reproduces the exact answer. Also they showed that, if the exact answer is unavailable, as say in the gamma-mean problem, the averaging method provides a simple approximation which is very close to that obtained from third order asymptotic theory. The general asymptotic accuracy, however, of the method has not been examined. In this paper, we show in a general asymptotic context that the averaging method is asymptotically a second order procedure for eliminating the effects of nuisance parameters.

Some keywords: Averaging; Confidence distribution function; Studentization.

1. Introduction

Studentization can be viewed as the general process of eliminating nuisance parameters from a pivotal quantity and has its origins in the Student (1908) investigation of the mean of the $N(\mu, \sigma^2)$ distribution: the measure of departure $(\bar{y} - \mu)/(\sigma/\sqrt{n})$ is standard normal but involves the nuisance parameter σ; the adjusted measure of departure $(\bar{y} - \mu)/(s_y/\sqrt{n})$ is free of σ and has the Student $(n-1)$ distribution.

More generally, many problems lead easily to a factorization of the density for a sufficient statistic in the form

$$f(\mathbf{x}; \psi, \lambda) = f_c(t|s; \psi, \lambda) f_m(s; \lambda) J(s, t) \ , \tag{1}$$

where ψ is the parameter of interest, λ is the nuisance parameter, and (s, t) is equivalent to $\mathbf{x}$ with Jacobian $J(s, t)$; for this, typically, s will be a statistic and t will be a pivotal quantity for assessing ψ, as in the above mentioned Student example. As ψ is completely contained in the first factor, that factor is commonly used for inference concerning ψ; however, this conditional density still depends on the nuisance parameter λ.

For the case of scalar ψ and λ, Fraser & Wong (1993) introduced a nuisance-averaging method as a device to eliminate λ from the conditional density $f_c(t|s;\psi,\lambda)$. Wong (1995) extended the averaging method to the case where λ is a vector. Fraser & Wong (1993) showed that for the location-scale problem, inference concerning the location parameter obtained via the averaging method agrees with the general Student-type analysis obtained from standard conditional theory. For other problems, such as the gamma-mean problem, the averaging method leads to a simple inference quantity that agrees closely with that obtained by nontrivial iterative and numerical calculations based on third-order asymptotic theory.

The averaging method is reviewed in Section 2. In Section 3 the averaging method is shown in a general asymptotic context to have second-order accuracy for the elimination of nuisance parameter effects. Some concluding remarks are given in Section 4.

2. The parameter-averaging method

Consider the factorization (1) with scalar parameters ψ and λ. As the nuisance parameter λ in the marginal density of s is isolated from ψ, it is typical (Kalbfleisch & Sprott, 1973) to use $f_m(s;\lambda)$ for inference concerning λ. The *confidence distribution function* $p(\lambda)$ for the parameter λ at a scalar data point s is given by the probability integral transformation

$$p(\lambda) = \int_{-\infty}^{s} f_m(u;\lambda)\, du \; . \tag{2}$$

This function inverts automatically to give standard confidence intervals; for example, in the stochastically increasing case, a $(1-\alpha)\times 100\%$ confidence interval for λ is $\{p^{-1}(1-\alpha/2)\, ,\, p^{-1}(\alpha/2)\}$. Thus $p(\lambda)$ gives a convenient and appropriate summary concerning λ based on the observed data.

As the parameter of interest ψ exists only in the conditional density of t given s, it is appropriate (e.g., Kalbfleisch & Sprott, (1973)) to use $f_c(t|s;\psi,\lambda)$ for inference concerning ψ. However, $f_c(t|s;\psi,\lambda)$ still depends on the nuisance parameter λ. A common and simple solution is to substitute an estimate of λ, such as the maximum likelihood estimate $\hat{\lambda}$, into $f_c(t|s;\psi,\lambda)$ to eliminate λ; for an example in the literature see Shuie & Bain (1983) who examine the gamma model with mean and shape parameter. Another solution proposed by Cox (1975) employs a prior distribution for λ which is combined with the $f_m(s;\lambda)$ density from (1) giving a posterior distribution for λ which is then used to eliminate λ from $f_c(t|s;\psi,\lambda)$. This approach is referred to as the partially-Bayes method.

