UNIFORM CONSISTENCY OF A CROSS-VALIDATED DENSITY ESTIMATOR

by

James Stephen Marron
University of North Carolina
Chapel Hill, North Carolina

AMS 1980 Subject Classification: Primary 62G05, secondary 62G20

Key Words and Phrases: Nonparametric density estimation, kernel estimator, cross-validation.

ABSTRACT

In the problem of nonparametric estimation of a probability density, kernel estimators are considered. Uniform consistency is established when the bandwidth is chosen by a version of cross-validation. This estimator has been shown in Marron (1983a) to have excellent mean integrated square error properties.

1. INTRODUCTION

Consider the problem of estimating a univariate probability density, f, using a sample $X_1,\ldots,X_n$ from f. A very popular estimator is the "kernel estimator" defined as follows. Given a "kernel function", K, and a "bandwidth", $h > 0$, let

$$(1.1) \qquad \hat{f}(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) .$$

There are many results giving various asymptotic properties of $\hat{f}(x,h)$. In particular, it is well known that if $h = h(n)$ is to be chosen deterministically in an optimal fashion, then h must depend on the unknown "smoothness" of f. In Marron (1983a), a data-based means of choosing h is proposed. This is seen to have the same optimal asymptotic properties as the deterministic choice of h, but does not make use of the smoothness of f.

The estimator of Marron (1983a) is defined as follows. For $j=1,\ldots,n$ define the "leave one out" estimators,

$$(1.2) \qquad \hat{f}_j(x,h) = \frac{1}{(n-1)h}\sum_{i\neq j} K\left(\frac{x-X_i}{h}\right) ,$$

$$\hat{f}_j^+(x,h) = \max(\hat{f}_j(x,h),0) .$$

Find an interval $[a,b]$ on which f is known to be bounded above 0 and define

$$(1.3) \qquad \rho(y) = \int_a^b \frac{1}{h} K\left(\frac{x-y}{h}\right) dx .$$

Now define the "estimated likelihood",

$$(1.4) \qquad \hat{L}(h) = \prod_{j=1}^{n} \hat{f}_j^+(X_j,h)^{1_{[a,b]}(X_j)} e^{-\rho(X_j)} ,$$

and take $\hat{h}$ to maximize $\hat{L}(h)$. This choice of $\hat{h}$ is a modification of the technique of cross-validation introduced by Habbema, Hermans and van den Broeck (1974). This particular modification is heuristically motivated in Marron (1983a).

The error criterion that is optimized by the estimator $\hat{f}(x,\hat{h})$ is a particular version of the Mean Integrated Square Error, defined by

$$\text{MISE} = E\int_a^b [\hat{f}(x,h)-f(x)]^2 f(x)^{-1}dx \ . \tag{1.5}$$

In Marron (1983a) the proofs that are given require that maximization of $\hat{L}(h)$ be performed over a restricted range of h's, which is determined asymptotically. This can be disturbing to the experimenter who has a fixed sample size and hence does not know the appropriate range of h's. In this paper, it is seen that the cross-validated $\hat{h}$ lies in this range. Hence, for sufficiently large sample size, the experimenter need only maximize $\hat{L}(h)$ over $h > 0$.

Another result of this paper is that the estimator $\hat{f}(x,h)$ is consistent uniformly over the entire real line. This is comforting to the experimenter because, unlike the integral norm (1.5), the uniform norm guarantees that the estimator $\hat{f}(x,\hat{h})$ is well-behaved both at each individual point and also for x not in the interval [a,b].

2. ASSUMPTIONS AND THEOREMS

It will be assumed that the underlying density, f, satisfies

(f.1) f is bounded,

(f.2) f is bounded above 0 on [a,b]

(f.3) there are constants M, $\gamma > 0$ so that for all $x,y \in \mathbb{R}$,

$$|f(x)-f(y)| \le M|x-y|^{\gamma} \ .$$

It will also be assumed that the kernel function, K, satisfies:

(K.1) $\int K(x)dx = 1$,

(K.2) K is bounded and supported inside [-1,1],

(K.3) K is of bounded variation

(K.4) K is uniformly continuous, and letting w(u) denote the square root of the modulus of continuity,

$$\int_0^1 [-\log u]^{\frac{1}{2}} dw(u) < \infty .$$

The theorems of this paper can now be stated.

