Journal of Multivariate Analysis 141 (2015) 179–196

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Nonparametric and semiparametric compound estimation in multiple covariates

Richard Charnigo *, Limin Feng [1], Cidambi Srinivasan

Department of Statistics, 725 Rose Street, University of Kentucky, Lexington KY 40536, United States

ARTICLE INFO

Article history:
Received 28 August 2014
Available online 21 July 2015

AMS 2010 subject classifications:
62G08
62H12
62P10
62G20

Keywords:
Derivative
Parkinson's disease
Random effects
Regression
Repeated measures
Telemonitoring

ABSTRACT

We consider the problem of simultaneously estimating a mean response function and its partial derivatives, when the mean response function depends nonparametrically on two or more covariates. To address this problem, we propose a "compound estimation" approach, in which differentiation and estimation are interchangeable: an estimated partial derivative is exactly equal to the corresponding partial derivative of the estimated mean response function. Compound estimation yields essentially optimal convergence rates and may exhibit substantially smaller squared error in finite samples compared to local regression. We also explain how to employ compound estimation under more general circumstances, when the mean response function depends parametrically on some additional covariates and the observations are not statistically independent. In a case study, we apply compound estimation to examine how the progression of Parkinson's disease may relate to a subject's age and the signal fractal scaling exponent of the subject's recorded voice. Especially among those intermediate in age, an abnormal signal fractal scaling exponent may portend greater symptom progression.

1. Introduction

Let J and D be positive integers. Suppose that we observe data from the nonparametric regression model

$$Y_i = \mu(\mathbf{x}_i) + \epsilon_i \quad \text{for } i \in \{1, \dots, n\}, \tag{1}$$

where the mean response function $\mu(\mathbf{x})$ is defined on a (nonempty) compact rectangle $\mathcal{X} \subset \mathbb{R}^D$ and has continuous derivatives of order $(J + 1)$, the vectors of observed covariate values $\mathbf{x}_1, \dots, \mathbf{x}_n$ belong to $\mathcal{X}$, and $\epsilon_1, \dots, \epsilon_n$ are independent zero-mean errors with variances bounded above by a constant $M \in (0, \infty)$. Without loss of generality, we assume that $\mathcal{X} = [-1, 1]^D$ except where otherwise indicated. Also, for the purpose of developing theory in this work, we regard covariate values as fixed; if they were realizations of a random process, then our analysis is conditional on those realizations.

Our goal for model (1) is to simultaneously estimate $\mu(\mathbf{x})$ and its derivatives of up to order J, in such a way that differentiation and estimation are interchangeable: a partial derivative of $\widehat{\mu}(\mathbf{x})$ estimates the corresponding partial derivative of $\mu(\mathbf{x})$, with the same tuning parameters used in the estimation of both $\mu(\mathbf{x})$ and its derivatives. In Section 2 we propose a "compound estimator" that accomplishes this objective for arbitrary J and D while enjoying essentially optimal convergence

Abbreviations: BLUE, best linear unbiased estimator; UPDRS, Unified Parkinson's Disease Rating Scale.

* Corresponding author.
E-mail address: RJCharn2@aol.com (R. Charnigo).

[1] Present address: Intel Corporation, 2501 NW 229th Ave, Hillsboro, OR 97124, United States.

http://dx.doi.org/10.1016/j.jmva.2015.07.005

rates for both $\mu(\mathbf{x})$ and its derivatives, in that the attained convergence rate differs from what is optimal, as identified by Stone [30,31], by only a logarithmic factor.

This problem is of interest because of the numerous practical applications. These range from modeling biological processes such as human growth [26] and disease progression [25] to identifying the structure of nanoparticles from light scattering [4] and the composition of bulk materials from Raman spectroscopy [5]. Further applications of derivative estimation include the analysis of particle velocimetry data, in particular the estimation of vorticity in a turbulent flow [11], and the examination of acoustic Doppler velocimetry data, in particular the replacement of spurious information and detection of spikes [12]. In applications related to human disease, derivative estimation permits greater understanding of the marginal benefits or harms associated with changes in risk factors, with implications for the levels of effort and investment directed toward altering modifiable risk factors and monitoring non-modifiable risk factors. For example, one may wonder at which ages of a patient and previous values of a diagnostic test for Parkinson's disease are deteriorations most worrisome.

Derivative estimation is also important to economic and other time series data analyses. For instance, Park et al. [23] study nonparametric regression for a response variable bounded by a known constant at one end, as may occur for an economic index. Yang et al. [36] examine derivative estimation for generalized additive models, indicating its importance for inferences regarding elasticities and returns to scale. Lu [20] notes a relationship between derivative estimation and chaos, indicating that estimation of Lyapunov exponents is connected to inferences regarding partial derivatives of an autoregression function. There is also literature on estimating derivatives of probability density functions, rather than of mean response functions. We do not survey it except to mention the recent paper by Chacón and Duong [3], which has a substantial bibliography and mentions applications in clustering, image analysis, and object tracking.

