*The Canadian Journal of Statistics*
Vol. 7, No. 2, 1979, Pages 159–167
*La Revue Canadienne de Statistique*



# Recursive estimation of the mode of a multivariate density

Luc DEVROYE*

*McGill University*

*Key words and phrases:* Density estimation, multivariate density, recursive estimation, strong uniform consistency, estimation of the mode.
*AMS* 1980 *subject classifications:* Primary 60F15; secondary 62G05.

## ABSTRACT

Let $f$ be an unknown possibly multimodal density on $\mathbb{R}^d$ and let $X_1, X_2, \ldots$ be a sequence of independent random vectors with density $f$. Several recursive estimates of the mode of $f$ are proposed, and sufficient conditions ensuring their weak and strong consistency are established.

## 1. INTRODUCTION

In this paper we are concerned with estimating the mode of a density $f$ on $\mathbb{R}^d$ from a sample $X_1, X_2, \ldots, X_n$ of independent identically distributed random vectors with density $f$. Estimates of the mode can be classified as "direct" (when there is a simple recipe to obtain the estimate $Z_n$ from the data) or "indirect" (when first $f$ is estimated by $f_n$ and then $Z_n$ is taken to be any point for which $f_n(Z_n) = \max_x f_n(x)$).

Direct estimates for $d = 1$ were proposed by Grenander (1965), Dalenius (1965), Venter (1967), Ekblom (1972), Robertson and Cryer (1974), Sager (1975, 1978) and Chernoff (1964). Dalenius takes the midpoint or the median of the shortest inteval containing at least $k_n$ points; Venter theoretically and Ekblom experimentally study its properties; Robertson and Cryer robustize the estimate by iterative computation; and Sager proposes for $d > 1$ to pick some point inside the smallest set in a certain class of sets (e.g., spheres, rectangles) that contains at least $k_n$ of the data points. The estimates of Chernoff and Grenander also use the concept of search for the "best" interval but their criteria are different.

Most authors give conditions on $f$ and $k_n$ that ensure the almost sure convergence of $Z_n$ to $z$, the mode, when $z$ is the unique point for which $f(z) = \max_x f(x)$. For the most general theorems, and weakest conditions on $f$, the reader is referred to Sager (1978).

Consider now indirect estimates where

$$f_n(Z_n) = \max_x f_n(x) \tag{1}$$

and $f_n$ is some density estimate. In view of

$$|f(Z_n) - f(z)| \leq 2\sup|f_n(x) - f(x)| \tag{2}$$

* Research supported in part by USAF Grants 72-2371 and 77-3385 at the University of Texas at Austin.

it is clear that $f(Z_n) \xrightarrow{n} f(z)$ almost surely whenever

$$\sup_x |f_n(x) - f(x)| \xrightarrow{n} 0 \quad \text{almost surely.} \tag{3}$$

Here $\xrightarrow{n}$ should be read "as $n \to \infty$". This observation is applied to the Parzen-Rosenblatt kernel estimate (Rosenblatt 1957; Parzen 1962) by Parzen (1962), Nadaraya (1965) and Van Ryzin (1969) and to histogram estimates by Kim and Van Ryzin (1975). The kernel estimate is given by

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right), \tag{4}$$

where $K$ is a given density on $\mathbb{R}^d$ and $h_n$ is a positive number. Sufficient conditions for its strong uniform consistency, cf. (3), are given by Parzen (1962), Nadaraya (1965), Van Ryzin (1969), Deheuvels (1974), Földes and Révész (1974), Silverman (1978) and Devroye and Wagner (1980). Schuster (1970) showed that $f$ must necessarily be uniformly continuous for (3) to hold. The estimate defined by (1) is of small practical value because a time-consuming search is necessary. Also, classical search methods perform satisfactorily only when $f_n$ is sufficiently "regular" (continuous, unimodal, etcetera). A simpler and more direct estimate picks $Z_n$ among $X_1$, $X_2, \ldots, X_{\lambda_n}$ such that

$$f_n(Z_n) = \max_{1 \le j \le \lambda_n} f_n(X_j). \tag{5}$$

Here the integer $\lambda_n \le n$ is chosen by the statistician. Thus, to find $Z_n$, $f_n$ must be computed $\lambda_n$ times.