Theorem 1: Assume (f.1)-(f.3) and (K.1)-(K.4). If $\hat{h} = \hat{h}(n)$ is any sequence of maxima of $\hat{L}(h)$, then

$$\lim_{c\to 0} \overline{\lim_{n}} P[\hat{h} < cn^{\frac{-1}{2\gamma+1}}] = 0 .$$

Theorem 2: Under the assumptions of Theorem 1,

$$\hat{h} \to 0 \quad \text{a.s.}$$

Note that a consequence of Theorem 1, Theorem 2, and Theorem A of Silverman (1978) is the uniform consistency result:

Theorem 3: Under the assumptions of Theorem 1,

$$\sup_{x\in\mathbb{R}} |\hat{f}(x,\hat{h}) - f(x)| \to 0 \qquad \text{in probability.}$$

3. PROOF OF THEOREM 1

Using the familiar (see, for example, Rosenblatt (1971) or (3.6) of Marron (1983a)) variance and bias^2 decomposition of (1.5) it is easily seen that

$$\text{MISE} = \frac{b-a}{nh}(\int K(u)^2 du) + o(\frac{1}{nh}) + \int_a^b [\int K(u) f(y-hu)du - f(y)]^2 \frac{1}{f(y)} dy .$$

It will be convenient to denote the bias^2 term by

$$s_f(h) = \int_a^b [\int K(u) f(y-hu)du - f(y)]^2 f(y)^{-1} dy .$$

Note that by (f.2), (f.3), (K.1) and (K.2), as $h \to 0$

$$s_f(h) = O(h^{2\gamma}) .$$

The above may be summarized as

$$\text{(3.1)} \qquad \text{MISE} = \frac{b-a}{nh}\left(\int K(u)^2 du\right) + o\left(\frac{1}{nh}\right) + s_f(h) = \frac{C}{nh} + o\left(\frac{1}{nh}\right) + O(h^{2\gamma}) ,$$

for a constant C.

For sequences $\{a_n\}$ and $\{b_n\}$ it will be convenient to let the phrase "$h = h(n)$ is between a_n and b_n" mean:

$$\lim_{n\to\infty} a_n h^{-1} = 0 \quad \text{and} \quad \lim_{n\to\infty} b_n h^{-1} = \infty .$$

It will also be useful to define, for $j=1,\ldots,n$,

$$\text{(3.2)} \qquad \Delta_j = \frac{\hat{f}_j(X_j,h) - f(X_j)}{f(X_j)} , \qquad \Delta_j^+ = \frac{\hat{f}_j^+(X_j,h) - f(X_j)}{f(X_j)} .$$

By Theorem A of Silverman (1978), for h between $n^{-1}\log n$ and 1,

$$\text{(3.3)} \qquad \sup_x |\hat{f}(x,h) - f(x)| \to 0 \quad \text{a.s.}$$

But, by (1.1), (1.2) and (K.2), uniformly in x and j,

$$\text{(3.4)} \qquad \hat{f}_j(x,h) - \hat{f}(x,h) = \frac{1}{n(n-1)h}\sum_{i\neq j} K\left(\frac{x-X_i}{h}\right) - \frac{1}{nh}K\left(\frac{x-X_j}{h}\right) = O\left(\frac{1}{nh}\right) .$$

Hence, by (f.2), for h between $n^{-1}\log n$ and 1,

$$\sup_{j=1,\ldots,n} |1_{[a,b]}(X_j)\Delta_j^+| \le \sup_{j=1,\ldots,n} |1_{[a,b]}(X_j)\Delta_j| \to 0 ,$$

in probability. Now, for $n=1,2,\ldots$ define the event

$$\text{(3.5)} \qquad U_n = \{1_{[a,b]}(X_j)\Delta_j^+ = 1_{[a,b]}(X_j)\Delta_j \quad \text{for } j = 1,\ldots,n\} .$$

It follows from the above that for h between $n^{-1}\log n$ and 1,

$$\lim_{n\to\infty} P[U_n] = 1 .$$

It will be convenient to define (analogous to (1.4))

$$L = \prod_{j=1}^{n} (f(X_j)e^{-1})^{1_{[a,b]}(X_j)} .$$

Note that maximizing $\hat{L}(h)$ by choice of h is the same as maximizing $n^{-1}\log(\hat{L}(h)/L)$. From the above it follows that for h between $n^{-1}\log n$ and 1, on the event U_n,

$$n^{-1}\log(\hat{L}(h)/L) = n^{-1}\sum_{j=1}^{n} [1_{[a,b]}(X_j)\log(1+\Delta_j) - \rho(X_j) + 1_{[a,b]}(X_j)] =$$

$$= n^{-1}\sum_{j=1}^{n} [1_{[a,b]}(X_j)(1+\Delta_j - \tfrac{1}{2}\Delta_j^2 + o_p(\Delta_j^2)) - \rho(X_j)] =$$

$$= n^{-1}\sum_{j=1}^{n} [1_{[a,b]}(X_j)(1+\Delta_j - \rho(X_j))] - (2n)^{-1}\sum_{j=1}^{n} 1_{[a,b]}(X_j)\Delta_j^2 + \tag{3.6}$$

$$+ o_p(n^{-1}\sum_{j=1}^{n} 1_{[a,b]}(X_j)\Delta_j^2) .$$

The terms of this expansion may be handled by the following Lemmas:

<u>Lemma 1.1</u>: For h between n^{-1} and 1,

$$n^{-1}\sum_{j=1}^{n} [1_{[a,b]}(X_j)(1+\Delta_j - \rho(X_j))] = o_p(\text{MISE}) .$$

<u>Lemma 1.2</u>: For h between n^{-1} and 1,

$$n^{-1}\sum_{j=1}^{n} 1_{[a,b]}(X_j)\Delta_j = \text{MISE} + o_p(\text{MISE}) .$$

Lemma 1.1 is a consequence of (3.1) and Lemma 1 of Marron (1983a). Lemma 1.2 is Theorem 2 of Marron (1983b) in the special case d=1, $w(x) = 1_{[a,b]}(x)f(x)^{-2}$.

Theorem 1 would follow from Lemma 1.1, Lemma 1.2 and (3.1) if $\hat{h}$ were restricted to be between $n^{-1} \log n$ and ∞. The fact that this is really no restriction at all is a consequence of (3.1) and the following lemmas.

<u>Lemma 1.3</u>: There exists $\epsilon > 0$ so that, for $h \le \epsilon n^{-1}\log n$,

$$\lim_{n\to\infty} P[\hat{L}(h)=0] = 1 .$$

Lemma 1.4: There exists $\delta > 0$ so that for $\epsilon n^{-1}\log n \le h \le n^{\frac{-1}{2\gamma+1}}$,

$$\lim_{n\to\infty} P[n^{-1}\log(\hat{L}(h)/L) \le -\delta \cdot \mathrm{MISE}] = 1.$$

These lemmas will now be established.

Proof of Lemma 1.3

This proof is based on an order statistics result of Cheng (1983) and is very similar to the proof of (ii) in Lemma 1.1 of Chow, Geman and Wu (1983).

It will be convenient to define the set

$$A = \{j=1,\ldots,n: X_j \in [a,b]\} ,$$

and to let N denote the cardinality of A. Note that N is Binomial(n,p) where

(3.7) $$p = P[X_i \in [a,b]] = \int_a^b f(x)dx.$$

Next, let $X_{(1)},\ldots,X_{(N)}$ denote the order statistics of $\{X_j : j\in A\}$. Define $X_{(0)} = a$, $X_{(N+1)} = b$. By (1.2), (1.4) and (K.2) a sufficient condition for $\hat{L}(h) = 0$ is: for some $j=1,\ldots,N$, both $X_{(j)}-X_{(j-1)} > h$ and $X_{(j+1)}-X_{(j)} > h$. Hence, for $\hat{L}(h) = 0$, it is sufficient that

$$\max_{j=1,\ldots,N} \min(X_{(j)}-X_{(j-1)}, X_{(j+1)}-X_{(j)}) > h .$$

Now let $\bar{F}$ denote the c.d.f. of X_j conditioned on $X_j \in [a,b]$. By (f.1) and (f.2) there is a constant $\sigma > 1$ so that, for $x\in(a,b)$,

$$\sigma^{-1} < \bar{F}'(x) < \sigma .$$

Thus, for $\hat{L}(h) = 0$ it is sufficient that

$$\max_{j=1,\ldots,N} \min(\bar{F}(X_{(j)}) - \bar{F}(X_{(j-1)}), \bar{F}(X_{(j+1)}) - \bar{F}(X_{(j)})) > \sigma h$$

But conditioned on A, $\{\bar{F}(X_j) : j\in A\}$ have the same distribution as N independent Uniform (0,1) random variables. With this in mind, let $U_{(1)},\ldots,U_{(m)}$ denote the order statistics of an iid Uniform (0,1) sample. Let $U_{(0)} = 0$, $U_{(m+1)} = 1$. Define

$$M_m = \max_{j=1,\ldots,m} \min(U_{(j)}-U_{(j-1)}, U_{(j+1)}-U_{(j)}) .$$

Theorem 4.8 of Cheng (1983) is:

$$P[\lim_{m\to\infty} \frac{2mM_m}{\log m} = 1] = 1.$$

It follows from the above that there is an $\epsilon > 0$ so that for $h \le \epsilon n^{-1}\log n$,

$$\lim_{n\to\infty} P[\hat{L}(h)=0|N\ge np/2] \ge \lim_{n\to\infty} P[M_N > \sigma h|N\ge np/2] = 1.$$

Lemma 1.3 is now an easy consequence of the fact that N is a Binomial(n,p) variable.

