Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v7n2p315

Kernel estimation of the regression mode for fixed design model

By Salha R.

Published: 14 October 2014

Electronic Journal of Applied Statistical Analysis
Vol. 07, Issue 02, 2014, 315-325
DOI: 10.1285/i20705948v7n2p315

Kernel estimation of the regression mode for fixed design model

Raid B. Salha*

The Islamic university, Department of mathematics, Gaza, Palestine

Published: 14 October 2014

In this paper, we study the problem of estimating nonparametrically the regression mode for fixed design model. We suppose the error random variables are independent. The joint asymptotic normality of the regression mode estimator at different fixed design points is established under some regularity conditions. The performance of the proposed estimator is tested via two applications using a simulation and real life data.

keywords: Fixed design model, kernel estimation, regression mode, asymptotic normality.

1 Introduction

Parzen (1962) considered the problem of estimating the mode of a univariate pdf. Parzen (1962) and Nadaraya (1965) have shown that under some regularity conditions the estimator of the population mode obtained by maximizing a kernel estimator of the pdf is strongly consistent and asymptotically normally distributed. Samanta (1973) has given multivariate versions of Parzen's results. For independent and identically distributed data, Samanta and Thavanesmaran (1990) considered the problem of estimating the mode of a conditional pdf, for random design model, and they have shown under regularity conditions that the estimator of the population conditional mode is strongly consistent and asymptotically normally distributed. Salha and Ioanides (2004) generalized their work for the multivariate case. Salha and Ioanides (2007) considered the estimation of the conditional mode under dependence conditions.

We assume that $(x_1, Y_1), (x_2, Y_2), \dots (x_n, Y_n)$ are i.i.d. random variables with conditional pdf $f(y|x)$ where the variable Y depending upon a fixed design predictor x

*Corresponding author: rbsalha@iugaza.edu.ps

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

through a regression function $m(x)$. In this paper, we consider the regression model $Y_i = m(x_i) + \epsilon_i$. for $i = 0, 1, 2, \ldots, n$ and $m(x)$ is an unknown regression function. The design points $x_1, x_2, \ldots x_n$ determined by the experimenter, are ordered, i.e. we have $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1$. In the absence of other information, we can take $x_i = \frac{i}{n}, i = 1, 2, \ldots, n$. ϵ_i are independent random variables with mean 0 and variance σ^2.

Most investigations are concerned with regression mean function $m(x)$, where $m(x) = \int y f(y|x) dy$. However, new insights about the underling structures can be gained by considering other aspects of the conditional distribution function $f(y|x)$ of Y at a given value $x \in [0, 1]$ such as the regression quantile and the regression mode. In this paper, we consider the problem of estimating the mode of the unknown function $f(y|x)$.

We consider $f_n(y|x)$ as an estimator of $f(y|x)$, and it is given by

$$f_n(y|x) = h_n^{-1} \sum_{i=1}^{n} u_{ni}(x) L(\frac{y - Y_i}{h_n}),$$

where, K and L are two kernel functions satisfying some condition and h_n is a sequence of positive numbers converging to zero and satisfies some specific conditions. $u_{ni}(x) = h_n^{-1} \int_{s_{i-1}}^{s_i} K(\frac{x-u}{h_n}) du$ is the Gasser-Mueller function. $s_1, s_2, \ldots s_n$ is a sequence of interpolating points such that $x_{i-1} \leq s_i \leq x_i$, for $i = 1, 2, \ldots, n$. In general, we take $s_i = \dfrac{x_{i-1} + x_i}{2}$.

Consequentially, there is a random variable $M_n(x)$ which is called the sample regression mode, such that

$$f_n(M_n(x)|x) = \max_{-\infty < y < \infty} f_n(y|x).$$

In this paper, for distinct points $x_1, x_2, \ldots, x_k$ we will establish conditions under which $(nh_n^4)^{\frac{1}{4}} (M_n(x_1), M_n(x_2), \ldots, M_n(x_k))^T$, where T denotes the transpose, is asymptotically multivariate normally distributed random variable.

This paper is organised as follow, in Section 2, we state the conditions under which the results of this paper are established. In Section 3, we mention some preliminaries lemmas and notations that will help us in the remaining of this paper. The main results of this paper are stated and proved in Section 4. Finally, Section 5 contains two applications to test the performance of the proposed estimator.