We have

Property 1. *If $f$ is uniformly continuous, if $\lambda_n \xrightarrow{n} \infty$ and if*

$$\sup_x |f_n(x) - f(x)| \xrightarrow{n} 0 \text{ in probability (almost surely),}$$

*then $f(Z_n) \xrightarrow{n} \max_x f(x)$ in probability (almost surely) for estimate* (5).

For the kernel estimate, the strong version of Property 1 is valid when

$$h_n \xrightarrow{n} 0, \qquad nh_n^d/\log n \xrightarrow{n} \infty, \tag{6}$$

and when $K$ is a Riemann integrable bounded density with compact support (Devroye and Wagner 1980). We show in Section 4 that it remains true for *all* bounded densities $K$.

Property 2. *Let $Z_n$ be the estimate defined by* (5). *If $f$ is uniformly continuous, if $\lambda_n \uparrow \infty$, if $K$ is a bounded density on $\mathbb{R}^d$, and if*

$$h_n \xrightarrow{n} 0, \qquad nh_n^d/\log \lambda_n \xrightarrow{n} \infty,$$

*then $f(Z_n) \xrightarrow{n} \max_x f(x)$ in probability. If also*

$$\sum_{n=1}^{\infty} \exp(-\alpha n h_n^d) < \infty \quad \textit{for all} \quad \alpha > 0,$$

*then $f(z_n) \xrightarrow{n} \max_x f(x)$ almost surely.*

There are situations in which for practical or economical reasons one cannot keep $X_1, X_2, \ldots, X_n$ in memory. Thus, $Z_n$ must be recursively computed as a function of $X_n$ and the memory contents at time $n - 1$. The stochastic approximation algorithms of Fritz (1973) and Mizoguchi and Shimura (1976) are capable of locating one of the local maxima of $f$. If $f$ is known to be unimodal, their algorithms are very useful. In general however, even if $f$ is uniformly continuous, $f$ can have a countably infinite number of local peaks. In the next section we define a simple recursive estimate of the mode and we show that $f(Z_n) \xrightarrow{n} \max_x f(x)$ almost surely under no conditions on the number of modes or their location.

## 2. RECURSIVE ESTIMATION OF THE MODE

Let $t_1, t_2, \ldots$ be a sequence of positive integers and let $h_1, h_2, \ldots$ and $\varepsilon_1, \varepsilon_2, \ldots$ be positive number sequences. With $s_0 = 0$, $s_k = t_1 + \cdots + t_k$, we define the sequences of random variables $Z_0, Z_1, \ldots$ and $W_0, W_1, \ldots$ by

$$Z_0 = X_1,$$

$$W_k = X_{s_k+1},\ k \geq 0$$

$$Z_k = \begin{cases} W_{k-1} & \text{if } f_k(W_{k-1}) > f_k(Z_{k-1}) + \varepsilon_k, \\ Z_{k-1} & \text{otherwise,} \end{cases} \tag{7}$$

where $f_k$ is the Parzen estimate of $f$ with $X_{s_{k-1}}, \ldots, X_{s_k}$:

$$f_k(x) = \frac{1}{t_k h_k^d} \sum_{i=s_{k-1}+1}^{s_k} K\left(\frac{x - X_i}{h_k}\right). \tag{8}$$

The computation of $f_k(W_{k-1})$ and $f_k(Z_{k-1})$ can be done recursively; $Z_k$ can be regarded as the estimate of the mode after $k$ iterations and $W_k$ can be considered as a candidate estimate at the $k$th iteration. In spite of its simplicity, the following is true for the sequence $\{Z_n\}$ defined by (7) and (8).

THEOREM 1. *If $f$ is Lipschitz, that is,*

$$\sup_{x,y} |f(x) - f(y)| \leq C\|x - y\| \quad \textit{for some} \quad C > 0, \tag{9}$$

*if $K$ is a bounded density on $\mathbb{R}^d$ with*

$$\int \|x\| K(x)\, dx < \infty, \tag{10}$$

*and if*

$$\varepsilon_n \xrightarrow{n} 0,\ h_n/\varepsilon_n \xrightarrow{n} 0, \tag{11}$$

*and*

$$t_n h_n^d \varepsilon_n^2 \xrightarrow{n} \infty$$

*then $f(Z_n) \xrightarrow{n} \max_x f(x)$ in probability for the estimate defined by* (7) *and* (8). *If also*

$$\sum_{n=1}^{\infty} \exp(-\alpha t_n h_n^d \varepsilon_n^2) < \infty \quad \textit{for all} \quad \alpha > 0, \tag{12}$$

*then $f(Z_n) \xrightarrow{n} \max_x f(x)$ almost surely.*

*Remarks.* (1) The threshold $\varepsilon_n$ measures the statistician's conservative nature. It is the handicap given to $W_{n-1}$ in the decision rule (7).

