Statistics & Probability Letters 8 (1989) 425–433
North-Holland
October 1989

# ON THE NON-CONSISTENCY OF THE $L_2$-CROSS-VALIDATED KERNEL DENSITY ESTIMATE

Luc DEVROYE

*School of Computer Science, McGill University, Montreal, Canada H3A 2K6*

Received September 1988
Revised December 1988

*Abstract*: Let $f_n$ be the $L_2$ cross-validated kernel estimate of a univariate density $f$. We show that

$$\liminf_{n\to\infty} E\int|f_n-f|\geqslant 1$$

when $K$ is a symmetric bounded unimodal density, $f$ is a monotone density on $[0,\infty)$ and $x^{3/4}f(x)\to\infty$ as $x\downarrow 0$.

*AMS 1980 Subject Classifications*: 62G05, 62E20.

*Keywords*: density estimation, consistency, cross-validated choice, data-based smoothing, kernel estimate.

## 1. Introduction

When a certain density estimate forms part of a software package, it is usually there because of some desirable features. The kernel density estimate

$$f_n(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\triangleq\frac{1}{n}\sum_{i=1}^{n}K_h(x-X_i)$$

(where $X_1,\ldots,X_n$ are i.i.d. observations drawn from an unknown density $f$, $K$ is a fixed density called the *kernel*, and $h>0$ is the *smoothing factor*; see Parzen, 1962, or Rosenblatt, 1956) is thus eligible for inclusion if we can find acceptable ways of selecting both $K$ and $h$. The criteria used in such selections are related to the closeness of $f_n$ to $f$ in some general sense, where it is often assumed that $f$ belongs to a certain class of "nice" densities. Many choices guarantee good or even optimal asymptotic behavior with respect to some measure of closeness of $f_n$ to $f$. Although the choice of $K$ seems vitally important, especially if we allow functions $K$ with negative values (while still $\int K=1$), most methods fix $K$ beforehand, and specify $h$ as a function of the data. Partial surveys can be found in Devroye and Györfi (1985, Chapter 6) and Silverman (1986).

Any method for choosing $h$ is necessarily non-consistent (i.e., $\int|f_n-f|$ does not tend to zero in probability for some $f$) if either $h$ does not tend to zero in probability or $nh$ does not tend to zero in probability (Devroye, 1987; Broniatowski, Devroye and Deheuvels, 1989). And indeed, some methods are easily seen to be non-consistent via this device. Of course, one could always guard against such mishaps by defining two deterministic sequences $\alpha_n$ and $\beta_n$ with $n\alpha_n\to\infty$, $\beta_n\to 0$ and $0<\alpha_n\leqslant\beta_n<\infty$, and truncate the random variable $h$ to the interval $[\alpha_n, \beta_n]$. But this is hardly acceptable in a universally applicable

Research of the author was funded by NSERC grant A3456.

0167-7152/89/$3.50 

piece of software, since we can not specify universal choices for $\alpha_n$ and $\beta_n$. Without the truncation safeguard, many methods of choosing $h$ lead to non-consistency. We have e.g.:

(A) The two-step methods, in which $f$ is estimated by $\hat{f}$ in some manner, and $\hat{f}$ is used in the formula $h_{\text{opt}}(\hat{f}, n)$, where $h_{\text{opt}}(f, n)$ is the asymptotically optimal choice for $h$ for a given $f$ and $n$. (See e.g. Nadaraya, 1974; Watson and Leadbetter, 1963; Woodroofe, 1970; Scott and Factor, 1981; Bretagnolle and Huber, 1979; and Hall and Wand, 1987.) Most are non-consistent for certain $f$.

(B) The maximum likelihood cross-validation method, in which $h$ is selected to maximize

$$\prod_{i=1}^{n} f_{ni}(X_i)$$

and $f_{ni}(X_i)$ is the kernel estimate of $f$ based upon $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ (Duin, 1976; Habbema, Hermans and Vandenbroek, 1974). For consistency considerations, see Chow, Geman and Wu (1983). The maximum likelihood cross-validation method is non-consistent when the extreme values of the sample drawn from $f$ are not stable, i.e. loosely speaking, when the tails of $f$ are at least as big as the tails of some Laplace density (Schuster and Gregory, 1981; Devroye and Györfi, 1985), and this condition is virtually necessary and sufficient (Broniatowski, Devroye and Deheuvels, 1989).