2 Conditions

We consider the fixed equally spaced design regression model

$$Y_i = m(x_i) + \epsilon_i, i = 1, 2, \ldots, n,$$

where

$$x_i = \frac{i}{n}, i = 1, 2, \ldots, n.$$

We assume the following conditions are satisfied.

Condition 1

1. $m(x)$ is an bounded function on $[0, 1]$.
2. ϵ_i are independent random variables with mean 0 and variance σ^2 .

Condition 2
The partial derivatives

$$f^{(j)}(x, y) = \frac{\partial^j f(y|x)}{\partial y^j}, j = 1, 2$$

exist and are bounded.

Condition 3

1. The kernel function $K(.)$ has support $[-1, 1]$.
2. $\int_{-1}^{1} K(u)du = 1$ and $\int_{-1}^{1} uK(u)du = 0$.
3. The kernel function $L(.)$ is asymmetric pdf.

Condition 4
The bandwidth are chosen such that $h_n = cn^{-\delta}, c > 0, 0 < \delta < 1$ such that

$$\lim_{n \longrightarrow \infty} nh_n^8 = \infty, \quad \lim_{n \longrightarrow \infty} nh_n^{10} = 0.$$

3 Preliminaries Lemmas

The proof of the following four lemmas can be obtained by a modification of the proof of the lemmas in Samanta and Thavanesmaran (1990).

Lemma 1 (Bochner Lemma) Suppose $K_1(u)$ and $K_2(u)$ are real valued Borel measurable functions satisfying the following conditions:

1. $\sup_{u \in R} |K_i(u)| < \infty, \ i = 1, 2.$
2. $\int_{-\infty}^{\infty} |K_i(u)| \, du < \infty, \ i = 1, 2.$
3. $\lim_{|u| \to \infty} u^2 \, |K_i(u)| = 0, \ i = 1, 2.$

If $f(x, y) \in C(f)$, the set of continuity points of f, then for any $\eta \geq 0$,
$\lim_{n \to \infty} \left[h_n^{-2} \int_{-\infty}^{\infty} |K_1(\frac{u}{h_n}) K_2(\frac{v}{h_n})|^{1+\eta} f(y - v|x - u) dv \, du \right] = f(y|x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |K_1(u)K_2(v)|^{1+\eta} dv du.$

Define,

$$f_n^{(j)}(y|x) = h_n^{-(j+1)} \sum_{i=1}^{n} u_{ni}(x) L^{(j)}(\frac{y-Y_i}{h_n}), \; j = 1, 2,$$

$$V_{ni}(y|x) = h_n^{-2} u_{ni}(x) L^{(1)}(\frac{y-Y_i}{h_n}), \; i = 1, 2, \ldots, n.$$

Lemma 2 Under the conditions (1) through (4), if $f(x,y) \in C(f)$, we have

1. $\lim_{n\to\infty} nh_n^2 \left[Var\left(f_n^{(1)}(y|x)\right)\right] = f(y|x) \int_{-\infty}^{\infty} \int_{-1}^{1} \left(K(u)L^{(1)}(v)\right)^2 dudv.$
2. $(nh_n^2)^{\frac{1}{2}} \left[Ef_n^{(1)}(y|x) - f^{(1)}(y|x)\right] = o(1).$

Lemma 3 Under the conditions (1) through (4), we have

$$\lim_{n\to\infty} (n^{-1}h_n^2)^{1+\frac{\delta}{2}} \left[\sum_{i=1}^{n} E|V_{ni} - EV_{ni}|^{2+\delta}\right] = 0.$$

Lemma 4 Under the conditions (1) through (4), if $g(x) > 0$, then $f_n^{(2)}(M_n^*(x)|x)$ converges in probability to $f^{(2)}(M(x)|x)$ as n tends to infinity, where $|M_n^*(x) - M(x)| < |M_n(x) - M(x)|$.

4 Main Results

The main results of this paper are stated and proved in this section.

