

### On the Kernel Estimation of the Conditional Mode\*

<sup>1</sup>Raid Salha and <sup>2</sup>Hazem El Shekh Ahmed

<sup>1</sup>Department of Mathematics, The Islamic University of Gaza, Palestine

<sup>2</sup>Department of Mathematics, Al Quds Open University, Palestine

---

**Abstract:** The estimator of the conditional mode obtained by maximizing the Nadaraya Watson (NW) kernel estimator of the conditional density function has disadvantages of producing rather large bias and boundary effects. The aim of this study is to overcome these disadvantages by proposing a modified estimator of the conditional mode obtained by maximizing the Reweighted Nadaraya Watson (RNW) kernel estimator of the conditional density function. The asymptotic normality and consistency of the proposed estimator are established and its efficiency is examined by two applications for both simulation and real life data.

**Key words:** Kernel estimation, conditional distribution, conditional mode, reweighted nadaraya watson estimator

---

#### INTRODUCTION

The problem of estimating the mode of a probability density function (pdf) is a matter of both theoretical and practical interest. Parzen (1962) considered the problem of estimating the mode of a univariate pdf. Parzen (1962) and Nadaraya (1965) have shown that under 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. Samanta and Thavaneswaran (1990) considered the problem of estimating the mode of a conditional pdf and they have shown under regularity conditions that the estimator of the population conditional mode is strongly consistent and asymptotically normally distributed. Salha and Ioannides (2004) generalized these results by considering the conditional mode evaluated at distinct conditional points. Vieu (1996) presented and compared four mode estimation procedures. Recently, for random design models, Ziegler (2002) proposed a kernel estimator of the mode and its asymptotic normality has been shown by Ziegler (2003). In addition, Ziegler (2004) presented an adaptive kernel estimator for the mode.

Assume that  $(X_1, Y_1, \dots, X_n, Y_n)$  are i.i.d. random variables with joint pdf  $f(x, y)$  and a conditional pdf  $f(y|x)$  of  $Y_1$  given  $X_1 = x$ . We assume that for each  $x$ ,  $f(y|x)$  is uniformly continuous in  $y$  and it follows that  $f(y|x)$  possesses a uniquely conditional mode  $M(x)$  defined by:

$$f(M(x)|x) = \max_{-\infty < y < \infty} f(y|x)$$

Samanta and Thavaneswaran (1990) considered the problem of estimating the conditional mode and they use the Nadaraya Watson (NW) estimator of the conditional density function, but this estimator has disadvantages of producing rather large bias and boundary effects. To overcome these difficulties, Hall *et al.* (1999) proposed the Reweighted Nadaraya Watson (RNW) estimator as a weighted version of the NW estimator, which combines the better sides of the Local Linear (LL) estimators such as bias reduction and no boundary effects to preserve the property of the NW estimator is always a distribution function.

---

**Corresponding Author:** Raid Salha, Department of Mathematics, The Islamic University of Gaza, Palestine

\*Originally Published in Asian Journal of Mathematics and Statistics, 2009

Let  $\tau_i(x)$  denote the probability like weights with properties that  $\tau_i(x) \geq 0$ ,  $\sum_{i=1}^n \tau_i(x) = 1$  and  $\sum_{i=1}^n \tau_i(x)(X_i - x)K_h(x - X_i) = 0$ , where  $K(\cdot)$  is a kernel function,  $K_h(\cdot) = 1/hK(\frac{\cdot}{h})$  and  $h = h_n > 0$  is the bandwidth. The role of  $\tau_i(x)$  is to adjust the NW weights such that the resulting conditional density estimator resembles that from the LL estimators. The RNW conditional density estimator is defined as follows:

$$f_n(y|x) = \frac{\sum_{i=1}^n \tau_i(x)(X_i - x)K_h(y - Y_i)}{\sum_{i=1}^n \tau_i(x)K_h(x - X_i)}$$

If  $K(u)$  is chosen such that  $K(u)$  tends to zero as  $u$  tends to  $\pm\infty$ , then for every sample sequence and for each  $x$ ,  $f_n(y|x)$  is a continuous function of  $y$  and tends to zero as  $y$  tends to  $\pm\infty$ . Consequentially, there is a random variable  $M_n(x)$ , which is called the sample conditional mode, such that:

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

In this study, the conditional mode will be estimated using the RNW estimator of the conditional pdf and the asymptotic normality of this estimator will be proved and its performance will be examined by two applications.

