ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Confidence distributions from likelihoods by median bias correction

Pierpaolo De Blasi [a,*], Tore Schweder [b]

[a] University of Torino and Collegio Carlo Alberto, Italy
[b] Department of Economics, University of Oslo, Norway

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Asymptotic expansion
Confidence curve
Confidence distribution
Exponential family
Modified directed likelihood
Normal transformation family

ABSTRACT

By the modified directed likelihood, higher order accurate confidence limits for a scalar parameter are obtained from the likelihood. They are conveniently described in terms of a confidence distribution, that is a sample dependent distribution function on the parameter space. In this paper we explore a different route to accurate confidence limits via tail-symmetric confidence curves, that is curves that describe equal tailed intervals at any level. Instead of modifying the directed likelihood, we consider inversion of the log-likelihood ratio when evaluated at the median of the maximum likelihood estimator. This is shown to provide equal tailed intervals, and thus an exact confidence distribution, to the third-order of approximation in regular one-dimensional models. Median bias correction also provides an alternative approximation to the modified directed likelihood which holds up to the second order in exponential families.

1. Introduction

The level of reported confidence intervals are most often 95%, with equal probability of missing the target at both sides. Sometimes other levels are used, but rarely are several intervals at their different levels reported in applied work. Instead of only reporting one confidence interval we suggest to report a family of nested confidence intervals for parameters of primary interest. The family is indexed by the confidence level $1-\alpha$ for $\alpha \in (0, 1)$ and is conveniently represented by what is called a *confidence curve*, a quantity introduced by Birnbaum (1961) to give a complete picture of the estimation uncertainty. As an example, take $\hat\theta \sim \mathrm{N}(\theta, \sigma^2)$ for σ known. It yields the curve $cc(\theta) = |1 - 2\Phi((\theta-\hat\theta)/\sigma)|$ for $\Phi(z)$ the cumulative distribution function of a N(0, 1). This is a confidence curve since, for all $\alpha \in (0, 1)$, $\{\theta : cc(\theta) \le 1-\alpha\} = (\hat\theta + \sigma\Phi^{-1}(\alpha/2),\ \hat\theta + \sigma\Phi^{-1}(1-\alpha/2))$ is the respective confidence interval of level $1-\alpha$. In the example the confidence curve has its minimum at $\hat\theta$ which is a point estimate of θ. The normal confidence curve is tail-symmetric, i.e. the probability of missing the parameter to the left equals that to the right and is $\alpha/2$ at level $1-\alpha$. A tail-symmetric confidence curve represents uniquely a confidence distribution, that is confidence curves that describe upper confidence limits. Confidence distribution is a term coined by Cox (1958) and formally defined in Schweder and Hjort (2002). For scalar parameters the fiducial distributions developed by Fisher (1930) are confidence distributions. Neyman (1934) saw that the fiducial distribution leads to confidence intervals. Cox (2013) sees confidence distributions as "simple and interpretable summaries of what can reasonably be learned from the data (and an assumed model)". Confidence distributions are reviewed by Xie and Singh (2013), and more broadly and with more emphasis on confidence curves by Schweder and Hjort (2016). In location models and other simple models the confidence distribution is obtained from pivots, e.g. the normal pivot $(\hat\theta-\theta)/\sigma$ in the

* Corresponding author.
E-mail address: pierpaolo.deblasi@unito.it (P. De Blasi).

http://dx.doi.org/10.1016/j.jspi.2017.09.010
0378-3758/

Please cite this article in press as: De Blasi P., Schweder T., Confidence distributions from likelihoods by median bias correction. J. Statist. Plann. Inference (2017), http://dx.doi.org/10.1016/j.jspi.2017.09.010.

above example. A canonical pivot is $C(\theta) = C(\theta; \hat{\theta}) = 1 - G(\hat{\theta}, \theta)$ where $G(y, \theta) = \mathrm{P}(\hat{\theta} \le y; \theta)$ is the distribution function of the maximum likelihood estimator $\hat{\theta}$, assumed to be absolutely continuous with respect to the Lebesgue measure and non-increasing in θ. See Section 2 for precise definitions and notation. The confidence distribution $C(\theta)$ is a canonical pivot in the sense of being uniformly distributed on the unit interval when $\hat{\theta}$ is distributed according to θ. When $\hat{\theta}$ is a sufficient statistic with monotone likelihood ratio, $C(\theta)$ is also optimal in the Neyman–Pearson sense, that is it describes smaller confidence intervals at a given level when compared to any other confidence distribution for the parameter θ (Schweder and Hjort, 2016 Section 5.4). An equal tailed confidence curve is readily obtained from $C(\theta)$ by $cc(\theta) = |1 - 2C(\theta)|$.