Theorem 1 Suppose that $x_1, x_2, \ldots, x_k$ are distinct points, where $f(y|x_i) > 0$, $i = 1, 2, \ldots, k$, then under the conditions (1) through (4),

$$(nh_n^2)^{\frac{1}{2}} \left(f_n^{(1)}(y|x_1), f_n^{(1)}(y|x_2), \ldots, f_n^{(1)}(y|x_k)\right)^T$$

is asymptotically multivariate normal with mean vector zero and diagonal covariance matrix $\Gamma = [\gamma_{ij}]$, with

$$\gamma_{ii} = f(y|x_i) \int_{-\infty}^{\infty} \int_{-1}^{1} \left(K(u)L^{(1)}(v)\right)^2 dudv.$$

Without loss of generality, we consider the special case $k = 2$. The method of the proof remains valid for more general case. Firstly, we define for $(i = 1, 2, \ldots, n)$ and $(s = 1, 2)$ the following notations:

$$W_{ni}(x_s) = h_n(V_{ni}(y|x_s) - EV_{ni}(y|x_s)), \; W_n(x_s) = \sum_{i=1}^{n} W_{ni}(x_s),$$

$$\mathbf{Z}_{ni} = (W_{ni}(x_1), W_{ni}(x_2))^T, \; \mathbf{Z}_n = n^{-\frac{1}{2}} (W_n(x_1), W_n(x_2))^T.$$

$$\mathbf{Z}_n = (nh^4)^{\frac{1}{2}} \left(f_n^{(1)}(y|x_1) - Ef_n^{(1)}(y|x_1), f_n^{(1)}(y|x_2) - Ef^{(1)}(y|x_2) \right)^T \quad (1)$$

Let $A = [a_{rs}]$ be a 2×2 diagonal matrix with

$$a_{ss} = f(y|x_s) \int_{-\infty}^{\infty} \int_{-1}^{1} \left(K(u)L^{(1)}(v) \right)^2 du\, dv.$$

Let $\mathbf{Z}$ be a bivariate normally distributed random variable with mean zero vector and covariance matrix A.

First we will show that $\mathbf{Z}_n$ converges in distribution to $\mathbf{Z}$. To do that, we will use the multivariate version of Cramér - World theorem, see Pranab and Julio (1993). It will be sufficient to prove that $\mathbf{C}\mathbf{Z}_n^T$ converges in distribution to $\mathbf{C}\mathbf{Z}^T$ for any constant $\mathbf{C} = (c_1, c_2) \in \mathbf{R}^2$, $\mathbf{C} \neq \mathbf{0}$.

Note that, $\mathbf{C}\mathbf{Z}_n^T = \sum_{i=1}^{n} n^{-\frac{1}{2}} \mathbf{C}\mathbf{Z}_{ni}^T$, $E(n^{-\frac{1}{2}}\mathbf{C}\mathbf{Z}_{ni}^T) = 0$.

Let $\rho_{ni}^{2+\delta} = E|n^{-\frac{1}{2}}\mathbf{C}\mathbf{Z}_{ni}|^{2+\delta}$, $\rho_n^{2+\delta} = \sum_{i=1}^{n} \rho_{ni}^{2+\delta}$, and $\sigma_n^2 = Var(\mathbf{C}\mathbf{Z}_n^T)$.

Using Liapounov's condition, it will be sufficient to show that,

$$\lim_{n\to\infty} \frac{\rho_n^{2+\delta}}{\sigma_n^2} = 0. \quad (2)$$

Now, the proof of the theorem will be satisfied via the following two lemmas.

Lemma 5 Under the conditions (1) through (4), if $f(x, y) \in C(f)$, then for $s = 1, 2,\ r = 1, 2$, the following are true:

a. $\lim_{n\to} EW_{ni}^2(x_s) = f(y|x_s) \int_{-\infty}^{\infty} \int_{-1}^{1} \left(K(u)L^{(1)}(v) \right)^2 du\, dv.$

b. $\lim_{n\to\infty} EW_{ni}(x_s)EW_{ni}(x_r) = 0,\ (r \neq s).$

Proof:

a. From the definition of $W_{ni}(x_s)$, we have

$$EW_{ni}^2(x_s) = h_n^2 \left[EV_{ni}^2(y|x_s) - (EV_{ni}(y|x_s))^2 \right]. \quad (3)$$

$$\begin{aligned} h_n^2 EV_{ni}^2(y|x_s) &= h_n^2 \left[h_n^{-4} \int_{-\infty}^{\infty} \int_{-1}^{1} \left(K(\frac{x_s - u}{h_n}\,) L^{(1)}(\frac{y - v}{h_n}) \right)^2 f(v|u)\, du\, dv \right] \\ &= h_n^{-2} \left[\int_{-\infty}^{\infty} \int_{-1} \left(K(\frac{u}{h_n})\, L^{(1)}(\frac{v}{h_n}) \right)^2 f(y - v|x_s - u)\, du\, dv \right]. \end{aligned}$$

