Computational Statistics and Data Analysis 56 (2012) 2410–2420

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Parametric bootstrap under model mis-specification

H.Y. Kevin Lu, G. Alastair Young *

Department of Mathematics, Imperial College London, London, SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 25 January 2011
Received in revised form 19 January 2012
Accepted 19 January 2012
Available online 30 January 2012

Keywords:
Asymptotic approximation
Model mis-specification
Non-parametric inference
Parametric bootstrap
Resampling
Signed root likelihood ratio statistic

ABSTRACT

Under model correctness, highly accurate inference on a scalar interest parameter in the presence of a nuisance parameter can be achieved by several routes, among them considering the bootstrap distribution of the signed root likelihood ratio statistic. The context of model mis-specification is considered and inference based on a robust form of the signed root statistic is discussed in detail. Stability of the distribution of the statistic allows accurate inference, outperforming that based on first-order asymptotic approximation, by considering the bootstrap distribution of the statistic under the incorrectly assumed distribution. Comparisons of this simple approach with alternative analytic and non-parametric inference schemes are discussed.

1. Introduction

Let $Y = \{Y_1, \ldots, Y_n\}$ be a random sample of size n, from a distribution assumed to have probability density function $f(y; \eta)$, with $\eta = (\psi, \lambda)$, where ψ is a scalar interest parameter and λ a nuisance parameter, possibly vector-valued. Consider testing the null hypothesis $H_0 : \psi = \psi_0$, with ψ_0 specified, against a one-sided alternative of the form $H_1 : \psi < \psi_0$ or $H_1 : \psi > \psi_0$.

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

Inference may be based on the signed square root likelihood ratio statistic, defined by

$$R \equiv R(\psi_0) = \mathrm{sgn}(\hat{\psi} - \psi_0)[2\{l(\hat{\psi}, \hat{\lambda}) - l(\psi_0, \hat{\lambda}_0)\}]^{1/2},$$

where $\mathrm{sgn}(x) = -1$ if $x < 0$, $= 0$ if $x = 0$ and $= 1$ if $x > 0$, and $\hat{\lambda}_0 = \hat{\lambda}_{\psi_0}$. Under H_0, $R(\psi_0)$ is asymptotically distributed according to the standard normal distribution $N(0, 1)$, provided the assumed parametric distribution is correct. The level of error of the $N(0, 1)$ approximation to the sampling distribution of $R(\psi_0)$ is of the first-order, $O(n^{-1/2})$ in the sample size n.

Two main approaches emerge (Young, 2009) to reduce this level of error, to the third-order, $O(n^{-3/2})$: analytic adjustment of the statistic R, and replacement of the $N(0, 1)$ approximation by a bootstrap estimate of the distribution of the statistic. A key form of the analytically adjusted statistic is Barndorff–Nielsen's R^* statistic (Barndorff-Nielsen, 1986), which is of the form $R^* = R + \log(U/R)/R$, in terms of an analytic adjustment quantity U. DiCiccio et al. (2001) and Lee and Young (2005) considered inference based on the bootstrap distribution obtained by considering the distribution of $R(\psi_0)$ under sampling from the density $f(y; \psi_0, \hat{\lambda}_0)$. Both of these third-order accurate inference procedures are observed in many situations to achieve spectacularly low levels of error even in small sample settings.

* Corresponding author. Tel.: +44 0 20 7594 8560.
E-mail addresses: kevin.lu02@imperial.ac.uk (H.Y.K. Lu), alastair.young@imperial.ac.uk (G.A. Young).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.01.018

Concern here is with the stated inference problem in circumstances when Y is a random sample from a density $g(y)$ which does not belong to the assumed parametric family of densities $f(y; \psi, \lambda)$. Specific consideration is given to the following formulation of the inference problem under model mis-specification, as described, for example, by Kent (1982) and Stafford (1996). Let $\eta(g) = \{\psi(g), \lambda(g)\}$ minimise the Kullback–Leibler distance between $g(y)$ and $f(y; \eta)$, given by $\int \log\{g(y)/f(y; \eta)\}g(y)dy$. Then consider testing $H_0 : \psi = \psi_0$, with $\psi_0 = \psi(g)$. Such an inference problem is natural whenever $\psi(g)$ has a direct interpretation under the true $g(y)$, for example as an expected value. In these circumstances, the statistic $R(\psi_0)$ is asymptotically distributed as $N(0, v)$, where $v \equiv v(g) \neq 1$ in general. It has been suggested (Stafford, 1996: see also related work by Viraswami and Reid, 1996, 1998) that the signed root likelihood ratio statistic $R(\psi_0)$ be 'robustified' by rescaling, through construction of a statistic of the form $R' \equiv R'(\psi_0) = R/\sqrt{\hat{v}}$, where $\hat{v}$ is an empirical estimate, constructed from Y, of the asymptotic variance v. Such a modified statistic $R'(\psi_0)$ is asymptotically distributed as standard normal under H_0. Again, the level of error of a $N(0, 1)$ approximation to the sampling distribution is of the order $O(n^{-1/2})$.

