Australian & New Zealand Journal of Statistics

Aust. N. Z. J. Stat. **51**(2), 2009, 115–126 doi: 10.1111/j.1467-842X.2009.00548.x

ROUTES TO HIGHER-ORDER ACCURACY IN PARAMETRIC INFERENCE

G. ALASTAIR YOUNG[1]

Imperial College London

Summary

Developments in the theory of frequentist parametric inference in recent decades have been driven largely by the desire to achieve higher-order accuracy, in particular distributional approximations that improve on first-order asymptotic theory by one or two orders of magnitude. At the same time, much methodology is specifically designed to respect key principles of parametric inference, in particular conditionality principles. Two main routes to higher-order accuracy have emerged: analytic methods based on 'small-sample asymptotics', and simulation, or 'bootstrap', approaches. It is argued here that, of these, the simulation methodology provides a simple and effective approach, which nevertheless retains finer inferential components of theory. The paper seeks to track likely developments of parametric inference, in an era dominated by the emergence of methodological problems involving complex dependences and/or high-dimensional parameters that typically exceed available data sample sizes.

Key words: analytic methods; ancillary statistic; bootstrap; conditionality; full exponential family; likelihood; likelihood ratio statistic; nuisance parameter; objective Bayes; signed root likelihood ratio statistic; simulation.

1. Introduction

The primary purpose of this paper is to review key aspects of frequentist parametric inference methodology, as it has developed over the last 25 years or so. Developments have been driven largely by the desire to achieve higher-order accuracy, in particular distributional approximations that improve on first-order asymptotic theory by one or two orders of magnitude.

Two main routes to achieving this higher-order accuracy have emerged: analytic methods based on the techniques of small-sample asymptotics, such as saddlepoint approximation, and simulation, or bootstrap, approaches. Objective Bayes methods provide a further route to higher-order frequentist accuracy in many circumstances.

A personal evaluation of these different routes is provided, with the aim of justifying the assertion that parametric bootstrap procedures, with appropriate handling of nuisance parameters, provide satisfactory, simple approaches to inference, yielding the desired higher-order accuracy while retaining finer inferential components of statistical theory, in particular those associated with conditional inference. Parametric bootstrap methods might, therefore, be viewed as the inferential approach of choice in many settings.

[1]Department of Mathematics, Imperial College London, London SW7 2AZ, UK.
e-mail: alastair.young@imperial.ac.uk

Acknowledgments. This paper is based on a keynote address given at the Australian Statistical Conference, Melbourne, July 2008. The author is grateful to the organizers of the conference, in particular Michael Martin and Steven Roberts, for the invitation to speak. Many of the ideas expressed in the paper have emerged from collaborative work with Tom DiCiccio.

A further aim is to flag issues and theoretical questions worth development in a statistical environment where demands are evolving as a result of the emergence of methodological problems involving both complex dependences and high-dimensional parameters that typically exceed available data sample sizes.

2. Driving forces

A number of forces has driven the development of statistical theory and methodology in recent decades. Among the main such forces are the following.

(i) Developments have often been targetted specifically at achieving accurate inference with small sample sizes, n. This aim is typically achieved through the construction of inference procedures with error rate converging to zero rapidly with n, such low error rates generally translating to satisfactory error properties in finite, even very small, sample-size contexts.

(ii) Inference procedures have typically been explicitly constructed to respect key principles of inference, especially conditionality principles.

(iii) Statistical theory has focused primarily on situations in which the parameter of interest is of low dimension p, typically $p = 1$. This focus stems (Cox 2006, section 1.1) from the typical desire to break a substantive research problem of interest into simple components corresponding to strongly focused and incisive research questions. It should be noted, however, that very often the practical usefulness of an inference methodology is validated in terms of satisfactory accommodation of a nuisance parameter of high dimension $\simeq n$. In the immediate future, application areas such as microarray analysis and finance are likely to demand a focus on inference in problems where n, although large by conventional standards, is very much smaller than the parameter dimension d in an appropriate parametric representation. In such situations, inference proceeds by seeking a sparse representation, where the *effective* parameter dimension *is* less than n.

(iv) No explicit optimality criteria are invoked in much of statistical theory.

(v) Instead, theory has predominantly considered fairly general classes of models, where inferential subtlety is important. The key classes considered are multi-parameter exponential families and transformation models, or more generally models that admit ancillary statistics.

In the next section a particular inference problem, which serves as a convenient vehicle for subsequent discussion, is defined. It should be noted, however, that much of what is described relates more broadly, for instance to inference based on general likelihood-based statistics. The intention here is not to be definitive, or exhaustive, but rather telegraphic.