Now, by an application of Lemma 1, we get that

$$\lim_{n\to\infty} h_n^2 EV_{ni}^2(y|x_s) = f(y|x_s) \int_{-\infty}^{\infty} \int_{-1}^{1} \left(K(u)\, L^{(1)}(v) \right)^2 du\, dv. \tag{4}$$

$$h_n^2 (EV_{ni}(y|x_s))^2 = h_n^2 \left[h_n^{-2} \int_{-\infty}^{\infty} \int_{-1}^{1} K(\frac{u}{h_n})\, L^{(1)}(\frac{v}{h_n})\, f(y-v|x_s-u)\, du\, dv \right].$$

By another application of Lemma 1, we get that

$$\lim_{n\to\infty} h_n^2 (EV_{ni}(y|x_s))^2 = 0. \tag{5}$$

Now, a combination of (3), (4) and (5), (a) is satisfied.

b. From the definition of $W_{ni}(x)$, we have

$$EW_{ni}(x_1)W_{ni}(x_2) = h_n^2 \left(EV_{ni}(y|x_1)V_{ni}(y|x_2) - EV_{ni}(y|x_1)EV_{ni}(y|x_2) \right). \tag{6}$$

Suppose that without loss of generality, $x_2 > x_1$, let $\delta = x_2 - x_1$ and $\delta_n = \frac{\delta}{h_n}$.

$$\begin{aligned} h_n^2 EV_{ni}(y|x_1)V_{ni}(y|x_2) &= h_n^{-2} \int_{-\infty}^{\infty} \int_{-1}^{1} K(\frac{x_1-u}{h_n}) K(\frac{x_2-u}{h_n}) \left(L^{(1)}(\frac{y-v}{h_n}) \right)^2 f(v|u)\, du\, dv \\ &= \int_{-\infty}^{\infty} \int_{-1}^{1} K(u)K(\delta_n+u) \left(L^{(1)}(v) \right)^2 f(y-h_n v|x_1-h_n u)\, du\, dv \\ &= \int_{-\infty}^{\infty} \left(L^{(1)}(v) \right) f(y-h_n v|x_1-h_n u)\, dv \left[\int_{-1}^{1} K(u)K(\delta_n+u) \right] du. \end{aligned} \tag{7}$$

Next, note that

$$\begin{aligned} \int_{-1}^{1} K(u)K(\delta_n+u)\, du &= \int_{|u|<\frac{\delta_n}{2}} K(u)K(\delta_n+u)\, du + \int_{|u|\geq\frac{\delta_n}{2}} K(u)K(\delta_n+u)\, du \\ &\leq \sup_{|u|<\frac{\delta_n}{2}} K(\delta_n+u) \int_{-1}^{1} K(z)\, dz + \sup_{|u|\geq\frac{\delta_n}{2}} K(u) \int_{-1}^{1} K(\delta_n+z)\, dz \\ &\leq 2 \sup_{|u|\geq\frac{\delta_n}{2}} K(u).O(1) \leq \frac{4}{\delta_n} \sup_{|u|\geq\frac{\delta_n}{2}} |uK(u)|.O(1) \\ &= \frac{4h_n}{\delta} \sup_{|u|\geq\frac{\delta_n}{2}} |uK(u)|.O(1) = O(h_n). \end{aligned} \tag{8}$$

