Journal of Statistical Planning and Inference 140 (2010) 2875–2884

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

On nonparametric comparison of images and regression surfaces

Xiao-Feng Wang [a,b,*], Deping Ye [c]

[a] Department of Quantitative Health Sciences/Biostatistics, Cleveland Clinic Foundation, 9500 Euclid Avenue/JJN3, Cleveland, OH 44195, USA
[b] Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA
[c] Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106, USA

ARTICLE INFO

Article history:
Received 26 September 2008
Accepted 9 November 2009
Available online 7 March 2010

MSC:
62G10
62G08

Keywords:
Multivariate local regression
Images
Equality
Regression surfaces
Wild bootstrap
Heteroscedastic model
Spatial correlated errors

ABSTRACT

Multivariate local regression is an important tool for image processing and analysis. In many practical biomedical problems, one is often interested in comparing a group of images or regression surfaces. In this paper, we extend the existing method of testing the equality of nonparametric curves by Dette and Neumeyer (2001) and consider a test statistic by means of an $\mathcal{L}^2$–distance in the multi-dimensional case under a completely heteroscedastic nonparametric model. The test statistic is also extended to be used in the case of spatial correlated errors. Two bootstrap procedures are described in order to approximate the critical values of the test depending on the nature of random errors. The resulting algorithms and analyses are illustrated from both simulation studies and a real medical example.

1. Introduction

Medical images are increasingly used in health care and biomedical research and a wide range of imaging modalities are now available. Statistical image analysis, hence, becomes an active research area. Nonparametric regression techniques have been broadly applied to image analysis, including image reconstruction, denoising and interpolation. An image can be considered as a surface of the image intensity at each pixel. A regression surface from a noisy image is often fitted by local smoothing procedures (Qiu, 1998; Takeda et al., 2007). Based on the nonparametric modeling framework, one object of primary interest is to compare a set of smoothed images or regression surfaces. For example, one often exams the equality of two or more images under different clinical conditions in medical applications.

Nonparametric comparison of a set of regression curves has been paid considerable attentions in both theoretical and applied regression analysis. Much effort has been devoted to this problem in the literature. Hall and Hart (1990) and King et al. (1991) had early considerations of the problem, where they discussed a completely nonparametric homoscedastic model in the case of equal design points. While Kulasekera (1995) proposed several alternative tests in the case of unequal design points, Young and Bowman (1995) generalized the one-way analysis of variance (ANOVA) to the nonparametric regression setting. Hardle and Mammen (1990) considered a method based on a weighted $\mathcal{L}^2$–distance for semiparametric comparison of regression curves. Dette and Neumeyer (2001) discussed three methods using nonparametric estimators of

* Corresponding author at: Department of Quantitative Health Sciences/Biostatistics, Cleveland Clinic Foundation, 9500 Euclid Avenue/JJN3, Cleveland, OH 44195, USA.
E-mail addresses: wangx6@ccf.org (X.-F. Wang), deping.ye@case.edu (D. Ye).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2010.03.011

the regression function in the problem of testing the equality of k regression curves from independent samples. Their first test was based on a linear combination of variance estimators; the second approach was an ANOVA-type method; and the third test compared the differences between the estimates of the individual regression curves by means of an $\mathcal{L}^2$−distance. Recent contributions on this problem include Pardo-Fernandez et al. (2007) and Pardo-Fernandez (2007) among others. Several new testing statistics have been discussed such as Kolmogorov–Smirnov type and Cramer–von Mises type statistics. A good and recent review on this topic can be found in Neumeyer and Dette (2003).

Nonparametric comparison of different images or regression surfaces relates to nonparametric analysis of covariance with multiple covariates. It is important to extend the methods for nonparametric curve comparison to the multi-dimensional case with applications to image analysis. Recently, Bowman (2006) suggested a generalization of the ANOVA-type test by Young and Bowman (1995) to compare regression surfaces. Under the assumptions of equal homoscedastic variances in all groups and normal distributed errors, Bowman (2006) proposed a χ^2−approximation of the corresponding test statistic under the null hypothesis. In this paper, we discuss the comparison of regression surfaces under a general model which does not require any additional assumptions (such as homoscedasticity or normality of errors, equal design points). In Section 2, we review the classic framework of local regression for image data and suggest a generalization of Dette and Neumeyer (2001)'s test, a test procedure based on an $\mathcal{L}^2$−distance of regression surfaces under a general heteroscedastic model. The asymptotic results of the proposed test statistic are presented here. We also extend the test statistic to be used in the case of spatial correlated errors. Two bootstrap procedures are described in order to approximate the critical values of the test. In Section 3, we present numerical examples. Simulation studies are conducted to investigate the finite sample properties of the proposed test. A real data analysis is performed to illustrate the use of our method. The paper ends with the concluding remarks in Section 4.

