LECTURE 3

Global Measures of Deviation

There is by now a large literature on lower bounds attainable in nonparametric density or regression estimation. Discussion of these questions can be found in the works of Farrell (1972), Stone (1980, 1982) and Hall (1989). We shall follow an exposition given by Hall (1989) when one estimates an unknown scalar. Consider models f from a family $\mathscr{F}$. L_f will denote the likelihood under model f. Given $f_0, f_1 \in \mathscr{F}$ let

$$d = \begin{cases} 0 & \text{if } L_{f_0}/L_{f_1} \geq 1, \\ 1 & \text{otherwise.} \end{cases}$$

Further let

$$c = \tfrac{1}{2} \inf_{n \geq n_0} \{P_{f_0}(d = 1) + P_{f_1}(d = 0)\} \tag{3.1}$$

or

$$c = \left\{4 \sup_{n \geq n_0} E_{f_0}\left(L_{f_1}/L_{f_0}\right)^2\right\}^{-1}. \tag{3.2}$$

The interest is in an estimate $\hat{\theta}$ of the unknown scalar $\theta = \theta(f)$. Set $b_n = \frac{1}{2}|\theta(f_0) - \theta(f_1)|$. Θ is the set of all nonparametric estimators of θ.

Two models f_0 and f_1 are selected from $\mathscr{F}$. It is usually the case that f_0 is fixed and f_1 converges to f_0 at an appropriate rate. A lower bound for the convergence rate of $\hat{\theta}$ to θ is given by b_n in the sense given by (3.3). In effect the basic issue centers on the ability to discriminate between f_1 and f_0.

THEOREM. *One can show that*

$$\inf_{\hat{\theta} \in \Theta} \sup_{f \in \mathscr{F}} P_f\left(|\hat{\theta} - \theta| \geq b_n\right) \geq c \tag{3.3}$$

for all $n \geq n_0$.

It should be understood that we are interested in what happens as $n \to \infty$. The theorem has a minimax character typical of many results in this direction.

The proof of the theorem runs as follows: Set

$$\tilde{d} = \begin{cases} 0 & \text{if } |\hat{\theta} - \theta(f_0)| \le |\hat{\theta} - \theta(f_1)|, \\ 1 & \text{otherwise.} \end{cases}$$

It then follows that if $\tilde{d} = 1$, then

$$|\hat{\theta} - \theta(f_0)| \ge \tfrac{1}{2}\{|\hat{\theta} - \theta(f_0)| + |\hat{\theta} - \theta(f_1)|\} \ge \tfrac{1}{2}|\theta(f_0) - \theta(f_1)|$$

and if $\tilde{d} = 0$, then $|\hat{\theta} - \theta(f_1)| \ge \frac{1}{2}|\theta(f_0) - \theta(f_1)|$. Consequently

$$\begin{aligned} \max_{f \in \{f_0, f_1\}} & P_f\{|\hat{\theta} - \theta(f)| \ge \tfrac{1}{2}|\theta(f_0) - \theta(f_1)|\} \\ & \ge \max\left\{P_{f_0}(\tilde{d} = 1), P_{f_1}(\tilde{d} = 0)\right\} \\ & \ge \tfrac{1}{2}\left\{P_{f_0}(\tilde{d} = 1) + P_{f_1}(\tilde{d} = 0)\right\} \\ & \ge \tfrac{1}{2}\left\{P_{f_0}(d = 1) + P_{f_1}(d = 0)\right\} = \alpha, \end{aligned}$$

where the last inequality follows by the Neyman–Pearson lemma. The theorem follows if c is given by (3.1). If c is given by (3.2), let $p = P_{f_0}(d = 1)$, $q = P_{f_1}(d = 0)$ and $\beta = E_{f_0}(L_{f_1}/L_{f_0})^2 \ge 1$. Then

$$\begin{aligned} 1 - q = P_{f_1}(L_{f_0}/L_{f_1} < 1) &= E_{f_0}\left\{I(L_{f_0}/L_{f_1} < 1)L_{f_1}/L_{f_0}\right\} \\ &\le (p\beta)^{1/2} \end{aligned}$$

by the Schwarz inequality. Then

$$\begin{aligned} 1 &= 1 - q + q \le (p\beta)^{1/2} + q \le \beta^{1/2}(p^{1/2} + q^{1/2}) \le (2\beta)^{1/2}(p+q)^{1/2} \\ &= (4\beta s)^{1/2} \end{aligned} \tag{3.4}$$

with $s = (p + q)/2$ and the result follows from (3.4).

We now apply the theorem in two simple contexts. Let $v > 0$ and $c > 1$. $[v]$ denotes the largest integer less than v. The function g on R^k is said to be (v, C)-smooth if

1. g is zero outside $(0, 1)^k$;
2. the derivatives

   $$g^{(j)}(x) = D_{x_1}^{j_1} \cdots D_{x_k}^{j_k} g(x)$$

   with $j = (j_1, \dots, j_k)$ and $|j| = j_1 + \cdots + j_k \le [v]$ exist and are bounded by C in absolute value;
3.

   $$|g^{(j)}(x) - g^{(j)}(y)| \le C\|x - y\|^{v - [v]}$$

   for $x, y \in R^k$ and $|j| = [v]$.