Finally, from (7) and (8), we have that

$$\lim_{n\to} h_n^2 \left(EV_{ni}(y|x_1)V_{ni}(y|x_2)\right) = 0. \tag{9}$$

$$\begin{aligned} h_n^2 EV_{ni}(y|x_1)EV_{ni}(y|x_2) &= h_n^2 \left[h_n^{-2} \int_{-\infty}^{\infty}\int_{-1}^{1} K(\frac{u}{h_n}) \left(L^{(1)}(\frac{v}{h_n}) \right) f(y-v|x_1-u)\,du\,dv \right] \\ &\times \left[h_n^{-2} \int_{-\infty}^{\infty}\int_{-1}^{1} K(\frac{u}{h_n}) \left(L^{(1)}(\frac{v}{h_n}) \right) f(y-v|x_1-u)\,du\,dv \right] \\ &\to 0. \end{aligned} \tag{10}$$

By an application of Lemma 1. Hence a combination of (6), (9), and (10), gives the desired result.

Lemma 6 Under the conditions of Lemma 5, we have that $\lim\limits_{n\to\infty} \sigma_n^2 = \mathbf{C}A\mathbf{C}^T$.

Proof: $\sigma_n^2 = Var(\mathbf{C}\mathbf{Z}_n^T)$ and by the definition of $\mathbf{Z}_n^T$, we have

$$\begin{aligned} \sigma_n^2 &= Var(n^{-\frac{1}{2}}c_1W_n(x_1) + n^{-\frac{1}{2}}c_2W_n(x_2)) \\ &= n^{-1}c_1^2Var(W_n(x_1)) + n^{-1}c_2^2Var(W_n(x_2)) + 2n^{-1}c_1c_2Cov(W_n(x_1), W_n(x_2)) \\ &= n^{-1}c_1^2\sum_{i=1}^{n} Var(W_{ni}(x_1)) + n^{-1}c_2^2\sum_{i=1} Var(W_{ni}(x_2)) \\ &+ 2n^{-1}c_1c_2Cov\left(\sum_{i=1}^{n} W_{ni}(x_1), \sum_{i=1}^{n} W_{ni}(x_2))\right) \\ &= c_1^2Var(W_{ni}(x_1)) + c_2^2Var(W_{ni}(x_2)) + 2n^{-1}c_1c_2E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ni}(x_1)W_{nj}(x_2)\right). \end{aligned}$$

An application of Lemma 5 implies that,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty}\int_{-1}^{1} \left(K(u)L^{(1)}(v)\right)\,du\,dv \left[c_1^2f(y|x_1) + [c_2^2f(y|x_2)\right] = \mathbf{C}A\mathbf{C}^T > 0.$$

Next, we have

$$\begin{aligned} \rho_{ni}^{2+\delta} &\leq n^{-(1+\frac{\delta}{2})}|\mathbf{C}|^{2+\delta}E|\mathbf{Z}_{ni}|^{2+\delta} = n^{-(1+\frac{\delta}{2})}|\mathbf{C}|^{2+\delta}E|(W_{ni}(x_1), W_{ni}(x_2))|^{2+\delta} \\ &\leq n^{-(1+\frac{\delta}{2})}|\mathbf{C}|^{2+\delta}2^{2+\delta}\max\{E|W_{ni}(x_1)|^{2+\delta}, E|W_{ni}(x_2)|^{2+\delta}\}. \end{aligned}$$

Without loss of generality assume that $E|W_{ni}(x_1)|^{2+\delta} \geq E|W_{ni}(x_2)|^{2+\delta}$.

$$\begin{aligned}
\rho_{ni}^{2+\delta} &\leq n^{-(1+\frac{\delta}{2})}|\mathbf{C}|^{2+\delta}2^{2+\delta}E|W_{ni}(x_1)|^{2+\delta} \\
&= n^{-(1+\frac{\delta}{2})}|\mathbf{C}|^{2+\delta}2^{2+\delta}E|(h_n^2(V_{ni}(x_1) - EV_{ni}(x_1))|^{2+\delta} \\
&= 2^{2+\delta}|\mathbf{C}|^{2+\delta}(n^{-1})^{(1+\frac{\delta}{2})}(h_n^2)^{2+\delta}E|V_{ni}(x_1) - EV_{ni}(x_1)|^{2+\delta} \\
&= 2^{2+\delta}|\mathbf{C}|^{2+\delta}(n^{-1}h_n^4)^{(1+\frac{\delta}{2})}E|V_{ni}(x_1) - EV_{ni}(x_1)|^{2+\delta}
\end{aligned}$$

