Statistical Methodology 31 (2016) 20–42

Contents lists available at ScienceDirect

Statistical Methodology

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

On testing local hypotheses via local divergence

G. Avlogiaris [a], A. Micheas [b], K. Zografos [a,*]

[a] University of Ioannina, Department of Mathematics, 451 10 Ioannina, Greece
[b] University of Missouri-Columbia, Department of Statistics, Columbia, USA

ARTICLE INFO

Article history:
Received 29 December 2014
Received in revised form
12 September 2015
Accepted 11 January 2016
Available online 29 January 2016

MSC:
62F03
62F05
62B10

Keywords:
Cressie and Read power divergence
Local divergence
Local Fisher information matrix
Local Rao and Wald tests
Local goodness of fit test
Local homogeneity test
Local ϕ-divergence test

ABSTRACT

The aim of this paper is to propose procedures that test statistical hypotheses locally, that is, assess the validity of a model in a specific domain of the data. In this context, the one and two sample problems will be discussed. The proposed tests are based on local divergences which are defined in such a way as to quantify the divergence between probability distributions locally, in a specific area of the joint domain of the underlined models. The theoretical results are exemplified using simulations and two real datasets.

1. Introduction

Statistical hypotheses tests have been developed as general procedures that help the experimenter assess the applicability of an entertained model. Typically, the test statistic is a quantity that is computed over the whole domain of the data (global test) and is based on a divergence or discrepancy measure between the estimated model and the model that is specified by the null hypothesis. This paper has been motivated by the fact that the test statistic, namely, the divergence between the estimated and the null model, may be considerably different in some areas of the domain of the data.

* Corresponding author.
E-mail address: kzograf@uoi.gr (K. Zografos).

http://dx.doi.org/10.1016/j.stamet.2016.01.003
1572-3127/

This means that although a null model is accepted or rejected over the whole domain of the data by some test, the conclusion is quite different if the interest is focused on a specific area of the domain of the data.

Statistical information theory provides with tools for measuring the divergence between two or more statistical models. As such, over the last few decades statistical information theory became the main tool in developing methodologies in mathematical statistics. The monograph by L. Pardo and the references therein (cf., [17]) provides an exhaustive discussion on the use of divergences in testing statistical hypotheses, in several disciplines and contexts. Recent advances on the subject can be found in [3]. Global homogeneity tests between two or more distributions were studied in [23].

Starting with two probability distributions f_1 and f_2, a divergence measure $D(f_1, f_2)$ quantifies the discrepancy or dissimilarity between the underlined models f_1 and f_2, if the divergence D satisfies the key property $D(f_1, f_2) \geq 0$, with equality if and only if models f_1 and f_2 coincide. Consequently, an empirical version of a divergence measure $D(f_1, f_2)$ can serve as a test statistic for testing coincidence of the models f_1 and f_2. In a similar manner, an empirical version of a divergence measure $D(f, f_0)$, serves as a test statistic for testing if the true but unknown model f can be approximated well by the model f_0, which is specified by the null hypothesis, and therefore, an empirical version of $D(f, f_0)$ can be used in a goodness of fit setting.

There is a vast literature on divergence measures, starting with the pioneering divergence measure introduced in the paper by Kullback and Leibler [12]. This measure has been extensively studied and illustrated with applications to several fields in the book by Kullback [11]. However, the most broad family of divergence measures was proposed by Csiszár [5,6] and independently by Ali and Silvey [1]. Csiszár's ϕ-divergence, in a fully parametric framework, is defined as follows; let $(\mathcal{X}, \mathcal{A}, P_\theta)$, $\theta \in \Theta$, be a probability space, with Θ an open subset of $\mathcal{R}^M$, with $M \geq 1$. For $\theta_1, \theta_2 \in \Theta$, denote by f_{θ_i}, the Radon–Nikodym derivatives $f_{\theta_i} = \frac{dP_{\theta_i}}{d\mu}$, with μ a σ-finite measure on $(\mathcal{X}, \mathcal{A})$ and $P_{\theta_i} \ll \mu$, for $i = 1, 2$. In this framework, Csiszár's ϕ-divergence between f_{θ_1} and f_{θ_2} is defined by

$$D_\phi(f_{\theta_1}, f_{\theta_2}) = D_\phi(\theta_1, \theta_2) = \int_{\mathcal{X}} f_{\theta_2}(x)\phi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) d\mu(x), \tag{1}$$

where ϕ is a real valued convex function, satisfying appropriate conditions which ensure the existence of the above integral (cf., [5,6,17]). Csiszár's ϕ-divergence has been axiomatically characterized and studied extensively by Liese and Vajda [14,15], Vajda [22], and Stummer and Vajda [21], among many others. It can be thought of as a similarity measure between f_{θ_1} and f_{θ_2}, since it satisfies the property

$$D_\phi(f_{\theta_1}, f_{\theta_2}) \geq 0, \text{ with equality, if and only if } f_{\theta_1} = f_{\theta_2}, \tag{2}$$

in the class of the strictly convex functions ϕ at 1, with $\phi(1) = 0$. Particular choices of the convex functions ϕ, lead to important measures of divergence including Kullback and Leibler [12], Renyi [18] and Cressie and Read [4] λ-power divergence, to name a few.

Consider two independent random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ from the populations described by models f_{θ_1} and f_{θ_2}, $\theta_1, \theta_2 \in \Theta$, respectively. In such a statistical framework, testing the null hypothesis of homogeneity of the two populations, $H_0 : \theta_1 = \theta_2$, is of primary interest. A test procedure for testing H_0 can be based on the test statistic $D_\phi(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$, obtained from (1), where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ denote the MLE's (Maximum Likelihood Estimators) of θ_1 and θ_2, based on the two random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ from f_{θ_1} and f_{θ_2}, respectively. Based on (2), small values of $D_\phi(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ are in favor of H_0, while large values of $D_\phi(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ suggest rejection of H_0. This approach is supported intuitively, since large values of $D_\phi(f_{\widehat{\theta}_1}, f_{\widehat{\theta}_2})$ suggest that the empirical models $f_{\widehat{\theta}_1}$ and $f_{\widehat{\theta}_2}$ are not the same and the same is expected for the respective theoretic models f_{θ_1} and f_{θ_2}, since $D_\phi(f_{\theta_1}, f_{\theta_2})$ quantifies the degree of divergence of f_{θ_1} and f_{θ_2} in the whole domain $\mathcal{X}$.

However, there are examples of real datasets where an elementary descriptive analysis of the data indicates that there are areas of the joint domain $\mathcal{X}$ of the data where the hypothesis of homogeneity is violated, even if homogeneity is accepted in the whole domain $\mathcal{X}$ using any standard test of homogeneity. Such an example will be presented in the application section at the end of this paper. Motivated by this example, Csiszár's ϕ-divergence $D_\phi(f_{\theta_1}, f_{\theta_2})$ informs about the discrepancy or the