Application 1. Density estimation. Let $\mathscr{F}$ be the class of (v, C)-smooth densities f on R^k. Set $\theta(f) = f(x_0)$ with x_0 a fixed point in $(0, 1)^k$. $X_1, \dots, X_n$ are assumed to be independent with common density f. One wishes to show that $n^{-v/(2v+k)}$ is a lower bound for the rate of convergence of any estimator $\hat{\theta} = \hat{f}(x_0)$ of $\theta(f) = f(x_0)$. This will be established if (3.3) holds with $b_n =$ const. $n^{-v/(2v+k)}$ and $c > 0$. Choose $f_0 \in \mathscr{F}$ with f_0 taking on a constant value $k_1 \in (0, C)$ in a neighborhood of x_0. The function ω is assumed to be differentiable up to order ($[v] + 2$), nonzero at the origin, zero outside $(-1, 1)^k$ and with integral zero. Let $\delta = \delta(n) \to 0$ and set

$$f_1(x) = f_0(x) + k_2\delta^v\omega\{\delta^{-1}(x - x_0)\}.$$

If k_2 is small enough, $f_1 \in \mathscr{F}$ for all $n \geq n_0$. If the true density is f, the likelihood of the X_i's is $L_f = \prod_i f(X_i)$. Then

$$\begin{aligned} E_{f_0}(L_{f_1}/L_{f_0})^2 &= \left(\int f_1(x)^2 f_0(x)^{-1}\,dx\right)^n \\ &= \left\{1 + \int (f_1(x) - f_0(x))^2 f_0(x)^{-1}\,dx\right\}^n \\ &= \{1 + O(\delta^{2v+k})\}^n. \end{aligned}$$

If one takes $\delta = n^{-1/(2v+k)}$, the c of the theorem is strictly positive. With this choice of δ, it is clear that $b_n = \frac{1}{2}k_2|\omega(0)|n^{-v/(2v+k)}$ and the lower bound is demonstrated.

One can show that this convergence rate is realized by kernel estimators which we shall write

$$\hat{f}(x) = (n\lambda^k)^{-1}\sum_{i-1}^{n}\omega((x - X_i)\lambda^{-1}).$$

Here λ is the bandwidth, $\omega(x) = \prod_j L(x_j)$ with L a bounded function of a real variable of finite support satisfying

$$\int x^j L(x)\,dx = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{for } 1 \leq j \leq [v]. \end{cases}$$

If $\lambda = \lambda(n) = n^{-1/(2v+k)}$, the variance and squared bias of $\hat{f}(x)$ are each of order $n^{-2v/(2v+k)}$ for all $f \in \mathscr{F}$ and $x \in (0, 1)^k$.

APPLICATION 2. REGRESSION. As before $\mathscr{F}$ is the set of (v, C)-smooth functions on R^k. The observations are pairs (X_i, Y_i), $i = 1, \ldots, n$, that are independent with

$$Y_i = f(X_i) + \eta_i. \tag{3.5}$$

The X_i's are assumed to be independent of the η_i's. The X_i's are distributed on $(0, 1)^k$ with a common density p and are independent. Let the η distribution be $N(0, 1)$ with density ϕ. Consider $\theta(f) = f(x_0)$ at a point $x_0 \in (0, 1)^k$ with $f \in \mathscr{F}$. It is claimed that the convergence rate of an estimator $\hat{\theta}$ of $f(x_0)$ is $n^{-v/(2v+k)}$.

To prove this, one takes f_0 and f_1 as in the previous application. The likelihood is

$$L_f = \prod_i [p(X_i)\phi\{Y_i - f(X_i)\}]$$

and so

$$E_{f_0}\left(L_{f_1}^2 / L_{f_0}^2\right) = \{1 + O(\delta^{2v+k})\}^n.$$

Thus setting $\delta = n^{-1/(2v+k)}$ yields the result. If the density p is bounded away from zero in the vicinity of x_0, one can obtain the convergence rate with a kernel estimator.

It is clear that the argument just given can be adapted to any open neighborhood rather than just $(0, 1)^k$. Stone (1980) has also considered the estimation of a linear functional

$$Lf(x_0) = \sum_{|j| \le m} c_j D^{(j)} f(x_0)$$

for some integer $0 \le m \le [v]$ with $c_j \ne 0$ for some j with $|j| = m$ in the applications noted above. Under appropriate conditions he has shown that the lower bound for the rate of convergence of any estimator is

$$n^{-(v-m)/(2v+k)}$$

if $v > m$. Further this rate can be attained.