Proof of Lemma 1.4

Suppose $\epsilon > 0$ is given. Using the proof of Theorem A of Silverman (1978), it can easily be shown that there is a constant $B > 0$ so that, for $\epsilon n^{-1}\log n \le h \le n^{\frac{-1}{2\gamma+1}}$

$$\overline{\lim_n} \sup_{x\in[a,b]} |\hat{f}(x,h)-f(x)| < B \quad \text{a.s.}$$

Hence by (f.2), (3.2) and (3.4), there is a constant $B' > 0$ so that, for $\epsilon n^{-1}\log n \le h \le n^{\frac{-1}{2\gamma+1}}$,

$$\overline{\lim_n} \sup_{j=1,\ldots,n} |1_{[a,b]}(X_j)\Delta_j| < B' \quad \text{a.s.}$$

Next observe that, by (3.5), (3.6) and Lemma 1.1, for h between n^{-1} and 1, on the event U_n,

$$\frac{1}{n}\log(\frac{\hat{L}(h)}{L}) = \frac{1}{n}\sum_{j=1}^{n} 1_{[a,b]}(X_j)[\log(1+\Delta_j)-\Delta_j] + o_p(\text{MISE}) .$$

By calculus it is readily verified that there is a $\delta > 0$ so that, for $y\in(-1,B')$,

$$\log(1+y)-y \le -\delta y^2 .$$

Thus, for h between n^{-1} and 1, on the event $\{\sup_{j=1,\ldots,n} |1_{[a,b]}(X_j)\Delta_j|<B'\} \cap U_n$,

$$\frac{1}{n}\log(\frac{\hat{L}(h)}{L}) \le -\delta n^{-1}\sum_{j=1}^{n} 1_{[a,b]}(X_j)\Delta_j^2 + o_p(\text{MISE}) .$$

Lemma 1.4 is now a consequence of Lemma 1.2 and the fact that $\hat{L}(h) = 0$ on the complement of U_n. This completes the proof of theorem 1.

4. PROOF OF THEOREM 2

This proof uses techniques developed by Chow, Geman and Wu (1983). The details of the proof here are quite different for two reasons. First, assumption (f.2) avoids many of the difficulties encountered by Chow, Geman and Wu. Second, complications arise here from allowing the kernel, K, to assume negative values.

Note that theorem 2 will be established when it is shown that, for any $h_o > 0$,

$$\sup_{h > h_o} \overline{\lim_n}\, n^{-1} \log \hat{L}(h) < \lim_{h \to 0} \underline{\lim_n}\, n^{-1} \log \hat{L}(h), \quad \text{a.s.} \tag{4.1}$$

As in Chow, Geman and Wu, measurability difficulties are avoided by assuming that the probability measure, P, is complete and noting that all statements are made with probability 0 or 1.

Given $\alpha > 0$, it will be convenient to define

$$\begin{aligned} f_h(x) &= E\hat{f}(x,h) = \int \tfrac{1}{h} K(\tfrac{x-y}{h}) f(y)\,dy = \int K(u) f(x-hu)\,du\ , \\ f_h^+(x) &= \max(f_h(x),\ 0)\ , \\ f_h^*(x) &= \max(f_h(x),\ \alpha)\ , \end{aligned} \tag{4.2}$$

and for j=1,...,n to define

$$\hat{f}_j^*(x,h) = \max(\hat{f}_j(x,h),\alpha)\ . \tag{4.3}$$

In the same spirit as (3.7) define

$$\begin{aligned} p_h &= \int_a^b f_h(x)\,dx, \\ p_h^* &= \int_a^b f_h^*(x)\,dx\ . \end{aligned} \tag{4.4}$$

Now given $h_o > 0$, define:

$$H_1 = \{h \geq h_o : \inf_{x\in[a,b]} f_h(x) \geq \alpha\} ,$$

(4.5) $$H_2 = \{h \geq h_o : -\alpha \leq \inf_{x\in[a,b]} f_h(x) < \alpha\} ,$$

$$H_3 = \{h \geq h_o : \inf_{x\in[a,b]} f_h(x) < -\alpha\} .$$

Note that (4.1), and hence theorem 2, is a consequence of the following lemmas.

