IUG Journal of Natural and Engineering Studies
Peer-reviewed Journal of Islamic University-Gaza

Vol 23, No 2, 2015, pp 12-20, ISSN 2409-4587
IUGNES

Estimating the Conditional Mode Using the Symmetrized Nearest Neighbor Kernel Estimator

Raid B. Salha[1], Bisher M. Iqelan[1,*]

[1]Department of Mathematics, Faculty of Science, Islamic University of Gaza, Gaza Strip, Palestine

Received on (26-11-2014) Accepted on (17-2-2015)

Abstract

In this paper, the kernel estimation of the mode of a conditional probability density function is studied. We propose the Symmetrized Nearest Neighbor (SNN) kernel estimator to estimate the conditional mode. We study the asymptotic properties of the proposed estimator. Also, we derive its asymptotic normality under some conditions much weaker than that needed for the Nadaraya-Watson (NW) kernel estimator. The performance of the SNN kernel estimator is tested using three simulated and real data which indicate that the proposed estimator is reasonably good. In addition, a comparison between the proposed estimator and the NW estimator is given.

Keywords Kernel Estimation, Conditional Mode, Nearest Neighbor Estimator, Nadaraya-Watson Estimator, Asymptotic Properties.

تقدير المنوال المشروط باستخدام مقدر النواة المتماثل للجوار الأقرب

ملخص

في هذا البحث ندرس تقدير النواة لمنوال دالة الكثافة الاحتمالية المشروطة. واقترحنا استخدام مقدر النواة المتماثل للجوار الأقرب لتقدير المنوال المشروط. ودرسنا خصائص المقدر وكذلك اثبتنا أن توزيعه الاحتمالي يتقارب إلى التوزيع الطبيعي تحت شروط أضعف من تلك التي نحتاجها في حالة مقدر ندارايا– واتسن. ولقد تم اختبار أداء المقدر المقترح باستخدام ثلاث دراسات محاكاة وبيانات حقيقية و لقد أشارت هذه الدراسات أن أدائه جيد. بالإضافة لذلك تم مقارنته بمقدر ندارايا – واتسن.

كلمات مفتاحية: تقدير النواة، المنوال المشروط، تقدير الجوار الأقرب، تقدير ندارايا – واتسن، الخواص التقاربية.

* Corresponding author e-mail address: rbsalha@iugaza.edu.ps

1. Introduction:

Nonparametric estimation of the population mode obtained by maximizing a kernel estimator of the probability density function (pdf) is a matter of both theoretical and practical interest in statistics. [5] considered the problem of estimating the mode of a univariate pdf. [5] and [4] have shown that under some regularity conditions, the estimator of the mode is strongly consistent and asymptotically normally distributed. [11] has given multivariate versions of Parzen's results. [10] considered the problem of estimating the mode of a conditional pdf and they have shown under some regularity conditions, that the estimator of the population conditional mode is strongly consistent and asymptotically normally distributed. [9] generalized these results by considering the conditional mode evaluated at distinct conditional points. [12] presented and compared four mode estimation procedures. Recently, for random design models, [16] proposed a kernel estimator of the mode and its asymptotic normality has been shown in [15]. In addition, [14] presented an adaptive kernel estimator for the mode.

[7, 8] considered two modified kernel estimators of the conditional mode based on the Rewighted Nadaraya-Watson estimator and the local variable kernel estimator of the conditional density function.

[13] used the SNN kernel estimator to estimate the conditional mean. Recently, new articles studied the estimation of the conditional mode. [3] established the asymptotic normality of a new kernel estimator of the conditional mode for left-truncated and dependent data exhibit some kind of dependence. [2] proposed a semi-parametric mode regression estimator for the case in which the dependent variable has a continuous conditional density with a well-defined global mode. They showed that the proposed estimator is consistent and has a tractable asymptotic distribution. [1] studied the nonparametric estimation of the conditional density and mode under α-mixing dependence, based on the single-index structure.

In this paper, we propose a new kernel estimator of the conditional mode based on the SNN kernel estimator of the conditional probability density function. The asymptotic consistency and normality of the proposed estimator has been obtained under some conditions much weaker than that needed for the NW estimator. The performance of the SNN kernel estimator is tested using three simulated and real data, which indicate that the performance of the proposed estimator is reasonably good.