The remarks above relate to optimal rates of convergence locally. Stone (1980) also considered optimal rates of convergence globally relative to regression. Above, the case of a regression model in which the independent variable X was random was considered. However, one can obtain essentially the same type of result in terms of order of magnitude whether X is random or deterministic under appropriate conditions. The set S on which convergence is considered is assumed to be compact with a nonempty interior. U is taken to be an open subset of R^k containing S. The conditional distribution of Y given X is assumed to be given by $g(y|x, t)\phi(dy)$ with ϕ a measure on R, t belonging to an open interval I and such that

$$\int y g(y|x, t)\phi(dy) = t \quad \text{for } x \in R^k,\ t \in I.$$

As before it will be assumed that the possible regression $t = \theta(x)$ is (v, C)-smooth. For $x \in U$, $\theta(x) \in I$, g is assumed to be strictly positive and continuously differentiable in t. Further it is assumed that interchange of integration and differentiation is feasible so that

$$\int g(y|x, t)\phi(dy) = 1$$

leads to

$$\int g'(y|x, t)\phi(dy) = 0$$

and

$$\int g''(y|x,t)\phi(dy) = 0.$$

Set $l(y|x,t) = \log g(y|x,t)$. There are given constants ε_0, $K > 0$ and a function $M(y|x,t)$ so that

$$|l''(y|x, t + \varepsilon)| \le M(y|x,t) \quad \text{for } |\varepsilon| \le \varepsilon_0$$

and

$$\int M(y|x,t)g(y|x,t)\phi(dy) \le K.$$

Also let there be an $s > 0$ so that

$$\int e^{s|y-t|}g(y|x,t)\phi(dy)$$

is bounded for $x \in U$ and $t \in I$.

Last of all, one assumes that for each $\lambda \in (0, 1/k)$ and $c > 0$, there is a $c' > 0$ so that

$$\lim_n P\left(\#\{i: 1 \le i \le n \text{ and } |X_i - x| \le cn^{-\lambda}\} \ge c'n^{1-\lambda k} \text{ for all } x \in U\right) = 1.$$

Stone considers as before the estimation of a linear functional

$$L\theta(x) = \sum_{|j| \le m} c_j D^{(j)}\theta(x)$$

for some integer $0 \le m \le [v]$ with $c_j \ne 0$ for some j with $|j| = m$ over S. He studies the L^q norm

$$\|g\| = \left(\int_C |g(x)|^q \, dx\right)^{1/q}, \qquad 0 < q < \infty,$$

and the L^∞ norm

$$\|g\|_\infty = \sup_{x \in C} |g(x)|.$$

The following result is obtained by a discrimination argument like that in the case of local convergence, except that it is now necessary to discriminate between a large number of alternative models instead of just two and the number of these models increases with the sample size. Here $r = (v - m)/(2v + k)$.

THEOREM. *Under the assumptions noted above, if* $0 < q < \infty$, $\{n^{-r}\}$ *is the optimal rate of convergence. If* $q = \infty$, $\{n^{-1} \log n\}^r$ *is the optimal rate of convergence.*

There are two limited aspects to the interesting global results of Stone on regression estimation as noted by Nussbaum (1986) in his paper. First of all, the global convergence is not attained in, say $\overline{U}$, but rather in the compact subset S of U. The modifications that may have to be made in, say kernel estimation, to obtain the same rate of convergence at a boundary point have already been noted earlier in Lecture 2. The second remark is that the estimators that Stone employs to show that the optimal rate can be obtained have a piecewise polynomial character. Thus they may not be as smooth as the function being estimated. Nussbaum wishes to approximate by estimators that are at least as smooth as the function estimated. Nussbaum's objective is to remedy these two deficiencies he sees in Stone's result. In other ways his results are more special than Stone's are. His basic model is that given in (3.5). The regression function f is assumed given on a simply-connected domain U in R^k whose closure is compact [say contained in $(-\frac{1}{2},\frac{1}{2})^k$]. The Sobolev space

$$W_p^\beta(\Omega) = \{g \in L^p(\Omega);\, D^{(j)}g \in L^p(\Omega),\, j \in Z_+^k,\, |j| = \beta\}$$

with the norm

$$\|g\|_{\beta,p}(\Omega) = \|g\|_p(\Omega) + |g|_{\beta,p}(\Omega),$$

where

$$|g|_{\beta,p}(\Omega) = \sum_{|j|=\beta} \|D^{(j)}g\|_p(\Omega).$$

A condition referred to as a *Lipschitz boundary condition* is imposed on the domain Ω. A cone is defined as a set of the form

$$\{x \in R^k;\, \|x\| < \|a\|,\, x'a > \tau\|x\|\,\|a\|\}$$

for some $a \in R^k - \{0\}$ and $\tau \in (0,1)$. One says that the domain Ω has a Lipschitz boundary if for some positive integer m, there are open subsets U_i and cones A_i, $i = 1,\dots,m$, such that

$$\overline{\Omega} \subset \bigcup_{i=1}^m U_i, \qquad (U_i \cap \overline{\Omega}) + A_i \subset \Omega, \qquad i = 1,\dots,m.$$

The information on the regression function is of the form

$$f \in W_p^\beta(\Omega, L) = \{f \in W_p^\beta(\Omega);\, |f|_{\beta,p}(\Omega) \le L\}.$$

The argument is given under the strong assumption that the points x_i, $i = 1,\dots,n$ are nonrandom and on a rectangular grid. However, it is claimed that a modification of the argument will still yield the same result if the maximal distance between any point $x \in \Omega$ and the design set x_i is $O(n^{-1/k})$. The errors in Nussbaum's argument are also taken to be $N(0,1)$ though it is claimed this can be relaxed.