Now, by an application of Lemma 3, we get that

$$\begin{aligned}
\rho_n^{2+\delta} &= n^{-(1+\frac{\delta}{2})}\sum_{i=1}^{n}\rho_{ni}^{2+\delta} \\
&\leq 2^{2+\delta}|\mathbf{C}|^{2+\delta}(n^{-1}h_n^4)^{(1+\frac{\delta}{2})}\sum_{i=1}^{n}E|V_{ni}(x_1)) - EV_{ni}(x_1)|^{2+\delta} \to 0.
\end{aligned}$$

Hence the Liapounov's condition (2) is satisfied. Therefore, we have that, $\mathbf{C}\mathbf{Z}_n^T$ is asymptotically normally distributed with mean zero and variance $\mathbf{C}A\mathbf{C}^T$ and by the multivariate version of Cramr - World theorem, we have that $\mathbf{Z}_n$ converges in distribution to $\mathbf{Z}$. Now an application of the second part of Lemma 2 in conjunction with the last convergence gives the proof of the theorem.

Theorem 2 Supposet $x_1, x_2, \ldots, x_k$ are distinct points, where $f(y|x_i) > 0$, $i = 1, 2, \ldots, k$, then under the conditions (1) through (4), the random vector variable

$$(nh_n^2)^{\frac{1}{2}}\,(M_n(x_1) - (M(x_1), \ldots, M_n(x_k) - (M(x_k))^T$$

is asymptotically multivariate normally distributed with mean vector zero and a diagonal covariance matrix $B = [b_{ij}]$, with

$$b_{ii} = \frac{f(M(x_i)|x_i)}{(f^{(2)}(M(x_i)|x_i))^2}\int_{-\infty}^{\infty}\int_{-1}^{1}\Big(K(u)L^{(1)}(v)\Big)^2\,dudv.$$

Proof: For a fixed x, taking the Taylor expansion of $f_n^{(1)}(M_n(x)|x)$ around $M(x)$, implies

$$0 = f_n^{(1)}(M_n(x)|x) \approx f_n^{(1)}(M(x)|x) + (M_n(x) - M(x))f_n^{(2)}(M_n^*(x)|x),$$

where

$$|M_n^*(x) - M(x)| < |M_n(x) - M(x)|.$$

This implies that,

$$M_n(x) - M(x) \approx -\frac{f_n^{(1)}(M(x)|x)}{f_n^{(2)}(M_n^*(x)|x)},$$

and therefore

$$(nh_n^4)^{\frac{1}{2}}(M_n(x_1) - M(x_1), \ldots, M_n(x_k) - M(x_k))^T \approx$$

$$-(nh_n^4)^{\frac{1}{2}} \left[\frac{f_n^{(1)}(M(x_1)|x_1)}{f_n^{(2)}(M_n^*(x_1)|x_1)}, \ldots, \frac{f_n^{(1)}(M(x_k)|x_k)}{f_n^{(2)}(M_n^*(x_k)|x_k)} \right],$$

where

$$|M_n^*(x_i) - M(x_i)| < |M_n(x_i) - M(x_i)|,\ i = 1, 2, \ldots, k.$$

An application of Theorem 1 and Lemma 4 completes the proof of the theorem.

5 Applications

In this section, we test the performance of the proposed estimator by considering two applications. The first one deals with a simulation data, while the other one deals with a real life data. The results of the two applications indicate that the proposed estimator of the regression mode is reasonably good.

5.1 Simulation study

In this subsection, we consider an application using a simulation data. A sample of size 400 from the model $y = sin 2\pi(1-x^2)+xe$ where x has standard normal distribution and e has a uniform $[0, 1]$ distribution. A perfect smooth would recapture the original single $y = sin 2\pi(1 - x^2)$ exactly. We used the following two kernel functions, the rectangle kernel

$$K(x) = \frac{1}{2}, |x| < 1$$

and the Gaussian kernel

$$L(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty.$$

For a direct comparison of the perfect smooth and the regression mode estimation, a scatter plot of the original data, the perfect smooth and the estimated regression mode curve are plotted in Figure 1. The performance of the estimator can be tested using $R^2_{y,\hat{y}}$ (the correlation coefficients between y the predicated values and $\hat{y}$ the actual values).

