Computer-intensive conditional inference

G. Alastair Young and Thomas J. DiCiccio

Abstract. Conditional inference is a fundamental part of statistical theory. However, exact conditional inference is often awkward, leading to the desire for methods which offer accurate approximations. Such a methodology is provided by small-sample likelihood asymptotics. We argue in this paper that simple, simulation-based methods also offer accurate approximations to exact conditional inference in multiparameter exponential family and ancillary statistic settings. Bootstrap simulation of the marginal distribution of an appropriate statistic provides a conceptually simple and highly effective alternative to analytic procedures of approximate conditional inference.

Key words: analytic approximation, ancillary statistic, Bartlett correction, bootstrap, conditional inference, exponential family, likelihood ratio statistic, stability

1 Introduction

Conditional inference has been, since the seminal work of Fisher [16], a fundamental part of the theory of parametric inference, but is a less established part of statistical practice.

Conditioning has two principal operational objectives: (i) the elimination of nuisance parameters; (ii) ensuring relevance of inference to an observed data sample, through the conditionality principle, of conditioning on the observed value of an ancillary statistic, when such a statistic exists. The concept of an ancillary statistic here is usually taken simply to mean one which is distribution constant. The former notion is usually associated with conditioning on sufficient statistics, and is most transparently and uncontroversially applied for inference in multiparameter exponential family models. Basu [7] provides a general and critical discussion of conditioning to eliminate nuisance parameters. The notion of conditioning to ensure relevance, together with the associated problem, which exercised Fisher himself (Fisher [17]), of recovering information lost when reducing the dimension of a statistical problem (say, to that of the maximum likelihood estimator, when this is not sufficient), is most transparent in transformation models, such as the location-scale model considered by Fisher [16].

Mantovan, P., Secchi, P. (Eds.): Complex Data Modeling and Computationally Intensive Statistical Methods
© Springer-Verlag Italia 2010

In some circumstances issues to do with conditioning are clear cut. Though most often applied as a slick way to establish independence between two statistics, Basu's Theorem (Basu [4]) shows that a boundedly complete sufficient statistic is independent of every distribution constant statistic. This establishes the irrelevance for inference of any ancillary statistic when a boundedly complete sufficient statistic exists.

In many other circumstances however, we have come to understand that there are formal difficulties with conditional inference. We list just a few. (1) It is well understood that conflict can emerge between conditioning and conventional measures of repeated sampling optimality, such as power. The most celebrated illustration is due to Cox [11]. (2) Typically there is arbitrariness on what to condition on. In particular, ancillary statistics are often not unique and a maximal ancillary may not exist. See, for instance, Basu [5, 6] and McCullagh [23]. (3) We must also confront the awkward mathematical contradiction of Birnbaum [9], which says that the conditionality principle, taken together with the quite uncontroversial sufficiency principle, imply acceptance of the likelihood principle of statistical inference, which is incompatible with the common methods of inference, such as calculation of p-values or construction of confidence sets, where we are drawn to the notion of conditioning.

Calculating a conditional sampling distribution is also typically not easy, and such practical difficulties, taken together with the formal difficulties with conditional inference, have led to much of modern statistical theory being based on notions of inference which automatically accommodate conditioning, at least to some high order of approximation. Of particular focus are methods which respect the conditionality principle without requiring explicit specification of the conditioning ancillary, and which therefore circumvent the difficulties associated with non-uniqueness of ancillaries.

Much attention in parametric theory now lies, therefore, in inference procedures which are stable, that is, which are based on a statistic which has, to some high order in the available data sample size, the same repeated sampling behaviour both marginally and conditional on the value of the appropriate conditioning statistic. The notion is that accurate approximation to an exact conditional inference can then be achieved by considering the marginal distribution of the stable statistic, ignoring the relevant conditioning. This idea is elegantly expressed for the ancillary statistic context by, for example, Barndorff-Nielsen and Cox [2, Section 7.2], Pace and Salvan [24, Section 2.8] and Severini [26, Section 6.4]. See also Efron and Hinkley [15] and Cox [12].

A principal approach to approximation of an intractable exact conditional inference by this route lies in developments in higher-order small-sample likelihood asymptotics, based on saddle point and related analytic methods. Book length treatments of this analytic approach are given by Barndorff-Nielsen and Cox [2] and Severini [26]. Brazzale *et al.* [10] demonstrate very convincingly how to apply these developments in practice. Methods have been constructed which automatically achieve, to a high order of approximation, the elimination of nuisance parameters which is desired in the exponential family setting, though focus has been predominantly on ancillary statistic models. Here, a key development concerns construction of adjusted forms of the signed root likelihood ratio statistic, which require specification of the ancillary statistic, but are distributed, conditionally on the ancillary, as $N(0, 1)$ to third order, $O(n^{-3/2})$, in the data sample size n. Normal approximation to the sampling distribution of the adjusted statistic therefore provides third-order approximation to exact conditional inference: see Barndorff-Nielsen [1]. Approximations which yield second-order conditional accuracy, that is, which approximate the exact conditional inference to an error of order $O(n^{-1})$, but which avoid specification of the ancillary statistic, are possible: Severini [26, Section 7.5] reviews such methods.