(2) Condition (12) holds if $t_n h_n^d \varepsilon_n^2/\log n \overset{n}{\to} \infty$.

(3) The classical tail condition on $K$, namely $\| x \|^d K(x) \to 0$ as $\| x \| \to \infty$, does not follow from (10). Just let

$$K(x) = \sum_{k=1}^{\infty} \frac{ck^2}{2^k} \mathscr{I}_{[2^k/k^2, 2^k/k^2 + 1/k^2]}(x),$$

where $c$ is a normalization constant and $\mathscr{I}$ is the indicator function.

(4) The scheme defined by (7) and (8) can be generalized by considering $L$ challengers $W_k(1), \ldots, W_k(L)$ for $Z_k$ instead of just one challenger $W_k$. Let $t_k \geq L$ for all $k$ and consider:

$$Z_0 = X_1,$$

$$W_k(i) = X_{s_k+i}, \qquad 1 \leq i \leq L,\ k \geq 0,$$

$$Z_k = \begin{cases} W_{k-1}(i) & \text{if } f_k(W_{k-1}(i)) = \max_{i \leq j \leq L} f_k(W_{k-1}(j)) > f_k(Z_{k-1}) + \varepsilon_k, \\ Z_{k-1} & \text{otherwise.} \end{cases} \tag{13}$$

Then Theorem 1 remains valid for (13) as well.

## 3. A GENERALIZATION

In (7) and (8) we have no control over the choice of the candidate points $W_k$. Rather than a totally random selection (i.e., $W_k = X_{s_k+1}$), a careful choice of $W_k$ (e.g., $W_k$ close to $Z_k$) may accelerate local search towards the maximum of $f$ and thus increase the accuracy of the estimate. Formally, let $T_k$ be an $\mathbb{R}^d$-valued random variable independent of $X_{s_k+2}, X_{s_k+3}, \ldots$ and let $\alpha_0, \alpha_1, \alpha_2, \ldots$ be a sequence of numbers from [0, 1]. Replace (7) with

$$Z_0 = X_1,$$

$$W_k = \begin{cases} X_{s_k+1} & \text{with probability } \alpha_k, \\ T_k & \text{otherwise, } k \geq 0, \end{cases}$$

$$Z_k = \begin{cases} W_{k-1} & \text{if } f_k(W_{k-1}) > f_k(Z_{k-1}) + \varepsilon_k, \\ Z_{k-1} & \text{otherwise.} \end{cases} \tag{14}$$

THEOREM 2. *Let $f$ be a density satisfying* (9), *let $K$ be a bounded density on $\mathbb{R}^d$ satisfying* (10), *let* (11) *hold and assume that*

$$t_n h_n^d \varepsilon_n^2 + \log \alpha_n \overset{n}{\to} \infty \tag{15}$$

*and*

$$\sum_{n=1}^{\infty} \alpha_n = \infty. \tag{16}$$

*Then $f(Z_n) \overset{n}{\to} \max_x f(x)$ in probability for scheme* (14). *If* (15) *is replaced by* (12), *then $f(Z_n) \overset{n}{\to} \max_x f(x)$ almost surely.*

*Examples.* (1) $T_k$ has a normal distribution with centre $Z_k$ and variance $\sigma_k^2$ in all directions.

(2) We may let $\sigma_k > 0$ be a small radius and let $T_k$ be the point of gravity of all those $X_i$ with $s_{k-1} < i \leq s_k$ and $\| X_i - Z_{k-1} \| \leq t_k$. If $t_k$ grows large, $T_k - Z_{k-1}$ will roughly follow the gradient of $f$ at $Z_{k-1}$.

(3) $T_k$ is the outcome of a local search started at $Z_{k-1}$ with the data $X_{s_{k-1}+1}, \ldots, X_{s_k}$. The algorithms of Fritz (1973) or of Mizoguchi and Shimura (1976) can be used for this purpose.

