Statistical Methodology 33 (2016) 114–130

Contents lists available at ScienceDirect

Statistical Methodology

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

Large sample convergence diagnostics for likelihood based inference: Logistic regression

Michael Brimacombe

Department of Biostatistics, KUMC, United States

ARTICLE INFO

Article history:
Received 27 January 2016
Received in revised form
3 August 2016
Accepted 5 August 2016
Available online 24 August 2016

Keywords:
Conditional distribution
Directional distribution
Likelihood diagnostics
Logistic regression
Nonlinear models

ABSTRACT

A general diagnostic approach to the evaluation of asymptotic approximation in likelihood based models is developed and applied to logistic regression. The expected asymptotic and observed log-likelihood functions are compared using a chi distribution in a directional Bayesian setting. This provides a general approach to assessing and visualizing non-convergence in higher dimensional models. Several well-known examples from the logistic regression literature are discussed.

1. Introduction

The importance of the likelihood function to statistical modeling and applications in the context of parametric statistical inference is well known, from both frequentist and Bayesian perspectives. From the frequentist perspective the likelihood function yields minimal sufficient statistics, if they exist, as well as providing a tool for the generation of pivotal quantities and measures of information on which to base estimation and hypothesis testing procedures. For researchers employing a Bayesian perspective the likelihood function is the major source of information regarding the data. It is modulated into a probability distribution directly on the parameter space through the use of a prior density and Bayes theorem. The Bayesian context preserves the whole of the likelihood function and allows for the use of probability calculus on the parameter space Ω itself. This usually takes the form of averaging out unwanted parameters in order to obtain marginal distributions for parameters of interest. Bernardo and Smith [3] and Box and Tiao [4] are general references.

E-mail address: mbrimacombe@kumc.edu.

http://dx.doi.org/10.1016/j.stamet.2016.08.001

Both approaches to inference may employ approximation, typically based on larger sample sizes, to evaluate required tail areas or central estimation regions both directly and comparatively. These are often based on the large sample normal forms of the likelihood function in many standard modeling situations. Assessing the convergence of the likelihood function to this form, at least locally, is of importance in assessing the accuracy of asymptotically justified calculations or simply the regularity of the likelihood in the presence of large samples. With the advent of Markov Chain Monte Carlo (MCMC) based methods exact calculations in the Bayesian setting are possible. See [11]. Here we use the Bayesian setting with non-informative or flat priors to examine likelihood properties.

In the case of nonlinear regression models, unusually shaped likelihood or posterior surfaces may result in standard asymptotic results being improperly applied. Standard asymptotic confidence regions can have very poor coverage when the likelihood is of non-standard form [9,8]. In the logistic regression model, which is nonlinear in its original parameters of interest, the effects of such nonlinearity, as measured in [13,15,8], may be pronounced leading to poor coverage and misleading inferences. While some of these effects can be reduced by employing specific reparameterizations of the likelihood function, see [8], these model adjustments are typically local in nature and often non-unique. Their application and interpretation is often not standard practice.

Standard methods of likelihood related diagnostic assessment, often in generalized linear models, typically cover a variety of model based issues. The overall predictive accuracy of the model-data combination, assessment of model fit through residual analysis, outlier detection and identifying influence points based on Cook's likelihood based distance are among the most common. See [14] for a review. Note that the lack of convergence of the distribution of the *m.l.e.* to its asymptotic normal distribution may also affect these model related issues.

In the presence of small samples frequentist pivotal statistics, especially those based on the Wald statistic, will often not achieve their expected large sample approximate distributions when the likelihood is non-normal or non standard in shape [18]. As noted in [12], the standard errors involved with the Wald statistic reflect the local curvature of the log-likelihood about the null value θ_0. If this curvature is much less than the curvature at the m.l.e. $\widehat{\theta}$, which may occur for example if the cell counts are highly unbalanced, the Wald statistic tends to underestimate changes in the log-likelihood and report inappropriate non-significant results.

An approach to assessing the local curvature of the likelihood surface is through the profiling of the likelihood. This is typically done along the axis in Ω corresponding to the specific coordinate or parameter of interest, say θ_1. The intersection of the surface defined by the remaining $p-1$ normal equations related to $\theta_2, \ldots, \theta_p$ all set equal to their maximum values and the original p-dimensional likelihood surface gives a one-dimensional curve or profile along the likelihood surface parallel to the θ_1 coordinate axis. If this profile follows a fairly normal shape, the Wald statistic tends be more stable and accurate [18], though in general, test statistics based on profiles of the likelihood function can give inefficient or inconsistent estimators. This typically occurs if the number of nuisance parameters depends on the size of the sample, but may be corrected by adjusting the first derivative of the log-likelihood [16].

The assessment of local curvature properties and anomalies in the likelihood is also useful in purely Bayesian settings. Correlations in the joint posterior often correspond to nonlinear aspects of the likelihood and the resulting posterior density surface, affecting the accuracy and convergence rates of posterior sampling approaches used to obtain marginal posterior and predictive distributions. Diagnostic approaches to assessing poorly behaved likelihood functions and resulting inaccuracies for frequentist-likelihood and Bayesian settings have been suggested in [1,13,15,19,17].

Directional considerations from a frequentist diagnostic perspective are given in [13] where the directional aspect is defined over the sample space rather than Ω. A directional tail area of interest is defined to locally compare the actual density to an approximating normal density by varying a directional unit vector over the sample space. There is often no easy way to relate a direction of interest defined in the sample space with directions in Ω over which the likelihood function is defined.

In settings with higher dimensional likelihood surfaces, directional approaches offer a practical insight to the difficult problem of assessing higher dimensional surface properties. Rather than likelihood profiles which may be unstable [18], a directional Bayesian approach provides a stable probability based scale for reference. In the diagnostic approach developed here, the entire actual