3. An inference problem

Let $Y = \{Y_1, \ldots, Y_n\}$ be a random sample from an underlying distribution $F(y; \eta)$, indexed by a d-dimensional parameter η, where each Y_i may be a random vector. Let $\theta = g(\eta)$ be a (possibly vector) parameter of interest, of dimension p. Without loss we may assume that $\eta = (\theta, \lambda)$, with θ the p-dimensional interest parameter and λ a q-dimensional nuisance parameter.

To be specific, suppose that we wish to test a null hypothesis of the form $H_0 : \theta = \theta_0$, with θ_0 specified, or, through the familiar duality between tests of hypotheses and confidence sets, to construct a confidence set for the interest parameter θ. If $p = 1$, we may wish to allow one-sided inference, for instance a test of H_0 against a one-sided alternative of the form $\theta > \theta_0$ or $\theta < \theta_0$, or construction of a one-sided confidence limit.

Let $l(\eta) = l(\eta; Y)$ be the log-likelihood for η based on Y. Furthermore, denote by $\widehat{\eta} = (\widehat{\theta}, \widehat{\lambda})$ the overall maximum likelihood estimator of η, and by $\widehat{\lambda}_\theta$ the constrained maximum likelihood estimator of λ, for a given fixed value of θ.

Inference on θ may be based on the likelihood ratio statistic,

$$w(\theta) = 2\{l(\widehat{\eta}) - l(\theta, \widehat{\lambda}_\theta)\}.$$

If $p = 1$, inference uses the signed square root likelihood ratio statistic

$$r(\theta) = \mathrm{sgn}(\widehat{\theta} - \theta)w(\theta)^{1/2},$$

where sgn $(x) = -1$ if $x < 0$, $= 0$ if $x = 0$, and $=1$ if $x > 0$.

In cases, typically involving complex dependences, where $l(\eta)$ is intractable, use is often made of a 'composite likelihood', as an approximation to the true likelihood, in the construction of the likelihood ratio statistic or its signed square root. A very accessible review of composite likelihood methods is given by Varin (2008). Among possible composite likelihoods is an independence likelihood, constructed by assuming that the components of each Y_i are independent, and a pairwise likelihood, obtained from all bivariate marginal distributions of pairs of components of Y_i. Although a first-order theory of inference based on composite likelihood is generally available (see, for instance, the overview provided by Varin 2008), inference procedures that achieve higher-order accuracy are undeveloped.

In a first-order theory of inference, the two key distributional results are that the likelihood ratio statistic $w(\theta)$ is distributed as χ^2_p, to error $O(n^{-1})$, whereas the signed square root $r(\theta)$ is distributed as $N(0, 1)$, to error $O(n^{-1/2})$. A slightly more elaborate first-order theory operates for the composite likelihood, with, for instance, the composite likelihood ratio statistic asymptotically distributed as a weighted sum of chi-squared variates.

4. The two key model classes

Here we review the main features of inference in the key exponential family and ancillary statistic model classes. Very useful and detailed accounts of inference in these models are given by Pace & Salvan (1997) and Lehmann & Romano (2005).

Suppose that the log-likelihood is of the form

$$l(\eta) = \theta s_1(Y) + \lambda s_2(Y) - k(\theta, \lambda) - d(Y),$$

so that θ is a natural parameter of a multi-parameter exponential family. Then the conditional distribution of s_1 given s_2 depends only on θ, so that conditioning on s_2 is indicated as a means of eliminating the nuisance parameter. The appropriate inference on θ is based on the distribution of s_1, given the observed data value of s_2. This is, in principle, known, as it is completely specified, once θ is fixed. In fact, this conditional inference has the unconditional (repeated sampling) optimality property of yielding a uniformly most powerful unbiased test. In practice, however, the exact inference may be difficult to construct: the relevant conditional distribution typically requires awkward analytic calculations, numerical integrations etc., and may even be completely intractable.

In the ancillary statistic context, inferential subtlety, also involving conditioning, is provided by the Fisherian proposition, which asserts that inference about θ should be based not on the original specified model $F(y;\eta)$, but instead on the derived model obtained by conditioning on an ancillary statistic, when this exists.

Formally, suppose that the minimal sufficient statistic for η can be written as $(\widehat{\eta}, a)$, with a (at least approximately) distribution-constant; that is, a has a sampling distribution that does not depend on the parameter η. Then a is said to be ancillary, and the Fisherian proposition or conditionality principle dictates that, for inference on θ to be relevant to the observed data sample this should be made conditional on the observed value of a.