### CONDITIONS

Consider the following conditions:

#### Condition 1

The kernel function  $K(u)$  is a symmetric and bounded probability density function such that

- The first two derivatives of  $K(u)$ ,  $(K^{(l)}(u))$ ,  $l = 1, 2$  are functions of bounded variations.
- $|u|K(u) \rightarrow 0$ , as  $|u| \rightarrow \infty$
- $\int u^2 K(u) du < \infty$

#### Condition 2

The marginal density  $g(x)$  is uniformly continuous and is bounded from below by a positive constant.

#### Condition 3

The partial derivatives  $f^{(i,j)}(x, y) = \frac{\partial^{i+j} f(x, y)}{\partial x^i \partial y^j}$  exist and are bounded for  $1 \leq i+j \leq 3$ .

#### Condition 4

The bandwidth satisfying the following:

- $\lim_{n \rightarrow \infty} h_n = 0$ ,
- $\lim_{n \rightarrow \infty} nh_n^3 = \infty$  and  $\lim_{n \rightarrow \infty} nh_n^{10} = 0$

**MAIN RESULTS**

Here, the main two theorems of this study theorem 1 and 2 will be presented and proved. For proving these theorems, the following lemmas are required.

**Lemma 1**

Under the conditions 1, 2 and 4 (I) the following is true:

$$\tau_i(x) = n^{-1}b_i(x)(1 + o_p(1))$$

Where:

$$b_i(x) = \left(1 + \frac{hk_i g^{(0)}(x)}{g(x)}(X_i - x)K_h(x - X_i)\right)^{-1} \text{ and } k_i = \int u^2 K(u) du$$

**Poof**

The proof of this lemma is a part of the proof of theorem 1 by De Gooijer and Zerom (2003).

$$f_n^{(0,1)}(y|x) - f^{(0,1)}(y|x) = \frac{n^{-1} \sum_{i=1}^n (h^{-1}K_h^{(1)}(y - Y_i) - f^{(0,1)}(y|x))b_i(x)K_h(x - X_i)}{n^{-1} \sum_{i=1}^n b_i(x)K_h(x - X_i)}$$

Let

$$m(X_i, y) = E(h^{-1}K_h^{(1)}(y - Y_i)|X_i)$$

$$f_n^{(0,1)}(y|x) - f^{(0,1)}(y|x) = \frac{n^{-1} \sum_{i=1}^n (h^{-1}K_h^{(1)}(y - Y_i) - m(X_i, y) + m(X_i, y) - f^{(0,1)}(y|x))b_i(x)K_h(x - X_i)}{n^{-1} \sum_{i=1}^n b_i(x)K_h(x - X_i)} + \{1 + o_p(1)\} = \{(nh^4)^{-1} J_1 + J_2\}(J_3^{-1})\{1 + o_p(1)\}$$

Where:

$$J_1 = h^2 n^{-1} \sum_{i=1}^n b_i(x) \varepsilon_i K_h(x - X_i), \text{ where } \varepsilon_i = h^{-1}K_h^{(1)}(y - Y_i) - m(X_i, y)$$

$$J_2 = n^{-1} \sum_{i=1}^n (m(X_i, y) - f^{(0,1)}(y|x))b_i(x)K_h(x - X_i)$$

$$j_3 = n^{-1} \sum_{i=1}^n b_i(x)K_h(x - X_i)$$

**Lemma 2**

Under the conditions 1, 3 and 4, the following holds:

- $J_2 = g(x)\text{Bias}(f_n^{(0,1)}(y|x)) + o_p(h^2)$
- $J_3 = g(x) + o_p(1)$

**Proof**

Using Taylor expansion and integration by parts,

$$\begin{aligned}
 m(X_i, y) &= h^{-1} \int K_h^{(0)}(y - Y_i) [f(y|X_i) + (x - X_i) f^{(1,0)}(Y_i|x) + (y - Y_i) f^{(0,1)}(Y_i|x) + \\
 &\quad (x - X_i)(y - Y_i) f^{(1,1)}(Y_i|x) + \frac{1}{2} (x - X_i)^2 (y - Y_i)^2 f^{(0,2)}(Y_i|x)] dy_i \\
 &= f^{(0,1)}(y|x) + h^{-1} \int K_h^{(1)}(y - Y_i) (x - X_i) f^{(1,0)}(Y_i|x) dY_i + h^{-1} \int K_h^{(1)}(x - X_i) (y - Y_i) f^{(0,1)}(Y_i|x) dY_i + \\
 &\quad h^{-1} \int (x - X_i)(y - Y_i) f^{(1,1)}(Y_i|x) K_h^{(0)}(y - Y_i) dY_i + \frac{1}{2} \int (x - X_i)^2 f^{(2,0)}(Y_i|x) K_h^{(0)}(y - Y_i) dY_i \\
 &\quad + \frac{1}{2} \int (y - Y_i)^2 K_h^{(0)}(y - Y_i) f^{(0,2)}(Y_i|x) dY_i
 \end{aligned}$$

Then,

$$\begin{aligned}
 J_2 &\approx \int [m(X_i, y) - f^{(1,0)}(y_i|x)] b_i(x) K_h(x - X_i) g(x_i) dX_i \\
 &= \iint (x - X_i) f^{(1,0)}(Y_i|x) h^{-1} K_h^{(1)}(y - Y_i) dY_i b_i(x) K_h(x - X_i) g(X_i) dX_i \\
 &\quad + \iint (y - Y_i) f^{(0,1)}(Y_i|x) h^{-1} K_h^{(1)}(y - Y_i) dY_i (b_i(x) K_h(x - X_i) g(X_i)) dX_i \\
 &\quad + \iint (x - X_i)(y - Y_i) f^{(1,1)}(Y_i|x) h^{-1} K_h^{(1)}(y - Y_i) dY_i (b_i(x) K_h(x - X_i) g(x_i)) dx_i \\
 &\quad + \frac{1}{2} \iint (x - X_i)^2 f^{(2,0)}(Y_i|x) h^{-1} K_h^{(1)}(y - Y_i) (b_i(x) K_h(x - X_i) g(X_i)) dY_i dX_i \\
 &\quad + \frac{1}{2} \iint (y - Y_i)^2 f^{(0,2)}(Y_i|x) h^{-1} K_h^{(1)}(y - Y_i) (b_i(x) K_h(x - X_i) g(X_i)) dY_i dX_i \\
 &= \frac{1}{2} \int \left( \int (h^{-1} K_h^{(1)}(y - Y_i) f^{(2,0)}(Y_i|x) dY_i) (x - X_i)^2 K_h(x - X_i) g(X_i) \right) dX_i \\
 &\quad \approx \frac{1}{2} f^{(2,1)}(y|x) (x - X_i)^2 K_h(x - X_i) g(x_i) dX_i \\
 &\quad \approx \frac{1}{2} h^2 g(x) f^{(2,0)}(y|x) \int u^2 K(u) du + o_p(h^2)
 \end{aligned} \tag{1}$$

Then,

$$\begin{aligned}
 &\frac{1}{2} \iint (y - Y_i)^2 h^{-1} K_h^{(1)}(y - Y_i) f^{(0,2)}(Y_i|x) K_h(x - X_i) g(X_i) dY_i dX_i \\
 &\approx \frac{1}{2} g(x) \iint (y - Y_i)^2 h^{-1} K_h^{(1)}(y - Y_i) f^{(0,2)}(Y_i|x) dY_i = \frac{1}{2} h \int u^2 K(u) f^{(0,2)}(y - hu|x) du \\
 &= \frac{1}{2} f^{(0,3)} h^2 \int u^2 K(u) du + o_p(h^2)
 \end{aligned} \tag{2}$$

From Eq. 1 and 2, we get:

- $J_2 = \frac{1}{2} h^2 g(x) \int u^2 K(u) du \{f^{(2,1)}(y|x) + f^{(0,3)}(y|x)\} + o_p(h^2) = g(x) \text{Bias}(f_n^{(0,1)}(y|x)) + o_p(h^2)$
- $J_3 = \int b_i(x) K(x - X_i) g(x_i) dX_i = g(x) \int K(u) du + o_p(1) = g(x) + o_p(1)$

This completes the proof of the lemma.