*Remarks.* (1) If $f(Z_n) \xrightarrow{n} \max_x f(x)$ in probability, if $\alpha_n \xrightarrow{n} 0$, if $f$ is uniformly continuous, and if $\| T_n - Z_{n-1} \| \xrightarrow{n} 0$ in probability, then $f(W_n) \xrightarrow{n} \max_x f(x)$ in probability. (In Example 1, it suffices to let $\sigma_n \xrightarrow{n} 0$.) This follows from the inequality

$$P\{f(W_n) < \max_x f(x) - \varepsilon\} \leq \alpha_n + P\{f(T_n) < \max_x f(x) - \varepsilon\}$$

$$\leq \alpha_n + P\{f(Z_{n-1}) < \max_x f(x) - \varepsilon/2\} + P\{\| T_n - Z_{n-1} \| > \delta\}$$

where $\delta > 0$ is chosen small enough.

(2) The estimate due to Loftsgaarden and Quesenberry (1965), when used in (5), leads to the following mode estimate: pick $Z_n = X_i$ among $X_1, \ldots, X_{\lambda_n}$ such that

$$D_{ni} = \min_{1 \leq j \leq \lambda_n} D_{nj}, \tag{17}$$

where $D_{ni}$ is the distance from $X_i$ to its $k_n$th nearest neighbour among $X_1, \ldots, X_n$. From Property 1 and a result due to Devroye and Wagner (1977) (see also Deheuvels 1974 and Moore and Yackel 1977) it is readily seen that $f(Z_n) \xrightarrow{n} \max_x f(x)$ almost surely whenever $f$ is uniformly continuous, $\lambda_n \xrightarrow{n} \infty$, $k_n/n \xrightarrow{n} 0$ and $k_n/\log n \xrightarrow{n} \infty$. Recursive versions of (17) in the sense of (7) would be less practical because at the $n$th stage we would need memory depth $k_n$ but $k_n \xrightarrow{n} \infty$ is needed to ensure the consistency of the estimate.

## 4. PROOFS

*Proof of Property 1.* Pick any $z \in \mathbb{R}^d$ with $f(z) = \max_x f(x)$. For $\varepsilon > 0$ find $\delta > 0$ such that $\| y - z \| \leq \delta$ implies $f(y) > f(z) - \varepsilon/3$. If $c = P\{\| X_1 - z \| \leq \delta\}$, then

$$P\{\min_{1 \leq i \leq \lambda_n} \| X_i - z \| > \delta\} \leq (1 - c)^{\lambda_n} \xrightarrow{n} 0,$$

and thus

$$P\{f(Z_n) < \max_x f(x) - \varepsilon\} \leq P\{\sup_x | f_n(x) - f(x) | > \varepsilon/3\}$$

$$+ P\{\min_{1 \leq i \leq \lambda_n} \| X_i - z \| > \delta\} \xrightarrow{n} 0.$$

The almost sure convergence part follows from

$$P\{\bigcup_{k \geq n} \{f(Z_k) < \max_x f(x) - \varepsilon\}\}$$

$$\leq P\{\bigcup_{k \geq n} \{\sup_x | f_k(x) - f(x) | > \varepsilon/3\}\} + P\{\min_{1 \leq i \leq \min(\lambda_n, \lambda_{n+1}, \ldots)} \| X_i - z \| > \delta\}$$

and the given assumptions. Q.E.D.

*Proof of Property 2.* We will use the notation of the previous proof. For arbitrary $\varepsilon > 0$ and for all $n$ so large that $nh_n^d \geq 12\, M_K/\varepsilon$ $\left(M_K = \sup_x K(x)\right)$ we have

$$
\begin{aligned}
P\{f(Z_n) < \max_x f(x) - \varepsilon\} &\le e^{-c\lambda_n} + P\{\sup_{1\le i\le\lambda_n} |f_n(X_i) - f(X_i)| > \varepsilon/3\} \\
&= e^{-c\lambda_n} + \lambda_n P\{|f_n(X_1) - f(X_1)| > \varepsilon/3\} \\
&\le e^{-c\lambda_n} + \lambda_n P\{|f_n(X_{n+1}) - f(X_{n+1})| > \varepsilon/6\} \\
&\le e^{-c\lambda_n} + \lambda_n \sup_x P\{|f_n(x) - f(x)| > \varepsilon/6\}
\end{aligned}
$$

and

$$
P\{\bigcup_{k\ge n} \{f(Z_k) < \max_x f(x) - \varepsilon\}\} \le \exp(-c\lambda_n) + \sum_{k=n}^{\infty} \lambda_k \sup_x P\{|f_k(x) - f(x)| > \varepsilon/6\}.
$$

Clearly, for $n$ large enough, $\sup_x |f(x) - \mathscr{E}\{f_n(x)\}| < \varepsilon/12$, cf. Nadaraya (1965), Van Ryzin (1969), Devroye and Wagner (1978), in view of the uniform continuity of $f$ and $h_n \xrightarrow{n} 0$.