The approximation to the regression function is carried out in terms of multidimensional splines of order d. These splines are represented in terms of a basis of multidimensional B-splines of order d. The multidimensional B-splines of order d are simply products of one-dimensional B-splines of order d. These, of course, have the pleasant property that their one-dimensional projections are nonzero only on d successive one-dimensional intervals. The basic grid mesh for the splines is $\Delta = \Delta(n)$, where $\Delta^{-k}/n \to 0$ as $n \to \infty$. The argument is carried out in part by patching together local least squares fits. The smoothness is retained by carrying this out in terms of the splines. The boundary condition on the domain Ω is used to insure that a cube straddling the boundary can be replaced by a closed cube totally within Ω. The following result is obtained.

THEOREM. *If $f \in \mathscr{F} = W_p^\beta(\Omega, L)$ and Δ is chosen so that*

$$\tilde{n} \simeq \Delta^{-1/(2\beta+k)}$$

($\tilde{n} = n$ if $1 \le p < \infty$ and $\tilde{n} = n/\log n$ if $p = \infty$), then there is a $\delta > 0$ such that

$$\sup_n \sup_{f \in \mathscr{F}} E \exp\left(\delta \|\hat{f} - f\|_p(\Omega) \tilde{n}^{\beta/(2\beta+k)}\right) < \infty.$$

An interesting discussion of some results on optimal rates of convergence can be found in the paper of Kiefer (1982).

We give a sketch of an argument presented in Bickel and Rosenblatt (1973). Let $X_1, \dots, X_n$ be independent identically distributed random variables with distribution function $F(x)$ and positive continuous probability density $f(x) = F'(x)$. Let $F_n(x)$ be the sample distribution function

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} I_x(Y_j)$$

of $Y_1, \dots, Y_n$, independent random variables uniformly distributed on $[0, 1]$, where

$$I_x(u) = \begin{cases} 1 & \text{if } u \le x, \\ 0 & \text{otherwise}. \end{cases}$$

Then the kernel probability density estimate of $f(x)$ with kernel $\omega(\cdot)$ can be written as

$$f_n(x) = \frac{1}{b(n)} \int \omega\left(\frac{x-u}{b(n)}\right) dF_n(F(u))$$

with $Y_j = F(X_j)$. We already know that under appropriate conditions

$$[nb(n)]^{1/2} f(x)^{-1/2} \{f_n(x) - Ef_n(x)\}$$

$(n \to \infty,\ b(n) \downarrow 0,\ nb(n) \to \infty)$ will be asymptotically normally distributed as $n \to \infty$. But this last expression is

$$Y_n(x) = b(n)^{-1/2} f(x)^{-1/2} \int \omega\left(\frac{x-u}{b(n)}\right) dZ_n^0(F(u))$$

with

$$Z_n^0(u) = n^{1/2}\{F_n(u) - F(u)\}.$$

We shall show that the process $Y_n(x)$ can be approximated under appropriate conditions in a natural way by a Gaussian process when n is large. Let $Z^0(u)$ be the Brownian bridge process, that is, the Gaussian process with mean zero

$$EZ^0(u) \equiv 0$$

and covariance function

$$EZ^0(u)Z^0(v) = \min(u,v) - uv, \qquad 0 \le u, v \le 1.$$

We shall make use of a remarkable result of Komlós, Major and Tusnády (1975) that states the following. Given independent uniformly distributed random variables (on [0, 1]) $Y_1, Y_2, \ldots$ on a sufficiently rich probability space, one can construct a sequence of Brownian bridge processes ${}_{(n)}Z^0(u)$ on the space such that

$$P\left\{ \sup_{0 \le u \le 1} |Z_n^0(u) - {}_{(n)}Z^0(u)| > n^{-1/2}(c \log n + x) \right\} \le Le^{-\lambda x}$$

for appropriate positive constants c, L, λ. Thus the difference between $Z_n^0(u)$ and the approximating Gaussian process ${}_{(n)}Z^0(u)$ is uniformly (in u)

$$O(\log n / n^{1/2}).$$

Let us now make the following assumptions on ω and f.

ASSUMPTION A1. (a) ω vanishes outside $[-L, L]$ (L finite) and is absolutely continuous with the derivative ω' inside or (b) ω is absolutely continuous on $(-\infty, \infty)$ with $\int |\omega'(u)|^k \, du$ finite, $k = 1, 2$.

ASSUMPTION A2. $f^{1/2}$ is absolutely continuous with bounded derivative $\frac{1}{2} f'/f^{1/2}$. Also assume that

$$\int |u|^{3/2} |\log \log(u)|^{1/2} [|\omega'(u)| + |\omega(u)|] \, du < \infty$$

and that $f(x)$ is not zero.