This paper is organized as follows. In Section 1, we present some preliminaries and the definition of the proposed estimator. In Section 2, we study the asymptotic properties of the proposed estimator. In this section, we prove two theorems that present the main theoretical results of this paper. The performance of the SNN estimator is tested by applications to simulated and real data in Section 3. Finally, Section 4 contains some conclusion remarks.

1. Preliminaries:

Assume that $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$ are independent and identically distributed two dimensional random variables with a joint density function $f(x,y)$. The conditional density function of Y given $X=x$ is

$$f(y|x)=\frac{f(x,y)}{g(x)},$$

where $g(x)$ is the marginal density function of X, and defined by

$$g(x)=\int_{-\infty}^{\infty} f(x,y)\,dy.$$

Assume that $f\left(y\,|x\right)$ is uniformly continuous in y, for each x. It follows that $f\left(y\,|x\right)$ possesses a mode $q(x)$, which defined by

$$q(x)=\max_{y\in\mathbf{R}} f\left(y\,|x\right).$$

Definition 1 (Empirical Distribution Function)
Given an observed random sample $x_1, x_2, ..., x_n$, an empirical distribution function $\hat{G}\left(x\right)$ is the fraction of sample observations less than or equal to the value x. More specifically, if $y_1 < y_2 < ... < y_n$ are the order statistics of the observed random sample, with no two observations being equal, then the empirical distribution function is defined as:

$$\hat{G}\left(x\right)=\begin{cases} 0, & x < y_1. \\ \frac{k}{n}, & y_k \le x < y_{k+1}, k = 1,2,\mathbf{K}, n-1. \\ 1, & x \ge y_n. \end{cases}$$

Definition 2 (SNN Estimator of the Conditional Mode)
The SNN estimator of the conditional mode $q(x)$ is given by

$$\hat{q}(x)=\max_{y\in\mathbf{R}} f_n\left(y\,|x\right),$$

where

$$f_n\left(y\,|\,x\right)=\frac{\sum_{i=1}^{n} K_{h_n}\left(\hat{G}_n(x)-\hat{G}_n(X_i)\right)K_{h_n}\left(\hat{G}_n(y)-\hat{G}_n(Y_i)\right)}{\sum_{i=1}^{n} K_{h_n}\left(\hat{G}_n(x)-\hat{G}_n(X_i)\right)},$$

where $K\left(u\right)$ is a symmetric Borel function which tends to zero as $u \to \infty$, and $h_n \to 0$ as $n \to \infty$.

2. Main Results:

This section contains the two main theoretical results of this paper. In the first one, we study the strong consistency of the proposed estimator, while on the second one, we derive the asymptotic normality of the SNN estimator.

Firstly, we consider the following conditions which will be used to prove the main results in this paper.

1. $f\left(y\,|x\right)$ is uniformly continuous in y.

2 $f^{(i,j)}(x,y)=\dfrac{\partial^{i+j} f\left(x,y\right)}{\partial x^i \partial y^j}$ exists and bounded for $1 \le i + j \le 3$.

3. $K\left(u\right), K'(u), K''(u)$ are functions of bounded variation.

4. $\lim_{|u|\to\infty}\left|u^2 K\left(u\right)\right| = 0$ and $\lim_{|u|\to\infty}\left|u^2 K'\left(u\right)\right| = 0.$

5. $\int_{-\infty}^{\infty} K(u)\,du = 1$ and $\int_{-\infty}^{\infty} u^i K(u)\,du = 0 \quad (\text{for } i = 1,2)$

6. $\int_{-\infty}^{\infty} |u|^3 K\left(u\right) du < \infty.$

7. $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} n h_n = \infty$, where $h_n = n^{-d}, \frac{1}{10} < d < \frac{1}{8}$.

To prove Theorem 1, which contains the first result, we need the following two lemmas.

Lemma 1
Under the conditions 2, 5, and 6, the following holds

$$p\left(\lim_{n\to\infty}\sup_{y\in}\left|f_n\left(x,y\right)-f\left(x,y\right)\right|=0\right)=1.$$

Proof.