Notice next that $f_n(x)$ is the average of $n$ independent identically distributed random variables

$$
Y_i = \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right)
$$

with $\mathscr{E}\{Y_i^2\} \le M_K M_f / h_n^d$, $M_f = \sup_x f(x)$, $|Y_i| \le M_K/h_n^d$. By the inequality (Bennett 1962)

$$
P\{|f_n(x) - \mathscr{E}\{f_n(x)\}| > \varepsilon/12\} \le 2\exp\left(-c_2 n h_n^d (\varepsilon/12)^2\right),
$$

where $c_2 = [2(M_K M_f + M_f \varepsilon/12)]^{-1}$. For large $n$ this is also an upper bound for $P\{|f_n(x) - f(x)| > \varepsilon/6\}$. Since the bound is uniform in $x$, we see that $f(Z_n) \xrightarrow{n} \max_x f(x)$ in probability when $\lambda_n \xrightarrow{n} \infty$ and $\lambda_n \exp(-\alpha n h_n^d) \xrightarrow{n} 0$ for all $\alpha > 0$. The convergence is almost sure if $\lambda_n \xrightarrow{n} \infty$ and the sequence $\{\lambda_n \exp(-\alpha n h_n^d)\}$ is summable for all $\alpha > 0$.

LEMMA 1. *Let $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ be nonnegative number sequences with $b_n \le 1$ and $a_{n+1} \le a_n(1 - b_n) + c_n$ for all $n \ge 0$.*

(i) *If $\sum_{n=1}^{\infty} b_n = \infty$ and $\sum_{n=1}^{\infty} c_n < \infty$, then $a_n \xrightarrow{n} 0$.*

(ii) *If $\sum_{n=1}^{\infty} b_n = \infty$ and $c_n/b_n \xrightarrow{n} 0$, then $a_n \xrightarrow{n} 0$.*

*Proof.* By iterative computation, for $1 < \rho_n < n$,

$$
\begin{aligned}
a_{n+1} &\le \sum_{i=0}^{n} c_i \prod_{j=i+1}^{n} (1 - b_j) + a_0 \prod_{j=0}^{n} (1 - b_j) \\
&\le (\sum_{i=0}^{\infty} c_i) \prod_{j=\rho_n+1}^{n} (1 - b_j) + \sum_{i=\rho_n+1}^{n} c_i + a_0 \exp(-\sum_{j=0}^{n} b_j) \\
&\le (\sum_{i=0}^{\infty} c_i) \exp(-\sum_{j=\rho_n+1}^{n} b_j) + \sum_{i=\rho_n+1}^{\infty} c_i + a_0 \exp(-\sum_{j=0}^{n} b_j)
\end{aligned}
$$

which tends to 0 as $n \to \infty$ if $\rho_n \xrightarrow{n} \infty$ and $\sum_{j=\rho_n+1}^{n} b_j \xrightarrow{n} \infty$. Such a sequence can be found since the $b_n$ are not summable. This proves part (i). Part (ii) is due to Braverman and Rozonoer (1969). Q.E.D.

*Proof of Theorem 1.* Consider $n$ so large that $\varepsilon_n < \frac{1}{4}\varepsilon$, $Ch_n \int \|x\| K(x)\, dx < \frac{1}{8}\varepsilon_n$ and $t_n h_n^d \varepsilon_n > 8M_K$ where $\varepsilon > 0$ is arbitrary and $C$ is the Lipschitz constant. If