5. Refinements: higher-order accuracy

Heading the list of desiderata for refinement of the inference procedures furnished by first-order asymptotic theory is the achievement of higher-order accuracy in distributional approximation, while respecting the conditioning indicated in the two problem classes.

As has already been sketched, two main routes to higher-order accuracy emerge from contemporary statistical theory. The more developed route is that which utilizes analytic procedures based on 'small-sample asymptotics', such as saddlepoint approximation and related methods, to refine first-order distribution theory. The second route involves simulation or bootstrap methods, which aim to obtain refined distributional approximations directly, without analytic approximation.

A potentially valuable third way to higher-order accuracy lies in the use of objective Bayes procedures. These are Bayes inference methods with a prior distribution explicitly specified so that the (marginal) posterior distribution for the interest parameter θ yields confidence limits with the correct frequentist interpretation, to high order: such a prior distribution is termed a probability matching prior distribution. A detailed account is given by Datta & Mukerjee (2004). Formally, we would require that

$$\Pr_\eta\{\theta \le \theta^{(1-\alpha)}(\pi, Y)\} = 1-\alpha + O(n^{-r/2}),$$

for $r = 2$ or 3, and for each $0 < \alpha < 1$. Here n is, as before, the sample size; $\theta^{(1-\alpha)}(\pi, Y)$ is the $(1-\alpha)$ quantile of the marginal posterior distribution, given data Y, of the interest parameter θ, under a prior density $\pi(\theta, \lambda)$; and $\Pr_\eta$ denotes the frequentist probability, under repeated sampling of Y, under parameter value η. If the condition holds with $r = 2$, we speak of $\pi(\theta, \lambda)$ as being a first-order probability matching prior density, and if the condition holds with $r = 3$, we speak of $\pi(\theta, \lambda)$ as being a second-order probability matching prior density. Such a second-order prior density yields confidence limits that are frequentist limits of coverage error $O(n^{-3/2})$; as we shall see, such 'third-order' frequentist accuracy is not necessarily straightforward to obtain by the other two routes. However, although the objective Bayes route is conceptually simple, the marginalization required to obtain the posterior distribution of θ will typically make it awkward with a high-dimensional nuisance parameter. Furthermore, crucially, this route is not always open, in that higher-order accuracy is not necessarily obtainable by this means. Datta & Mukerjee (2004) give a thorough account of circumstances in which a second-order probability matching prior may be constructed.

6. Analytic methods: the highlights

A detailed account of analytic methods for distributional approximation that yield higher-order accuracy is given by Barndorff-Nielsen & Cox (1994). Two particular highlights of a sophisticated and intricate theory stand out. These are Bartlett correction of the likelihood ratio statistic $w(\theta)$ and the construction of analytically modified forms of the signed root likelihood ratio statistic $r(\theta)$, specifically designed to offer conditional validity, to high (asymptotic) order, in both multi-parameter exponential family and ancillary statistic contexts. Particularly central to the analytic approach to higher-order accurate inference is Barndorff-Nielsen's r^* statistic (Barndorff-Nielsen 1986).

In some generality, the expectation of $w(\theta)$ under parameter value η may be expanded as

$$\mathrm{E}_\eta\{w(\theta)\} = p\left\{1 + \frac{b(\eta)}{n} + O(n^{-2})\right\}.$$

The basis of the Bartlett correction is to modify $w(\theta)$, through a scale adjustment, to a new statistic

$$w'(\theta) = w(\theta)/\{1 + b(\eta)/n\}.$$

Then it turns out that $w'(\theta)$ is distributed as χ^2_p, to error $O(n^{-2})$, rather than the error $O(n^{-1})$ for the raw statistic $w(\theta)$. Remarkably, and crucially for inference in the presence of nuisance parameters, this same reduction in the order of error of a χ^2_p approximation is achievable if the scale adjustment is made using the quantity $b(\theta, \widehat{\lambda}_\theta)$; see Barndorff-Nielsen & Hall (1988). Note that this result may be re-expressed as saying that the statistic

$$w''(\theta) = [p/\mathrm{E}_{(\theta,\widehat{\lambda}_\theta)}\{w(\theta)\}]w(\theta)$$

is distributed as χ^2_p to error $O(n^{-2})$. The quantity $\mathrm{E}_{(\theta,\widehat{\lambda}_\theta)}\{w(\theta)\}$ may be approximated by simulation, allowing the Bartlett correction to be carried out purely empirically, without analytic calculation. This approach is seen to be applied quite routinely in applications of higher-order asymptotic methodology: see, for instance, the numerical illustrations given by Brazzale, Davison & Reid (2007).