$$E\left(f_n(x,y)\right)=$$

$$\left(nh_n^2\right)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{\hat{G}(x)-u}{h_n}\right)K\left(\frac{\hat{G}(y)-v}{h_n}\right)f(u,v)\,du\,dv$$

$$=\left(nh_n^2\right)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(u)K(v)\,f\left(\hat{G}(x)-h_nu,\hat{G}(y)-h_nv\right)du\,dv$$

$$=\left(nh_n^2\right)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(u)K(v)\left\{f\left(\hat{G}(x),\hat{G}(y)\right)\right.$$

$$+h_nuf^{(1,0)}\left(\hat{G}(x),\hat{G}(y)\right)+h_nvf^{(0,1)}\left(\hat{G}(x),\hat{G}(y)\right)$$

$$+h_n^2u^2f^{(2,0)}\left(\hat{G}(x),\hat{G}(y)\right)+h_n^2v^2f^{(0,2)}\left(\hat{G}(x),\hat{G}(y)\right)$$

$$\left.+\left[h_n^2u^2+h_n^2v^2\right]f^{(1,1)}\left(\hat{G}(x),\hat{G}(y)\right)+o\left(h^2\right)\right\}du\,dv.$$

Following the same techniques of [13], we have

$$E\left(f_n(x,y)\right)=f(x,y)+o\left(h_n^2\right)$$

Therefore,

$$\sup_{(x,y)\in\mathbf{R}^2}\left|Ef_n(x,y)-f(x,y)\right|\le Ch_n^2,$$

where C is a constant.

This implies that,

$$\sup_{(x,y)\in\mathbf{R}^2}\left|Ef_n(x,y)-f(x,y)\right|=o(1). \tag{1}$$

Now, by following the same techniques as in [10], we obtain that

$$\sup_{(x,y)\in\mathbf{R}^2}\left|f_n(x,y)-f(x,y)\right|\le\sup_{(x,y)\in\mathbf{R}^2}\left|f_n(x,y)-Ef_n(x,y)\right| + \sup_{(x,y)\in\mathbf{R}^2}\left|Ef_n(x,y)-f(x,y)\right|. \tag{2}$$

Then an application of Borel – Cantelli lemma in conjunction with Equation 1 and Equation 2, the proof of the lemma is completed.

Lemma 2

Under the conditions 1, 2, 5 and 6, the following holds

$$P\left(\lim_{n\to\infty}\sup_{y\in\mathbf{R}}\left|f_n(y|x)-f(y|x)\right|=0\right)=1\cdot$$

Proof.

Let $L(x)$ be an upper bound for $\frac{f(x,y)}{g(x)}$, $y\in\mathbf{R}$.

$$\sup_{y\in\mathbf{R}}\left|f_n(y|x)-f(y|x)\right|=\sup_{y\in\mathbf{R}}\left|\frac{f_n(x,y)}{g_n(x)}-\frac{f(x,y)}{g(x)}\right|$$

$$=\sup_{y\in\mathbf{R}}\left|\frac{f_n(x,y)}{g_n(x)}-\frac{f(x,y)}{g_n(x)}+\frac{f(x,y)}{g_n(x)}-\frac{f(x,y)}{g(x)}\right|$$

$$\le\sup_{y\in\mathbf{R}}\left|\frac{f_n(x,y)}{g_n(x)}-\frac{f(x,y)}{g_n(x)}\right|+\sup_{y\in\mathbf{R}}\left|\frac{f(x,y)}{g_n(x)}-\frac{f(x,y)}{g(x)}\right|$$

$$\le\sup_{y\in\mathbf{R}}\left|\frac{f_n(x,y)}{g_n(x)}-\frac{f(x,y)}{g_n(x)}\right|+\sup_{y\in\mathbf{R}}\left|\frac{f(x,y)}{g(x)}\right|\left|\frac{g(x)}{g_n(x)}-1\right|$$

$$\le\frac{\sup_{y\in\mathbf{R}}\left|f_n(x,y)-f(x,y)\right|}{g_n(x)}+L(x)\left|\frac{g(x)}{g_n(x)}-1\right|. \tag{3}$$

The proof of the Lemma is completed by an application of Lemma 1 and Equation 3.

Theorem 1:

Under the conditions of Lemma 2, the following holds

$\hat{q}(x)\to q(x)$, with probability one.

Proof.

