Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal. Vol. 2, Issue 1 (2009), 11 – 21
ISSN 2070-5948, DOI 10.1285/i20705948v2n1p11

VARIABLE BANDWIDTH KERNEL ESTIMATOR OF THE MODE

Raid B. Salha*

Department of Mathematics,Science Faculty,
The Islamic University, Rimal, Gaza, Palestine

Received 19 April 2009; Accepted 12 August 2009
Available online 15 August 2009

Abstract: *In this paper, the problem of estimating the mode of a probability density function has been studied. Parzen (1962) proposed a kernel estimator of the mode depends on a single bandwidth. In this paper, the Parzen estimator has been improved by proposing a kernel estimator with variable bandwidth for the mode of the density function. Proceeding as in Parzen (1962), the consistency and asymptotic normality of the proposed estimator are shown. Moreover, the good performance of the proposed estimator is tested via simulation study and it is shown that the proposed estimator is more efficient than the Parzen estimator.*

***Keywords**: Kernel estimator, variable bandwidth, asymptotic normality, consistency, Parzen estimator.*

1. Introduction

A mode of probability density function $f(x)$ is a value of x which maximizes f. Therefore, any method for estimating the mode must estimate the density function, either explicitly or implicitly.

Here, attention is focused on the class of kernel estimators introduced by Rosenblatt (1956). Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function $f(x)$. Rosenblatt proposed estimating $f(x)$ by:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right), \quad (1)$$

* Email: rbsalha@iugaza.edu.ps

where the kernel K is a bounded measurable function satisfying $\lim_{|x|\to\infty} K(x) = 0,$ and the bandwidth $h = h_n$ is a sequence of positive numbers satisfying $\lim_{n\to\infty} h = 0.$ This means that the observations which are far from x will have a little influence on $\hat{f}(x).$
Parzen (1962) maximized the estimator in Equation (1) to propose a kernel estimator $\hat{\theta}$ of the mode θ, which is defined as:

$$\hat{\theta} = \max_x \hat{f}(x). \tag{2}$$

Parzen (1962, Theorem 3A) gave conditions under which $\hat{\theta}$ is a consistent estimator of θ. Nadarya (1965) derived stronger consistency results. Parzen (1962, Theorem 5A) gave conditions under which $\hat{\theta}$ has asymptotic normal distribution. Samanta (1973) has given multivariate versions of Parzen's results.
The bandwidth in Equation (1) remains constant. This means that it depends neither on the location of x nor on the data X_i. Such an estimator does not fully incorporate the information provided by the density function of the data points. Furthermore, a constant bandwidth is not flexible enough for estimating curves with a complicated shape. All these considerations lead to introduce new estimators of the density function that depend on different bandwidths.
In this paper, a Kernel estimator with variable bandwidth for the mode of the density function is proposed. Proceeding as in Parzen (1962) the consistency and asymptotic normality of the proposed estimator are shown. Moreover, the performance of the proposed estimator is tested via simulation study.
The paper is organized as follows. In the next section, the proposed mode estimator is presented. In Section 3, the conditions that are needed to derive the main results in the paper are stated. In Section 4, the main results of this paper are presented and proved, while in Section 5, the performance of the proposed estimator is investigated through simulation study. The last section draws some concluding remarks.

2. Variable bandwidth kernel estimator of the mode

Since the mode is a local feature of the shape of the density function, it is natural to use location adaptive bandwidths. Vieu (1996) used this idea and presented a mode estimator based on local kernel density estimation.
A quite different idea from local kernel density estimation is that of variable kernel density estimation. In a variable kernel density estimation, the single bandwidth h is replaced by n different bandwidths depend on X_i, $i = 1, 2, ..., n$. The basic idea is to construct a kernel estimator consisting of kernels placed at the observed data points, but allows the bandwidth of the kernels to vary from point to another.
The variable bandwidth kernel density estimator $f_n(x)$ of $f(x)$ is defined as:

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i h}K\left(\frac{x - X_i}{\lambda_i h}\right), \tag{3}$$

where $\lambda_i = \tilde{f}(X_i)^{-\alpha}$, $\tilde{f}(x)$ is a pilot estimator that satisfies $\tilde{f}(X_i) > 0$ for all i, $0 \le \alpha \le 1$. Abramson (1982) shows that taking $\alpha = \frac{1}{2}$ is a good choice since one can achieve a bias of order h^4 rather than h^2. For more details, see Silverman (1986, pp. 100-102), Wand and Jones (1995, pp. 40-43) and Fan and Gijbels (1992).