Now,

$$(nh^4)^{\frac{1}{2}} \left( f_n^{(0,1)}(y|x) - f^{(0,1)}(y|x) - \text{Bias}(f_n^{(0,1)}(y|x) + o_p(h^2)) \right) = g^{-1}(x) J_1 + o_p(1)$$

$$\text{Since, } (nh^8) \rightarrow 0, (nh^4)^{\frac{1}{2}} o_p(h^2) = o_p(1).$$

This implies that:

$$(nh^4)^{\frac{1}{2}} \left( f_n^{(0,1)}(y|x) - f^{(0,1)}(y|x) - \text{Bias}(f_n^{(0,1)}(y|x)) \right) = g^{-1}(x) J_1 + o_p(1) \tag{3}$$

**Lemma 3**

Under the conditions 1, 3 and 4, the following holds:

$$J_1 \xrightarrow{d} N(0, f(x, y) \iint (K(u)K^{(1)})^2 du dv)$$

**Proof**

Let  $\Delta_i = h^2 \varepsilon_i b_i(x) K_h(x - X_i)$ .  $J_1 = n^{-\frac{1}{2}} \sum_{i=1}^n \Delta_i$ . Since,  $E(\varepsilon_i | x) = 0$ , then  $E(\Delta) = 0$  which implies that  $E(J_1) = 0$ .

$$\begin{aligned} \Delta_i^2 &= h^4 \varepsilon_i^2 b_i^2(x) K_h^2(x - X_i) \text{ therefore, } E(\Delta_i^2) = h^4 E(\varepsilon_i^2 b_i^2(x) K_h^2(x - X_i)) + o_p(1) \\ \varepsilon_i^2 &= h^{-2} K_h^{(1)}(y - Y) - 2h^{-1} K_h^{(1)}(y - Y) m(X_i, y) + m^2(X_i, y) \\ h^4 E(\varepsilon_i^2 K_h^2(x - X_i)) &= h^4 \iint h^{-2} (K_h^{(1)}(y - Y_i) K_h(x - X_i))^2 f(X_i, Y_i) dX_i dY_i \\ &= h^2 \iint (K_h(x - X_i) K_h^{(1)}(y - Y_i))^2 f(X_i, Y_i) dX_i dY_i = \iint K(u) K^{(1)}(v))^2 f(x - hu, y - hv) dudv \\ &= f(x, y) \iint K(u) K^{(1)}(v))^2 dudv + o_p(1) \end{aligned}$$

This implies that,

$$\text{Var}(J_1) = E\Delta_1^2 = f(x, y) \iint K(u) K^{(1)}(v))^2 dudv + o_p(1)$$

To show that,  $J_1 \xrightarrow{d} N(0, \text{Var}(J_1))$  we will use Liapunov's Theorem, (Pranab and Julio Singer, 1993). It is sufficient to show that:

$$\rho_n = \frac{\sum_{i=1}^n E|\Delta_i - E\Delta_i|^{2+6}}{\left[\sum_{i=1}^n \text{Var}(\Delta_i)\right]^{-1+6/2}} \rightarrow 0$$

Since,  $E\Delta_i = 0$ ,  $\sum_{i=1}^n \text{Var}(\Delta_i) = n\text{Var}(J_1)$ . Therefore, the following holds:

$$\frac{\sum_{i=1}^n E|\Delta_i - E\Delta_i|^{2+6}}{\left[\sum_{i=1}^n \text{Var}(\Delta_i)\right]^{-1+6/2}} = \frac{\sum_{i=1}^n E|\Delta_i|^{2+6}}{[n \text{Var}(\Delta_i)]^{1+6/2}} = \frac{(n^{-1})^{1+\frac{6}{2}} \sum_{i=1}^n E|\Delta_i|^{2+6}}{[n \text{Var}(\Delta_i)]^{1+6/2}} \rightarrow 0$$

Since,

$$(n^{-1})^{1+\frac{6}{2}} \sum_{i=1}^n E|\Delta_i|^{2+6} = n^{-\frac{6}{2}} \iint |\Delta_i|^{2+6} f(X_i, Y_i) dX_i dY_i \rightarrow 0, \text{ as } n \rightarrow \infty$$

Therefore,  $\rho_n \rightarrow \infty$ .

