Computational Statistics and Data Analysis 56 (2012) 4366–4380

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Improved confidence regions based on Edgeworth expansions

Christopher S. Withers [a], Saralees Nadarajah [b,*]

[a] Applied Mathematics Group, Industrial Research Limited, Lower Hutt, New Zealand
[b] School of Mathematics, University of Manchester, Manchester M13 9PL, UK

ARTICLE INFO

Article history:
Received 11 October 2011
Received in revised form 26 March 2012
Accepted 26 March 2012
Available online 1 April 2012

Keywords:
Cumulants
Edgeworth expansions
Ellipsoidal confidence regions

ABSTRACT

Let $\widehat{\mathbf{w}}$ be a consistent estimate of $\mathbf{w}$ in $\mathbb{R}^p$ satisfying the standard cumulant expansion in powers of n^{-1} with asymptotic covariance $n^{-1}\mathbf{V}$. Then $n^{1/2}(\widehat{\mathbf{w}}-\mathbf{w})$ has the standard Edgeworth expansion about $\mathcal{N}_p(\mathbf{0},\mathbf{V})$. We obtain from this the Edgeworth expansions for $T_n(\mathbf{V}) = n(\widehat{\mathbf{w}}-\mathbf{w})'\mathbf{V}^{-1}(\widehat{\mathbf{w}}-\mathbf{w})$ about χ^2_p and for its Studentized version, $T_n(\widehat{\mathbf{V}})$. So, we obtain a confidence region for $\mathbf{w}$ of level $\alpha + O(n^{-2})$.

1. Introduction and summary

The need for confidence regions for a population parameter vector arises in many areas of statistics. We mention: response surface methodology, generalized regression models, the statistical decision theory, Bayesian statistics, classification, reliability, statistical inference, multivariate statistics, nonlinear models, change-point problems, regression graphics, sequential analysis, probability inequalities, nonparametric statistics, time series analysis, boundary value problem methods and information theory. For excellent accounts of known confidence regions, we refer the readers to: confidence regions for linear models with a standard covariance structure (Sengupta and Jammalamadaka, 2003, Chapters 4 and 5); confidence regions for probability matching priors (Datta and Mukerjee, 2004); confidence regions for the direction of steepest ascent in process improvement and confidence regions for the location of the stationary point in general analysis of second order response surfaces (Myers et al., 2009, Chapters 5 and 6); profile likelihood-based confidence regions for general linear models and nonlinear regression models (Uusipaikka, 2009, Chapters 3 and 4).

The number of papers proposing confidence regions is too many to cite. Confidence regions can be constructed using many different methods. Some of these involve: the exact distribution theory, asymptotic chi-square distributions, asymptotic normal distributions, record values, bootstrapping, empirical likelihood method, likelihood ratio statistics, Bayesian test statistics, sign based test statistics, rank based test statistics, differential geometric frameworks, jackknife, least square statistics, sequential methods, Monte Carlo likelihood methods, adaptive estimation methods, the M estimation theory, empirical Bayes, calibration methods, score statistics, quadratic approximations, subsampling methods, Gibbs sampling methods, smoothing methods (for example, polynomial splines, kernel type smoothing, local adaptive smoothing and Granulometric smoothing), Kolmogorov–Smirnov statistics, highest posterior density methods, R estimation methods, stochastic inequalities (for example, Hajek–Rényi type inequalities), multiple comparison methods, maximum modulus methods, minimum area confidence methods, U statistics, Rényi statistics and statistics based on other entropy measures, generalized p values, James–Stein type statistics, power divergence statistics, importance sampling methods, Uusipaikka's method, genetic algorithms and uniform minimum variance unbiased statistics.

* Corresponding author.
E-mail address: snadarajah2@unl.edu (S. Nadarajah).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.03.019

Often exact confidence regions cannot be found especially when the data are not normal. So, approximations are needed. Approximations are usually based on distributional assumptions (normal, chi-square, etc.) or computationally intensive methods (bootstrap, empirical likelihood, etc.). The distributional approximations may not be accurate. The computationally intensive methods are expensive, time consuming and may not be accessible to everyone.

The aim of this paper is to provide an accessible method to construct confidence regions of any accuracy for any distribution, normal or non-normal. This method is based on Edgeworth expansions. These expansions have been used before for constructing confidence regions. But their use has not been accessible or comprehensive.