<u>Lemma 2.1</u>: For α sufficiently small,

$$\lim_{h\to 0} \underline{\lim}_{n} n^{-1}\log \hat{L}(h) = \int_a^b f(x)\log f(x)dx-p \quad \text{a.s.} ,$$

where p was defined in (3.7).

<u>Lemma 2.2</u>

$$\sup_{h \in H_1} \overline{\lim}_{n} n^{-1}\log \hat{L}(h) < \int_a^b f(x)\log f(x)dx-p \quad \text{a.s.}$$

<u>Lemma 2.3</u>: For α sufficiently small,

$$\sup_{h \in H_2} \overline{\lim}_{n} n^{-1}\log \hat{L}(h) < \int_a^b f(x)\log f(x)dx-p \quad \text{a.s.}$$

<u>Lemma 2.4</u>

$$\sup_{h \in H_3} \overline{\lim}_{n} n^{-1}\log \hat{L}(h) < \int_a^b f(x)\log f(x)dx-p \quad \text{a.s.}$$

Before these lemmas are proved, three lemmas which will be useful in several of the proofs will be stated.

<u>Lemma 2.5</u>: Given $h_1 > 0$, as $n \to \infty$

$$\sup_{h > h_1} |n^{-1} \sum_{j=1}^{n} 1_{[a,b]}(X_j)\log \hat{f}_j^*(X_j,h) - \int_a^b f(x)\log f_h^*(x)dx| \to 0 \quad \text{a.s.}$$

<u>Lemma 2.6</u>: As $n \to \infty$,

$$\sup_{h > 0} |n^{-1} \sum_{j=1}^{n} \rho(X_j) - p_h| \to 0 \quad \text{a.s.}$$

Lemma 2.7: Given a measurable set S (on which f is bounded above 0) and an integrable, nonnegative function g(x),

$$\int_S f(x)\log g(x)dx - \int_S g(x)dx \le \int_S f(x)\log f(x)dx - \int_S f(x)dx,$$

with equality if and only if f(x) = g(x) a.e. on S. Furthermore, if there is a constant $\xi > 0$, so that

$$\int_S g(x)dx \le \int_S f(x)dx - \xi ,$$

then the above inequality may be sharpened to:

$$\int_S f(x)\log g(x)dx - \int_S g(x)dx \le \int_S f(x)\log f(x)dx - \int_S f(x)dx - \xi^2[2\int_S f(x)dx]^{-1}.$$

Since Lemmas 2.5, 2.6 and 2.7 are used in the proof of the other Lemmas, they will be proved first.

Proof of Lemma 2.5

First, for j=1,...,n define the empirical distribution functions

$$\hat{F}(x) = n^{-1}\sum_{i=1}^{n} 1_{(-\infty,X_i]}(x) ,$$

$$\hat{F}_j(x) = (n-1)^{-1}\sum_{i\neq j} 1_{(-\infty,X_i]}(x) ,$$

and define

$$K^*(y) = \frac{1}{h}K(\frac{X_j-y}{h}) .$$

Note that, by (K.3), K* is of bounded variation uniformly over $h > h_1$ and over j=1,...,n. Using (1.2) and integration by parts, for $h > h_1$ and j=1,...,n,

$$\begin{aligned}|\hat{f}_j(X_j,h) - f_h(X_j)| &= |\int\frac{1}{h}K(\frac{X_j-y}{h})d(\hat{F}_j-F)(y)| = \\ &= |\int[\hat{F}_j(y)-F(y)]dK^*(y)| \le \\ &\le \int|dK^*(y)|\cdot \sup_x |\hat{F}_j(x)-F(x)| \le \\ &\le \int|dK^*(y)|[n^{-1} + \sup_x|\hat{F}(x)-F(x)|].\end{aligned}$$

It follows from the above that

$$(4.6)\qquad \sup_{h > h_1}\ \sup_{j=1,\dots,n} |\hat{f}_j(X_j,h)-f_h(X_j)| \to 0 \quad \text{a.s.},$$

and hence that

$$\sup_{h > h_1} \sup_{j=1,\ldots,n} |\hat{f}_j^*(X_j,h)-f_h^*(X_j)| \to 0 \quad \text{a.s.},$$

and so,

$$(4.7) \qquad \sup_{h > h_1} |n^{-1} \sum_{j=1}^{n} 1_{[a,b]}(X_j) \log \hat{f}_j^*(X_j,h) - n^{-1} \sum_{j=1}^{n} 1_{[a,b]}(X_j) \log f_h^*(X_j)| \to 0 \quad \text{a.s.}$$