Let

$${}_0Y_n(x) = b(n)^{-1/2} f(x)^{-1/2} \int \omega\left(\frac{x-u}{b(n)} \right) dZ^0(F(u)).$$

We shall carry through the argument under Assumptions A1 and A2, but with alternative A1(b). The derivation under the other alternative A1(a) is quite similar. The result of Komlós, Major and Tusnády tells us that by an integration by parts

$$\begin{aligned} Y_n(x) - {}_0Y_n(x) &= b(n)^{-1/2} f(x)^{-1/2} \\ &\quad \times \int [Z^0(F(u)) - Z_n^0(F(u))] \omega'\left(\frac{x-u}{b(n)} \right) du \, b(n)^{-1} \\ &= O\left(\frac{\log n}{(b(n)n)^{1/2}} \right) \end{aligned}$$

and so the difference tends to zero uniformly in x if $\log n/(b(n)n)^{1/2} \to 0$ as $n \to \infty$. Now

$$Z^0(u) = Z(u) - uZ(1), \qquad 0 \le u \le 1,$$

where $Z(u)$ is the Wiener process, the normal process with mean zero

$$EZ(u) \equiv 0$$

and covariance function

$$EZ(u)Z(v) = \min(u, v).$$

If

$${}_1Y_n(x) = b(n)^{-1/2} f(x)^{-1/2} \int \omega\left(\frac{x-u}{b(n)}\right) dZ(F(u)),$$

then

$${}_0Y_n(x) - {}_1Y_n(x) = O\big(b(n)^{1/2}\big).$$

Notice that

$${}_2Y_n(x) = (b(n)f(x))^{-1/2} \int \omega\left(\frac{x-u}{b(n)}\right) f(u)^{1/2}\, dZ(u)$$

is just another way of writing the process ${}_1Y_n(x)$. The final approximation is that given by

$${}_3Y_n(x) = b(n)^{-1/2} \int \omega\left(\frac{x-u}{b(n)}\right) dZ(u).$$

The difference

$$\begin{aligned}
&{}_2Y_n(x) - {}_3Y_n(x)\\
&\quad = b(n)^{-1/2} \int \omega(u)\left\{\left(\frac{f(x-b(n)u)}{f(x)}\right)^{1/2} - 1\right\} dZ(x - b(n)u)\\
&\quad = -b(n)^{-1/2} \int Z(x - b(n)u)\left[\omega(u)\left\{\left(\frac{f(x-b(n)u)}{f(x)}\right)^{1/2} - 1\right\}\right]' du.
\end{aligned}$$

The law of the iterated logarithm for the Wiener process tells us that

$$\limsup_{|u|\to\infty} |Z(u)|\{2|u|\log\log|u|\}^{-1/2} = 1.$$

If we use the law of the iterated logarithm for the Wiener process and the conditions of A2, it is seen that

$${}_2Y_n(x) - {}_3Y_n(x) = O\big(b(n)^{1/2}\big)$$

as $n \to \infty$. Therefore

$$Y_n(x) - {}_3Y_n(x) = O\left(\frac{\log n}{(nb(n))^{1/2}} + b(n)^{1/2}\right)$$

under Assumptions A1 and A2.

This last approximation allows one to get limit theorems for the maximal weighted deviation of a probability density estimate from the probability density as well as the integrated squared weighted deviation of a kernel probability density estimate from the probability density. The first example of a maximal deviation is of theoretical interest but the rate of convergence and the normalization are so slowly varying that there is no practical application. It is worthwhile looking at the quadratic statistic

$$T_n = nb(n)\int [f_n(x) - f(x)]^2 a(x)\,dx \tag{3.6}$$

with $a(x)$ a bounded piecewise smooth integrable function. Before looking at this expression let us first consider

$$\tilde{T}_n = nb(n)\int [f_n(x) - Ef_n(x)]^2 a(x)\,dx = \int L_n^2(x) a(x)\,dx,$$

where

$$L_n(x) = f^{1/2} Y_n. \tag{3.7}$$

Let ${}_3L_n(x)$ be the expression obtained by replacing Y_n by ${}_3Y_n$ in (3.7). It is clear that

$$\left|\tilde{T}_n - \int {}_3L_n^2(x) a(x)\,dx\right| = O\left(b(n)^{1/2}\right)$$

if

$$b(n) \to 0, \qquad \log n = O(n^{1/2} b(n)).$$

By directly looking at the characteristic function of

$$\int L_3(x)^2 a(x)\,dx,$$

one can show that its mean is

$$\int f(x) a(x)\,dx \int \omega(z)^2\,dz + O(b(n)),$$

its variance to the first order is

$$2b(n)\int [\omega * \bar{\omega}(u)]^2\,du \int a^2(x) f^2(x)\,dx, \qquad \bar{\omega}(u) = \omega(-u), \tag{3.8}$$

and the kth order cumulants ($k > 2$) are $O(b^{k-1}(n))$ as $n \to \infty$. Thus

$$b(n)^{-1/2}\left[\tilde{T}_n - \int f(x) a(x)\,dx \int \omega(z)^2\,dz\right]$$

is asymptotically normally distributed with mean zero and variance (3.8) divided by $b(n)$. Now T_n can be expanded as

$$\tilde{T}_n + 2nb(n)\int [f_n(x) - Ef_n(x)][Ef_n(x) - f(x)]a(x)\,dx$$
(3.9)
$$+ nb(n)\int [Ef_n(x) - f(x)]^2 a(x)\,dx.$$