A further important remark is that (Barndorff-Nielsen & Hall 1988) the above results on improving a χ^2_p approximation to the distribution of the likelihood ratio statistic continue to hold under conditioning on ancillary statistics. The Bartlett correction is therefore a device for obtaining higher-order conditional accuracy in inference on a vector interest parameter.

The adjusted signed root statistic r^* has the form

$$r^*(\theta) = r(\theta) + r(\theta)^{-1}\log\{u(\theta)/r(\theta)\}.$$

Here, the adjustment quantity $u(\theta)$ necessitates explicit specification of the ancillary statistic a in the ancillary statistic context and awkward analytic calculations, in both the ancillary statistic and exponential family situations. A useful and clear account of the r^* machinery is given by Severini (2000, chapter 7). This account contains a detailed description of the adjustment term $u(\theta)$ and an account of other analytic adjustments of the signed root statistic $r(\theta)$, including those that avoid explicit specification of the ancillary statistic at the price of lower conditional accuracy.

The sampling distribution of $r^*(\theta)$ is $N(0, 1)$, to error $O(n^{-3/2})$, conditionally on a, and therefore also unconditionally. Standard normal approximation to the sampling distribution of $r^*(\theta)$ therefore yields third-order (in fact, relative) conditional accuracy, in the ancillary statistic setting, and inference that respects that of exact conditional inference in the exponential family setting to the same third order. The analytic route therefore achieves the goal of improving on the error $O(n^{-1/2})$ obtained from the asymptotic distribution of $r(\theta)$ by two orders of magnitude, $O(n^{-1})$, while respecting the conditional inference desired in the two problem classes.

Although in general requiring awkward analytic calculations, analytic procedures of the kind described here have been successfully packaged (see, for example, Brazzale *et al.* 2007) for certain classes of model, such as nonlinear heteroscedastic regression models. Versions of the r^* statistic for a vector interest parameter are available (see the summary given by Brazzale *et al.* 2007, section 8.8), but the slight evidence that exists would indicate that these should perhaps be seen as less effective than in the case $p = 1$, or than Bartlett correction.

7. Parametric bootstrap

The simple idea behind the bootstrap or simulation alternative to analytic methods of inference is estimation of the sampling distribution of the statistic of interest by its sampling distribution under a member of the parametric family $F(y; \eta)$, fitted to the available sample data. The key to obtaining higher-order accuracy from this route lies, at least for one-sided inference on a scalar interest parameter, in appropriate handling of the nuisance parameter λ. The first detailed analysis of this issue was given by DiCiccio, Martin & Stern (2001). For the case of bootstrapping the distribution of the signed root statistic $r(\theta)$, DiCiccio *et al.* (2001) suggested that the bootstrap calculation should be based on the fitted model $F(y; (\theta, \widehat{\lambda}_\theta))$, that is, on the model with the nuisance parameter taken as the constrained maximum likelihood estimator, for any given value of θ. The analysis was extended to more general statistics by Lee & Young (2005), who analysed in detail the higher-order accuracy gains of this 'constrained bootstrap' handling of the nuisance parameter compared with alternatives, such as substitution of the global maximum likelihood estimator $\widehat{\lambda}$. See also Stern (2006), who extended this analysis to a broader class of inference problems based on M-estimators.

From a repeated sampling perspective, this bootstrap scheme is strikingly effective in producing highly accurate inference with small sample sizes. For instance, Young & Smith (2005, chapter 11) contains much evidence that, in many settings, the scheme produces essentially exact inference. Formally, by this procedure we may estimate the true sampling distribution of the signed root statistic $r(\theta)$ to error $O(n^{-3/2})$, the same order of error as obtained by normal approximation to Barndorff-Nielsen's r^* statistic. This reduction of error by an order of magnitude $O(n^{-1})$ holds for a general, asymptotically $N(0, 1)$ statistic, not just for $r(\theta)$. Other schemes, such as those that substitute the global maximum likelihood estimator of the nuisance parameter, are in general less effective than the constrained bootstrap. For instance, bootstrap inference based on fixing the nuisance parameter λ as its global maximum likelihood value $\widehat{\lambda}$ enables estimation of the true sampling distribution of $r(\theta)$ only to error of order $O(n^{-1})$. Lee & Young (2005) consider the effects of iterating the constrained bootstrap scheme, to accelerate the error reduction properties of successive refinements in estimation of a sampling distribution of interest.

TABLE 1

Repeated sampling coverages (%) of confidence limits of different nominal coverage, obtained by normal approximation to the sampling distributions of the statistics r and r^* and by bootstrap estimation of the distribution of r, sample size n – Example 1.

