Computational Statistics and Data Analysis 96 (2016) 120–132

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Frequentist nonparametric goodness-of-fit tests via marginal likelihood ratios

Jeffrey D. Hart [a], Taeryon Choi [b,*], Seongbaek Yi [c]

[a] Department of Statistics, Texas A&M University, USA
[b] Department of Statistics, Korea University, Republic of Korea
[c] Department of Statistics, Pukyong National University, Republic of Korea

ARTICLE INFO

Article history:
Received 30 June 2015
Received in revised form 23 October 2015
Accepted 28 October 2015
Available online 7 November 2015

Keywords:
Bandwidth parameter
Empirical null distribution
Goodness-of-fit tests
Kernel density estimation
Marginal likelihoods

ABSTRACT

A nonparametric procedure for testing the goodness of fit of a parametric density is investigated. The test statistic is the ratio of two marginal likelihoods corresponding to a kernel estimate and the parametric model. The marginal likelihood for the kernel estimate is obtained by proposing a prior for the estimate's bandwidth, and then integrating the product of this prior and a leave-one-out kernel likelihood. Properties of the kernel-based marginal likelihood depend importantly on the kernel used. In particular, a specific, somewhat heavy-tailed, kernel K_0 yields better performing marginal likelihood ratios than does the popular Gaussian kernel. Monte Carlo is used to compare the power of the new test with that of the Shapiro–Wilk test, the Kolmogorov–Smirnov test, and a recently proposed goodness-of-fit test based on empirical likelihood ratios. Properties of these tests are considered when testing the fit of normal and double exponential distributions. The new test is used to establish a claim made in the astronomy literature concerning the distribution of nebulae brightnesses in the Andromeda galaxy. Generalizations to the multivariate case are also described.

1. Introduction

We consider the classical problem of testing the goodness-of-fit of a parametric model for a distribution. Our approach is nonparametric in that our test can consistently detect virtually any departure from the null hypothesis. The proposed test statistic is based on the ratio of two marginal likelihoods, in which the alternative "model" corresponds to a kernel density estimator. There exist a number of well-known goodness-of-fit tests including Neyman's smooth test (Neyman, 1937), the Kolmogorov–Smirnov test (Smirnov, 1944), the Anderson–Darling test (Anderson and Darling, 1954) and, for testing normality, the Shapiro–Wilk test (Shapiro and Wilk, 1965). Additionally, several recent developments have been made in the literature of nonparametric goodness-of-fit testing, see, e.g., Claeskens and Hjort (2004) for a goodness-of-fit test based on nonparametric likelihood ratios, and, more recently, Vexler and Gurevich (2010) for a goodness-of-fit test based on sample entropy and empirical likelihood ratios. The reader is also referred to the monographs of Hart (1997), Ingster and Suslina (2003), Rayner et al. (2009) and Thas (2010) and references therein for general theory and methods in the treatment of goodness-of-fit and lack-of-fit problems.

* Corresponding author. Tel.: +82 2 3290 2245; fax: +82 2 924 9895.
E-mail address: trchoi@gmail.com (T. Choi).

http://dx.doi.org/10.1016/j.csda.2015.10.013
0167-9473/

Suppose that we have independent and identically distributed observations $\boldsymbol{X}_n = (X_1, \ldots, X_n)$, and wish to compare the fit of two candidate models $\mathcal{M}_0$ and $\mathcal{M}_1$ for the distribution of X_i:

$$\mathcal{M}_0 = \{f_0(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\} \quad \text{and} \quad \mathcal{M}_1 = \{f_1(\cdot|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}, \tag{1}$$

where f_0 and f_1 are known up to their respective parameter vectors.

The marginal likelihood $m(\boldsymbol{X}_n)$ is obtained by integrating the product of the likelihood and a prior over the parameter space. Given prior distributions π_0 and π_1 for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, respectively, the marginal likelihoods $m_0(\boldsymbol{X}_n)$ and $m_1(\boldsymbol{X}_n)$ are given by

$$m_0(\boldsymbol{X}_n) = \int \prod_{i=1}^{n} f_0(X_i|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}$$

and

$$m_1(\boldsymbol{X}_n) = \int \prod_{i=1}^{n} f_1(X_i|\boldsymbol{\lambda})\pi_1(\boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

The ratio $m_0(\boldsymbol{X}_n)/m_1(\boldsymbol{X}_n)$ is known as a Bayes factor (see, e.g., Jeffreys, 1961 and Kass and Raftery, 1995), and represents the ratio of the prior and posterior odds of the two models. On the other hand, it may also be used in frequentist fashion to test the fit of one model versus another, an idea which appears to be due to Good (1957), who referred to the idea as a *Bayes/non-Bayes synthesis(compromise)*. Good (1967) proposed using the null distribution of a Bayes factor as a significance criterion. An extensive review of such approaches is given by Good (1992). Recently (Aerts et al., 2004; Hart, 2009) proposed Bayesian-motivated frequentist tests for lack of fit in the context of regression.

