Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 6229–6241
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Semiparametric regression for assessing agreement using tolerance bands

Pankaj K. Choudhary*

Department of Mathematical Sciences EC 35, University of Texas at Dallas, P.O. Box 830688, Richardson, TX 75083-0688, USA

Received 18 August 2006; received in revised form 6 January 2007; accepted 6 January 2007
Available online 26 January 2007

Abstract

This article describes a Bayesian semiparametric approach for assessing agreement between two methods for measuring a continuous variable using tolerance bands. A tolerance band quantifies the extent of agreement in methods as a function of a covariate by estimating the range of their differences in a specified large proportion of population. The mean function of differences is modelled using a penalized spline through its mixed model representation. The covariance matrix of the errors may also depend on a covariate. The Bayesian approach is straightforward to implement using the Markov chain Monte Carlo methodology. It provides an alternative to the rather ad hoc frequentist likelihood-based approaches that do not work well in general. Simulation for two commonly used models and their special cases suggests that the proposed Bayesian method has reasonably good frequentist coverage. Two real data sets are used for illustration, and the Bayesian and the frequentist inferences are compared.

Keywords: Limits of agreement; Method comparison; Mixed model; Penalized spline; Tolerance interval; Total deviation index

1. Introduction

In this article, we discuss inference procedures for the following problem: we have a scalar response y_x, which conditional on a scalar continuous covariate $x \in \mathfrak{X}$, follows a $\mathcal{N}(\mu_x \equiv f(x, \beta, b), \sigma_x^2)$ distribution. The mean function f is modelled nonparametrically via penalized splines regression. A pth degree spline model is

$$f(x, \beta, b) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{k=1}^{K} b_k (x - c_k)_+^p = X_x \beta + Z_x b, \tag{1}$$

where β is the $(p + 1) \times 1$ vector $(\beta_0, \ldots, \beta_p)$; K is the number of knots; b is the $K \times 1$ vector $(b_1, \ldots, b_K)$; $c_1 < \cdots < c_K$ are the knot locations; $b_1, \ldots, b_K$ are the coefficients of the truncated polynomial basis functions $(x - c_1)_+^p, \ldots, (x - c_K)_+^p$; $(x - c)_+ = \max\{0, x - c\}$; X_x is the $1 \times (p + 1)$ vector $(1, x, \ldots, x^p)$; and Z_x is the $1 \times K$ vector $((x - c_1)_+^p, \ldots, (x - c_K)_+^p)$. See Ruppert et al. (2003) for an excellent introduction to the methodology of penalized splines regression. We use the mixed model representation of the spline f, where the coefficients $b_1, \ldots, b_K$ are treated as independently distributed $\mathcal{N}(0, \sigma_b^2)$ variables. In the mixed model terminology, β is called a fixed-effect and b is

* Tel.: +1 972 883 4436; fax: +1 972 883 6622.
E-mail address: pankaj@utdallas.edu.

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.01.006

called a random-effect. Our main interest lies in the parameter function q_x—the p_0th quantile of $|y_x|$ for a given large probability p_0. It is defined as

$$q_x = \sigma_x \{\chi_1^2(p_0, \mu_x^2/\sigma_x^2)\}^{1/2}, \tag{2}$$

where $\chi_1^2(p_0, \Delta)$ denotes the p_0th quantile of a noncentral chisquare distribution with one degree of freedom and noncentrality parameter Δ. This function is random under the mixed model representation of (1). Our goal is to obtain a simultaneous upper bound U_x such that

$$\Pr(q_x \leqslant U_x, x \in \mathfrak{X}) = 1 - \alpha. \tag{3}$$

In practice, $\mathfrak{X}$ is a finite interval representing the range of observed values of x.