Assume that the probability density function $f(x)$ is uniformly continuous and it has a unique mode θ defined as:

$$f(\theta) = \max_{-\infty < x < \infty} f(x).$$

Since $f_n(x)$ is continuous and tends to 0 as x tends to $\pm\infty$, there is a random variable θ_n such that:

$$f_n(\theta_n) = \max_{-\infty < x < \infty} f_n(x).$$

θ_n is the variable bandwidth kernel estimator of the mode θ.

3. Conditions

The main results in the paper are derived under the following conditions.

Condition 1.

(i) The density function $f(x)$ is uniformly continuous,

(ii) $\lim_{x \to \pm\infty} f(x) = 0$,

(iii) $f(x)$ has a continuous second derivative.

Condition 2. $K(x)$ is a Borel function satisfying the conditions

(i) $K(x)$ is twice differentiable,

(ii) $\sup_{-\infty < x < \infty} |K(x)| < \infty$,

(iii) $\int_{-\infty}^{\infty} |K(x)|\, dx < \infty$,

(iv) $\lim_{x \to \infty} |x\, K(x)| = 0$,

(v) $\int_{-\infty}^{\infty} K(x)\, dx = 1$.

Condition 3. The bandwidth $h = h_n$ is a function of n such that

(i) $\lim_{n\to\infty} h = 0$,
(ii) $\lim_{n\to\infty} nh = \infty$,
(iii) $\lim_{n\to\infty} nh^2 = \infty$,

4. Main Results

In this section, the main results are stated and proved. This section consists of two subsections. In the first subsection, the consistency of the proposed mode estimator θ_n is shown, while the asymptotic normality is shown in the second subsection.

4.1 Consistency

Firstly, the following two lemmas are proved, because they play an important rule in proving the consistency of the proposed mode estimator.

Lemma 1. Under Conditions 1, 2 and 3, the following holds.

$$P\left[\lim_{n\to\infty} \sup_{x} |f_n(x) - f(x)| = 0\right] = 1.$$

Proof. First, we show:

$$\sup_{x} |Ef_n(x) - f(x)| = o(1). \tag{4}$$

Let $y = \dfrac{u-x}{h}$, then by using Taylor expansion, we have:

$$Ef_n(x) = \int_{-\infty}^{\infty} h^{-1} f^{\frac{1}{2}}(u) K\left((x-u)h^{-1} f^{\frac{1}{2}}(u)\right) f(u)\,du = \int_{-\infty}^{\infty} f^{\frac{3}{2}}(x+hy) K(y\,f^{\frac{1}{2}}(x+hy))\,dy$$

$$\approx \int_{-\infty}^{\infty} \left(f^{\frac{3}{2}}(x) + \frac{3}{2} hy f^{\frac{1}{2}}(x) f'(x)\right) K(y\,f^{\frac{1}{2}}(x+hy))\,dy$$

$$+ \int_{-\infty}^{\infty} \left\{\frac{h^2 y^2}{2}\left(\frac{3}{4} f^{-\frac{1}{2}}(x) f'(x) + \frac{3}{2} f^{\frac{1}{2}}(x) f''(x)\right) + o(h^2)\right\} K(y\,f^{\frac{1}{2}}(x+hy))\,dy$$

$$\approx f^{\frac{3}{2}}(x)\int_{-\infty}^{\infty} K(y\, f^{\frac{1}{2}}(x+hy))dy + \frac{3}{2}hf^{\frac{1}{2}}(x)\, f'(x)\int_{-\infty}^{\infty} yK(y\, f^{\frac{1}{2}}(x+hy))\, dy$$
$$+\frac{h^2}{2}\left(\frac{3}{4}\, f^{-\frac{1}{2}}(x)\, f'(x)+\frac{3}{2}\, f^{\frac{1}{2}}(x)\, f''(x)\right)\int_{-\infty}^{\infty} y^2K(y\, f^{\frac{1}{2}}(x+hy))\, dy$$

let $t = yf^{\frac{1}{2}}(x+hy)$, and by using Taylor expansion of $f^{\frac{1}{2}}(x+hy)$ about x, we get:

$Ef_n(x) \approx f(x) + o(h^2)$.

This implies that $\sup_{x} |Ef_n(x) - f(x)| \le Ch^2$ which implies that Equation (4) is satisfied.