Our paper likewise proposes a Bayes/non-Bayes synthesis for nonparametric goodness-of-fit testing in which we use as test statistic a marginal likelihood ratio comparing a parametric model to essentially arbitrary alternatives. The alternative is estimated by a kernel density estimate, whose only parameter is its bandwidth. We therefore propose a prior distribution for the bandwidth and obtain a nonparametric marginal likelihood by integrating the product of this prior and a leave-one-out type of kernel likelihood. A traditional marginal likelihood is computed for the parametric null model. It is shown that the proposed nonparametric goodness-of-fit test using the marginal likelihood ratio is easily computed and powerful in detecting departures from the null hypothesis.

The remainder of the paper proceeds as follows. Our basic approach is described in Section 2. This includes how we choose kernels and prior distributions for the bandwidth, and a statement of conditions guaranteeing that the nonparametric marginal likelihood exists finite. Monte Carlo studies of power when testing normality and double exponentiality, and an application of our approach to astronomy data are presented in Section 3. Extensions to multivariate distributions are described in Section 4, and concluding remarks given in Section 5.

2. Basic approach: marginal likelihood ratio

Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ be a random sample of size n from an unknown probability density f. We wish to test the following hypotheses:

$$H_0 : f \in \mathcal{F}_0 = \{f_0(\cdot|\boldsymbol{\theta})|\boldsymbol{\theta} \in \Theta\} \quad \text{vs.} \quad H_1 : f \notin \mathcal{F}_0, \tag{2}$$

where $f_0(\cdot|\boldsymbol{\theta})$ is known up to the vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, which is unknown. Note that $\mathcal{F}_0$ is a completely specified parametric family of distributions, e.g., the normal distribution with mean θ_1 and variance θ_2, $\boldsymbol{\theta} = (\theta_1, \theta_2)$. The alternative is nonparametric, in that it contains densities not in $\mathcal{F}_0$.

As indicated previously, our test statistic is a marginal likelihood ratio that consists of a parametric marginal likelihood for the null model and a nonparametric marginal likelihood based on kernel density estimates for the alternative model. In particular, we consider the following cross-validated likelihood under H_1:

$$L(h|\boldsymbol{X}_n) = \prod_{i=1}^{n} \widehat{f}_i(X_i|h), \qquad \widehat{f}_i(x|h) = \frac{1}{(n-1)h}\sum_{j\neq i} K\left(\frac{x - X_j}{h}\right), \tag{3}$$

where $h > 0$ is a bandwidth and K a suitable kernel function, a popular choice of which is the Gaussian kernel, i.e., $K \equiv \phi$, with $\phi(z) = (2\pi)^{-1/2}\exp\left(-\frac{1}{2}z^2\right)$. The kernel estimates $\widehat{f}_i(x|h)$, $i = 1, \ldots, n$, Marron (1985) are known as *cross-validated* or *leave-one-out* estimates, inasmuch as $\widehat{f}_i(x|h)$ is computed without X_i. It is well-known that $\widehat{f}_i(x|h)$ is consistent for $f(x)$ (Chow et al., 1983) under mild conditions, including $nh \to \infty$ and $h \to 0$ as $n \to \infty$. Typically $L(h|\boldsymbol{X}_n)$ is used as a bandwidth selection criterion, but we use it to construct a nonparametric marginal likelihood under the alternative hypothesis H_1.

Let π_{bw} be a prior for the bandwidth h and π_0 a prior for $\boldsymbol{\theta}$, both of which are assumed to be proper. Define the following marginal likelihood ratio (MLR_n) based on n observations:

$$\mathrm{MLR}_n \equiv \mathrm{MLR}(\boldsymbol{X}_n) = \frac{\int_0^\infty L(h|\boldsymbol{X}_n)\pi_{\mathrm{bw}}(h)dh}{\int_\Theta \Pi_{i=1}^n f_0(X_i|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{4}$$

As noted in the previous section, MLR_n may be used in frequentist fashion to test H_0. Thus, one must approximate the null distribution of MLR_n (which is not analytically available in general), and then reject H_0 at level α if the observed value of MLR_n exceeds the $(1-\alpha)$ percentile of the null distribution.