Suppose that ω is symmetric about zero with

$$c = \int \omega(u)u^2\,du \neq 0.$$

Let us also assume that f has a continuous bounded second derivative. The second term of (3.9) can directly be shown to be asymptotically normal with mean zero and variance

$$nb(n)^4 c^2\left[\int f''(x)^2 a(x)^2 f(x)\,dx - \left\{\int f''(x) f(x) a(x)\,dx\right\}^2\right]$$

to the first order. The last term of (3.9) is

$$nb(n)^5 c^2 \int f''(x)^2 a(x)\,dx \quad \text{to the first order.}$$

One should note that the first and second terms of (3.9) are asymptotically independent as $n \to \infty$. We have the following result.

THEOREM. *Let $b(n) \to 0$, $nb(n) \to \infty$, $\log n = O(n^{1/2}b(n))$ and set*

$$\alpha(n) = \begin{cases} n^{-1/2}b(n)^{-2} & \text{if } nb(n)^5 \to \infty, \\ nb(n)^{1/2} & \text{if } nb(n)^5 \to 0, \\ n^{9/10} & \text{if } nb(n)^5 \to \lambda,\ 0 < \lambda < \infty. \end{cases}$$

Further let

$$\beta(n) = E\int [f_n(x) - f(x)]^2 a(x)\,dx.$$

Then, under the conditions on f, ω, and a mentioned above

$$\alpha(n)\left[\int [f_n(x) - f(x)]^2 a(x)\,dx - \beta(n)\right]$$

$$\to \begin{cases} c\sigma_3 Z & \text{if } nb(n)^5 \to \infty, \\ 2^{1/2}\sigma_1 Z & \text{if } nb(n)^5 \to 0, \\ \left(c^2\sigma_3^2\lambda^{4/5} + 2\sigma_1^2\lambda^{1/5}\right)^{1/2} Z & \text{if } nb(n)^5 \to \lambda,\ 0 < \lambda < \infty, \end{cases}$$

in distribution with Z a standard normal variable and

$$\sigma_3^2 = \int f''(x)^2 a^2(x) f(x)\, dx - \left(\int f''(x) f(x) a(x)\, dx \right)^2,$$

$$\sigma_1^2 = \int [\omega * \bar{\omega}(u)]^2\, du \int a^2(x) f^2(x)\, dx.$$

Also

$$\beta(n) = (nb(n))^{-1} \int f(x) a(x)\, dx \int \omega(z)^2\, dz$$

$$+ b(n)^4 c^2 \int f''(x)^2 a(x)\, dx$$

$$+ o\left((nb(n))^{-1} + b(n)^4\right).$$

The previous result does not directly deal with the case in which the density is only given on a finite interval and $a(x) \equiv 1$ on that interval. If the weight function ω is zero outside $[-1, 1]$ we shall have to modify the weight function as indicated in Chapter 2 whenever we get within a bandwidth $b(n)$ of the left or right boundary of the interval. If this is not done and f is positive in the neighborhood of the boundary, the mean $\beta(n)$ of $\int [f_n(x) - f(x)]^2\, dx$ will be perturbed by a term of the magnitude of $b(n)$. In many circumstances [if $n^{-1} = o(b(n))^2$] this will dominate $\beta(n)$. On the other hand, if the boundary adjustment is made, the asymptotic result cited in the previous theorem will still be valid.

A more effective way of obtaining the asymptotic distribution of (3.6), particularly in the multidimensional case, has been given by Hall (1984). We shall try to outline basic aspects of the procedure without going into too much detail. For convenience the case $a(x) \equiv 1$ is considered. Now

$$I_n = \int [f_n(x) - f(x)]^2\, dx = \int [\{f_n(x) - Ef_n(x)\} + \{Ef_n(x) - f(x)\}]^2\, dx$$

$$= I_{n1} + I_{n2} + I_{n3} + I_{n4},$$

where

$$I_{n1} = \sum\sum_{1 \le i < j \le n} H_n(X_i, X_j) 2\left(nb(n)^k\right)^{-2}$$

with

$$H_n(x, y) = \int A_n(u, x) A_n(u, y)\, du$$

and

$$A_n(u,x) = \omega((u-x)/b(n)) - E\{\omega((u-X)/b(n))\},$$

$$I_{n2} = \left(nb(n)^k\right)^{-2} \sum_{i=1}^{n} \int A_n(u, X_i)^2\, du,$$

$$I_{n3} = 2\int \{f_n(x) - Ef_n(x)\}\{Ef_n(x) - f(x)\}\, dx,$$

$$I_{n4} = \int [Ef_n(x) - f(x)]^2\, dx.$$

Here k is the dimension of the space. The terms I_{n2}, I_{n3} are sums of independent random variables and so can be handled in standard ways. The term