In the computer age, an attractive alternative approach to approximation of conditional inference uses marginal simulation, or 'parametric bootstrapping', of an appropriately chosen statistic to mimic its conditional distribution. The idea may be applied to approximate the conditioning that is appropriate to eliminate nuisance parameters in the exponential family setting, and can be used in ancillary statistic models, where specification of the conditioning ancillary statistic is certainly avoided.

Our primary purpose in this article is to review the properties of parametric bootstrap procedures in approximation of conditional inference. The discussion is phrased in terms of the inference problem described in Section 2. Exponential family and ancillary statistics models are described in Section 3. Key developments in analytic approximation methods are described in Section 4. Theoretical properties of the parametric bootstrap approach are described in Section 5, where comparisons are drawn with analytic approximation methods. A set of numerical examples are given in Section 6, with concluding remarks in Section 7.

2 An inference problem

We consider the following 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 $d-p$-dimensional nuisance parameter. Suppose 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, construct a confidence set for the parameter of interest 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. Also, denote by $\hat{\eta} = (\hat{\theta}, \hat{\lambda})$ the overall maximum likelihood estimator of η, and by $\hat{\lambda}_\theta$ 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, $W = w(\theta) = 2\{l(\hat{\eta}) - l(\theta, \hat{\lambda}_\theta)\}$. If $p = 1$, one-sided inference uses the signed square root likelihood ratio statistic $R = r(\theta) = \text{sgn}(\hat{\theta} - \theta) w(\theta)^{1/2}$, where $\text{sgn}(x) = -1$ if $x < 0$, $= 0$ if $x = 0$ and $= 1$ if $x > 0$. In a first-order theory of inference, the two key distributional results are that W is distributed as χ^2_p, to error of order $O(n^{-1})$, while R is distributed as $N(0, 1)$, to an error of order $O(n^{-1/2})$.

3 Exponential family and ancillary statistic models

Suppose the log-likelihood is of the form $l(\eta) = \theta s_1(Y) + \lambda^T s_2(Y) - k(\theta, \lambda) - d(Y)$, with θ scalar, so that θ is a natural parameter of a multiparameter exponential family. We wish to test $H_0 : \theta = \theta_0$ against a one-sided alternative, and do so using the signed root statistic R.

Here the conditional distribution of $s_1(Y)$ given $s_2(Y) = s_2$ depends only on θ, so that conditioning on the observed value s_2 is indicated as a means of eliminating the nuisance parameter. So, the appropriate inference on θ is based on the distribution of $s_1(Y)$, given the observed data value of s_2. This distribution is, in principle, known, since 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: see, for example, Young and Smith [29, Section 7.2]. 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 modern convention, ancillarity in the presence of nuisance parameters is generally defined in the following terms. Suppose the minimal sufficient statistic for η may be written as $(\hat{\eta}, A)$, where the statistic A has, at least approximately, a sampling distribution which does not depend on the parameter η. Then A is said to be ancillary and the conditionality principle would argue that inference should be made conditional on the observed value $A = a$.

McCullagh [22] showed that the conditional and marginal distributions of signed root statistics derived from the likelihood ratio statistic W for a vector interest parameter, but with no nuisance parameter, agree to an error of order $O(n^{-1})$, producing very similar p-values whether one conditions on an ancillary statistic or not. Severini [25] considered similar results in the context of a scalar interest parameter without nuisance parameters; see also Severini [26, Section 6.4.4]. Zaretski *et al.* [30] establish stability of the signed root statistic R, in the case of a scalar interest parameter and a general nuisance parameter. The key to their analysis is that the first two cumulants of the signed root statistic $r(\theta)$ are of the form

$$E\{r(\theta)\} = n^{-1/2}m(\eta) + O(n^{-3/2}), \quad var\{r(\theta)\} = 1 + n^{-1}v(\eta) + O(n^{-3/2}),$$

where $m(\eta)$ and $v(\eta)$ are of order $O(1)$. The third- and higher-order cumulants of $r(\theta)$ are of order $O(n^{-3/2})$; see Severini [26, Section 5.4]. This cumulant structure also holds conditionally given a statistic A assumed to be second-order ancillary; see McCullagh [22] for details of approximate ancillarity. Under conditions required for valid Edgeworth expansions, if the conditional and marginal expectations of the signed root statistic agree to an error of order $O(n^{-1})$ given the ancillary statistic A, then the conditional and marginal distributions agree to the same order of error. An intricate analysis shows that the conditional and marginal versions of $m(\eta)$ coincide, to order $O(n^{-1})$. This methodology may be readily extended to the case of a vector interest parameter θ to establish stability of signed root statistics derived from the likelihood ratio statistic W in the presence of nuisance parameters. Stability of W is immediate: the marginal and conditional distributions are both χ^2_p to error $O(n^{-1})$.