Nominal (%)		1.0	5.0	10.0	90.0	95.0	99.0
$n = 3$	r	7.6	16.3	22.5	77.6	83.7	92.4
	r^*	3.3	9.9	15.9	83.3	89.4	96.1
	bootstrap	1.1	5.1	10.2	89.9	94.8	99.0
$n = 5$	r	3.3	9.9	15.7	84.3	90.2	96.7
	r^*	1.9	7.2	12.8	87.3	92.7	98.0
	bootstrap	1.0	5.0	10.0	90.1	95.0	99.0
$n = 10$	r	1.8	7.0	12.5	87.6	92.9	98.1
	r^*	1.3	5.9	11.1	88.9	93.9	98.6
	bootstrap	1.0	5.1	10.0	90.0	94.8	98.9

With a vector interest parameter, the bootstrap scheme estimates the true sampling distribution of the likelihood ratio statistic $w(\theta)$ to error $O(n^{-2})$, the same order of error as achieved by Bartlett correction. However, an analysis similar to that of Lee & Young (2005) would appear to show that there is no particular advantage to fixing the nuisance parameter λ as $\widehat{\lambda}_\theta$, rather than $\widehat{\lambda}$, which achieves the same higher-order accuracy.

8. Two simple examples

Example 1: 'Behrens–Fisher'

This numerical illustration concerns an interest parameter of dimension $p = 1$, in the presence of a nuisance parameter of dimension $q = 20$.

Let $Y_{ij}, i = 1, \dots, n_g, j = 1, \dots, n_i$ be independent normal random variables, $Y_{ij} \sim N(\mu, \sigma_i^2)$. Suppose the interest parameter is the common mean μ, with the nuisance parameter consisting of the variances $(\sigma_1^2, \dots, \sigma_{n_g}^2)$.

In the illustration we consider the case $n_g = 20, n_i = n$, $\sigma_i^2 = i$, with varying n. In Table 1 we compare the repeated sampling coverages of one-sided (upper) confidence limits for a true $\mu = 0$, obtained by different methods: (i) $N(0, 1)$ approximation to the distribution of the statistic $r(\mu)$, (ii) $N(0, 1)$ approximation to the distribution of the statistic $r^*(\mu)$, and (iii) (constrained) bootstrap estimation of the distribution of $r(\mu)$. Each bootstrap calculation is based on the drawing of 10 000 bootstrap samples from a given fitted model, and the coverage figures are based on 50 000 Monte Carlo replications from the underlying model. High coverage accuracy of the bootstrap limits, relative to the analytic approach, is very evident, even for very small n.

Example 2: two-dimensional mean

Our second illustration concerns a $p = 2$ dimensional interest parameter, in the presence of a $q = 10$ dimensional nuisance parameter, and represents a slight extension of the model considered in Example 1.

Now, let $Y_{1ij}, Y_{2ij}, i = 1, \dots, n_g, j = 1, \dots, n_i$ be independent normal random variables, $Y_{1ij} \sim N(\mu_1, \sigma_i^2)$, $Y_{2ij} \sim N(\mu_2, \sigma_i^2)$. Now we take the interest parameter as (μ_1, μ_2), with nuisance parameter $(\sigma_1^2, \dots, \sigma_{n_g}^2)$.

TABLE 2

Repeated sampling coverages (%) of confidence limits of different nominal coverage, obtained by chi-squared approximation to the sampling distributions of the statistic w and its empirical Bartlett corrected version (BCw) and by bootstrap estimation of the distribution of w, sample size n, Example 2.

Nominal (%)		1.0	5.0	10.0	90.0	95.0	99.0
$n = 5$	w	0.8	4.1	8.0	83.9	90.7	97.5
	BCw	1.1	5.1	10.1	90.0	95.0	99.1
	bootstrap	0.9	5.0	9.9	89.9	94.9	99.0
$n = 10$	w	0.8	4.5	9.0	87.7	93.4	98.5
	BCw	1.1	5.1	9.9	90.0	95.0	99.1
	bootstrap	1.0	4.9	9.8	90.0	95.0	99.1
$n = 20$	w	0.9	4.6	9.6	89.2	94.3	98.7
	BCw	1.0	4.9	10.1	90.3	95.0	99.0
	bootstrap	1.0	5.0	9.6	90.0	95.1	99.1