$$U_n = 2^{-1}\left(nb(n)^k\right)^2 I_{n1}$$

is a U-statistic with H_n symmetric and the X_i's independent identically distributed random variables (which are k-vectors). Further $E\{H_n(X_1, X_2)\} = 0$ and U_n is degenerate in the sense that

$$E\{H_n(X_1, X_2)|X_1\} = 0 \tag{3.10}$$

with probability 1. A central limit theorem is obtained for such degenerate U statistics by basing it on a martingale central limit theorem. If one introduces

$$Y_i = \sum_{j=1}^{i-1} H_n(X_i, X_j),$$

it can be seen that $E(Y_i|X_1, \dots, X_{i-1}) = 0$ because of the degeneracy (3.10) and so $\{S_i = \sum_{j=2}^{i} Y_j,\ 2 \le i \le n\}$ is a martingale with $S_n = U_n$. The central limit theorem for degenerate U statistics is now stated.

THEOREM. *Let H_n be symmetric with finite second moment and $E\{H_n(X_1, X_2)|X_1\} = 0$ with probability 1 for each n. Then if*

$$G_n(x,y) = E\{H_n(X_1,x)H_n(X_1,y)\}$$

and

$$\left[E\{G_n^2(X_1, X_2)\} + n^{-1}E\{H_n^4(X_1, X_2)\}\right]/\left(E\{H_n^2(X_1, X_2)\}\right)^2 \to 0$$

as $n \to \infty$, U_n is asymptotically normal with mean zero and variance $\frac{1}{2}n^2E\{H_n^2(X_1, X_2)\}$.

The result just stated is obtained by applying the following martingale central limit theorem [see Hall and Heyde (1980)].

THEOREM. *Let $\{S_{ni}, \mathscr{F}_{ni}, 1 \le i \le m_n, n \ge 1\}$ be a triangular martingale array with σ-fields $\mathscr{F}_{n,i}$ satisfying $\mathscr{F}_{n,i} \subset \mathscr{F}_{n,i+1}$. Further let differences Y_{ni}*

have mean zero and finite second moment. Suppose $s_n^2 = ES_{nm_n}^2$,

$$s_n^{-2} \sum_i E[Y_{ni}^2 I(|Y_{ni}| > \varepsilon s_n) | \mathscr{F}_{n,i-1}) \to_p 0 \tag{3.11}$$

for every $\varepsilon > 0$ *and that*

$$s_n^{-2} V_{n,m_n}^2 = s_n^{-2} \sum E(Y_{ni}^2 | \mathscr{F}_{n,i-1}) \to_p \eta^2 \tag{3.12}$$

with η^2 *a random variable finite with probability* 1. *Then* $s_n^{-1} S_{nm_n} = s_n^{-1} \Sigma_i X_{ni}$ *converges in distribution to a random variable Z with characteristic function* $E \exp(-\frac{1}{2}\eta^2 t^2)$.

In the situation dealt with here, η^2 is the constant 1 and the limiting distribution is a standard normal distribution. One can show that

$$\sum_{i-2}^{n} E(Y_{ni}^4) \le \text{const.}\, n^3 E\{H_n^4(X_1, X_2)\}$$

and this implies that

$$s_n^{-4} \sum_{i-2}^{n} E(Y_{ni}^4) \to 0.$$

This last relation implies that (3.11) is satisfied. It can also be shown that

$$\begin{aligned} E(V_n^4) = {} & 2 \sum_{i-2}^{n} (i-1)(i-2)(2n-2i+1) E\{G_n^2(X_1, X_2)\} \\ & + \sum_{i-2}^{n} (i-1)(2n-2i+1) \operatorname{var}\{G_n(X_1, X_1)\} \\ & + \left[\tfrac{1}{2} n(n-1) E\{G_n^2(X_1, X_2)\}\right]^2 \end{aligned}$$

and that

$$s_n^2 = \tfrac{1}{2} n(n-1) E\{H_n^2(X_1, X_2)\}.$$

Thus

$$E(V_n^2 - s_n^2)^2 = EV_n^4 - s_n^4 \le \text{const.}\left[n^4 E\{G_n^2(X_1, X_2)\} + n^3 E\{H_n^2(X_1, X_2)\}\right]$$

and so $s_n^{-4} E(V_n^2 - s_n^2)^2 \to 0$, implying (3.12).

Hall assumes that ω is bounded and nonnegative and satisfies

$$\int \omega(z)\, dz = 1, \qquad \int z_i \omega(z)\, dz = 0,$$

$$\int z_i z_j \omega(z)\, dz = 2c\delta_{ij} < \infty.$$

The density f and its second order derivatives are assumed bounded and uniformly continuous.