The nuisance-averaging method is to use the confidence distribution function for λ given by (2) to eliminate λ from $f_c(t|s;\psi,\lambda)$; we then obtain

$$f_A(t|s;\psi) = \int_\Lambda f_c(t|s;\psi,\lambda)\ |dp(\lambda)| \tag{3}$$

as an averaged conditional density for t given s, this provides an appropriate basis for inference concerning ψ. In some problems $p(\lambda)$ may not exist in a closed form: Fraser & Wong (1993) suggested that the asymptotic distribution of the maximum likelihood estimate for λ be used to approximate $p(\lambda)$. Some applications of the averaging method are discussed in Fraser & Wong (1993) for scalar ψ and λ, and in Wong (1995) for scalar ψ and vector λ. In the next section, we show that the parameter-averaging method is a second order method for eliminating the nuisance parameter from the conditional density. In contrast, the estimate substitution method as in Shuie & Bain (1983) is of first order. For the partially Bayes method, different priors can give different results; accordingly we do not examine it here.

3. Second order accuracy of the parameter-averaging method

Consider a statistical model with asymptotic properties as discussed, for example, in DiCiccio, Field & Fraser (1990), and in Fraser & Reid (1993); typically these properties hold in the sampling case or more generally in the case with increasing data bearing on the same parameter. The use of an approximate third order ancillary has been discussed by Barndorff-Nielsen (1991); various construction methods for such third order ancillaries have been given by Fraser & Reid (1995). For the present analysis, we assume that our model is the conditional model obtained from such an ancillary reduction; we also assume for simplicity that the interest parameter ψ and the nuisance parameter λ are scalars. For the analysis we will use large-sample approximations and present these in terms of various location scale standardized variables: these approximations use the normal density together with cubic correction terms.

Now suppose the model $f(s,t;\psi,\lambda)$ can be factored as in (1) so that the parameter of interest ψ is isolated in the conditional distribution. In addition we assume that the conditioning is on the nuisance parameter score at the maximum likelihood value, at least to the first order; in particular this ensures that the conditional distribution is first order independent of the nuisance parameter. As a special case we can have conditioning on the nuisance parameter maximum likelihood estimate using a nuisance parameter that is orthopgonal to the interest parameter.

We consider testing a particular value for ψ; thus ψ can be considered now as a fixed value. For the asymptotic analysis we follow patterns in the references just given and assume that t and s with λ have been location scale standardized and that λ is orthogonal to ψ; then to the first order we have that t is $N(0,1)$ and s is $N(\lambda,1)$, and to the second order we include cubic corrections.

First consider the marginal distribution for s; following Abebe *et. al.* (1992), we have then the availability of the location model approximation

$$\log f_m(s;\lambda) = a - \frac{(s-\lambda)^2}{2} + a_3 \frac{(s-\lambda)^3}{6n^{1/2}} + O_p(n^{-1}) \tag{4}$$

to the second order; for some related discussion see also Fraser & Reid (1993).

Now consider the conditional density for t given s; following Fraser & Reid, (1993), for the fixed ψ (under the null hypothesis), we then have that the log conditional density for t given s for say $\lambda = 0$ can be standardized to have the asymptotic form

$$\log f_c(t|s;\lambda=0) = a - \frac{t^2}{2} + b_3 \frac{t^3}{6n^{1/2}} + O_p(n^{-1}). \tag{5}$$

The more general form for $\lambda \neq 0$ would have mean and variance adjustments of order $O(n^{-1/2})$:

$$\log f_c(t|s;\lambda) = a + c_1 \frac{t\lambda}{n^{1/2}} - \frac{t^2}{2}\left(1 + c_2 \frac{\lambda}{n^{1/2}}\right) + b_3 \frac{t^3}{6n^{1/2}} + O_p(n^{-1}); \tag{6}$$

this follows from the requirement on the conditioning variable.