We consider the case $n_g = 10, n_i = n, \sigma_i^2 = i$, again with varying n. In Table 2 we compare the repeated sampling coverages of confidence regions for true $(\mu_1, \mu_2) = (1, 2)$, obtained by: (i) χ_2^2 approximation to the distribution of the likelihood ratio statistic w, (ii) empirical Bartlett correction of w(BCw), and (iii) (constrained) bootstrap estimation of the sampling distribution of w. Again, all bootstrap calculations are based on drawing 10 000 bootstrap samples from the appropriate fitted model, and all coverage figures are based on 50 000 Monte Carlo replications. The empirical Bartlett correction simulates the analytic correction, as described in Section 6; this is done using the 10 000 samples drawn by the bootstrap scheme, although in practice a much smaller simulation might be used to make the empirical correction, as the simulation is only being used to estimate the mean of the likelihood ratio statistic w. Extensive numerical studies show that, in this and similar examples, quite comparable performance is seen if the empirical Bartlett correction is based on a very limited simulation, involving, say, the simulation of 50 or 100 bootstrap samples. In this situation, by contrast with Example 1, the bootstrap and the analytic procedure based on Bartlett correction display very similar levels of high accuracy, even for n small. The Bartlett correction is arguably simpler to apply.

9. Other properties of bootstrap, $p = 1$

In the examples of the last section, we were concerned only with the repeated sampling properties of the competing inference procedures. Recall that a key characteristic of the analytic approach, in particular use of the adjusted signed root statistic $r^*(\theta)$, is that procedures are explicitly constructed to yield the appropriate conditional validity to higher order. The bootstrap approach is explicitly applied in an unconditional manner, and the issue arises of the extent to which this approach respects the demands of conditional inference. In fact, the conditional properties of the unconditional (constrained) bootstrap scheme are quite startling, in the cases analysed to date, which consider a scalar interest parameter, $p = 1$. Conditionality properties of bootstrap procedures for inference on a vector interest parameter through bootstrapping statistics, such as the likelihood ratio statistic, which are asymptotically chi-squared, are as yet unknown.

In the multi-parameter exponential family context, the bootstrap scheme yields inference agreeing with exact conditional inference to relative error of third order, $O(n^{-3/2})$; see DiCiccio & Young (2008). Note that this is the same conditional accuracy as obtained through normal approximation to the sampling distribution of the r^* statistic.

In the same context, the constrained bootstrap automatically reproduces the appropriate objective ('conditional second-order probability matching') Bayesian inference to $O(n^{-3/2})$ in many circumstances: see DiCiccio & Young (2009). Now the goal of the objective Bayes method is *conditional* second-order probability matching, with the Bayes limit satisfying the repeated sampling requirement that

$$\Pr_\eta\{\theta \le \theta^{(1-\alpha)}(\pi, Y)|s_2(Y) = s_2\} = 1 - \alpha + O(n^{-3/2}),$$

with s_2 the observed value of the conditioning statistic in the appropriate conditional frequentist inference. It actually turns out (DiCiccio & Young 2009) that any second-order probability matching prior distribution in this exponential family context automatically achieves this second-order conditional probability matching.

In ancillary statistic models, the constrained bootstrap yields inference agreeing with the conditional inference indicated by the Fisherian proposition to second order, $O(n^{-1})$ – this to be compared to third-order conditional accuracy of analytic procedures based on r^*. However, it may be argued forcefully (see, for instance, DiCiccio *et al.* 2001) that demanding third-order conditional accuracy in this setting is actually unwarranted. The conditioning ancillary statistic is typically not unique, and different choices of an ancillary statistic induce differences in conditional sampling distributions at second order, $O(n^{-1})$, so demanding third-order conditional accuracy means ignoring the effects of choice of ancillary statistic. Furthermore, substantial empirical evidence exists that demonstrates that, in practice, the conditional accuracy achieved by bootstrapping in this context is actually impressive, quite competitive with that obtained by the analytic approach. Young & Smith (2005, section 11.5), for instance, give a number of numerical examples.

10. An example of conditional inference

Let $Y_1, \ldots, Y_n$ be independent, identically distributed observations from the inverse Gaussian density

$$f(y;\theta,\lambda) = \sqrt{\frac{\theta}{2\pi}}\exp(\sqrt{\theta\lambda})y^{-3/2}\exp\left\{-\frac{1}{2}(\theta y^{-1} + \lambda y)\right\}.$$

The interest parameter θ is the shape parameter of the distribution, which constitutes a two-parameter exponential family.

Let $S = n^{-1}\sum_i Y_i^{-1}$ and $C = n^{-1}\sum_i Y_i$. Then, the appropriate conditional frequentist inference is based on the conditional distribution of S, given $C = c$, the observed data value of C. Exact conditional inference is simple in this problem and is equivalent to inference being based on the marginal distribution of $V = \sum(Y_i^{-1} - \overline{Y}^{-1})$. We have $\theta\ V \sim \chi^2_{n-1}$.