4 Analytic approximations

A detailed development of analytic methods for distributional approximation which yield higher-order accuracy in approximation of an exact conditional inference is described by Barndorff-Nielsen and Cox [2]. A sophisticated and intricate theory yields two particularly important methodological contributions. These are Bartlett corrections of the likelihood ratio statistic W and the development of analytically adjusted forms of the signed root likelihood ratio statistic R, which are specifically constructed to offer conditional validity, to high asymptotic order, in both the multiparameter exponential family and ancillary statistic contexts. Particularly central to the analytic approach to higher-order accurate conditional inference is Barndorff-Nielsen's R^* statistic (Barndorff-Nielsen, [1]).

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

$$w(\theta)/\{1 + b(\eta)/n\},$$

which turns out to be distributed as χ^2_p, to an error of order $O(n^{-2})$, 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 an χ^2_p approximation is achievable if the scale adjustment is made using the quantity $b(\theta, \hat{\lambda}_\theta)$; see Barndorff-Nielsen and Hall [3]. Note that this result may be re-expressed as saying that the statistic

$$w^*(\theta) = \frac{p}{\mathrm{E}_{(\theta,\hat{\lambda}_\theta)}\{w(\theta)\}} w(\theta)$$

is distributed as χ^2_p to an error of order $O(n^{-2})$. Here the quantity $\mathrm{E}_{(\theta,\hat{\lambda}_\theta)}\{w(\theta)\}$ may be approximated by simulation, allowing the Bartlett correction to be carried out purely empirically, without analytic calculation.

The adjusted signed root statistic R^* has the form

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

Write $\eta = \left(\eta^1, \ldots, \eta^d\right)$, so that $\theta = \eta^1$ is the scalar parameter of interest, with $\lambda = \left(\eta^2, \ldots, \eta^d\right)$ a vector nuisance parameter. Let $l_{rs}(\eta) = \partial^2 l(\eta)/\partial\eta^r\eta^s$, and let $l_{\eta\eta} = \left(l_{rs}\right)$ be the $d \times d$ matrix with components $l_{rs}(\eta)$ and $l_{\lambda\lambda}$ be the $(d-1) \times (d-1)$ submatrix corresponding to the nuisance parameter λ. In the exponential family context, the adjustment quantity $u(\theta)$ takes the simple form

$$u(\theta) = \left(\hat{\theta} - \theta\right) \frac{\left| -l_{\eta\eta}(\hat{\theta}, \hat{\lambda})\right|^{1/2}}{\left| -l_{\lambda\lambda}(\theta, \hat{\lambda}_\theta)\right|^{1/2}}.$$

In the ancillary statistic context the adjustment necessitates explicit specification of the ancillary statistic A and more awkward analytic calculations. For details of its construction, see Barndorff-Nielsen and Cox [2, Section 6.6].

The sampling distribution of R^* is $N(0, 1)$, to an error of order $O(n^{-3/2})$, conditionally on $A = a$, and therefore also unconditionally. Standard normal approximation to the sampling distribution of R^* therefore yields third-order (in fact, relative) conditional accuracy, in the ancillary statistic setting, and inference which 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 of order $O(n^{-1/2})$ obtained from the asymptotic normal distribution of R by two orders of magnitude, $O(n^{-1})$, while respecting the conditional inference desired in the two problem classes.

5 Bootstrap approximations

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. A recent summary of the repeated sampling properties of such schemes is given by Young [28]. We are concerned here with an analysis of the extent to which the bootstrap methods, applied unconditionally, nevertheless achieve accurate approximation to conditional inference in the exponential family and ancillary statistic settings.

DiCiccio and Young [14] show that in the exponential family context, accurate approximation to the exact conditional inference may be obtained by considering the marginal distribution of the signed root statistic R under the fitted model $F(y; (\theta, \hat{\lambda}_\theta))$, that is the model with the nuisance parameter taken as the constrained maximum likelihood estimator, for any given value of θ. This scheme yields inference agreeing with exact conditional inference to a relative error of third order, $O(n^{-3/2})$. Specifically, DiCiccio and Young [14] show that

$$\text{pr}\{R \geq r; (\theta, \hat{\lambda}_\theta)\} = \text{pr}(R \geq r | s_2(Y) = s_2; \theta)\{1 + O(n^{-3/2})\},$$

when r is of order $O(1)$. Their result is shown for both continuous and discrete models. The approach therefore has the same asymptotic properties as saddle point methods developed by Skovgaard [27] and Barndorff-Nielsen [1] and studied by Jensen [18]. DiCiccio and Young [14] demonstrate in a number of examples that this approach of estimating the marginal distribution of R gives very accurate approximations to conditional inference even in very small sample sizes: further examples are discussed in Section 6 below. A crucial point of their analysis is that the marginal estimation should fix the nuisance parameter as its constrained maximum likelihood estimator: the same third-order accuracy is not obtained by fixing the nuisance parameter at its global maximum likelihood value $\hat{\lambda}$.