(C) The $L_1$ minimum distance method (Devroye, 1987), in which $h$ is selected so as to minimize $\int |f_n - g_n|$, where $g_n$ is the kernel estimate based upon the same data and with the same $h$ as $f_n$, and $g_n$ uses a higher order kernel $K^*$. This method is always consistent.

(D) The $L_2$ cross-validated choice, in which $h$ is selected to minimize

$$\int f_n^2 - \frac{2}{n} \sum_{i=1}^{n} f_{ni}(X_i)$$

(Rudemo, 1982; Bowman, 1982, 1984; see also Hall, 1983; Marron, 1987; Hall and Marron, 1987; Scott and Terrell, 1986). This method is asymptotically optimal for all bounded $f$ and all bounded compact support kernels $K$ (Stone, 1984), i.e.

$$\int (f_n - f)^2 \sim \inf_h \int (f_{nh} - f)^2 \quad \text{almost surely},$$

where $f_{nh}$ is the kernel density estimate with deterministic $h$. Also, it seems to be consistent whenever $\int f^2 < \infty$, but this won't be shown here.

In this note, we point out that the $L_2$ cross-validation method is non-consistent for some densities not in $L_2$, and that the non-consistency is due to the presence of one or more large infinite peaks, forcing $h$ to be so small that $nh \to 0$ in probability. However, we stop short of showing that the method is non-consistent when $\int f^2 = \infty$. Because the presence of big infinite peaks can not be checked beforehand, it seems necessary to modify the $L_2$ cross-validation method before its inclusion in a software package. For finite sample size, the $L_2$ cross-validated choice is probably much too small whenever the data show clustering around one or more points.

**Theorem.** *If $K$ is a symmetric bounded unimodal density, $f$ is a monotone density on $[0, \infty)$ and $x^{3/4} f(x) \to \infty$ as $x \downarrow 0$, then for the $L_2$ cross-validated kernel estimate,*

$$\liminf_{n \to \infty} E \int |f_n - f| \geq 1.$$

The $L_2$ cross-validation method is not designed to give good $L_1$ performance or even to assure $L_1$ consistency. Yet, it was considered as much more robust and reliable than most other methods, especially the maximum likelihood cross-validation method. The Theorem above shows that for a subclass of densities not in $L_2$, the method is not even consistent. For the class of counterexamples of the theorem, specially selected to provide some insight into the processes at work, we will show that $h = 1/n^2$, a ridiculously small smoothing parameter under any circumstances, yields a smaller value for the expression to be minimized than any $h$ in any interval of the form $[\varepsilon/n, 1/\varepsilon]$ when $n$ is large enough and $\varepsilon > 0$ is held constant. The same statement can be made for $h = 1/n^p$ for any real number $p \geqslant 2$, provided that we suitably restrict the class of densities by requiring that $x^b f(x) \to \infty$ as $x \downarrow 0$, where $b$ is a constant depending upon $p$ only. Let us also note that the Theorem covers all modes of convergence, since convergence in the $L_1$ sense is equivalent to $h \to 0$ and $nh \to \infty$, so that under the conditions of the theorem, we can't have consistency in the $L_p$ sense for any $p \in (0, \infty]$.

The remainder of the note forms the proof of the theorem. We take the liberty of cutting the proof up into many sections and to divide the work ahead into many small manageable pieces.

## 2. Two useful lemmas

**Lemma 1.** *Assume that K is an absolutely integrable function, and that $f_{nh}$ is the kernel estimate with kernel K and smoothing factor h. For every f, there exists a continuous function $\omega \geqslant 0$ with $\omega(0) = 0$, such that, deterministically,*

$$\inf_{nh \leqslant u} \int | f_{nh} - f | \geqslant 1 - \omega(u).$$

**Lemma 2.** *Assume that K is a symmetric unimodal density, and that $f_{nh}$ is the kernel estimate with kernel K and smoothing factor h. For every f, there exists a continuous function $\omega \geqslant 0$ with $\omega(0) = 0$, such that, deterministically,*

$$\inf_{h \geqslant u} \int | f_{nh} - f | \geqslant 1 - \omega\left(\frac{1}{u}\right).$$

In both lemmas, the functions $\omega$ depend both upon the (fixed) kernel $K$ and the density $f$.