2. Testing the equality of images and regression surfaces

2.1. Local surface model for images

An image can be represented as a function $m(\cdot)$ from a plane $\mathbb{R}^2$ to p-variate space $\mathbb{R}^p$, where the value of $m(\cdot)$ in the jth coordinate represents its intensity. The dimension of m$(\cdot)$, p, can be >1. For instance, $p=3$ for a natural color image, representing the three primary colors. For simplicity and clarity we will treat only the case $p=1$, while the methodology is similar in other cases. In image application, the function $m(\cdot)$ is assumed to be a smooth function of $\mathbf{X}\in\mathbb{R}^2$. We only have discrete data on the model with the $\mathbf{X}$ variable restricted to a regular or irregular grid and the intensity values observed are often measured with noise. In general, this model can be formalized as

$$Y_j = m(\mathbf{X}_j)+\sigma_i(\mathbf{X}_j)\varepsilon_j, \quad j=1,\ldots,n, \tag{2.1}$$

where ε_j are independent and identically distributed random variables, which represents random errors in the observations. We further assume that ε_j have zero mean and finite variance 1.

Using the data $(\mathbf{X}_j, Y_j)$, $j=1,\ldots,n$, we want to construct a "denoised" image, an estimator of the regression function $m(\cdot)$ which is the conditional expectation of the dependent variable Y given the independent variable $\mathbf{X}$,

$$m(\mathbf{x}) = \mathrm{E}(Y|\mathbf{X}=\mathbf{x}).$$

Local smoothing method is an important tool in image processing and analysis (Wand and Jones, 1995; Takeda et al., 2007). We now describe how to construct a smoothed image by local approximations of $m(\cdot)$, using a least-squares method with kernel weights.

The Taylor's expansion of the regression function at $\mathbf{X}_j$ implies that,

$$m(\mathbf{X}_j)\approx \beta_0+\boldsymbol{\beta}_1^T(\mathbf{X}_j-\mathbf{x})+\boldsymbol{\beta}_2^T \text{ vech } \{(\mathbf{X}_j-\mathbf{x})^T(\mathbf{X}_j-\mathbf{x})\}+\cdots,$$

where vech$(\cdot)$ returns the vector obtained by eliminating all supradiagonal elements of the square matrix and stacking the result one column above the other. $\beta_0=m(\mathbf{x})$ is the pixel value of interest and the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are

$$\boldsymbol{\beta}_1=\left[\frac{\partial m(\mathbf{x})}{\partial x_1}, \quad \frac{\partial m(\mathbf{x})}{\partial x_2}\right]^T, \quad \boldsymbol{\beta}_2=\frac{1}{2}\left[\frac{\partial^2 m(\mathbf{x})}{\partial x_1^2}, \quad 2\frac{\partial^2 m(\mathbf{x})}{\partial x_1\partial x_2}, \quad \frac{\partial^2 m(\mathbf{x})}{\partial x_2^2}\right]^T.$$

Estimating $\boldsymbol{\beta}=[\beta_0,\boldsymbol{\beta}_1^T,\boldsymbol{\beta}_2^T,\ldots]^T$ can be through solving the following least squares problem,

$$\min_{\boldsymbol{\beta}}\sum_{j=1}^{n}\{Y_j-\beta_0-\boldsymbol{\beta}_1^T(\mathbf{X}_j-\mathbf{x})-\boldsymbol{\beta}_2^T \text{ vech } \{(\mathbf{X}_j-\mathbf{x})^T(\mathbf{X}_j-\mathbf{x})\}-\cdots\}^2K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_j).$$

The weight function (kernel function) $K_{\mathbf{H}}(\mathbf{x})=det(\mathbf{H})^{-1}\ K(\mathbf{H}^{-1}\mathbf{x})$ is defined on the multivariate space, hence observations close to a fitting point $\mathbf{x}$ receive large weights. $\mathbf{H}$ is a bandwidth matrix which is symmetric positive-definite and $det(\mathbf{H})$ is the determinant of the matrix $\mathbf{H}$. The local least squares estimator of $m(\mathbf{x})$ is

$$\hat{m}_{\mathbf{H}}(\mathbf{x})=\mathbf{e}_1^T(\tilde{\mathbf{X}}^T\mathbf{W}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\mathbf{W}\mathbf{Y}, \tag{2.2}$$