Let $e>0$, since for each x, $f(y|x)$ is uniformly continuous with a unique mode $q(x)$. Then there is an $h(e)>0$ such that for every point x, $\left|y-q(x)\right|\ge e$ implies that

$$\left|f(y|x)-f\left(q(x)|x\right)\right|\ge h(e).$$

Now

$$\left|f\left(\hat{q}(x)|x\right)-f\left(q(x)|x\right)\right|\le\left|f\left(\hat{q}(x)|x\right)-f_n\left(\hat{q}(x)|x\right)\right|+\left|f_n\left(\hat{q}(x)|x\right)-f\left(q(x)|x\right)\right|$$
$$\le\sup_{y\in\mathbf{R}}\left|f\left(y|x\right)-f_n\left(y|x\right)\right|+\left|\sup_{y\in\mathbf{R}}f_n\left(y|x\right)-\sup_{y\in\mathbf{R}}f\left(y|x\right)\right|$$
$$\le2\sup_{y\in\mathbf{R}}\left|f_n\left(y|x\right)-f\left(y|x\right)\right|.$$

This implies that,

$$\sum_{n=1}^{\infty}P\left[\left|\hat{q}(x)-q(x)\right|\ge e\right]\le\sum_{n=1}^{\infty}P\left[\left|f_n(y\mid x)-f(y\mid x)\right|\ge h(e)\right]<\infty.$$

The proof of the theorem is completed by an application of the Borel – Cantelli Lemma.

To prove Theorem 2, which contains the second result, we need the following three lemmas.

Define for $j=1,2$

$$f_n^{(0,j)}(x,y)=\frac{1}{nh_n^{j+2}}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)K'\left(\frac{y-Y_i}{h_n}\right).$$

Lemma 3

Under the conditions 3, 4, and 5, if (x,y) is a point for which $f(x,y)$ is continuous then $\left(nh_n^4\right)\mathrm{Var}\left(f_n^{(0,1)}(x,y)\right)\to f(x,y)\iint K^2(u)K'^2(v)\,du\,dv$ as $n\to\infty$.

Proof.

$$\mathrm{Var}\left(f_n^{(0,1)}(x,y)\right)=E\left(f_n^{(0,1)}(x,y)\right)^2-\left(E\left(f_n^{(0,1)}(x,y)\right)\right)^2$$

$$nh_n^4\,\mathrm{Var}\left(f_n^{(0,1)}(x,y)\right)=\frac{1}{h_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-X_i}{h_n}\right)K'^2\left(\frac{y-Y_i}{h_n}\right)f(x,y)\,dxdy$$
$$-h_n^2\left\{\left(h_n^3\right)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-X_i}{h_n}\right)K'\left(\frac{y-Y_i}{h_n}\right)f(x,y)\,dxdy\right\}^2$$
$$=\frac{1}{h_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-u}{h_n}\right)K'^2\left(\frac{y-v}{h_n}\right)f(u,v)\,du\,dv$$
$$-\left\{\frac{\left(h_n^3\right)^{-1}}{h_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-u}{h_n}\right)K'\left(\frac{y-v}{h_n}\right)f(u,v)\,dudv\right\}^2.$$

The proof of the Lemma is completed by an application of Bochner Lemma.

Lemma 4

Under the conditions 2 through 7, the following is true

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{f_n^{(0,1)}(x,y)-f^{(0,1)}(x,y)\right\}\xrightarrow{d}N\left(0,\ f(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2(u)K'^2(v)\,dudv\right)$$

Proof.

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{E\left(f_n^{(0,1)}(x,y)\right)-f^{(0,1)}(x,y)\right\}=$$
$$\left\{\left(nh_n^4\right)^{\frac{1}{2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-X}{h_n}\right)K'^2\left(\frac{y-Y}{h_n}\right)f(x,y)\,dx\,dy-f^{(0,1)}(x,y)\right\}.$$

A substitution $u=\frac{x-X_J}{h_n},\ v=\frac{y-Y_J}{h_n}$ leads to the equation

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{E\left(f_n^{(0,1)}(x,y)\right)-f^{(0,1)}(x,y)\right\}=$$
$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2(u)K'^2(v)f(x-h_nu,y-h_nv)\,du\,dv-f^{(0,1)}(x,y)\right\}$$

Now, using Taylor expansion of the function $f(x-h_nu,y-h_nv)$ around the point (x,y) implies that,

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{E\left(f_n^{(0,1)}(x,y)\right)-f^{(0,1)}(x,y)\right\}=$$
$$\left(nh_n^4\right)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2(u)K'^2(v)\left\{f(x,y)+h_nuf^{(1,0)}(x,y)\right.$$
$$+h_nvf^{(0,1)}(x,y)+h_n^2(x)u^2f^{(2,0)}(x,y)$$
$$\left.+h_n^2v^2f^{(0,2)}(x,y)+\left[h_n^2u^2+h_n^2v^2\right]f^{(1,1)}(x,y)+o\left(h_n^2\right)\right\}.$$

Using conditions 5 and 7 implies that

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{E\left(f_n^{(0,1)}(x,y)\right)-f^{(0,1)}(x,y)\right\}=o(1).\qquad(4)$$

Define $B_{jn}=\frac{1}{h_n^3}K\left(\frac{x-X_j}{h_n}\right)K'\left(\frac{y-Y_j}{h_n}\right),$

$j=1,2,...,n.$

Note that, B_{jn} are i.i.d. random variables, and $f_n^{(0,1)}(x,y)=\frac{1}{n}\sum_{j=1}^{n}B_{jn}$.