**Proof of Lemma 1.** Take $\varepsilon > 0$ arbitrary. Find $\delta > 0$ such that $\int_{|y| > \delta} |K| \leqslant \varepsilon$. Let $A = \bigcup_{i=1}^{n} [X_i - \delta h, X_i + \delta h]$. We have

$$\int | f_{nh} - f | \geqslant \int_{A^c} | f_{nh} - f | \geqslant \int_{A^c} f - \int_{A^c} | f_{nh} | \geqslant \int_{A^c} f - \varepsilon.$$

Since $\int_A \mathrm{d}x \leqslant 2n\delta h \leqslant 2\delta u$ for $nh \leqslant u$, we know that $\int_A f \leqslant \varepsilon$ for $2\delta u$ small enough. For such $u$, we conclude that $\int | f_{nh} - f | \geqslant 1 - 2\varepsilon$. □

**Proof of Lemma 2.** Let $\delta > 0$ be arbitrary, and choose $\alpha > 0$ such that $\int_{-\alpha}^{\alpha} f \geqslant 1 - \delta$, and find $M$ so large that $\int_{-\alpha/M}^{\alpha/M} K \leqslant \delta$. For $h \geqslant M$, we then have

$$\int_{-\alpha}^{\alpha} K_h(x - X_i) \, \mathrm{d}x \leqslant \sup_y \int_{y-\alpha}^{y+\alpha} K_h(x) \, \mathrm{d}x = \int_{-\alpha}^{\alpha} K(x) \, \mathrm{d}x = \int_{-\alpha/h}^{\alpha/h} K(x) \, \mathrm{d}x$$
$$\leqslant \int_{-\alpha/M}^{\alpha/M} K(x) \, \mathrm{d}x \leqslant \delta.$$

Thus, $\int_{-\alpha}^{\alpha} f_{nh} \leqslant \delta$, and thus $\int | f_{nh} - f | \geqslant 1 - 2\delta$. □

### 3. The criterion that is minimized

We choose $h$ such as to minimize

$$C_n(h) \triangleq \int f_n^2 - \frac{2}{n} \sum_{i=1}^{n} f_{ni}(X_i).$$

It is useful to rewrite this criterion in different ways, in order to identify its active components. First, we begin by defining a kernel obtained from $K$ by splicing (see Stuetzle and Mittal, 1979),

$$L(z) = 2K(z) - \int K(u)K(z-u)\,\mathrm{d}u = 2K(z) - K*K(z).$$

Also, we define

$$U_n(h) \triangleq \frac{1}{n(n-1)} \sum_{i \neq j} L_h(X_i - X_j),$$

$$V_n(h) \triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \left(EL_h(X_i - X_j) - L_h(X_i - X_j)\right),$$

and

$$W_n(h) \triangleq \int K^2/(nh) + \left(\frac{1}{n^2} - \frac{1}{n(n-1)}\right) \sum_{i \neq j} \int K_h(x - X_i) K_h(x - X_j)\,\mathrm{d}x.$$

We are now ready for:

**Lemma 3.** *We have the following representations, when $K$ is a symmetric square integrable kernel*:

$$C_n(h) = W_n(h) - U_n(h) = W_n(h) + V_n(h) - EL_h(X_1 - X_2).$$

*When $K \geqslant 0$, we have*

$$0 \leqslant W_n(h) \leqslant \int K^2/(nh);$$

*also*, $|U_n(h)| \leqslant \|L\|_\infty/h$.