In this paper we shall be concerned with confidence curves obtained from the log-likelihood ratio $w(\theta)$, and we shall study the properties of median bias correction. Median bias correction of a confidence curve, proposed by Schweder (2007), is a method to make the resulting confidence curve approximately tail-symmetric. In the normal example $w(\theta) = ((\hat{\theta} - \theta)/\sigma)^2$ and the confidence curve mentioned above is also given by $cc(\theta) = Q(w(\theta))$ where Q is the cumulative chi-square distribution function with one degree of freedom. This confidence curve is tail-symmetric, as mentioned, and the confidence interval of level 0 is the single point $\hat{\theta}$ which thus has median θ and is said to be median unbiased. In general $w(\theta)$ hits zero at the maximum likelihood estimator, which might not be median unbiased. Let $\hat{\theta}$ have median $b(\theta)$. The median bias corrected confidence curve is the confidence curve of the parameter $b(\theta)$. The idea is to probability transform the bias corrected log likelihood ratio $w^*(\theta) = w(b(\theta))$ rather than $w(\theta)$. With $F^*(y; \theta)$ denoting the sampling distribution of $w^*(\theta)$ when the data is distributed according to θ, the bias corrected confidence curve is $cc^*(\theta) = F^*(w^*(\theta); \theta)$. Since $\arg\min(cc^*(\theta)) = b^{-1}(\hat{\theta})$ is median unbiased, the level set at $\alpha = 0$ is typically the single point $b^{-1}(\hat{\theta})$ and, by continuity, $cc^*(\theta)$ is close to be equal tailed at low levels. We undertake a theoretical study of the asymptotic properties of $cc^*(\theta)$ by showing that $cc^*(\theta)$ is third-order tail-symmetric for $(\theta, \hat{\theta})$ in the normal deviation range in two important classes of parametric models with parameter dimension one. First, we consider parametric models that belong to the Efron's normal transformation family (Efron, 1982). Then, we extend the result to regular one dimensional exponential families, where we also discuss the relation between median bias corrected and modified directed likelihood of Barndorff-Nielsen (1986), thus providing an alternative approximation to the latter. Since median bias correction works so well in these cases, it is reasonable to expect the method to work well quite generally. However, when a canonical confidence distribution is available, as in the exponential family models, we do of course not advocate to use median bias correction rather than using the canonical confidence distribution.

The rest of the paper is organized as follows. In Section 2, we recast confidence estimation in terms of confidence curves and introduce the notation we use in the sequel. We also define the confidence curve based on inverting the median bias corrected version of the log-likelihood ratio. In Sections 3 and 4, we investigate its asymptotic properties in terms of tail symmetry in the Efron's normal transformation family and in one dimensional exponential families, respectively. Finally, in Section 5 some concluding remarks and lines of future research are presented, together with an example that provides a preliminary illustration of the use of median bias correction in the presence of nuisance parameters. Some proofs and a technical lemma are deferred to Appendix.

2. Likelihood-based confidence curves

Let $X = (X_1, \ldots, X_n)$ be a continuous random sample with density $f(x; \theta)$ depending on a real parameter $\theta \in \Theta \subset \mathbb{R}$ and let $\mathrm{P}(\,\cdot\,; \theta)$ indicate probabilities calculated under $f(x; \theta)$. The log-likelihood is $\ell(\theta) = \ell(\theta; x) = \log f(x; \theta)$, and the log-likelihood ratio is $w(\theta) = w(\theta; x) = 2(\ell(\hat{\theta}; x) - \ell(\theta; x))$, where $\hat{\theta}$ is the maximum likelihood estimate. We drop the second argument in sample-dependent functions like w and ℓ whenever it is clear from the context whether we refer to a random quantity or to its observed value. Unless otherwise specified, all asymptotic approximations are for $n \to \infty$ and stochastic term $O_p(\cdot)$ refers to convergence in probability with respect to $f(x; \theta)$. We assume that the model is sufficiently regular for the validity of first order asymptotic theory, cfr. Barndorff-Nielsen and Cox (1994, Chapter 3). In particular, $w(\theta)$ converges in distribution to a chi-squared random variable, hence, by contouring $w(\theta)$ with respect to this distribution we obtain intervals of θ values given by the level sets for the curve $Q(w(\theta))$ where Q is the distribution function of the chi-squared distribution with 1 degree of freedom. This curve depends on the sample x and has its minimum at $\hat{\theta}$. However its level sets are not in general exact confidence intervals since the chi-squared approximation for the distribution of $w(\theta)$ is valid only for n large and the coverage probabilities equal the nominal levels only in the limit. As a consequence, $Q(w(\theta))$ is not uniformly distributed on the unit interval under $P(\,\cdot\,; \theta)$, a property we require for a *regular* confidence curve as spelled in the following definition.

Definition 1. A function $cc : \Theta \times \mathbb{R}^n \to [0, 1)$ is a regular confidence curve when $\min_\theta cc(\theta; x) = 0$, the level sets $\{\theta : cc(\theta; x) \le 1 - \alpha\}$ are finite intervals for all $\alpha \in (0, 1)$, and $cc(\theta; X) \sim \mathrm{Unif}(0, 1)$ under $P(\,\cdot\,; \theta)$.

Confidence curves might be defined for parameters of higher dimension and also for irregular curves that even might have more than one local minimum or might have infinite level sets for $\alpha < 1$, see Schweder and Hjort (2016, Section 4.6). Note that, under Definition 1, $I = \{\theta : cc(\theta; x) \le 1 - \alpha\}$ is an exact confidence region of level $1 - \alpha$ since $\mathrm{P}(I \ni \theta; \theta) = \mathrm{P}(cc(\theta; X) \le 1 - \alpha; \theta) = 1 - \alpha$. Among confidence curves, of special importance are confidence distributions, which are confidence curves that describe upper confidence limits. The definition is as follows.

Definition 2. A function $C : \Theta \times \mathbb{R}^n \to [0, 1)$ is a confidence distribution when $C(\cdot; x)$ is a cumulative distribution function in θ for all x and $C(\theta; X) \sim \mathrm{Unif}(0, 1)$ under $P(\,\cdot\,; \theta)$.

Please cite this article in press as: De Blasi P., Schweder T., Confidence distributions from likelihoods by median bias correction. J. Statist. Plann. Inference (2017), http://dx.doi.org/10.1016/j.jspi.2017.09.010.