To prove the lemma, the Liapounov's condition must be satisfied, see [6]. That is for some $d>0$,

$$\lim_{n\to\infty}\frac{\sum_{j=1}^{n}E\left|B_{jn}-E\left(B_{jn}\right)\right|^{2+d}}{\left[\sum_{j=1}^{n}\mathrm{Var}\left(B_{jn}\right)\right]^{1+\frac{d}{2}}}=0.$$

$$\frac{\sum_{j=1}^{n}E\left|B_{jn}-E\left(B_{jn}\right)\right|^{2+d}}{\left[\sum_{j=1}^{n}\mathrm{Var}\left(B_{jn}\right)\right]^{1+\frac{d}{2}}}=\frac{\left(nh_n^4\right)^{1+\frac{d}{2}}\left[\sum_{j=1}^{n}E\left|B_{jn}-E\left(B_{jn}\right)\right|^{2+d}\right]}{\left[\left(nh_n^4\right)\mathrm{Var}\left\{f_n^{(0,1)}(x,y)\right\}\right]^{1+\frac{d}{2}}}. \quad (5)$$

Now, the denominator in Equation 5 will tends to $f(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-u}{h_n}\right)K'^2\left(\frac{y-v}{h_n}\right)du\,dv$, by using Lemma 3 .

The numerator can be spilt to the following two terms

$$2^{1+d}\left(n^{-1}h_n^4\right)^{1+\frac{d}{2}}\left[\sum_{j=1}^{n}E\left|B_{jn}\right|^{2+d}\right]+2^{1+d}\left(n^{-1}h_n^4\right)^{1+\frac{d}{2}}\left[\sum_{j=1}^{n}\left\{E\left|B_{jn}\right|\right\}^{2+d}\right] \quad (6)$$

$$E\left|B_{jn}\right|^{2+d}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{h_n^3}\right)^{2+d}\left|K\left(\frac{x-X_j}{h_n}\right)K'\left(\frac{y-Y_j}{h_n}\right)\right|^{2+d}f(x,y)\,dxdy$$

Let $u=x-X_j, v=y-Y_j$, this leads to the following equation

$$E\left|B_{jn}\right|^{2+d}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{h_n^3}\right)^{2+d}\left|K\left(\frac{u}{h_n}\right)K'\left(\frac{v}{h_n}\right)\right|^{2+d}f(x-u,y-v)\,du\,dv$$

Now, the first term of Equation 6 can be written as

$$2^{1+d}\left(n^{-1}h_n^4\right)^{1+\frac{d}{2}}\left[\sum_{j=1}^{n}E\left|B_{jn}\right|^{2+d}\right]=$$

$$2^{1+d}\left(n^{-1}h_n^4\right)^{1+\frac{d}{2}}\left[n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(h_n^3\right)^{-2-d}\left|K\left(\frac{u}{h_n}\right)K'\left(\frac{v}{h_n}\right)\right|^{2+d}f(x-u,y-v)\,du\,dv\right]$$

$$2^{1+d}n^{-\frac{d}{2}}\left(h_n^2\right)^{-1+\frac{d}{2}}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|K\left(\frac{u}{h_n}\right)K'\left(\frac{v}{h_n}\right)\right|^{2+d}f(x-u,y-v)\,du\,dv\right]\longrightarrow 0,$$

by Bochner Lemma and condition 7.

Also, the second term of the Equation 6 can be written as

$$2^{1+d}\left(n^{-1}h(x)h^3(y)\right)^{1+\frac{d}{2}}\left[\sum_{j=1}^{n}\left|E\left(B_{jn}\right)\right|^{2+d}\right]$$

$$=2^{1+d}\left(n^{-1}h_n^4\right)^{1+\frac{d}{2}}\left[\left(nh_n^3\right)^{-2-d}\left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{u}{h_n}\right)K'\left(\frac{v}{h_n}\right)\right|^{2+d}f(x-u,y-v)\,du\,dv\right]\to 0,$$

by another application of Bochner Lemma and condition 7.