Additionally, derivative estimation may be relevant to some problems in statistical inference. For instance, Yin and Li [37] rely on derivative estimation to implement their methodology for "sufficient dimension reduction", which seeks to reduce the number of covariates in a regression model without sacrificing predictive utility [15,10]. More specifically, letting Y and $\mathbf{X}$ denote a response variable and a $p \times 1$ predictor vector respectively, Yin and Li [37] wish to replace $\mathbf{X}$ by $\boldsymbol{\beta}^T\mathbf{X}$, where $\boldsymbol{\beta}$ is a $p \times d_0$ matrix with $d_0 < p$. To this end, they estimate $\partial E(Y|\mathbf{X} = \mathbf{x})/\partial \mathbf{x}$, which equals $\boldsymbol{\beta}\, \partial E(Y|\boldsymbol{\beta}^T\mathbf{X} = \mathbf{u})/\partial \mathbf{u}$ and thus allows them to recover (the span of) $\boldsymbol{\beta}$. For this purpose they use local regression, as their approach to sufficient dimension reduction does not necessitate interchangeability of differentiation and estimation in the sense indicated above. Even so, our simulation studies in Section 5 provide some suggestion that compound estimation may outperform local regression in terms of squared error.

We also consider a generalization of model (1) for scenarios with repeated measures on subjects and/or so many covariates that treating all of their contributions nonparametrically is undesirable. This generalization, a semiparametric model with random effects, has the form

$$Y_{ij} = \sum_{k=1}^{K} \beta_k z_{k,ij} + \mu(\mathbf{x}_{ij}) + \gamma_i + \epsilon_{ij} \quad \text{for } i \in \{1, \dots, n\},\ j \in \{1, \dots, m_i\}. \tag{2}$$

Above, Y_{ij} denotes the jth (of m_i) response(s) for subject i, $\sum_{k=1}^{K} \beta_k z_{k,ij}$ is the parametric fixed effects contribution from covariates in $(z_1, \dots, z_K)^T$, $\mu(\mathbf{x}_{ij})$ is the nonparametric fixed effects contribution from covariates in $\mathbf{x}$, γ_i is the random effect for subject i, and ϵ_{ij} is an error term.

Our goal for model (2) is to simultaneously estimate $\mu(\mathbf{x})$, its derivatives of up to order J, and the coefficients $\beta_1, \dots, \beta_K$, in such a way that differentiation and estimation are interchangeable. In Section 3 we show how this objective can be accomplished using compound estimation, and we identify circumstances where essentially optimal convergence rates are obtained.

Model (2) and variants thereof are of considerable practical utility, but they have typically been studied assuming that the covariate $\mathbf{x}$ is a scalar (e.g., [38]; [27], and references therein; [16]). Moreover, there has been little emphasis on derivative estimation. The present paper is partly motivated by – and will use model (2) with a vector covariate $\mathbf{x}$ and derivative estimation to address – the question of how Parkinson's disease symptoms progress in relation to the vocal characteristics of patients, which constitute a sort of diagnostic test [17,32]. Section 4 provides a fuller description of the publicly available Parkinson's telemonitoring data set, including definitions of the response variable and covariates. For now we mention that $D = 2$ (number of covariates treated nonparametrically) and $K = 7$ (number of covariates treated parametrically), which should suffice for a basic understanding of Fig. 1.

The top panel of Fig. 1 is a close-up near a local minimum of an estimate of $\mu(x_1, x_2)$ obtained from local regression [9,18], an existing method that may be regarded as a benchmark in light of previously established optimality results [30,31]. The middle and bottom panels of Fig. 1 show estimates of $\frac{\partial}{\partial x_1}\mu(x_1, x_2)$ and $\frac{\partial}{\partial x_2}\mu(x_1, x_2)$. The "+" symbol identifies the local minimum, while the "×" symbol identifies the location at which both partial derivative estimates are zero. These two locations being different is not a consequence of numerical error but rather a fundamental problem with local regression, namely that the operations of differentiation and estimation are not interchangeable. Compound estimation overcomes this problem, thereby eliminating the logical inconsistencies among inferences about $\mu(x_1, x_2)$ and how this function varies with respect to x_1 and x_2.

However, compound estimation does more than eliminate logical inconsistencies. While we contend that compound estimation produces more credible results for the Parkinson's telemonitoring data set than does local regression, we further validate compound estimation in Section 5 through simulation studies. More specifically, we demonstrate that compound