Third-order accuracy can also be achieved, in principle, by estimating the marginal distributions of other asymptotically standard normal pivots, notably Wald and score statistics. However, in numerical investigations, using R is routinely shown to provide more accurate results. A major advantage of using R is its low skewness; consequently, third-order error can be achieved, although not in a relative sense, by merely correcting R for its mean and variance and using a standard normal approximation to the standardised version of R. Since it is computationally much easier to approximate the mean and variance of R by parametric bootstrapping at $(\theta, \hat{\lambda}_\theta)$ than it is to simulate the entire distribution of R, the use of mean and variance correction offers substantial computational savings, especially for constructing confidence intervals. Although these savings are at the expense of accuracy, numerical work suggests that the loss of accuracy is unacceptable only when the sample size is very small.

In theory, less conditional accuracy is seen in ancillary statistic models. Since the marginal and conditional distributions of R coincide with an error of order $O(n^{-1})$ given $A = a$, it follows that the conditional p-values obtained from R are approximated to the same order of error by the marginal p-values. Moreover, for approximating the marginal p-values, the marginal distribution of R can be approximated to an error of order $O(n^{-1})$ by means of the parametric bootstrap; the value of η used in the bootstrap can be either the overall maximum likelihood estimator, $\eta = (\hat{\theta}, \hat{\lambda})$, or the constrained maximum likelihood estimator, $\eta = (\theta, \hat{\lambda}_\theta)$. For testing the null hypothesis $H_0 : \theta = \theta_0$, the latter choice is feasible; however, for constructing confidence intervals, the choice $\eta = (\hat{\theta}, \hat{\lambda})$ is computationally less demanding. DiCiccio *et al.* [13] and Lee and Young [21] showed that the p-values obtained by using $\eta = (\theta, \hat{\lambda}_\theta)$ are marginally uniformly distributed to an error of order $O(n^{-3/2})$, while those obtained by using $\eta = (\hat{\theta}, \hat{\lambda})$ are uniformly distributed to an error of order $O(n^{-1})$ only. Numerical work indicates that using $\eta = (\theta, \hat{\lambda}_\theta)$ improves conditional accuracy as well, although, formally, there is no difference in the orders of error to which conditional p-values are approximated by using the two choices. Though in principle the order of error in approximation of exact conditional inference obtained by considering the marginal distribution of R is larger than the third-order, $O(n^{-3/2})$, error obtained by normal approximation to the sampling distribution of the adjusted signed root statistic R^*, substantial numerical evidence suggests very accurate approximations are obtained in practice. Examples are given in Section 6, and further particular examples are considered by DiCiccio *et al.* [13], Young and Smith [29, Section 11.5] and Zaretzki *et al.* [30].

In the case of a vector interest parameter θ, both the marginal and conditional distributions of $W = w(\theta)$ are chi-squared to error $O(n^{-1})$, and hence, using the χ^2_p approximation to the distribution of W achieves conditional inference to an error of second-order. Here, however, we have noted that a simple scale adjustment of the likelihood ratio statistic improves the chi-squared approximation:

$$\frac{p}{\mathrm{E}_{(\theta,\lambda)}\{w(\theta)\}} w(\theta)$$

is distributed as χ^2_p to an error of order $O(n^{-2})$. Since $\mathrm{E}_{(\theta,\lambda)}\{w(\theta)\}$ is of the form $p + O(n^{-1})$, it follows that $\mathrm{E}_{(\theta,\hat{\lambda}_\theta)}\{w(\theta)\} = \mathrm{E}_{(\theta,\lambda)}\{w(\theta)\} + O_p(n^{-3/2})$. Thus, estimation of the marginal distribution of W by bootstrapping with $\eta = (\theta, \hat{\lambda}_\theta)$ yields an approximation having an error of order $O(n^{-3/2})$; moreover, to an error of order $O(n^{-2})$, this approximation is the distribution of a scaled χ^2_p random variable with scaling factor $\mathrm{E}_{(\theta,\hat{\lambda}_\theta)}\{w(\theta)\}/p$. The result of Barndorff-Nielsen and Hall [3], that

$$\frac{p}{\mathrm{E}_{(\theta,\hat{\lambda}_\theta)}\{w(\theta)\}}w(\theta)$$

is distributed as χ^2_p to an error of order $O(n^{-2})$, shows that confidence sets constructed by using the bootstrap approximation to the marginal distribution of W have marginal coverage error of order $O(n^{-2})$.