Hence, the proof of the lemma is completed.

Now, for a fixed x, use Taylor expansion of the function $f_n^{(0,1)}\left(\hat{q}(x)|x\right)$ around $q(x)$, to get

$$f_n^{(0,1)}\left(\hat{q}(x)|x\right)=f_n^{(0,1)}\left(q(x)|x\right)$$
$$+\left\{\hat{q}(x)-q(x)\right\}f_n^{(0,2)}\left(\hat{q}^*(x)|x\right)=0$$

where, $\left|\hat{q}^*(x)-q(x)\right|<\left|\hat{q}(x)-q(x)\right|$, for some $\hat{q}^*(x)$.

This leads to the following equation

$$\left\{\hat{q}(x)-q(x)\right\}=-\frac{f_n^{(0,1)}\left(\hat{q}(x)|x\right)}{f_n^{(0,2)}\left(\hat{q}^*(x)|x\right)}, \quad (7)$$

and hence the following equation holds

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{\hat{q}(x)-q(x)\right\}=-\left(nh_n^4\right)^{\frac{1}{2}}\frac{f_n^{(0,1)}\left(\hat{q}(x)|x\right)}{f_n^{(0,2)}\left(\hat{q}^*(x)|x\right)}. \quad (8)$$

Lemma 5
Under the conditions 1, 2, 3, 5 , 7 and if $g(x)>0$, then

$$f_n^{(0,2)}\left(\hat{q}^*(x)|x\right)\xrightarrow{p}f_n^{(0,2)}\left(q(x)|x\right),$$

as n tends to infinity.

proof.
Using the same technique as in the proof of Lemma 1, the following holds,

$$P\left[\lim_{n\to\infty}\left(\sup_{(x,y)\in\mathbf{R}^2}\left|f_n^{(0,2)}(y|x)-f^{(0,2)}(y|x)\right|=0\right)\right]=1. \quad (9)$$

Since,

$$\left| f_n^{(0,2)}\left(\hat{q}^*(x)|x\right) - f^{(0,2)}\left(q(x)|x\right)\right|$$

$$= \left| \frac{f_n^{(0,2)}\left(x,\hat{q}^*(x)\right)}{g(x)} - \frac{f^{(0,2)}\left(x,q^*(x)\right)}{g(x)} + \frac{f^{(0,2)}\left(x,q^*(x)\right)}{g(x)} - \frac{f^{(0,2)}\left(x,q(x)\right)}{g(x)} \right|$$

$$\leq \frac{\left| f_n^{(0,2)}\left(x,\hat{q}^*(x)\right) - f^{(0,2)}\left(x,q^*(x)\right)\right| + \left| f^{(0,2)}\left(x,q^*(x)\right) - f^{(0,2)}\left(x,q(x)\right)\right|}{g(x)}$$

$$\leq \sup_{(x,y)\in\mathbf{R}^2} \frac{\left| f_n^{(0,2)}(x,y) - f^{(0,2)}(x,y)\right| + \left| f^{(0,2)}\left(x,\hat{q}^*(x)\right) - f^{(0,2)}\left(x,q(x)\right)\right|}{g(x)},$$

with probability one.

Now, a combination of Equation 9 and Theorem 1 completes the proof of the lemma.

Theorem 2:
Under conditions 1 through 7, if $g(x) > 0$, then the sample conditional mode $\left(nh_n^4\right)^{\frac{1}{2}}\left\{\hat{q}(x) - q(x)\right\}$ is asymptotically normal distributed with mean zero and variance

$$\left(\frac{f\left(x,\hat{q}(x)\right)}{\left\{f^{(0,2)}\left(x,\hat{q}(x)\right)\right\}^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{K(u)K'(v)\right\}^2 dudv \right).$$

Proof
From Lemma 5 and Equation 8, the following is true

$$\left(nh_n^4\right)^{\frac{1}{2}}\left\{\hat{q}(x) - q(x)\right\} = -\left(nh_n^4\right)^{\frac{1}{2}} \frac{f_n^{(0,1)}\left(x,q(x)\right)}{f_n^{(0,2)}\left(x,\hat{q}(x)\right) + o_p(1)} \qquad (10)$$

Now, an application of Lemma 4 and Equation 10 completes the proof of the theorem.

3. Applications:

In this section, the performance of the SNN kernel estimator is tested using simulated and real data. We used the S-Plus program and the Epanechinkov kernel function.