But now, since $f_h^*(x)$ is bounded above and below uniformly over $h > h_1$,

$$\sup_{h > h_1} |n^{-1} \sum_{j=1}^{n} 1_{[a,b]}(X_j) \log f_h^*(X_j) - \int_a^b f(x) \log f_h^*(x) dx| =$$

$$(4.8) \qquad = \sup_{h > h_1} |\int_a^b \log f_h^*(x) \; d(\hat{F}-F)(x)| \to 0 \quad \text{a.s.}$$

Lemma 2.5 follows from (4.7) and (4.8).

<u>Proof of Lemma 2.6</u>

Note that by (1.3), (4.2) and (4.4), for $h > 0$,

$$n^{-1} \sum_{j=1}^{n} \rho(X_j) - p_h = n^{-1} \sum_{j=1}^{n} \int_a^b \frac{1}{h} K(\frac{x-X_j}{h}) dx - \int_a^b [\int \frac{1}{h} K(\frac{x-y}{h}) f(y) dy] dx =$$

$$= n^{-1} \sum_{j=1}^{n} \int_{\frac{a-X_j}{h}}^{\frac{b-X_j}{h}} K(u) du - \int [\int_{\frac{a-y}{h}}^{\frac{b-y}{h}} K(u) du] f(y) dy \; .$$

But, by (K.2), K has a bounded antiderivative, $K^{(-1)}$.

Hence,

$$\sup_{h > 0} |n^{-1} \sum_{j=1}^{n} \rho(X_j) - p_h| =$$

$$= \sup_{h > 0} |\int [K^{(-1)}(\frac{b-y}{h}) - K^{(-1)}(\frac{a-y}{h})] d(\hat{F}-F)(y)| \to 0 \quad \text{a.s.}$$

Lemma 2.6 follows easily from this.

<u>Proof of Lemma 2.7</u>

It will be convenient to define

$$Q = \int_S f(x) dx \; ,$$

$$Q' = \int_S g(x) dx \; .$$

Note that by Jensen's Inequality,

$$\int_S \frac{f(x)}{Q} \log\left[\frac{Qg(x)}{Q'f(x)}\right] dx \le \log\left[\int_S \frac{g(x)}{Q'} dx\right] = 0 ,$$

with equality if and only if

$$g(x)/Q' = f(x)/Q \qquad \text{a.e. on } S .$$

Hence,

$$\int_S f(x)\log g(x)dx - \int_S f(x)\log f(x)dx \le Q \log Q' - Q \log Q . \tag{4.9}$$

By calculus it is easily verified that, for $x,y > 0$,

$$y \log(x/y) \le x-y ,$$

with equality if and only if $x=y$.

Thus,

$$Q \log Q' - Q \log Q \le Q' - Q .$$

From the above it follows that

$$\int_S f(x)\log g(x)dx - Q' \le \int_S f(x)\log f(x)dx - Q ,$$

with equality if and only if $g(x) = f(x)$ a.e. on S.

The second part of Lemma 2.7 is established similarly. First, note that for $x < y$,

$$y \log(x/y) \le (x-y) - (x-y)^2/2y .$$

Thus for $Q' \le Q - \xi$,

$$Q \log Q' - Q \log Q \le Q' - Q - \xi^2/2Q ,$$

And so, by (4.9) ,

$$\int_S f(x)\log g(x)dx - Q' \le \int_S f(x)\log f(x)dx - Q - \xi^2/2Q .$$

This completes the proof of lemma 2.7.

Proof of Lemma 2.1

First note that, by (f.3) and (4.2), for all $x \in \mathbb{R}$,

$$|f_h(x) - f(x)| = |\int K(u)[f(x-hu) - f(x)]du| \le$$

$$\le Mh^\gamma \int |u^\gamma K(u)| du .$$

Hence, by (K.2),

$$\lim_{h\to 0} \sup_x |f_h(x) - f(x)| = 0 , \tag{4.10}$$

and so,

$$\lim_{h\to 0} \sup_x |f_h^+(x)-f(x)| = 0.$$

It now follows from (f.1), (f.2), (3.7), (4.4) and the dominated convergence theorem that

$$\lim_{h\to 0}[\int_a^b f(x)\log f_h^+(x)dx-p_h] = \int_a^b f(x)\log f(x)dx-p \ . \tag{4.11}$$