The preceding results continue to hold under conditioning on the ancillary statistic. In particular,

$$\frac{p}{\mathrm{E}_{(\theta,\lambda)}\{w(\theta)|A=a\}}w(\theta)$$

is conditional on $A = a$, also χ^2_p to an error of order $O(n^{-2})$. The conditional distribution of W is, to an error of order $O(n^{-2})$, the distribution of a scaled χ^2_p random variable with scaling factor $\mathrm{E}_{(\theta,\lambda)}\{w(\theta)|A=a\}/p$. Generally, the difference between $\mathrm{E}_{(\theta,\lambda)}\{w(\theta)\}$ and $\mathrm{E}_{(\theta,\lambda)}\{w(\theta)|A=a\}$ is of order $O(n^{-3/2})$ given $A = a$, and using the bootstrap estimate of the marginal distribution of W approximates the conditional distribution to an error of order $O(n^{-3/2})$. Thus, confidence sets constructed from the bootstrap approximation have conditional coverage error of order $O(n^{-3/2})$, as well as marginal coverage error of order $O(n^{-2})$.

Bootstrapping the entire distribution of W at $\eta = (\theta, \hat{\lambda}_\theta)$ is computationally expensive, especially when constructing confidence sets, and two avenues for simplification are feasible. Firstly, the order of error in approximation to conditional inference remains of order $O(n^{-3/2})$ even if the marginal distribution of W is estimated by bootstrapping with $\eta = (\hat{\theta}, \hat{\lambda})$, the global maximum likelihood estimator. It is likely that using $\eta = (\theta, \hat{\lambda}_\theta)$ produces greater accuracy, however, this increase in accuracy might not be sufficient to warrant the additional computational demands. Secondly, instead of bootstrapping the entire distribution of W, the scaled chi-squared approximation could be used, with the scaling factor $\mathrm{E}_{(\theta,\hat{\lambda}_\theta)}\{w(\theta)\}/p$ being estimated by the bootstrap. This latter approach of empirical Bartlett adjustment is studied in numerical examples in Section 6. Use of the bootstrap for estimating Bartlett adjustment factors was proposed by Bickel and Ghosh [8].

6 Examples

6.1 Inverse Gaussian distribution

Let $\{Y_1, \ldots, Y_n\}$ be a random sample from the inverse Gaussian density

$$f(y;\theta,\lambda) = \sqrt{\frac{\theta}{2\pi}}\exp(\sqrt{\theta\lambda})y^{-3/2}\exp\{-\frac{1}{2}(\theta y^{-1}+\lambda y)\},\ \ y>0,\ \ \theta>0, \lambda>0.$$

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

With $S = n^{-1}\sum_{i=1}^{n} Y_i^{-1}$ and $C = n^{-1}\sum_{i=1}^{n} Y_i$, the appropriate conditional inference is based on the conditional distribution of S, given $C = c$, the observed data value of C. This, making exact conditional inference simple in this problem, is equivalent to inference being based on the marginal distribution of $V = \sum_{i=1}^{n}(Y_i^{-1} - \overline{Y}^{-1})$. The distribution of θV is χ^2_{n-1}.

The signed root statistic $r(\theta)$ is given by $r(\theta) = \text{sgn}(\hat{\theta} - \theta)\{n(\log\hat{\theta} - 1 - \log\theta + \theta/\hat{\theta})\}^{1/2}$, with the global maximum likelihood estimator $\hat{\theta}$ given by $\hat{\theta} = n/V$. The signed root statistic $r(\theta)$ is seen to be a function of V, and therefore has a sampling distribution which does not depend on the nuisance parameter λ. Since $r(\theta)$ is in fact a monotonic function of V, and the exact conditional inference is equivalent to inference based on the marginal distribution of this latter statistic, the bootstrap inference will actually replicate the exact conditional inference without error, at least in an infinite bootstrap simulation. Thus, from a conditional inference perspective, a bootstrap inference will be exact in this example.

6.2 Log-normal mean

As a second example of conditional inference in an exponential family, suppose $\{Y_1, \ldots, Y_n\}$ is a random sample from the normal distribution with mean μ and variance τ, and that we want to test the null hypothesis that $\psi \equiv \mu + \frac{1}{2}\tau = \psi_0$, with τ as nuisance parameter. This inference problem is equivalent to that around the mean of the associated log-normal distribution.

The likelihood ratio statistic is

$$w(\psi) = n[-\log\hat{\tau} - 1 + \log\hat{\tau}_0 + \frac{1}{4}\hat{\tau}_0 + \{\hat{\tau} + (\bar{Y} - \psi_0)^2\}/\hat{\tau}_0 + (\bar{Y} - \psi_0)],$$

with $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$, $\hat{\tau} = n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, and where the constrained maximum likelihood estimator of the nuisance parameter τ under the hypothesis $\psi = \psi_0$ is given by $\hat{\tau}_0 = 2[\{1 + \hat{\tau} + (\bar{Y} - \psi_0)^2\}^{1/2} - 1]$.