$$R^2_{y,\hat{y}} = 1 - \frac{SSE}{SSTO} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{28.4857}{153.1601} = 0.8140,$$

where $\bar{y}$ denotes the mean of actual values. Also the mean squared error,

$$MSE = \frac{SSE}{n} = 0.0712.$$

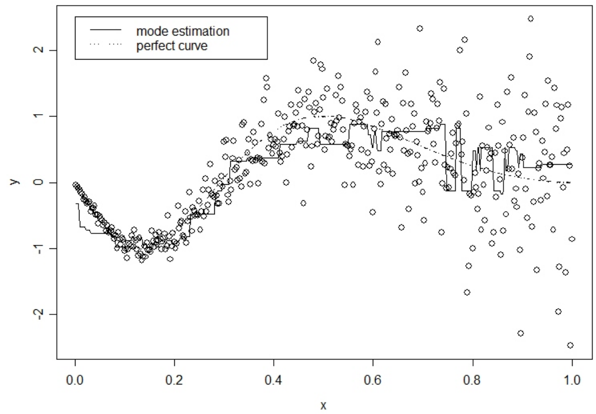

Figure 1: Regression mode estimation

5.2 Real life data study

In this subsection, we consider an application using a real life data which is built in S-Plus program. We consider the ethanol data which records 88 measurements from an experiment in which ethanol was burned in a single cylinder auto-mobile test engine. We used the first 71 observation from the variable E, which indicates the measure of the richness of the air/ethanol mixture to estimate the last 17 observations. The mean squared error is computed, $MSE = 0.0436$. Figure 2 shows the plot of the data and the regression mode estimation.

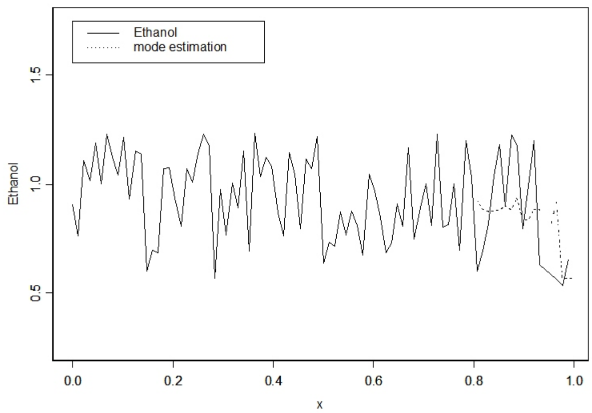

Figure 2: Regression mode estimation of the ethanol data

6 Conclusion

In this paper, the problem of estimating the regression mode for fixed design model has been considered. The joint asymptotic normality of the regression mode estimator at different fixed design points has been established under some regularity conditions. The performance of the proposed estimator is tested via two applications using a simulation and real life data. The results of the two applications indicate that the performance of the estimator is reasonably good. The results of the applications can be slightly improved if the rectangle kernel is replaced by the Epanchinkov kernel which is the optimal kernel. We used the rectangle kernel for computational reasons. The new estimator can be modified by considering a new bandwidth selection technique that uses a variable bandwidth that depends on the points at which the mode function is estimated rather than a constant bandwidth.

By considering some conditions, the results of this paper can be generalized to the case of dependent data under some mixing conditions and for the case of time series data.

References

Nadaraya, E. A. (1965). On nonparametric estimation of density functions and regression curves. *Theory of Probability and their Applications*, 10:186–190.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076.

Pranab, K. and Julio, M. (1993). *Large sample methods in statistics*. Champion and Hall.

Salha, R. and Ioanides, D. (2004). Joint asymptotic distribution of the estimated conditional mode at a finite number of distinct points. In *Proceedings of the 17th international conference of statistics*, pages 587–594. The Greek statistican association.

Salha, R. and Ioanides, D. (2007). The asymptotic distribution of the estimated conditional mode at a finite number of distinct points under dependence conditions. *The Islamic University Journal*, 15(2):199–214.

Samanta, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statistical Journal*, 7:109–117.

Samanta, M. and Thavanesmaran, A. (1990). Nonparametric estimation of the conditional mode. *Communications in Statistics Theory and Methods*, 19:4515–4524.