Another consequence of (4.10) is that, by (f.2), for h and α sufficiently small,

$$\inf_{x\in[a,b]} f_h(x) \geq \alpha \ ,$$

and so, from (4.2), $f_h \equiv f_h^*$. In a similar spirit, note that by (3.3) and

(3.4), for α sufficiently small,

$$\lim_{h\to 0} \lim_{n\to\infty} \inf_{j=1,\ldots,n} \inf_x \hat{f}_j(x,h) \geq 2\alpha \ ,$$

and so, for h sufficiently small, n sufficiently large, j=1,...,n,

$$\hat{f}_j^+(X_j,h) = \hat{f}_j^*(X_j,h) \ . \tag{4.12}$$

Hence by Lemma 2.5,

$$\lim_{h\to 0} \overline{\lim_n} |n^{-1}\sum_{j=1}^n 1_{[a,b]}(X_j)\log \hat{f}_j^+(X_j,h) - \int_a^b f(x)\log f_h(x)dx| = 0 \quad \text{a.s.}$$

Lemma 2.1 follows from this, (1.4), (4.11) and Lemma 2.6.

<u>Proof of Lemma 2.2</u>

First define the function

$$G(h) = \int_a^b f(x)\log f_h^+(x)dx - p_h \ . \tag{4.13}$$

Lemma 2.2 is a consequence of the following lemmas.

<u>Lemma 2.2.1</u>: As $n \to \infty$,

$$\sup_{h\in H_1} |n^{-1}\log \hat{L}(h)-G(h)| \to 0 \quad \text{a.s.}$$

<u>Lemma 2.2.2</u>

$$\lim_{h\to\infty} G(h) = -\infty \ .$$

Lemma 2.2.3: $G(h)$ restricted to H_1 is continuous.

Lemma 2.2.4: Given $h_2 > h_o$, the set $\{h \le h_2 : h \in H_1\}$ is compact.

Lemma 2.2.5: For all $h \in H_1$,

$$G(h) < \int_a^b f(x) \log f(x) - p .$$

These lemmas will now be established.

Proof of Lemma 2.2.1

By the fact that on H_1, $f_h \equiv f_h^*$ and by (4.6), it follows from Lemma 2.5 that, as $n \to \infty$,

$$\sup_{h \in H_1} |n^{-1} \sum_{j=1}^{n} 1_{[a,b]}(X_j) \log \hat{f}_j^+(X_j,h) - \int_a^b f(x) \log f_h(x)dx| \to 0 \quad \text{a.s.}$$

Lemma 2.2.1 is a consequence of this together with (1.4) and Lemma 2.6.

Proof of Lemma 2.2.2

Note that by (K.2) and (4.2),

$$\lim_{h\to\infty} \sup_x |f_h(x)| = \lim_{h\to\infty} \sup_x |\int \tfrac{1}{h} K(\tfrac{x-y}{h}) f(y)dy| = 0.$$

and so

$$\lim_{h\to\infty} \sup_x \log f_h^+(x) = -\infty .$$

Lemma 2.2.2 follows from this and (4.13).

Proof of Lemma 2.2.3

Note that by (4.2), (4.4) and (4.5), for $h,h' \in H_1$,

$$|[\int_a^b f(x) \log f_h^+(x)dx - p_h] - [\int_a^b f(x) \log f_{h'}^+(x)dx - p_{h'}]| \le$$

$$\le \int_a^b f(x) |\log f_h(x) - \log f_{h'}(x)| dx + \int_a^b |f_h(x) - f_{h'}(x)| dx .$$

But, by (f.3),

$$|f_h(x) - f_{h'}(x)| \le \int |K(u)| \cdot |f(x-hu) - f(x-hu')| du \to 0,$$

uniformly as $h \to h'$. Lemma 2.2.3 now follows from (4.13) and the dominated convergence theorem.

Proof of Lemma 2.2.4

Lemma 2.2.4 is an obvious consequence of (4.5), the continuity of $f_h(x)$ and the fact that the given set is contained in $[h_o,h_2]$.

Proof of Lemma 2.2.5

Lemma 2.2.5 is Lemma 2.7 in the special case $S = [a,b]$ and $g(x) = f_h^+(x)$.

This completes the proof of Lemma 2.2.