3.1 Simulation studies:

In this subsection, the performance of the SNN estimator is tested using simulated data and compared with the Nadaraya-Watson (NW) .

Three simulated data of sizes 50, 200 and 500 are simulated from three different models. In the first study, the data are simulated from the model $y = \sin 2p\,(1-x)^2 + xe$, where e has a $N(0,1)$ distribution and x has a $uniform[0,1]$ distribution, while in the second study, the data are simulated from the model $y = x \sin 2p\, x + e$, where e has a $N(0,0.1)$ distribution and x has a $uniform[0,1]$ distribution. In the third study, the data are simulated from the model $y = x + e^{-16(x-0.2)^2} + e^{-16(x-0.7)^2} + e$, where e has a $N(0, 0.05)$ distribution and x has a $uniform\,[0,1]$ distribution. For a direct comparison of these estimators, thee scatter plots of the simulated data of size 200 from the three simulated studies, the perfect smooth and their estimators using the SNN and NW estimators are plotted in Figure 1, Figure 2 and Figure 3 respectively.

Moreover, the performance of the estimators is tested using the mean squared error (MSE),

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n},$$

where, $\hat{y}$ is the predicted values, y is the actual values and n is the size of the data. The results of this comparison are shown in Table 1. This comparison indicates that the SNN estimator is reasonably good and it is much better than the NW estimator for the first and third simulated studies. On the other hand, the NW estimator is slightly better than the SNN estimator for the second estimator.

3.2 Real data:

In this subsection, we used the air data from S-Plus program to test the performance of the SNN estimator. It is a data frame with 111 observations and four variables, taken from an environmental study that measured the four variables ozone, solar radiation, temperature, and wind speed for 111 consecutive days.

We used the temperature (observed temperature, in degrees Fahrenheit) as a predictor variable and the ozone (surface concentration of ozone in New York, in parts per million) as the response variable. We used the first 106 observation to estimate the last five observations. The results are listed in Table 2. The mean squared error, MSE=0.0682677.

Table 1 *The mean squared error of the SNN and NW estimators.*

Simulation	Size	SNN	NW
Study 1	50	0.074893	0.114534
	200	0.044020	0.091580
	500	0.037230	0.078956
Study 2	50	0.007650	0.010654
	200	0.004992	0.003950
	500	0.003686	0.003054
Study 3	50	0.008348	0.016643
	200	0.006453	0.007585
	500	0.004930	0.005788

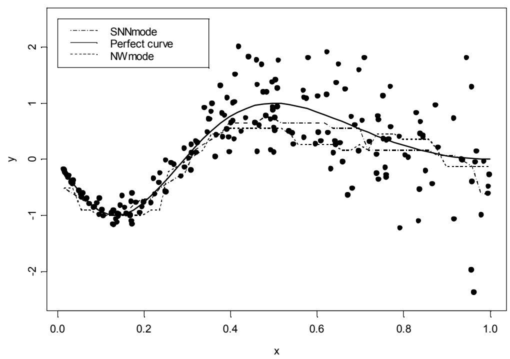

Figure 1 *(Simulation 1) Comparison between SNN and NW estimators*

Table 2 *The estimated values for the ozone data using the SNN estimator.*

Observation number	Estimated Value	True Value
107	2.661039	2.410142
108	2.651819	3.107233
109	2.670259	2.410142
110	2.642599	2.620741
111	2.661039	2.714418