Götze and Zitikis (1995) use Edgeworth expansions and bootstrap methods to obtain confidence regions for degenerate von Mises statistics that are of the order of $O(n^{-1})$, where n is the sample size. Amendola (1996a,b) uses Edgeworth expansions to construct likelihood functions for non-normal data. The likelihood functions are used to yield confidence regions. But the author makes no mention of accuracy. Sun et al. (2000) use Edgeworth expansions and a version of Skorohod's representation theorem to construct one-sided confidence regions of order $n^{-1/2}$ and two-sided confidence regions of order n^{-1} for generalized linear models. Guillou and Merlevéde (2003) use a mixture of Edgeworth expansions and bootstrap methods to provide second-order correct confidence regions for density estimators for continuous time processes. Xu and Gupta (2006) use Edgeworth expansions to propose two confidence regions for mean vector of non-normal distributions. The first confidence region uses the "Cornish–Fisher expansion of the percentile of T^2 based on its asymptotic null distribution under general condition". The second confidence region is based on a "normalization transformation". Both confidence regions have coverage error of the order of $o(n^{-1})$. But their use is limited for n moderate. Chiang et al. (2009) use bootstrap as well as an Edgeworth expansion with remainder term $o(n^{-1/2})$ to construct confidence regions for time-dependent area under receiver operating characteristic curves. Vrbik (2009) describes a procedure for constructing confidence regions based on Edgeworth expansions, but this procedure eliminates the beyond-normal terms by a polynomial transformation.

One could also use Hermite Gauss or Poisson Charlier expansions instead of Edgeworth expansions. But we have not been able to find any published paper using such expansions to construct confidence regions.

The method for constructing confidence regions presented in this paper is based on Edgeworth expansions. The method is simpler and more general than those in the cited papers. We derive confidence regions for any parameter, not just the mean vector, and for any distribution, normal or non-normal. The confidence regions apply for small, moderate and large n. Besides, our method can be applied to give confidence regions having error $O(n^{-k})$ for any $k \geq 1$, although the illustrations are limited for $k = 2$.

Suppose we seek a confidence region for $\mathbf{w}$, an unknown parameter in $\mathbb{R}^p$, given an estimate $\widehat{\mathbf{w}}$. We show how to improve on the accuracy of the traditional ellipsoidal confidence region obtained by Studentizing $\mathbf{w}$. The improved confidence regions derived here are as simple as the traditional version, so they would require the same amount of time.

Suppose $n^{1/2}(\widehat{\mathbf{w}} - \mathbf{w}) \to \mathcal{N}_p(\mathbf{0}, \mathbf{V})$ as $n \to \infty$ and $\widehat{\mathbf{V}} \stackrel{p}{\to} \mathbf{V}$, positive-definite. Then

$$T_n(\mathbf{V}) = n(\widehat{\mathbf{w}} - \mathbf{w})'\mathbf{V}^{-1}(\widehat{\mathbf{w}} - \mathbf{w})$$

satisfies

$$P(T_n(\widehat{\mathbf{V}}) < x) \to F_p(x) = P(\chi_p^2 < x)$$

as $n \to \infty$. This gives the traditional confidence region for $\mathbf{w}$ in $\mathbb{R}^p$. Typically the order of magnitude is $O(n^{-1})$.

In Section 2, we devise Edgeworth expansions for $n^{1/2}(\widehat{\mathbf{w}} - \mathbf{w})$ and for $T_n(\mathbf{V})$ under the usual Cornish–Fisher type assumption on the cumulants of $\widehat{\mathbf{w}}$. In Section 3, we consider parametric and nonparametric cases, where $\widehat{\mathbf{w}} = t(\widehat{\boldsymbol{\theta}})$ and $\widehat{\mathbf{V}} = \mathbf{V}(\widehat{\boldsymbol{\theta}})$ are functions of some estimate $\widehat{\boldsymbol{\theta}}$ in $\mathbb{R}^q$: we derive Edgeworth expansions for the Studentized statistics $n^{1/2}\widehat{\mathbf{V}}^{-1/2}(\widehat{\mathbf{w}} - \mathbf{w})$ and $T_n(\widehat{\mathbf{V}})$. Section 4 performs a simulation study to compare the derived confidence region of level $F_p(r) + O(n^{-2})$ versus the traditional version. The simulation is performed for non-normal data. The proofs of all results including technical lemmas are given in Appendix.

It is important to point out that Edgeworth expansions are formal expansions and they often diverge (Cramer, 1946; Kenney and Keeping, 1959; Kendall and Stuart, 1969; Petrov, 1995). So, the expansions derived in Sections 2 and 3 are formal. But we have given an analytical condition for all expansions in Sections 2 and 3 to be valid, see (2.7). The condition takes the form $n > n_0(\mathbf{V})$: this condition does not mean that the expansions are not valid for $n \leq n_0(\mathbf{V})$. The practitioner may choose to check this condition. In practice, however, $\mathbf{V}$ may be unknown.

An alternative use of the expansions is to choose the number of terms to yield as good a confidence interval/confidence region as possible. Suppose the confidence interval/confidence region based on the expansions takes the form:

$$\left(M_1 - M_2 \sum_{k=1}^{I} u_k, M_1 + M_3 \sum_{k=1}^{I} u_k \right),$$

where I determines the number of terms chosen. One could choose I to maximize the coverage probability and to keep their lengths as short as possible. In other words, I could be chosen so large that the coverage probability is as large as possible, but not too large that the length diverges. The choice should be a trade off between I being too large and I being too small. One possible choice is to take $I \geq 1$ as the largest integer for which $|u_k|$ decreases for all $k \leq I$. For more discussion on this, see Sections 2 and 3 of Withers and Nadarajah (in press).