This implies that:

$$\sum_{i=1}^n \Delta_i \xrightarrow{d} N\left(0, \sum_{i=1}^n \text{Var}(\Delta_i)\right)$$

which leads to:

$$J_1 = n^{-\frac{1}{2}} \sum_{i=1}^n \Delta_i \xrightarrow{d} N\left(0, n^{-1} \sum_{i=1}^n \text{Var}(\Delta_i)\right)$$

Since,

$$\text{Var}(J_1) = n^{-1} \sum_{i=1}^n \text{Var}(\Delta_i)$$

we get that:

$$J_1 = n^{-1} \sum_i \Delta_i \xrightarrow{d} N\left(0, f(x, y) \iint K(u)K^{(1)}(v)^2 du dv\right)$$

**Theorem 1**

Under the conditions 1, 3 and 4 the following is true:

$$(nh^4)^{1/2} [f_n^{(0,1)}(y|x) - f^{(0,1)}(y|x) - \text{Bias}(f_n^{(0,1)}(y|x))] \xrightarrow{d} N\left(0, \frac{f(x, y)}{g^2(x)} \iint (K(u)K^{(1)}(v))^2 du dv\right)$$

where;

$$\text{Bias}(f_n^{(0,1)}(y|x)) = \frac{1}{2} h^2 [f^{(2,1)}(y|x) + f^{(0,3)}(y|x)] \int u^2 K(u) du$$

**Proof**

A combination of lemma 3 and Eq. 3 completes the proof of theorem 1.

Now, using Taylor expansion:

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

This implies that:

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

where;

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

Therefore,

$$(nh^4)^{1/2} (M_n(x) - M(x)) = \frac{-(nh^4)^{1/2} f_n^{(0,1)}(M(x)|x)}{f_n^{(0,2)}(M_n^*(x)|x)} \tag{4}$$

**Lemma 4**

Under the conditions 1-4, the following holds:

$$f_n^{(0,2)}(M_n^*(x)|x) \xrightarrow{p} f^{(0,2)}(M(x)|x)$$

**Proof**

Using the same techniques of lemma 4 by Samanta and Thavaneswaran (1990).

**Theorem 2**

Under the conditions 1-4, the following is true:

$$(\text{nh}^4)^{\frac{1}{2}}(M_n(x) - M(x)) \xrightarrow{d} N\left(-(\text{nh}^4)^{\frac{1}{2}}\text{Bias}(f_n^{(0,1)}(M(x)|x)), \frac{f(x,y)\iint K(u)K^{(1)}(v))^2 du dv}{[f_n^{(0,2)}(x,M(x))]^2}\right)$$

**Proof**

The proof of the theorem follows directly by using Eq. 4, lemma 4 and theorem 1.

Note that  $\text{Bias}(f_n(M(x)|x)) \rightarrow 0$  if we assume that the second moment of the kernel function  $K$  vanishes, that is  $\int u^2 K(u) du = 0$ , as by Samanta and Thavaneswaran (1990).

**Applications**

The proposed method of RNW estimator is applied to find the conditional mode of different data sets. Standardized normal kernel function is used and the weights  $\tau_i(x)$  are calculated as described by De Gooijer and Zerom (2003) and Cai (2002).

**Example 1**

This application will depend on some simulation data. Simulate a sample of size 200 from the model  $y = \sin 2\pi(1-x)^2 + \epsilon$ , where  $x \sim N(0,1)$  and  $\epsilon \sim \text{uniform}[0,1]$ . A perfect smooth would recapture the original single  $y = \sin 2\pi(1-x)^2$ , exactly. For a direct comparison of the perfect smooth and the conditional mode estimation, a scatter plot of the original data, the perfect smooth and the estimated conditional mode curve is shown in Fig. 1. The performance of the estimator can be tested using  $R^2_{y,\hat{y}}$  (the correlation coefficient between  $\hat{y}$ , the predicted values and  $y$ , the actual values).

$$R^2_{y,\hat{y}} = 1 - \frac{\text{SSE}}{\text{SSTO}}$$

where,  $\text{SSE} = \sum (y_i - \bar{y})^2$  denotes the error sum of squares,  $\text{SSTO} = \sum (y_i - \bar{y})^2$  denotes the total sum of squares and  $\bar{y}$  denotes the mean of the actual values. For the current data,  $\text{SSE} = 1.3958$ , which is small relative to  $\text{SSTO} = 15.3209$  and  $R^2_{y,\hat{y}} = 0.9089$ , which is closed to 1 and indicates that the correlation between the actual and predicated values is very strong. This comparison indicates that the proposed estimator of the conditional mode is reasonably good.