Now, let $F_n(u) = \frac{1}{n}\sum_{i=1}^{n} I(u - X_i)$, $\quad I(x-y) = 1$ if $x \ge y$ and $I(x-y) = 0$ if $x < y$.

$$\sup_x |f_n(x) - Ef_n(x)| = \sup_x \left| \frac{1}{n}\sum_{i=1}^{n} \frac{1}{f^{-\frac{1}{2}}(X_i)h} K\left(\frac{x-X_i}{f^{-\frac{1}{2}}(X_i)h}\right) - \int_{-\infty}^{\infty} \frac{1}{f^{-\frac{1}{2}}(u)h} K\left(\frac{x-u}{f^{-\frac{1}{2}}(u)h}\right) f(u)\, du \right|$$
$$= \sup_x \left| \int_{-\infty}^{\infty} \frac{1}{f^{-\frac{1}{2}}(u)h} K\left(\frac{x-u}{f^{-\frac{1}{2}}(u)h}\right) \{dF_n(u) - dF(u)\} \right|$$
$$\le h^{-1} \sup_u |F_n(u) - F(u)|\, \mu,$$

where:

$$\mu = \int_{-\infty}^{\infty}\left\{h^{-1} f(u)\, K'\left((x-u)h^{-1} f^{-\frac{1}{2}}(u)\right) + \frac{1}{2}h^{-1}K'\left((x-u)h^{-1}f^{-\frac{1}{2}}(u)\right) + \frac{1}{2}h^{-1}f^{-\frac{1}{2}}(u)K\left((x-u)h^{-1}f^{-\frac{1}{2}}(u)\right)\right\}$$

Now, for every ε>0,

$$\sum_{n=1}^{\infty} P\{\sup_x |\, f_n(x) - f(x)| > \varepsilon\} \le \sum_{n=1}^{\infty} P\{\sup_x f^{\frac{1}{2}}(x)\,|F_n(x) - F(x)| > \frac{\varepsilon h}{\mu}\}$$
$$\le C.\sum_{n=1}^{\infty} \exp\left\{\frac{-2\varepsilon^2 nh^2}{\mu^2}\right\} < \infty, \text{ since } nh^2 \to \infty.$$

Since $\sup_x |f_n(x) - f(x)| \le \sup_x |f_n(x) - Ef_n(x)| + \sup_x |Ef_n(x) - f(x)|$, then by an application of Borel-Cantelli lemma, see Pranab and Seng (1993, pp. 55), in conjunction with Equation (4), the proof of the lemma is completed.

Now, define the following estimators of the first two derivatives of $f(x)$ as:

$$f_n'(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\lambda_i h)^2}K'\left(\frac{x - X_i}{\lambda_i h}\right),$$

$$f_n''(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\lambda_i h)^3}K''\left(\frac{x - X_i}{\lambda_i h}\right).$$

Lemma 2. Under the conditions 1, 2 and 3 the following is true:

(i) $P\left[\lim_{n\to\infty}\sup_{x}|f_n'(x) - f'(x)| = 0\right] = 1$.

(ii) $P\left[\lim_{n\to\infty}\sup_{x}|f_n''(x) - f''(x)| = 0\right] = 1$.

Proof. Following the same lines of the proof of Lemma1. Now, the consistency of the proposed mode estimator is verified in Theorem 1.

Theorem 1. Under the conditions 1, 2 and 3 the following is true:

$\theta_n \to \theta$, with probability 1.

Proof. Let $\varepsilon > 0$, proceeding as in Parzen (1962), because f is uniformly continuous with unique mode θ, there is an $\eta(\varepsilon) > 0$ such that for every point x,

$$|x - \theta| \geq \varepsilon \text{ implies } |f(x) - f(\theta)| \geq \eta(\varepsilon). \tag{5}$$

On the other hand, we have:

$$\begin{aligned} |f(\theta_n) - f(\theta)| &\leq |f(\theta_n) - f_n(\theta_n)| + |f_n(\theta_n) - f(\theta)| \\ &\leq \sup_x |f(x) - f_n(x)| + |\sup_x f_n(x) - \sup_x f(x)| \\ &\leq 2\sup_x |f_n(x) - f(x)|. \end{aligned} \tag{6}$$

Now, Equations (5) and (6) imply that:

$$P\left(|\theta_n - \theta| \geq \varepsilon\right) \leq P\left[|f(\theta_n) - f(\theta)| \geq \eta(\varepsilon)\right] \leq P\left[\sup_{x}|f_n(x) - f(x)| \geq \frac{1}{2}\eta(\varepsilon)\right].$$

Using Lemma 1 we obtain that:

$$\sum_{n=1}^{\infty} P\left(|\theta_n - \theta| \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} P\left[|f(\theta_n) - f(\theta)| \geq \eta(\varepsilon)\right]$$
$$\leq \sum_{n=1}^{\infty} P\left[\sup_x |f_n(x) - f(x)| \geq \frac{1}{2}\eta(\varepsilon)\right] < \infty.$$

Now, an application of Borel-Cantelli lemma implies that $P\left(|\theta_n - \theta| \geq \varepsilon \text{ i.o.}\right) = 0,$ where i.o. stands for infinitely often. This completes the proof of the theorem.

