C. R. Acad. Sci. Paris, Ser. I ●●● (●●●●) ●●●–●●●

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

Nonparametric trigonometric orthogonal regression estimation

Estimation non paramétrique de la fonction de régression par des séries trigonométriques

Nora Saadi, Smail Adjabi

LAMOS Laboratory, University of Bejaia, Algeria

A R T I C L E   I N F O

Article history:
Received 1 July 2015
Accepted 10 February 2016
Available online xxxx

Presented by the Editorial Board

A B S T R A C T

Eubank, Hart, and Speckman (1990) [2] have investigated the nonparametric trigonometric regression estimator. They assumed that the observation x_i points satisfy $\int_a^{x_i} \psi(s)\,ds = \frac{(1+i)}{n}$, $i = 1, \ldots, n$, where $\psi \in L^1[a,b]$ is a density satisfying certain smoothness conditions, and in a work by E. Rafajłowicz (1987) [3], the observation points coincide with knots of numerical quadratures. The aim of the present work is to introduce a new estimator of the regression function based on trigonometric series, for fixed point designs different from the ones considered so far, under milder restrictions on the observation points. This seems to be important since it may be numerically difficult to determine exactly the points x_i satisfying the recent condition or the knots of appropriate numerical quadratures, especially when their number is large.

R É S U M É

Eubank, Hart et Speckman (1990) [2] ont étudié l'estimation non paramétrique de la fonction de régression par des séries trigonométriques. Ils ont supposé que les observations x_i satisfont la condition $\int_a^{x_i} \psi(s)\,ds = \frac{(1+i)}{n}$, $i = 1, \ldots, n$, où $\psi \in L^1[a,b]$ est une densité vérifiant certaines conditions de régularité. Dans un travail de Rafajłowicz (1987) [3], les observations coincident avec les nœuds des fonctions numériques quadratiques. Ce travail a pour objectif d'introduire un nouvel estimateur de la fonction de régression basé sur un système trigonométrique. On supposera que les observations sont prises en des points équidistants, car il est difficile de déterminer numériquement avec précision les points x_i satisfaisant aux précédentes conditions, spécialement quand le nombre d'observations est grand.

E-mail addresses: saadi.nora@gmail.com (N. Saadi), adjabi@hotmail.com (S. Adjabi).

http://dx.doi.org/10.1016/j.crma.2016.02.013
1631-073X/© 2016 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

The field of nonparametric statistics has broadened its appeal in recent years with an array of new tools for statistical analysis. These new tools offer sophisticated alternatives to traditional parametric models for exploring large amounts of univariate or multivariate data without making specific distributional assumption. As one of these tools, nonparametric regression function estimation has become a prominent statistical research topic. In many areas of mathematical analysis, the smoothness of a function is more readily determined by the behaviour of its Fourier series. Our thesis in this paper is that the latter approach is natural and convenient when analyzing the properties of orthogonal trigonometric regression function estimators. Consider the partition of the interval $A = [a, b] \subset \mathbb{R}$ into n subintervals $A_1 = [a_0, a_1]$, $A_i = (a_{i-1}, a_i]$, $i = 1, \ldots, n$, where $a = a_0 < a_1 < \ldots < a_n$. Suppose that the observations $y_1, \ldots, y_n$ follow the model $y_i = R(x_i) + \eta_i$, where $R(x) : [a, b] \longrightarrow \mathbb{R}$ is an unknown function satisfying certain smoothness conditions specified below, $x_i \in A_i$, $i = 1, \ldots, n$, and η_i, $i = 1, \ldots, n$, are independent identically distributed random variables with mean zero and finite variance $\sigma^2 > 0$. Let the functions e_k, $k = 0, \ldots$, form an orthogonal system in the space $L^2[a, b]$ i.e.:

$$\int_a^b e_k(x) e_j(x)\,\mathrm{d}x = \delta_{kj}, 0 \le k, j < \infty. \tag{1}$$

Where δ_{kj} is the Kroncker delta. We assume that the regression function R is an element of this space and consequently it has a representation

$$R(x) = \sum_{k=0}^{\infty} c_k e_k(x), k = 0, \ldots, x \in [a, b]. \tag{2}$$

As an estimator of R, we take

$$\hat{R}_{d_n}(x) = \sum_{k=0}^{d_n} \hat{c}_k e_k(x), \tag{3}$$

where

$$\hat{c}_k = \sum_{i=1}^{n} y_i \int_{a_{i-1}}^{a_i} e_k(x)\,\mathrm{d}x, \tag{4}$$

and d_n it is a sequence of positive number chosen so that $d_n \longrightarrow \infty$ as $n \longrightarrow \infty$.

To cite a few specific works, Rutkowski [5] investigated sufficient conditions for almost sure convergence of the above estimator, constructed using the trigonometric functions and Legendre polynomials. Results concerning the convergence rates of the integrated mean-square error of the estimator constructed using the Fourier series were obtained in [2]. However, it is assumed that the observation points satisfy the condition $\int_a^{x_i} \psi(s)\,\mathrm{d}s = \frac{(1+i)}{n}$, $i = 1, \ldots, n$, where $\psi \in L^1[a, b]$ is a density satisfying certain smoothness conditions, and in [3] the observation points coincide with knots of numerical quadratures. The aim of the present work is to introduce a new estimator of the regression function based on trigonometric basis introduced by Saadi and Adjabi [6]. We obtain asymptotic results, in particular convergence rates for IMSE and the pointwise mean-square error of the estimator for fixed point designs different from the ones considered so far, under milder restrictions on the observation points. This seems to be important, since it may be numerically difficult to determine exactly the points x_i satisfying the recent condition or the knots of appropriate numerical quadratures, especially when their number is large.

2. Construction of the estimator

Consider the partition of the interval $A = [-\pi, \pi]$ into n subintervals $A_1 = [a_0, a_1]$, $A_i = (a_{i-1}, a_i]$, $i = 1, \ldots, n$, where $-\pi = a_0 < a_1 < \ldots < a_n = \pi$. Suppose that the observations $y_1, \ldots, y_n$ follow the model $y_i = R(x_i) + \eta_i$, where $R(x) : [-\pi, \pi] \longrightarrow \mathbb{R}$ is an unknown function satisfying certain smoothness conditions specified below, $x_i \in A_i$, $i = 1, \ldots, n$, and η_i, $i = 1, \ldots, n$, are independent identically distributed random variables with mean zero and finite variance $\sigma^2 > 0$. Let the functions $\{e_k, k = 0, \ldots\}$ form an orthogonal system in the space $L^2[-\pi, \pi]$ defined by Saadi and Adjabi [6]:

$$e_k(x) = \frac{1}{\sqrt{2\pi}} (\cos(kx) + \sin(kx)) 1_{[-\pi,\pi]}(x),\ k = 0, \ldots. \tag{5}$$

We assume that $R(x) \in L^2[-\pi, \pi]$ and consequently it has a representation

$$R(x) = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} c_k (\cos(kx) + \sin(kx)),\ \text{where } c_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} (\cos(kx) + \sin(kx)) R(x)\,\mathrm{d}x. \tag{6}$$

Please cite this article in press as: N. Saadi, S. Adjabi, Nonparametric trigonometric orthogonal regression estimation, C. R. Acad. Sci. Paris, Ser. I (2016), http://dx.doi.org/10.1016/j.crma.2016.02.013