In this example, calculation of the p-values associated with the exact conditional test is awkward, requiring numerical integration, but quite feasible: details of the test are given by Land [19]. We perform a simulation of 5000 datasets, for various sample sizes n, from the normal distribution with $\mu = 0, \tau = 1$, and consider one-sided testing of the hypothesis $H_0 : \psi = 1/2$, testing against $\psi > 1/2$. We compare the average absolute percentage relative error of different approximations to the exact conditional p-values over the 5000 replications in Table 1. Details of the methods are as follows: r is based on $N(0, 1)$ approximation to the distribution of $r(\psi)$; r^* is based on $N(0, 1)$ approximation to the distribution of $r^*(\psi)$; boot is based on bootstrap estimation of the marginal distribution of $r(\psi)$. All bootstrap results are based on 5,000,000 samples. The figures in parenthesis show the proportion of the 5000 replications where the corresponding method gave the smallest absolute percentage error. Bootstrapping the marginal distribution of the signed root statistic

Table 1. Log-normal mean problem: comparison of average absolute percentage relative errors in estimation of exact conditional p-values over 5000 replications

n	r	r^*	*boot*
5	6.718	0.476	0.367
	(0.3%)	(37.3%)	(62.4%)
10	4.527	0.154	0.136
	(0.1%)	(41.9%)	(58.0%)
15	3.750	0.085	0.077
	(0.0%)	(42.3%)	(57.7%)
20	3.184	0.054	0.050
	(0.0%)	(43.3%)	(56.7%)

is highly effective as a means of approximating the exact conditional inference for a small n, this procedure remaining competitive with the r^* approximation, which yields the same theoretical error rate, $O(n^{-3/2})$, as the sample size n increases.

6.3 Weibull distribution

As a simple illustration of an ancillary statistic model, suppose that $\{T_1, \ldots, T_n\}$ is a random sample from the Weibull density

$$f(t; \nu, \lambda) = \lambda\nu(\lambda t)^{\nu-1} \exp\{-(\lambda t)^{\nu}\},\ t > 0,\ \nu > 0, \lambda > 0,$$

and that we are interested in inference for the parameter ν: note that $\nu = 1$ reduces to the exponential distribution. If we take $Y_i = \log T_i$, then the Y_i are an independent sample of size n from an extreme value distribution $EV(\mu, \theta)$, a location-scale family, with scale and location parameters $\theta = \nu^{-1},\ \mu = -\log\lambda$. It is straightforward to construct exact inference for θ, conditional on the ancillary $a = (a_1, \ldots, a_n)$, with $a_i = (y_i - \hat{\mu})/\hat{\theta}$: see, for example, Pace and Salvan [24, Section 7.6].

Again, we perform a simulation of 5000 datasets, for various sample sizes n, from the Weibull density with $\nu = \lambda = 1$, and consider both one-sided and two-sided testing of the hypothesis $H_0 : \theta = 1$, in the one-sided case testing against $\theta > 1$. As before, we compare the average absolute percentage relative error of different approximations to the exact conditional p-values over the 5000 replications in Table 2. Details of the methods are as follows. For the one-sided inference: r is based on $N(0, 1)$ approximation to the distribution of $r(\theta)$; r^* is based on $N(0, 1)$ approximation to the distribution of $r^*(\theta)$; boot is based on bootstrap estimation of the marginal distribution of $r(\theta)$. For two-sided inference: w is based on χ^2_1 approximation to the distribution of $w(\theta)$; Bart is based on χ^2_1 approximation to the (empirically) Bartlett corrected $w^*(\theta)$; boot is based on bootstrap estimation of the marginal distribution of $w(\theta)$. As before, all bootstrap results are based on 5,000,000 samples, this same simulation being used for empirical Bartlett correction. Figures in parenthesis show the proportion of the 5000 replications where the corresponding method gave the smallest

Table 2. Weibull scale problem: comparison of average absolute percentage relative errors in estimation of exact conditional p-values over 5000 replications

		One-sided			*Two-sided*	
n	r	r^*	*boot*	w	*Bart*	*boot*
10	37.387	1.009	0.674	12.318	0.666	0.611
	(0.0%)	(17.1%)	(82.9%)	(0.0%)	(43.9%)	(56.1%)
20	25.473	0.388	0.397	6.118	0.185	0.227
	(0.0%)	(46.2%)	(53.8%)	(0.0%)	(63.4%)	(36.6%)
30	20.040	0.252	0.307	4.158	0.131	0.200
	(0.0%)	(60.9%)	(39.1%)	(0.0%)	(68.7%)	(31.3%)
40	17.865	0.250	0.273	3.064	0.117	0.177
	(0.0%)	(70.1%)	(29.9%)	(0.0%)	(69.7%)	(30.3%)

absolute percentage error. For both one-sided and two-sided inference, bootstrapping the marginal distribution of the appropriate statistic is highly effective for a small n, though there is some evidence that as n increases in the two-sided case, the simulation effort is better directed at estimation of the (marginal) expectation of $w(\theta)$, and the approximation to an exact conditional inference made via the chi-squared approximation to the scale-adjusted statistic $w^*(\theta)$.