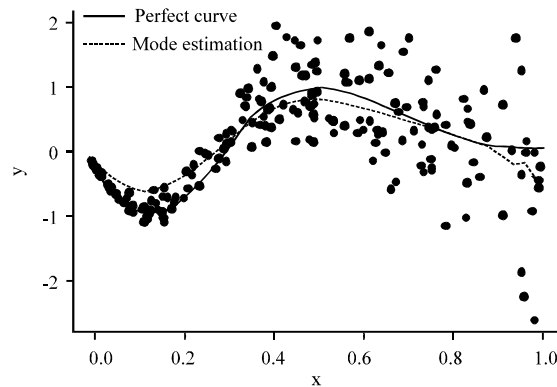


Fig. 1: Comparison between the mode estimation and the perfect curve

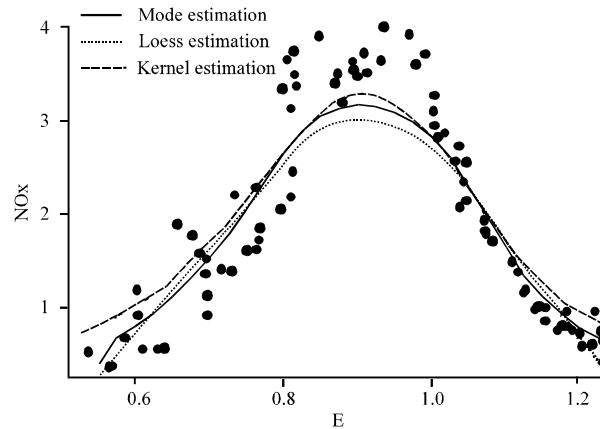


Fig. 2: Three different estimations for the ethanol data

### Example 2

Consider the ethanol data, which describes the relationship between the predictor E (ethanol) and the response NOx (Nitric Oxide). Clearly the relationship is not linear. The regression relation estimated by using three different estimators, the proposed estimator (mode estimator) and another two estimators from S-Plus program, the locally weighted regression (loess) estimator and the kernel estimator. A scatter plot of the data together with the graphs of the three estimators is shown in Fig. 2. It is clear that the proposed estimator is reasonably good.

### REFERENCES

- Cai, Z., 2002. Regression quintiles for time series. *Econometric Theory*, 18: 169-192.
- De Gooijer, J. and D. Zerom, 2003. On conditional density estimation. *Stat. Neerland.*, 57: 159-176.
- Hall, P., R.C.L. Wolf and Q. Yao, 1999. Methods for estimating a conditional distribution function. *J. Am. Stat. Associat.*, 94: 154-163.
- Nadaraya, E.A., 1965. On nonparametric estimation of density functions and regression curves. *Theory Prob. Appl.*, 10: 186-190.
- Parzen, E., 1962. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33: 1065-1076.
- Pranab Sen, K. and M. Julio Singer, 1993. *Large Sample Methods in Statistics*. 1st Edn., Chapman and Hall Inc., New York.
- Salha, R. and D. Ioannides, 2004. Joint asymptotic distribution of the estimated conditional mode at a finite number of distinct points. *Proceedings of the National Statistical Conference*, Apr. 14-18, Lefkada, Greece, pp: 587-594.
- Samanta, M., 1973. Nonparametric estimation of the mode of a multivariate density. *South Afr. Stat. J.*, 7: 109-117.
- Samanta, M. and A. Thavanesmaran, 1990. Nonparametric estimation of the conditional mode. *Commun. Stat. Theory Methods*, 19: 4515-4524.
- View, P., 1996. A note on density mode estimation. *Stat. Prob. Lett.*, 26: 297-307.
- Ziegler, K., 2002. On nonparametric kernel estimation of the mode of regression function in the random design model. *J. Nonparametric Stat.*, 14: 749-774.
- Ziegler, K., 2003. On the asymptotic normality of kernel regression estimators of the mode in the nonparametric random design model. *J. Stat. Plann. Inference*, 115: 123-144.