Trivial calculations show that the signed root statistic $r(\theta)$ is given by

$$r(\theta) = \operatorname{sgn}(\widehat{\theta} - \theta)\{n(\log\widehat{\theta} - 1 - \log\theta + \theta/\widehat{\theta})\}^{1/2},$$

with the global maximum likelihood estimator $\widehat{\theta}$ given by $\widehat{\theta} = n/V$. The signed root statistic $r(\theta)$ is seen to be a function of V, and therefore has a sampling distribution that does not

TABLE 3

Conditional frequentist coverages (%) of 5% and 95% posterior confidence limits obtained from two objective Bayes procedures OB1 and OB2 and from the r^* statistic, sample size n, inverse Gaussian shape example. Confidence limits are shown in parenthesis.

Procedure	n	5%		95%	
OB1	10	5.89	(0.210)	95.46	(1.033)
	20	5.39	(0.308)	95.27	(0.911)
	50	5.21	(0.409)	95.00	(0.796)
OB2	10	6.52	(0.217)	95.51	(1.035)
	20	5.39	(0.308)	95.03	(0.905)
	50	5.09	(0.408)	94.92	(0.795)
$r^*(\theta)$	10	5.12	(0.201)	95.10	(1.019)
	20	5.04	(0.304)	95.03	(0.905)
	50	4.99	(0.407)	95.00	(0.796)

depend on the nuisance parameter λ. From a repeated sampling perspective, the bootstrap scheme will, modulo any Monte Carlo simulation error (which we are not considering at all here), produce exact inference for θ. Furthermore, because $r(\theta)$ is a monotonic function of V, and the exact conditional inference is equivalent to inference based on the marginal distribution of V, the bootstrap inference will actually replicate the exact conditional inference without *any* error, at least in an infinite bootstrap simulation. Thus, from either a repeated sampling or a conditional inference perspective, a bootstrap inference will be exact in this example.

We consider objective Bayes inference on θ, based on two different prior densities. The first procedure (OB1) utilizes a joint prior density on (θ, λ) of the form $\pi(\theta, \lambda) \propto \theta^{-3/2}\, \lambda^{-1/2}$, whereas the second (OB2) uses a prior density $\pi(\theta, \lambda) \propto \theta^{-5/4}\, \lambda^{-3/4}$. Both these prior densities are second-order probability matching, and therefore (DiCiccio & Young 2009) should yield confidence limits of conditional frequentist coverage error $O(n^{-3/2})$; non-uniqueness of a second-order conditional probability matching prior distribution is typical of the objective Bayes route to higher-order frequentist accuracy. We consider the conditional frequentist confidence levels of posterior 5% and 95% quantiles, for data with observed values $s = 2.0$, $c = 3.0$ of the sufficient statistics, and for varying sample size n. These are compared with those confidence levels obtained by the analytic procedure involving normal approximation to the sampling distribution of the adjusted signed root statistic $r^*(\theta)$, which is easily constructed in this example. Table 3 shows the 5% and 95% confidence limits obtained by the three procedures in parentheses, with the corresponding exact conditional frequentist coverage levels; recall that the bootstrap applied to the signed root statistic $r(\theta)$ yields values exactly equal to the nominal coverage values in this example.

Some care is required in interpreting the conditional coverage levels, particularly of the objective Bayes procedures, shown in the Table. The posterior quantile $\theta^{(1-\alpha)}(\pi, Y) = \theta^{(1-\alpha)}(\pi, S, C)$. Fixing $C = c$, the quantile is a monotonic (decreasing) function of S. For fixed θ_0, we have

$$\left\{\theta_0 \le \theta^{(1-\alpha)}(\pi, Y)\right\} = \{S \le s_0\},$$

where s_0 has $\theta^{(1-\alpha)}(\pi, s_0, c) = \theta_0$. Then the conditional frequentist coverage under the true parameter value θ_0 is

$$\Pr\{\theta_0 \le \theta^{(1-\alpha)}(\pi, Y) \mid C = c; \theta_0) = \Pr\{S \le s_0 \mid C = c; \theta_0\}.$$

This latter probability is equivalent to

$$\Pr\{V \le n(s_0 - 1/c) \mid C = c; \theta_0) = \Pr\{V \le n(s_0 - 1/c); \theta_0\},$$

as V is independent of C. So, for example, conditional on $C = 3.0$, if $n = 10$ and $\theta = 0.210$, the posterior 5% quantile under objective Bayes procedure OB1 has conditional frequentist coverage 5.89%. It is clear that inference based on the r^* statistic yields greater conditional frequentist accuracy than that obtained by the objective Bayes route in this example. Different second-order conditional probability matching prior densities can yield noticeably different conditional frequentist properties.