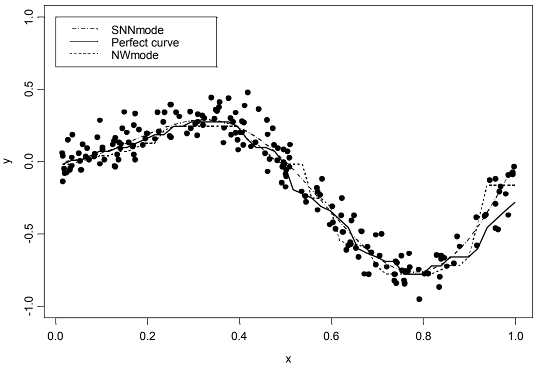

Figure 2 *(Simulation 2) Comparison between SNN and NW estimators*

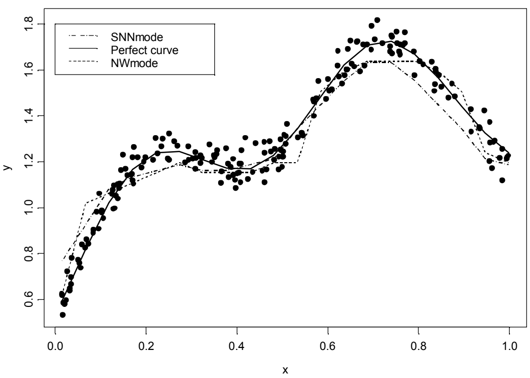

Figure 3 *(Simulation 3) Comparison between SNN and NW estimators*

4. Conclusion:

In this paper, we proposed the SNN kernel estimator to estimate the conditional mode. The strong consistency and the asymptotic normality of the estimator has been derived under some conditions much weaker than that needed for the NW kernel estimator. The performance of the SNN kernel estimator is tested using simulated and real data, which indicate that the proposed estimator is reasonably good. A comparison between the SNN estimator and the NW estimator indicates that the proposed estimator is better than the NW estimator.

The result of this paper can be generalized to the case of multivariate using the Cramer-Wold device by considering the convergence of $\left(nh_n^4\right)^{\frac{1}{2}}\left\{\hat{q}\left(x\right)-q\left(x\right)\right\}$ jointly at a finite number of points $x_1, x_2, \mathbf{K}, x_k$ with $\hat{q}(x_1), \hat{q}(x_2), \mathbf{K}, \hat{q}(x_k)$ being asymptotically independent. Also, the estimator can be improved by considering a varying bandwidth rather than the constant bandwidth.

References:

[1] Attaoui, S. On the Nonparametric Conditional Density and Mode Estimates in the Single Functional Index Model with Strongly Mixing Data. *The Indian Journal of Statistics*, ***76-A***, Vol. ***2***, (2014) 356-378.

[2] Kemp, G. C. R., and Santos Silva, J. M. C. Regression Towards the Mode. *Journal of Econometrics*, Vol. ***170***, (2012) 92-101.

[3] Liang, H. Y., and Àlvertaz, J. Asymptotic normality for estimator of Conditional Mode Left-Truncated and Dependent Observations. *Metrika*, Vol. ***72***, (2010) 1-19.

[4] Nadaraya, E. A. On Nonparametric Estimation of Density Functions and Regression Curves. *Theory of Probability and the Applications*, Vol. ***10***, (1965) 186-190.

[5] Parzen, E. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, Vol. ***33***, (1962) 1065-1076.

[6] Pranab, K., and Julio, M. *Large Sample Methods in Statistics.* Champion and Hall, (1993).

[7] Salha, R., and Shekh Ahmed, H. On the Kernel Estimation of the Conditional Mode. *Asian Journal for mathematics and Statistics*, Vol. ***2***, (2009) 1-8.

[8] Salha, R., El Shekh Ahmed, H. Local Variable Kernel Estimation of the Conditional Mode. *Rajasthan Academy of Physical Sciences*, ***8(3)***, (2009) 256-280.

[9] Salha, R., and Ioannides, D. Joint Asymptotic Distribution of the Estimated Conditional Mode at a Finite Number of Distinct Points. *The National Statistical Conference*, Lefkada, Greece, (14-18 April 2004) 587- 594.

[10] Samanta, M., Thavanesmaran, A. Nonparametric Estimation of the Conditional Mode. *Communications in Statistics, Theory and Methods*, Vol. ***19***, (1990) 4515-4524.

[11] Samanta, M. Nonparametric Estimation of the Mode of a Multivariate Density. *South African Statistical Journal*, Vol. ***7***, (1973) 109-117.

[12] Vieu, P. A Note on Density Mode Estimation. *Statistics and Probability Letters*, Vol. ***26***, (1996) 297-307.

[13] Stute, W. Asymptotic Normality of Nearest Neighbor Regression Function Estimates. *The Annals of Statistics*, ***12(3)***, (1984) 917-926.

[14] Ziegler, K. Adaptive Kernel Estimation of the Mode in the Nonparametric Random Design Regression Model. *Journal of Probability and Mathematical Statistics*, Vol. ***24***, (2004) 213-235.

[15] Ziegler, K. On the Asymptotic Normality of Kernel Regression Estimators of the Mode in the Nonparametric Random Design Model. *Journal of Statistical Planning and Inference*, Vol. ***115***, (2003) 123-144.

[16] Ziegler, K. On Nonparametric Kernel Estimation of the Mode of Regression Function in the Random Design Model. *Journal of Nonparametric Statistics*, Vol. ***14***, (2002) 749-774.