The ordinary Studentization procedure would involve working with the adjusted variable

$$T = \frac{t - c_1\hat{\lambda}/n^{1/2}}{1 - c_2\hat{\lambda}/2n^{1/2}}. \tag{7}$$

Substituting t obtained from (7) into the equation (6) gives the log conditional density

$$\log f_c(T|s;\lambda) = a + c_1 \frac{T(\lambda-\hat{\lambda})}{n^{1/2}} - \frac{T^2}{2}\left(1 + c_2 \frac{(\lambda-\hat{\lambda})}{n^{1/2}}\right) + b_3 \frac{T^3}{6n^{1/2}} + O_p(n^{-1}); \tag{8}$$

then with $\hat{\lambda}$ distributed as $N(\lambda,1)$ to order $O(n^{-1/2})$ we obtain the marginal log Studentized density for T

$$\log f_S(T) = a - \frac{T^2}{2} + b_3 \frac{T^3}{6n^{1/2}} + O_p(n^{-1}). \tag{9}$$

By contrast, the nuisance-averaging method (3) would use the $N(\hat{\lambda}, 1)$ distribution for λ to average (6) giving the marginal-type log density

$$\log f_A(t) = a + c_1 \frac{t\hat{\lambda}}{n^{1/2}} - \frac{t^2}{2}\left(1 + c_2 \frac{\hat{\lambda}}{n^{1/2}}\right) + b_3 \frac{t^3}{6n^{1/2}} + O_p(n^{-1}) \qquad (10)$$

for the variable t. Making the change of variable (7) to T gives the log nuisance-averaged density for T,

$$\log f_A(T) = a - \frac{T^2}{2} + b_3 \frac{T^3}{6n^{1/2}} + O_p(n^{-1}), \qquad (11)$$

which agrees with (9). We thus have the equivalence to the second order of the log densities: the log parameter-averaged density (11) is equivalent to the standard log Studentized density (9). Also we note that we only need the first order asymptotic distribution for the nuisance parameter to accomplish the averaging.

As a final note, if we just substitute $\hat{\lambda}$ for λ in the conditional density (6), we would obtain just a first order approximation.

4. Discussion

The parameter-averaging method discussed in Fraser & Wong (1993) is shown to be a second-order device for eliminating nuisance parameters. Alternatively, if we replace the nuisance parameter by its maximum likelihood estimate (as in Shuie & Bain 1983), we obtain just a first order method.

Acknowledgments

This research was partially supported by the Natural Science and Engineering Research Council of Canada. The authors wish to thank a referee for very helpful comments that improved the final version of the paper.

References

Abebe, F., Cakmak, S., Cheah, P.K., Fraser, D.A.S., Kuhn, J., Reid, N., and Tapia, A. (1992). Third order asymptotic model: exponential and location type approximations. *Parisankhyan Samikka* **1**, 1-5.

Barndorff-Nielsen O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557-564.

Cox, D.R. (1975). A note on partially Bayes inference and linear models. *Biometrika* **62**, 651-654.

DiCiccio, T., Field, C., and Fraser, D.A.S. (1990). Approximation of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77-95.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Royal Soc. A* **144**, 285-307.

Fraser, D.A.S., and Reid (1993). Simple asymptotic connections between densities and cumulant generating function leading to accurate approximations for distribution functions. *Statist. Sinica.* **3**, 67-82.

Fraser, D.A.S., and Reid (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

Fraser, D.A.S., and Wong, A. (1993). Approximation Studentization with marginal and conditional inference. *Canad. J. Statist.* **21**, 313-320.

Kalbfleisch, J.D., and Sprott, D. (1973). Marginal and conditional likelihood. *Sankyhā A* **35**, 311-328.

Shuie, W.K. and Bain, L.J. (1983). A two-sample test of equal gamma distribution scale parameters with unknown common shape parameter. *Technometrics* **25**, 377-381.

Student (1908). The probable error of a mean. *Biometrika* **6**, 1-25.

Wong, A. (1995). On approximate inference for the two-parameter gamma model. *Statist. Papers* **36**, 49-59.

D.A.S. Fraser
Department of Statistics
University of Toronto
Toronto, Ontario, Canada M5S 1A1

A.C.M. Wong
Department of Mathematics and Statistics
York University
North York, Ontario, Canada M3J 1P3