One can show that

$$E\{H_n^2(X_1, X_2)\} \sim b(n)^{3k}\sigma_1^2,$$

$$E\{H_n^4(X_1, X_2)\} = O\big(b(n)^{5k}\big),$$

$$E\{G_n^2(X_1, X_2)\} = O\big(b(n)^{7k}\big),$$

where

$$\sigma_1^2 = \left\{\int f^2(x)\,dx\right\}\left[\int\left\{\int \omega(u)\omega(u+v)\,du\right\}^2 dv\right].$$

By the central limit theorem for U-statistics, one can see that I_{n1} is asymptotically normal with mean 0 and variance $2n^{-2}b(n)^{-k}\sigma_1^2$. Also I_{n2} has a variance $O(n^3b(n)^{2k})^{-1}$ and a mean

$$EI_{n2} = \big(nb(n)^k\big)^{-1}\left\{\int \omega(u)^2\,du - b(n)^k \int\int \omega(u)\omega(u+v)\,du\,dv \right.$$
$$\left. \times \int f(x)\,f(x+b(n)v)\,dx\right\}.$$

A direction application of the Lindeberg central limit theorem implies that I_{n3} is asymptotically normal with mean zero and variance $n^{-1}b(n)^4c^2\sigma_3^2$, where

$$\sigma_3^2 = \int\{\nabla^2 f(x)\}^2 f(x)\,dx - \left[\int\{\nabla^2 f(x)\} f(x)\,dx\right]^2.$$

If $nb(n)^{k+4} \to \infty$, I_{n1} will be asymptotically negligible compared to I_{n3} while if $nb(n)^{k+4} \to 0$, the converse will hold. If $nb(n)^{k+4} \to \lambda$, $0 < \lambda < \infty$, I_{n1} and I_{n3} can be shown to be asymptotically independent. We then have the following result.

THEOREM. *Let $b(n) \to 0$, $nb(n)^k \to \infty$ as $n \to \infty$. Set*

$$\alpha(n) = \begin{cases} n^{-1/2}b(n)^{-2} & \text{if } nb(n)^{k+4} \to \infty, \\ nb(n)^{(1/2)k} & \text{if } nb(n)^{k+4} \to 0, \\ n^{(k+8)/2(k+4)} & \text{if } nb(n)^{k+4} \to \lambda,\ 0 < \lambda < \infty, \end{cases}$$

and

$$\beta(n) = \int\{Ef_n(x) - f(x)\}^2\,dx - EI_{n2}.$$

Then under the assumptions on ω and f,

$$\alpha(n)\left[\int\{f_n(x)-f(x)\}^2\,dx-\beta(n)\right]$$

$$\to\begin{cases}2^{1/2}\sigma_1 Z & \text{if } nb(n)^{k+4}\to 0,\\ c\sigma_3 Z & \text{if } nb(n)^{k+4}\to\infty,\\ \left(c^2\sigma_3^2\lambda^{4/(k+4)}+2\sigma_1^2\lambda^{-k/(k+4)}\right)^{1/2}Z & \text{if } nb(n)^{k+4}\to\lambda,\ 0<\lambda<\infty,\end{cases}$$

in distribution with Z a standard normal variable.

Without discussing the assumptions, we note that Csörgő and Horváth (1988) have obtained the asymptotic distribution of

$$I_n(p)=\int|f_n(t)-f(t)|^p a(t)\,dt,\qquad 1\le p<\infty,$$

in the one-dimensional case in certain circumstances. A crucial assumption is that the weight function is bounded, continuous, of finite support and with mean zero. Z represents a standard normal random variable. Let

$$m=E|Z|^p\left(\int\omega(u)^2\,du\right)^{p/2}\int f^{p/2}(t)a(t)\,dt,$$

$$r(t)=\int\omega(u)\omega(t+u)\,du\bigg/\int\omega(u)^2\,du,$$

and

$$\sigma^2=\sigma_1^2\int f^p(t)a(t)\,dt\left(\int\omega^2(t)\,dt\right)^p$$

with

$$\sigma_1^2=(2\pi)^{-1}\int\left\{\int\int|xy|^p(1-r^2(u))^{-1/2}\right.$$

$$\left.\times\exp\left(-\frac{1}{2(1-r^2(u))}(x^2-2r(u)xy+y^2)\right)dx\,dy-E|Z|^p\right\}du.$$

It is then claimed that if $b(n)\to 0$, $nb(n)\to\infty$, $nb(n)^4\to 0$ among other assumptions that

$$(b(n)\sigma^2)^{-1/2}\{(nb(n))^p I_n(p)-m\}$$

will converge in distribution to a standard normal variable. It is clear that one must require that

$$\int f(t)^p\,dt<\infty.$$

Heuristically the object is to replace $n^{1/2}[F_n(t) - F(t)]$ in the representation

$$f_n(t) - f(t) = \frac{1}{nb(n)} \int \omega\left(\frac{t-u}{b(n)}\right) d[F_n(t) - F(t)]$$

by the Brownian bridge process $Z^0(F(t))$ with error given by the result of Komlós, Major and Tusnády (1975). Then in turn one replaces the Brownian bridge by the Wiener process $Z(F(t))$ with a small error. In turn one can replace $dZ(F(t))$ by $f(t)^{1/2}\, dZ(t)$ in distribution. Because of the bandlimited character of the weight function ω, the resulting process obtained after all these replacements in $I_n(p)$ is seen to be $2b(n)$ dependent [if the support of ω is $(-1, 1)$]. One then makes use of a central limit theorem to establish asymptotic normality. We note that one of the earliest papers on uniform convergence of a kernel probability density estimate is that of Parzen (1962).