6.4 Exponential regression

Our final example concerns inference on a two-dimensional interest parameter, in the presence of a scalar nuisance parameter.

Let $Y_1, \ldots, Y_n$ be independent and exponentially distributed, where Y_i has mean $\lambda \exp(-\theta_1 z_i - \theta_2 x_i)$, where the z_i and x_i are covariates, with $\sum_{i=1}^n z_i = \sum_{i=1}^n x_i = 0$. The interest parameter is $\theta = (\theta_1, \theta_2)$, with λ nuisance. The log-likelihood function for (θ, λ) can be written as

$$l(\theta, \lambda) = -n \log \lambda - n\hat{\lambda}_\theta / \lambda.$$

Here $a = (a_1, \ldots, a_n)$ is the appropriate conditioning ancillary statistic, with $a_i = \log y_i - \log \hat{\lambda} + \hat{\theta}_1 z_i + \hat{\theta}_2 x_i$.

The likelihood ratio statistic $w(\theta)$ is easily shown to have the simple form

$$w(\theta) = 2 \log[\frac{1}{n} \sum_{i=1}^{n} \exp\{a_i + (\theta_1 - \hat{\theta}_1) z_i + (\theta_2 - \hat{\theta}_2) x_i\}].$$

Now we perform a simulation of 2000 datasets, for various sample sizes n, from this exponential regression model with $\theta = (0, 0)$, $\lambda = 1$, and consider testing of the hypothesis $H_0 : \theta = (0, 0)$. Let $z = (54, 52, 50, 65, 52, 52, 70, 40, 36, 44, 54, 59)$ and $x = (12, 8, 7, 21, 28, 13, 13, 22, 36, 9, 87)$: these are covariate values in a lung cancer survival dataset described by Lawless [20, Table 6.3.1]. In our simulations, for

Table 3. Exponential regression problem: comparison of average absolute percentage relative errors in estimation of exact conditional p-values over 2000 replications

n	w	*Bart*	*boot*
5	12.473	1.199	0.557
	(0.4%)	(23.8%)	(75.8%)
7	5.795	1.013	0.962
	(3.5%)	(42.7%)	(53.9%)
9	4.636	0.791	0.786
	(8.8%)	(43.1%)	(48.2%)
11	4.840	1.642	1.622
	(18.4%)	(33.3%)	(48.4%)

a given n, the covariate values $(z_1, \ldots, z_n)$ and $(x_1, \ldots, x_n)$ are taken as the first n members of z and x respectively, suitable centred to have $\sum_{i=1}^n z_i = \sum_{i=1}^n x_i = 0$. Exact conditional inference for this model is detailed by Lawless [20, Section 6.3.2]. Exact conditional p-values based on the likelihood ratio statistic W are, following Barndorff-Nielsen and Cox [2, Section 6.5], obtained by numerical integration of the exact conditional density of $\hat{\theta}$ given a , as described by Lawless [20, Section 6.3.2], over the appropriate set of values of $\hat{\theta}$ which give a value of W exceeding the observed value.

The average absolute percentage relative error of different approximations to the exact conditional p-values over the 2000 replications are given in Table 3. Now w is based on χ^2_2 approximation to the distribution of $w(\theta)$; Bart is based on χ^2_2 approximation to the (empirically) Bartlett corrected $w^*(\theta)$; boot is based on bootstrap estimation of the marginal distribution of $w(\theta)$. As before, all bootstrap results are based on 5,000,000 samples, this same simulation being used for empirical Bartlett correction, and the figures in parenthesis show the proportion of the 2000 replications where the corresponding method gave the smallest absolute percentage error. Now, exact conditional p-values appear to be effectively approximated by the marginal distribution of the likelihood ratio statistic, though the empirical Bartlett correction is quite comparable.

7 Conclusions

Marginal simulation approaches to approximation of an exact conditional inference have been shown to be highly effective, in both multiparameter exponential family and ancillary statistic models.

For inference on a scalar natural parameter in an exponential family, the appropriate exact one-sided conditional inference can be approximated to a high level of accuracy by marginal simulation of the signed root likelihood ratio statistic R. This procedure considers the sampling distribution of R under the model in which the interest parameter is fixed at its null hypothesis value and the nuisance parameter is specified as its constrained maximum likelihood value, for that fixed value of the interest parameter. The theoretical rate of error in approximation of the exact conditional inference is the same ($O(n^{-3/2})$) as that obtained by normal approximation to the distribution of the adjusted signed root statistic R^*, and excellent approximation is seen with small sample sizes. Similar practical effectiveness is seen with small sample sizes n in ancillary statistic models, though here the theoretical error rate of the marginal simulation approach, $O(n^{-1})$, is inferior to that of the analytic approach based on R^*.