$$a_n = P\{f(Z_n) < \max_x f(x) - \varepsilon\},$$

then

$$\begin{aligned}
a_n &\le P\{f(Z_{n-1}) < \max_x f(x) - \varepsilon,\ Z_n = Z_{n-1}, \\
&\qquad\qquad \text{or}\ \ f(W_{n-1}) < \max_x f(x) - \varepsilon,\ Z_n = W_{n-1}\} \\
&\le P\{f(Z_{n-1}) < \max_x f(x) - \varepsilon, \ \ \text{and}\ \ f(W_{n-1}) < \max_x f(x) - \tfrac{1}{2}\varepsilon \\
&\qquad \text{or}\ \ |f_n(Z_{n-1}) - f(Z_{n-1})| > \tfrac{1}{4}\varepsilon - \tfrac{1}{2}\varepsilon_n \ \ \text{or}\ \ |f_n(W_{n-1}) - f(W_{n-1})| > \tfrac{1}{4}\varepsilon - \tfrac{1}{2}\varepsilon_n\} \\
&\qquad + P\{f(Z_{n-1}) \ge \max_x f(x) - \varepsilon,\ \ f(W_{n-1}) < \max_x f(x) - \varepsilon,\ Z_n = W_{n-1}\} \\
&\le a_{n-1}\big(1 - \gamma\big(1 - P\{|f_n(Z_{n-1}) - f(Z_{n-1})| > \tfrac{1}{8}\varepsilon\} \\
&\qquad\qquad - P\{|f_n(W_{n-1}) - f(W_{n-1})| > \tfrac{1}{8}\varepsilon\}\big)\big) \\
&\qquad + P\{|f_n(Z_{n-1}) - f(Z_{n-1})| > \tfrac{1}{2}\varepsilon_n\} + P\{|f_n(W_{n-1}) - f(W_{n-1})| > \tfrac{1}{2}\varepsilon_n\}
\end{aligned}$$

by the independence of $W_{n-1}$ and $Z_{n-1}$. Here $\gamma = P\{f(X_1) > \max_x f(x) - \frac{1}{2}\varepsilon\}$ is strictly positive by the continuity of $f$. We will put this inequality in the form

$$a_n \le a_{n-1}\big(1 - \gamma(1 - \theta_n)\big) + \theta_n \tag{18}$$

where $\theta_n \xrightarrow{n} 0$. By Lemma 1, we may then conclude that $a_n \xrightarrow{n} 0$. By the arbitrariness of $\varepsilon > 0$ the weak convergence of $f(Z_n)$ to $\max_x f(x)$ then follows.

Clearly,

$$\begin{aligned}
P\{|f_n(Z_{n-1}) - f(Z_{n-1})| > \tfrac{1}{2}\varepsilon_n\} &+ P\{|f_n(W_{n-1}) - f(W_{n-1})| > \tfrac{1}{2}\varepsilon_n\} \\
&\le 2 \sup_x P\{|f_n(x) - f(x)| > \tfrac{1}{4}\varepsilon_n\} \\
&< 2 \sup_x P\{|f_n(x) - \mathscr{E}\{f_n(x)\}| > \tfrac{1}{8}\varepsilon_n\} \qquad (19)
\end{aligned}$$

if $t_n h_n^d \varepsilon_n > 8M_K$ and if for all $x$, $|f(x) - \mathscr{E}\{f_n(x)\}| < \frac{1}{8}\varepsilon_n$. Now,

$$\begin{aligned}
|f(x) - \mathscr{E}\{f_n(x)\}| &= \left|f(x) - \int h_n^{-d} K\left(\frac{x-y}{h_n}\right) f(y)\, dx\right| \\
&\le \int C\|x - y\| h_n^{-d} K\left(\frac{x-y}{h_n}\right) dy = Ch_n \int \|x\| K(x)\, dx < \tfrac{1}{8}\varepsilon_n.
\end{aligned}$$

We have already established an upper bound for (19). Thus, (18) is true with

$$\theta_n = 4 \exp\left\{-\frac{t_n(\frac{1}{8}\varepsilon_n)^2 h_n^d}{2M_K(M_f + \frac{1}{8}\varepsilon_n)}\right\}.$$

The almost sure convergence follows from $a_n \xrightarrow{n} 0$ for all $\varepsilon > 0$, $\sum \theta_n < \infty$ (for which (12) is needed) and

$$
\begin{aligned}
P\Big\{ \bigcup_{k\geq n} \{f(Z_k) < \max_x f(x) - \varepsilon\}\Big\} \\
&\leq P\{f(Z_n) < \max_x f(x) - \varepsilon\} + \sum_{k\geq n} P\{f(Z_{k+1}) < f(Z_k)\} \\
&\leq a_n + \sum_{k\geq n} \left(P\{|f_{k+1}(Z_k) - f(Z_k)| > \tfrac{1}{2}\varepsilon_{k+1}\} + P\{|f_{k+1}(W_k) - f(W_k)| > \tfrac{1}{3}\varepsilon_{k+1}\}\right) \\
&\leq a_n + \sum_{k\geq n} \theta_{k+1} \xrightarrow{n} 0. \qquad (20)
\end{aligned}
$$

Q.E.D.