11. Concluding remarks

This paper has provided a high-speed overview of statistical theory that is directed at achieving higher-order accuracy in parametric problems. An extensive and practically useful theory exists for the case of inference for an interest parameter of dimension $p = 1$. In that situation, it is the author's firm view that the bootstrap approach provides the simplest route to higher-order accuracy. Although quite effective and practical methodologies exist for the case of inference with $p > 1$, gaps remain in the underlying theory. An issue of importance is whether, from a theoretical perspective, there are any advantages in this context to a full-blown bootstrap approach to inference. It is argued that, in practice, the most effective results are seen through an empirical Bartlett correction of the likelihood ratio statistic, a procedure appealing for its potential computational efficiency relative to alternative bootstrap methods.

Although theoretical studies have been narrowly focused on inference procedures derived from a full likelihood for the model parameter, the practical effectiveness of parametric bootstrap procedures is also striking for inference based on composite likelihood statistics. Development of more elaborate theory in this direction is required.

Existing theories operate under the assumption of a fixed parameter dimension, d, and consider the asymptotic regime in which the sample size $n \to \infty$. Of some contemporary importance will be the development of theory for cases in which $d \to \infty$ with n, with the attendant issue of rates of increase in parameter dimension that will still allow analytic or simulation approaches to achieve higher-order accuracy.

Much of the discussion here has considered the extent to which higher-order accuracy is achievable, while respecting the requirements of conditional inference. In the future, focus is likely to be on problems of high dimension, and there is some question (Robins & Wasserman 2000) of the relevance of such considerations in high-dimensional contexts.

The discussion has proceeded under an assumption of correctness of the assumed parametric family, $F(y;\eta)$, of distributions. The behaviour of inference procedures under model misspecification is a serious issue. Robust forms of the likelihood ratio and signed root statistics are easily defined. Empirical investigations show that simulating the distribution of such a robust form of statistic under the wrongly assumed model is actually quite effective, compared with first-order asymptotics. But formal analysis of whether higher-order accuracy can reasonably be achieved in such a setting is still to be undertaken. Key to any such discussion is likely to be the centrality of multivariate normality assumptions in many contemporary settings.

Finally, the discussion has focused on the simple inference problems of confidence set construction and testing a single, simple, hypothesis. Procedures for multiple testing are of key contemporary importance, as is the issue of assessment of uncertainty after model selection. Routes to higher-order accuracy in such settings remain to be explored.

References

BARNDORFF-NIELSEN, O.E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73,** 307–322.

BARNDORFF-NIELSEN, O.E. & COX, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.

BARNDORFF-NIELSEN, O.E. & HALL, P.G. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75,** 374–378.

BRAZZALE, A.R., DAVISON, A.C. & REID, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge: Cambridge University Press.

COX, D.R. (2006). *Principles of Statistical Inference*. Cambridge: Cambridge University Press.

DATTA, G.S. & MUKERJEE, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. New York: Springer.

DICICCIO, T.J. & YOUNG, G.A. (2008). Conditional properties of unconditional parametric bootstrap procedures for inference in exponential families. *Biometrika* **95,** 747–758.

DICICCIO, T.J. & YOUNG, G.A. (2009). Objective Bayes and conditional inference in multi-parameter exponential families. *Biometrika*, in press.

DICICCIO, T.J., MARTIN, M.A. & STERN, S.E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. *Can. J. Statist.* **29,** 67–76.

LEHMANN, E.L. & ROMANO, J.P. (2005). *Testing Statistical Hypotheses*, 3rd edn. New York: Springer.

LEE, S.M.S. & YOUNG, G.A. (2005). Parametric bootstrapping with nuisance parameters. *Statist. Probab. Lett.* **71**, 143–153.

PACE, L. & SALVAN, A. (1997). *Principles of Statistical Inference: from a Neo-Fisherian Perspective*. Singapore: World Scientific.

ROBINS, J. & WASSERMAN, L. (2000). Conditioning, likelihood and coherence: a review of some foundational concepts. *J. Amer. Statist. Assoc.* **95,** 1340–1345.

SEVERINI, T.A. (2000). *Likelihood Methods in Statistics*. Oxford: Oxford University Press.

STERN, S.E. (2006). Simple and accurate one-sided inference based on a class of M-estimators. *Biometrika* **93,** 973–987.

VARIN, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis* **92,** 1–28.

YOUNG, G.A. & SMITH, R.L. (2005). *Essentials of Statistical Inference*. Cambridge: Cambridge University Press.