The purpose here is to examine closely the use of the statistic R' for inference, expanding on comments made by Lu and Young (2010). The objective is to stress the following methodological conclusions. When there is no model mis-specification, highly accurate inference can be achieved by bootstrapping the distribution of R': since R' is asymptotically distributed as $N(0, 1)$ it follows directly from Lee and Young (2005) that this procedure achieves the same third-order accuracy as inference based on normal approximation to the distribution of Barndorff–Nielsen's R^* statistic. Under model mis-specification, R^* is non-robust and $N(0, 1)$ approximation to its sampling distribution does not achieve a test with the correct asymptotic level. However, the parametric bootstrap procedure, which again samples from the (incorrect) density $f(y; \psi_0, \hat{\lambda}_0)$ is shown to yield accurate inference. Though in principle the level of error, $O(n^{-1/2})$, is no better than that offered by normal approximation to the sampling distribution, in practice the bootstrap procedure substantially outperforms normal approximation. The key to this property is that the distribution of R' typically does not depend much on the true density underlying the sample data Y, but converges rather slowly to its asymptotic limit. Therefore, using the sampling distribution of R' under the wrong density $f(y; \psi_0, \hat{\lambda}_0)$ as a surrogate for its distribution under the true density g is often a rather accurate estimation procedure, in particular for small n.

These observations, together with empirical comparisons between the parametric bootstrap procedure and non-parametric alternatives, suggest strongly that in the inference problem being considered the most effective approach is based on the distribution of the modified statistic R', under sampling from the density $f(y; \psi_0, \hat{\lambda}_0)$. This procedure yields highly accurate inference, with the same low levels of error as obtained by use of the R^* statistic, when the parametric assumption is correct, while protecting against model mis-specification. The analysis demonstrates, in particular, that in the latter setting the parametric bootstrap procedure, based on the wrong distribution, will often outperform the asymptotic method based on the $N(0, 1)$ approximation to the distribution of R'. This indicates that higher levels of accuracy than obtained by first-order asymptotic methods will often be achievable in this setting.

2. Asymptotic distribution of R'

Use is made of the notation, the stochastic expansion of R and the expression for $\hat{v}$ provided in Stafford (1996) to find the cumulants of the robust statistic R'. These cumulants are seen to be asymptotically the same as the cumulants of $N(0, 1)$, thereby showing that the asymptotic distribution of R' is standard normal.

Suppose $\eta = (\eta_1, \eta_2, \ldots, \eta_d)$, where $\eta_1 = \psi$ is the scalar interest parameter, $(\eta_2, \ldots, \eta_d) = \lambda$ is the vector nuisance parameter and d is the dimension of η. Let $l_i = l_i(\eta) = \log f(y_i; \eta)$ be the log-likelihood of the ith observation, and let $l_{i;s} = \frac{\partial}{\partial \eta_s} l_i(\eta)$, $l_{i;st} = \frac{\partial^2}{\partial \eta_s \partial \eta_t} l_i(\eta)$ be the partial derivatives of the log-likelihood for the i th observation, $s, t = 1, \ldots, d$.

Denote by $\eta_0 = (\psi_0, \lambda_0)$ the value of η which minimises the Kullback–Leibler distance, as described in the previous section.

In the following definitions and derivations, it is not necessary to assume whether the true distribution is mis-specified by f or not: the same asymptotic results hold true for both cases of $g = f$ and $g \neq f$.

Define

$$I_{st} \equiv I_{st}(\eta_0) = E_{g(y)}[l_{1;st}]|_{\eta=\eta_0},$$
$$I_{s,t} \equiv I_{s,t}(\eta_0) = E_{g(y)}[l_{1;s}l_{1;t}]|_{\eta=\eta_0},$$

and

$$Z_s \equiv Z_s(\eta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_{i;s}|_{\eta=\eta_0},$$

where $E_{g(y)}[.]$ denotes the expectation with respect to $g(y)$.

Further define a $d \times d$ matrix A with components I_{st} and denote the components of the $d \times d$ matrix $-A^{-1}$ by I^{st}, where A^{-1} is the usual matrix inverse of A. Denote by $\hat{I}_{st} \equiv \hat{I}_{st}(\hat{\eta}) = \frac{1}{n} \sum_{i=1}^{n} l_{i;st}|_{\eta=\hat{\eta}}$ and $\hat{I}_{s,t} \equiv \hat{I}_{s,t}(\hat{\eta}) = \frac{1}{n} \sum_{i=1}^{n} l_{i;s}l_{i;t}|_{\eta=\hat{\eta}}$ estimates of I_{st} and $I_{s,t}$ respectively. An estimate of A is $\hat{A}$, obtained by the replacement of I_{st} in A by $\hat{I}_{st}$. Then, $\hat{I}^{st}$, which is an estimate of I^{st}, can be read off from the corresponding (s, t)-entry of the matrix $-\hat{A}^{-1}$.