Proof of Lemma 2.3

Note that by (4.2), Lemma 2.5, Lemma 2.6, (4.4) and (4.5),

$$\sup_{h \in H_2} \overline{\lim_n}\, n^{-1}\hat{L}(h) = \sup_{h \in H_2} \overline{\lim_n}\, n^{-1} \sum_{j=1}^{n} [1_{[a,b]}(X_j)\log \hat{f}_j^+(X_j,h)-\rho(X_j)] \leq$$

$$\leq \sup_{h \in H_2} \overline{\lim_n}\, n^{-1} \sum_{j=1}^{n} [1_{[a,b]}(X_j)\log \hat{f}_j^*(X_j,h)- \rho(X_j)] =$$

$$(4.14) \qquad = \sup_{h \in H_2} [\int_a^b f(x)\log f_h^*(x)dx-p_h] \leq$$

$$\leq \sup_{h \in H_2} [\int_b^a f(x)\log f_h^*(x)dx-p_h^* + 2\alpha(b-a)] \quad \text{a.s.}$$

Now from (f.2) define

$$\beta = \inf_{x\in [a,b]} f(x) .$$

Given $h \in H_2$, define the set

$$V = \{x\in [a,b]: f_h^*(x) < 2\beta/3\} .$$

Note that, by (f.3) and (4.2), there is a constant $M' > 0$ so that for all $x,y \in \mathbb{R}$,

$$(4.15) \qquad |f_h(x) - f_h(y)| \leq M'|x-y|^\gamma ,$$

and hence,

$$|f_h^*(x) - f_h^*(y)| \leq M'|x-y|^\gamma .$$

Now take $\alpha \leq \beta/3$. If μ denotes Lebesgue measure, then it follows from the above that, for $h \in H_2$, (4.16) $$\mu(V) \geq 2\left(\frac{\beta}{3M'}\right)^{1/\gamma} .$$

Next let $V^c = [a,b]\backslash V$, and note that by Lemma 2.7,

(4.17) $$\int_{V^c} f(x)\log f_h^*(x)dx - \int_{V^c} f_h^*(x)dx \leq$$

$$\leq \int_{V^c} f(x)\log f(x)dx - \int_{V^c} f(x)dx .$$

Similarly, since

$$\int_V f_h^*(x)dx \leq \frac{2\beta}{3}\mu(V) = \beta\,\mu(V) - \frac{\beta}{3}\mu(V) \leq$$

$$\leq \int_V f(x)dx - \frac{\beta}{3}\mu(V) \quad ,$$

it follows from Lemma 2.7 that

$$\int_V f(x)\log f_h^*(x)dx - \int_V f_h^*(x)dx \leq$$

$$\leq \int_V f(x)\log f(x)dx - \int_V f(x)dx - \psi ,$$

for some constant, $\psi > 0$. Putting this together with (4.17) yields

$$\int_a^b f(x)\log f_h^*(x)dx - p_h^* \leq \int_a^b f(x)\log f(x) - p - \psi .$$

Lemma 2.3 follows from this together with (4.14) and (4.16).

Proof of Lemma 2.4

Given $h \in H_3$, let

$$V' = \{x \in [a,b]: f_h(x) \leq -\alpha/2\} \quad .$$

Note that, by (4.15)

$$\mu(V') \geq 2\left(\frac{\alpha}{2M'}\right)^{1/\gamma} .$$

It follows from (f.2) that for $h \in H_3$,

$$P[\{X_j \in V'\}\text{i.o.}] = 1,$$

where "i.o." means infinitely often. Hence

$$P[\{X_j \in [a,b] \text{ and } f_h(X_j) \leq -\alpha/2\}\text{i.o.}] = 1,$$

and so by (4.6), for $h \in H_3$,

$$P[\{X_j \in [a,b] \text{ and } \hat{f}_j^+(X_j,h) = 0\}\text{i.o.}] = 1$$

Lemma 2.4 follows easily from this and (1.4).

REFERENCES

Cheng, S.H. (1983). On a problem concerning spacings. Center for stochastic processes, technical report # 27.

Chow, Y.S., Geman, S. and Wu, L.D. (1983). Consistent cross-validated density estimation. Ann. Statist. 11, 25-38.

Habbema, J.D.F., Hermans, J. and van den Broek, K. (1974). A stepwise discrimination analysis program using density estimation. Compstat 1974: Proceedings in computational statistics. (G. Bruckman, ed.) 101-110, Vienna:Physica Verlag.

Marron, J.S. (1983a). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. North Carolina Institute of Statistics, Mimeo Series #1518.

Marron, J.S. (1983b). Convergence properties of an empirical error criterion for multivarite density estimation. North Carolina Institute of Statistics, Mimeo Series #1520.

Rosenblatt, M. (1971). Curve estimates. Ann. Math. Statist., 42, 1815-1842.

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