4.2 Asymptotic Normality

Firstly, the asymptotic normality of $f_n'(x)$ is shown in Theorem 2. Then it is used to derive the asymptotic normality of the proposed mode estimator θ_n .

Theorem 2. Under the conditions of Lemma 1, the following holds:

$$(nh^3)^{\frac{1}{2}}\{f_n'(x) - f'(x)\} \xrightarrow{d} N\left(0, \quad f^{\frac{5}{2}}(x) \int_{-\infty}^{\infty} K'^2(u)du\right).$$

Proof. Let $V_{ni} = \frac{1}{(\lambda_i h)^2} K'\left(\frac{x - X_i}{\lambda_i h}\right)$, then:

$$f_n'(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(\lambda_i h)^2} K'\left(\frac{x - X_i}{\lambda_i h}\right) = \frac{1}{n}\sum_{i=1}^{n} V_{ni}, \tag{7}$$

where $V_{ni}, i = 1, 2, ..., n$ are iid random variables and distributed as $V_n = \frac{1}{(\lambda h)^2} K'\left(\frac{x - X}{\lambda h}\right)$.

Now, we want to show that the Liapounov condition is satisfied, that is for some $\delta > 0$:

$$\lim_{n\to\infty} \frac{E\left(|V_n - E(V_n)|^{2+\delta}\right)}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)} = 0. \tag{8}$$

$$E\,|V_n|^{2+\delta} = \int_{-\infty}^{\infty} |\, h^{-2} f(u) K'\left((x-u) h^{-1} f^{\frac{1}{2}}(u) \right)|^{2+\delta} f(u)\, du$$

$$= \int_{-\infty}^{\infty} h^{-2(2+\delta)} f^{3+2\delta}(u) \left\{ K'\left((x-u) h^{-1} f^{\frac{1}{2}}(u) \right) \right\}^{2+\delta} du$$

$$= \int_{-\infty}^{\infty} h^{-3-2\delta} f^{3+2\delta}(x+hy) \left\{ K'\left(y f^{\frac{1}{2}}(x+hy) \right) \right\}^{2+\delta} dy$$

let $t = y f^{\frac{1}{2}}(x+hy)$, and by using Taylor expansion of $f^{\frac{1}{2}}(x+hy)$ and $f^{3+2\delta}(x+hy)$ about x, we get:

$$E\,|V_n|^{2+\delta} \approx h^{-3-2\delta} f^{\frac{5}{2}+2\delta}(x) \int_{-\infty}^{\infty} K'(t)^{\,2+\delta}\, dt.$$

Then by putting $\delta = 0$, we get that:

$$Var(V_n) \approx h^{-3} f^{\frac{5}{2}}(x) \int_{-\infty}^{\infty} K'^2(t)\, dt. \qquad (9)$$

This implies that $\dfrac{E\,|V_n - E(V_n)|^{2+\delta}}{n^{\frac{\delta}{2}} \sigma^{2+\delta}(V_n)} = \dfrac{(\lambda h)^{3+2\delta} E\,|V_n - E(V_n)|^{2+\delta}}{\lambda^{3+2\delta} (nh)^{\frac{\delta}{2}} h^{3+\frac{3}{2}\delta} \sigma^{2+\delta}(V_n)} \to 0,$

since

$$h^{3+\frac{3}{2}\delta} \sigma^{2+\delta}(V_n) = \left(h^3 \sigma^2(V_n) \right)^{\frac{2+\delta}{2}} \to \left(f^{\frac{5}{2}}(x) \int_{-\infty}^{\infty} K'^2(t) dt \right)^{\frac{2+\delta}{2}} < \infty, \text{ and } nh \to \infty.$$

Now, from Liapounov condition (8), we obtain that $V_{ni}, i = 1, 2, ..., n$ are asymptotically normally distributed as V_n. Therefore, from Lemma 2 (i), Equations (7) and (9), the proof of the theorem is completed.

Consider a probability density function $f(x)$ with a unique mode at θ. If $f(x)$ has a continuous second derivative, then $f'(\theta) = 0, \quad f''(\theta) < 0.$

If the kernel function $K(x)$ is chosen to be twice differentiable, then the estimated probability density function $f_n(x)$ is twice differentiable.