In ancillary statistic models, where interest is in a vector parameter, or in two-sided inference, based on the likelihood ratio statistic W, on a scalar interest parameter, two marginal simulation approaches compete. The first uses the simulation to approximate directly the sampling distribution of W and the second approximates the marginal expectation of W, this then being the basis of empirical Bartlett correction of W. The two methods are seen to perform rather similarly in practice, with direct approximation of the distribution of the stable statistic W yielding particularly good results in small sample size situations.

References

1. Barndorff-Nielsen, O.E.: Inference on full or partial parameters based on the standardized signed log likelihood ratio. Biometrika **73**, 307–22 (1986)
2. Barndorff-Nielsen, O.E., Cox, D.R.: Inference and Asymptotics. Chapman & Hall, London (1994)
3. Barndorff-Nielsen, O.E., Hall, P.G.: On the level-error after Bartlett adjustment of the likelihood ratio statistic. Biometrika **75**, 374–8 (1988)
4. Basu, D.: On statistics independent of a complete sufficient statistic. Sankhya **15**, 377–80 (1955)
5. Basu, D.: The family of ancillary statistics. Sankhya **21**, 247–56 (1959)
6. Basu, D.: Problems relating to the existence of maximal and minimal elements in some families of statistics (subfields). *Proc. Fifth Berk. Symp. Math. Statist. Probab.* **1**. University of California Press, Berkeley CA, 41–50 (1965)
7. Basu, D.: On the elimination of nuisance parameters. J. Amer. Statist. Assoc. **72**, 355–66 (1977)
8. Bickel, P.J., Ghosh, J.K.: A decomposition for the likelihood ratio statistic and the Bartlett correction – a Bayesian argument. Ann. Statist. **18**, 1070–90 (1990)
9. Birnbaum, A.: On the foundations of statistical inference (with discussion). J. Amer. Statist. Assoc. **57**, 269–306 (1962)
10. Brazzale, A.R., Davison, A.C., Reid, N.: Applied Asymptotics: Case Studies in Small-Sample Statistics. Cambridge University Press, Cambridge (2007)
11. Cox, D.R.: Some problems connected with statistical inference. Ann. Math. Stat. **29**, 357–72 (1958)
12. Cox, D.R.: Local ancillarity. Biometrika **67**, 279–86 (1980)
13. DiCiccio, T.J., Martin, M.A., Stern, S.E.: Simple and accurate one-sided inference from signed roots of likelihood ratios. Can. J. Statist. **29**, 67–76 (2001)
14. DiCiccio, T.J., Young, G.A.: Conditional properties of unconditional parametric bootstrap procedures for inference in exponential families. Biometrika **95**, 747–58 (2008)

15. Efron, B., Hinkley, D.V.: Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). Biometrika **65**, 457–87 (1978)
16. Fisher, R.A.: Two new properties of mathematical likelihood. Proc. R. Soc. Lond. A **144**, 285–307 (1934)
17. Fisher, R.A.: The logic of inductive inference. J.R. Statist. Soc. **98**, 39–82 (1935)
18. Jensen, J.L.: The modified signed likelihood statistic and saddlepoint approximations. Biometrika **79**, 693–703 (1992)
19. Land, C.E.: Confidence intervals for linear functions of normal mean and variance. Ann. Statist. **42**, 1187–205 (1971)
20. Lawless, J.: Statistical Models and Methods for Lifetime Data. Wiley, New York (1982)
21. Lee, S.M.S., Young, G.A.: Parametric bootstrapping with nuisance parameters. Stat. Prob. Letters **71**, 143–53 (2005)
22. McCullagh, P.: Local sufficiency. Biometrika **71**, 233–44 (1984)
23. McCullagh, P.: Conditional inference and Cauchy models. Biometrika **79**, 247–59 (1992)
24. Pace, L., Salvan, A.: Principles of Statistical Inference: from a Neo-Fisherian Perspective. World Scientific Publishing, Singapore (1997)
25. Severini, T.A.: Conditional properties of likelihood-based significance tests. Biometrika **77**, 343–52 (1990)
26. Severini, T.A.: Likelihood methods in Statistics. Oxford University Press, Oxford (2000)
27. Skovgaard, I.M.: Saddlepoint expansions for conditional distributions. J. Appl. Probab. **24**, 875–87 (1987)
28. Young, G.A.: Routes to higher-order accuracy in parametric inference. Aust. N.Z. J. Stat. **51**, 115–26 (2009)
29. Young, G.A., Smith, R.L.: Essentials of Statistical Inference. Cambridge University Press, Cambridge (2005)
30. Zaretzki, R., DiCiccio, T.J., Young, G.A.: Stability of the signed root likelihood ratio statistic in the presence of nuisance parameters. Submitted for publication (2010)