*Proof of Theorem 2.* Trivial calculations show that for $n$ large enough, we have

$$a_n \leq a_{n-1}\big(1 - \alpha_n \gamma(1 - \theta_n)\big) + \theta_n,$$

where $a_n$, $\theta_n$, $\gamma$ are defined in the proof of Theorem 1. Convergence in probability of $f(Z_n)$ to $\max_x f(x)$ follows whenever $\sum \alpha_n = \infty$ and $\theta_n/\alpha_n \xrightarrow{n} 0$, or if $\sum \alpha_n = \infty$ and $\sum \theta_n < \infty$ (Lemma 1). Under the latter conditions, we know that the convergence is almost sure as well, cf. (20). Q.E.D.


## RÉSUMÉ

Soit $f$ une densité inconnue possiblement multimodale définie dans $\mathbb{R}^d$ et soit $X_1, X_2, \ldots$ un échantillon aléatoire ayant une densité égale à $f$. On propose plusieurs estimateurs récursifs du mode de $f$, et on présente des conditions sous lesquelles ces estimateurs sont faiblement ou fortement consistents.


## REFERENCES


Asselin de Beauville, J.P. (1978). Estimation non paramétrique de la densité et du mode. Exemple de la distribution gamma. *Rev. Statist. Appl.*, 26, 47–69.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.*, 57, 33–45.

Braverman, E.M., and Rozonoer, L.E. (1969). Convergence of random processes in learning machines theory. *Automat. Remote Control*, 30, 57–77.

Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.*, 16, 31–41.

Dalenius, T. (1965). The mode-A neglected statistical parameter. *J. Roy. Statist. Soc. Ser. A.*, 128, 110–117.

Deheuvels, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C.R. Acad. Sci. Paris Sér. A*, 278, 1217–1220.

Devroye, L.P., and Wagner, T.J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.*, 5, 536–540.

Devroye, L.P., and Wagner, T.J. (1980). The strong uniform consistency of kernel density estimates. *Multivariate Analysis-V* (P.R. Krishnaiah, ed.). North-Holland, Amsterdam, 59-77.

Ekblom, H. (1972). A Monte-Carlo investigation of mode estimators in small samples. *Appl. Statist.*, 21, 177–184.

Földes, A., and Révész, P. (1974). A general method for density estimation. *Studia Sci. Math. Hungar.*, 8, 81–92.

Fritz, J. (1973). Stochastic approximation for finding local maxima of probability densities. *Studia Sci. Math. Hungar.*, 8, 309–322.

Grenander, U. (1965). Some direct estimates of the mode. *Ann. Math. Statist.*, 36, 131–138.

Kim, B.K., and Van Ryzin, J. (1975). Uniform consistency of a histogram density estimator and modal estimation. *Comm. Statist.*, 4, 303–315.

Loftsgaarden, C.P., and Quesenberry, D.O. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, 36, 1049–1051.

Mizoguchi, R., and Shimura, M. (1976). Nonparametric learning without a teacher based on mode estimation. *IEEE Trans. Comput.*, C-25, 1109–1117.

Moore, D.S., and Yackel, J.W. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.*, 5, 143–154.

Nadaraya, E.A. (1965). On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.*, 10, 186–190.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065–1076.

Robertson, T., and Cryer, J.D. (1974). An iterative procedure for estimating the mode. *J. Amer. Statist. Assoc.*, 69, 1012–1016.

Rosenblatt, M. (1957). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832–837.

Sager, T.W. (1975). Consistency in nonparametric estimation of the mode. *Ann. Statist*, 3, 698–706.

Sager, T.W. (1978). Estimation of a multivariate mode. *Ann. Statist.*, 6, 802–812.

Schuster, E.F. (1970). Note on the uniform convergence of density estimates. *Ann. Math. Statist.*, 41, 1347–1348.

Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, 6, 177–184.

Van Ryzin, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.*, 40, 1765–1772.

Venter, J.H. (1967). On estimation of the mode. *Ann. Math. Statist.*, 38, 1446–1455.

*Received 1 February 1977*
*Revised 10 May 1979*

*School of Computer Science*
*McGill University*
*Burnside Hall*
*805 Sherbrooke Street West*
*Montréal, Québec*
*H3A 2K6*