If θ_n is the mode of $f_n(x)$, then $f_n'(\theta_n) = 0, \quad f_n''(\theta_n) < 0.$ By Taylor expansion,

$$0 = f_n'(\theta_n) = f_n'(\theta_n) + (\theta_n - \theta) f_n''(\theta^*_n),$$

for some random variable θ_n^* between θ_n and θ. This implies that:

$$\theta_n - \theta = -\frac{f_n'(\theta)}{f_n''(\theta^*_n)}, \text{ if } f_n''(\theta^*_n) \neq 0. \tag{10}$$

Using Theorem 1, $\theta_n^* \xrightarrow{p} \theta$ and by Lemma 2 (ii), the following holds

$$f_n''(\theta^*_n) \xrightarrow{p} f''(\theta). \tag{11}$$

In the next theorem, the asymptotic normality of the sample mode θ_n is established.

Theorem 3. Under Conditions 1, 2 and 3, the following holds**.**

$$(nh^3)^{\frac{1}{2}}(\theta_n - \theta) \xrightarrow{d} N\left(0, \quad \frac{f^{\frac{5}{2}}(\theta)\int_{-\infty}^{\infty} K'^2(t)dt}{[f''(\theta)]^2}\right).$$

Proof. Using Theorem 2, we obtain that:

$$(nh^3)^{\frac{1}{2}} f_n'(\theta) \xrightarrow{d} N\left(0, \quad f^{\frac{5}{2}}(\theta)\int_{-\infty}^{\infty} K'^2(t)dt\right) \tag{12}$$

Now, the proof is completed by a combination of the Equations (10), (11) and (12).

5. Simulation Study

In this section, the performance of the variable bandwidth kernel mode estimator θ_n is investigated and a finite sample comparison with Parzen estimator $\hat{\theta}$ is given.
The behavior of the proposed mode estimator and Parzen estimator have been tested on 200 samples of sizes 100, 300, 500, 700 and 900 realizations of a standard normal distribution. The Gaussian kernel is used. The bandwidth h is computed using the following equation from Silverman (1986, pp.45-46):

$$h = 1.06\, s\, n^{-\frac{1}{5}}, \tag{13}$$

where n is the sample size and s is the sample standard deviation. The pilot estimator that we need to compute λ_i in Equation (3) is computed using the "density" function in S-Plus program.

A comparison between the two estimators was made by computing over 200 samples for the two estimators the simulated standard deviation and the mean squared error (MSE). The results are summarized in Table 1.

Table 1. A comparison between the proposed estimator θ_n and Parzen estimator $\hat{\theta}$

Sample Size		θ_n	$\hat{\theta}$
100	MSE Standard deviation	0.01211 0.01819	0.01709 0.02753
300	MSE Standard deviation	0.00490 0.00623	0.00866 0.01174
500	MSE Standard deviation	0.003228 0.00453	0.00664 0.01071
700	MSE Standard deviation	0.00277 0.00378	0.00575 0.01088
900	MSE Standard deviation	0.00192 0.00261	0.00662 0.01536

As expected proposed estimator gives interesting results and it's behavior is better than that of Parzen estimator. Moreover, from Table 1, we note that the performance of the proposed estimator is improved as the sample size is increased because the variable bandwidth depends on the sample data.

6. Conclusion

In this paper, using a kernel estimator with variable bandwidth for the estimating the mode of a density function is considered. The consistency and asymptotic normality of the proposed estimator are shown. Moreover, the performance of the proposed estimator is tested via a simulation study and compared with Parzen estimator. The simulation study indicated that the performance of the proposed estimator is more efficient than the Parzen estimator and it is improved as the sample size increased

Acknowledgements
The author thanks the reviewers for their valuable suggestions and comments.

References

[1]. Fan, J. and Gijbels, I., (1992). *Variable bandwidth and local regression smoothers.* The Annals of Statistics. 20. 2009-2036.

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

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

[4]. Pranab, K.Sen and Julio, M. Singer., (1993). *Large sample methods in statistics*. Champion and Hall. Inc.

[5]. Rosenblatt, M., (1956). *Remarks on some non-parametric estimates of a density function*. Annals of Mathematics and Statistics. 27. 832-837.

[6]. Samanta, M., (1973). *Nonparametric estimation of the mode of a multivariate density*. South African Statistical Journal. 7. 109-117.

[7]. Silverman, B. W., (1986). *Density estimation for statistics and data analysis*. Champion and Hall. Inc.

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

[9]. Wand, M. P. and Jones, M. C., (1995). *Kernel smoothing.* Champion and Hall. Inc.