We are interested in the application of this methodology to assess agreement between two methods of measuring a continuous response. In a typical method comparison study, one method serves as a test method and the other serves as a reference. Generally the test method provides a cheaper or less invasive alternative to the reference method. Sometimes the test method may well be more accurate and precise than the reference method. The goal of their comparison is to evaluate the extent of their agreement and judge whether it is high enough to warrant their interchangeable use in practice. In this context, x is a covariate and y_x represents the population of differences in paired measurements from the two methods at x. The quantile q_x measures the extent of agreement between the methods. Its small value implies a good agreement at x. With U_x defined by (3), the interval $[-U_x, U_x]$, $x \in \mathfrak{X}$, becomes a p_0 probability content simultaneous tolerance band for the distribution of y_x over $\mathfrak{X}$ in the sense that

$$\Pr\{F_x(U_x) - F_x(-U_x) \geqslant p_0, \quad \text{for all } \ x \in \mathfrak{X}\} = 1 - \alpha,$$

where F_x is the cumulative distribution function of y_x. This band estimates the range of p_0 proportion of population differences as a function of x. The practitioner uses it to infer regions of $\mathfrak{X}$ where the differences within the band are clinically unimportant. The agreement in these regions is considered good enough for interchangeable use of the two methods. This approach for agreement evaluation was introduced in Lin (2000), Lin et al. (2002) and Choudhary and Nagaraja (2007) for the case when the differences are independently and identically distributed. The agreement measure here—the p_0th quantile of absolute differences, is called the "total deviation index" in Lin (2000). Choudhary and Ng (2006) extended this approach for the case when the distribution of differences depends on x, and Choudhary (2007) generalized it to incorporate repeated measurements data. We now introduce two real data applications.

Oestradiol data: In this example from Hawkins (2002), the interest lies in comparing two assays for Oestradiol—a naturally occurring female hormone synthesized to treat estrogen deficiency. The data consist of pairs of measurements of Oestradiol concentration (in pg/ml) from the two assays. Here we take the difference (assay1 − assay2) in concentrations as the response y_x and the average concentration as the covariate x. This average serves as a proxy for the magnitude of the true concentration. Its choice as the covariate is motivated by an exploratory analysis of the data. Let (x_i, y_i) be the value of (x, y_x) on the ith unit in the sample, $i = 1, \ldots, m = 139$. We take $\mathfrak{X} = [\min x_i = 2, \max x_i = 12, 201]$. The scatterplot of $(\log x, y_x)$ in Fig. 1(a) reveals that the mean response and the variability in response depend on x. For these data, Choudhary and Ng (2006) consider a model of the form

$$y_i = f(x_i, \beta, b) + h^{1/2}(x_i, \beta, \lambda)\varepsilon_i, \tag{4}$$

where the mean function f is given by (1); the errors ε_i's follow independent $\mathcal{N}(0, \sigma_e^2)$ distributions and are mutually independent of the random-effects b_k's; and h is a variance function for modelling heteroscedasticity. It follows that the response $y_x \sim \mathcal{N}(\mu_x = f(x, \beta, b), \sigma_x^2 = \sigma_e^2\, h(x, \beta, \lambda))$. Based upon this model, the authors describe a likelihood-based methodology for computing U_x to satisfy (3). We analyze these data in Section 4.

Body fat data: In this Young Women's Health Study example from Chinchilli et al. (1996), we are interested in comparing two methods for measuring percentage body fat—skinfold calipers and dual energy X-ray absorptiometry (DEXA). The data consist of paired body fat measurements from the two methods taken over a course of about five years on a cohort of $m = 91$ adolescent girls. There were nine visits of the girls roughly six months apart with the first visit around age twelve. DEXA measurements are not available for the first visit. We have between four to eight complete pairs of measurements on each girl yielding a total of 654 pairs after excluding three outliers. Fig. 2 presents the scatterplots of these data for visits two through nine. The methods do not seem to be highly correlated. Here we take the covariate x as (age in years at the time of visit − 12) $\in \mathfrak{X} = [-0.80, 5.30]$. Let y_{ij} be the difference (calipers − DEXA)