**Proof.** We note that

$$C_n(h) = \int \left(\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)\right)^2 - \frac{2}{n} \sum_{i=1}^{n} \frac{1}{n-1} \sum_{j \neq i} K_h(X_i - X_j)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \int (K_h(x - X_i))^2 + \frac{1}{n^2} \sum_{i \neq j} \int K_h(x - X_i) K_h(x - X_j)\,\mathrm{d}x$$

$$- \frac{2}{n(n-1)} \sum_{j \neq i} K_h(X_i - X_j)$$

$$= W_n(h) - U_n(h) = W_n(h) + V_n(h) - EL_h(X_1 - X_2).$$

The upper bound for $U_n(h)$ is rather obvious. The upper bound for $W_n(h)$ for nonnegative $K$ is also trivial. Finally,

$$-W_n(h) \leqslant -\int K^2/(nh) + \frac{1}{n^2(n-1)} \sum_{i \neq j} \int K_h(x - X_i) K_h(x - X_j) \, dx$$

$$\leqslant \int K^2/(nh) + \frac{1}{n} \sup_z K_h^* K_h(z) \leqslant -\int K^2/(nh) + \frac{1}{n} \int (K_h)^2 = 0.$$

Here we used the Cauchy-Schwarz inequality in the observation that for an arbitrary density $g$, $\sup_z g^*g(z) \leqslant \int g^2$. □

## 4. Proof of the theorem

For fixed $\delta > 0$, we can find $\varepsilon > 0$ such that

$$\inf_{h \geqslant 1/\varepsilon \text{ or } h \leqslant \varepsilon/n} \int | f_{nh} - f | > 1 - \delta$$

(use Lemmas 1 and 2). Let $h^*$ be $h$ truncated to the interval $A \triangleq [\varepsilon/n, 1/\varepsilon]$. Then,

$$\int | f_{nh} - f | \geqslant (1 - \delta) I_{h \notin A} + \int | f_{nh^*} - f | I_{h \in A}.$$

Thus,

$$E \int | f_{nh} - f | \geqslant (1 - \delta) P(h \notin A).$$

The proof is complete if we can show that

$$P(h \in A) \to 0,$$

which in turn follows if we can show that

$$P\Big(C_n(1/n^2) \geqslant \inf_{h \in A} C_n(h)\Big) \to 0.$$

But, by Lemma 3,

$$\inf_{h \in A} C_n(h) \geqslant -2 \| L \|_\infty n/\varepsilon.$$

Also, $W_n(1/n^2) \leqslant n \int K^2$ (Lemma 3). Thus,

$$C_n(1/n^2) - \inf_{h \in A} C_n(h) \leqslant n\left(2 \| L \|_\infty + \int K^2\right)/\varepsilon + V_n(1/n^2) - EL_{1/n^2}(X_1 - X_2).$$

Hence, the Theorem is proved if we can shown:
(i) $\sqrt{h}\, EL_h(X_1 - X_2) \to \infty$ as $h \downarrow 0$.
(ii) For all $\theta > 0$, and $h = 1/n^2$, $P(V_n(h) > \theta EL_h(X_1 - X_2)) \to 0$.
Assume for the sake of the argument that there exists a function $G$ such that $\sqrt{u}\, G(u) \to \infty$ as $u \downarrow 0$, and that there are finite positive constants $c_1$ and $c_2$ such that

$$c_1 G(h) \leqslant EL_h(X_1 - X_2) \leqslant E| L_h(X_1 - X_2) | \leqslant c_2 G(h).$$

Then, obviously, condition (i) is fulfilled. To verify condition (ii), we apply Chebyshev's inequality:

$$P(V_n(h) > \theta EL_h(X_1 - X_2)) \leqslant E\big((V_n(h))^2\big)/\big(\theta^2 E^2 L_h(X_1 - X_2)\big)$$
$$\leqslant E\big((V_n(h))^2\big)/\big(\theta^2 c_1^2 G^2(h)\big).$$

We note that

$$V_n(h) = \frac{1}{n(n-1)} \sum_{i \neq j} Y_{ij},$$

where $Y_{ij} = L_h(X_i - X_j) - EL_h(X_1 - X_2)$. Thus, since $E(Y_{ij}Y_{kl}) = 0$ when $\{i, j\} \cap \{k, l\}$ is empty, we have

$$E\big((V_n(h))^2\big) \leqslant \mathrm{O}(n^{-4})\big(\mathrm{O}(n^2)EY_{12}^2 + \mathrm{O}(n^3)E\,|Y_{12}|\,|Y_{13}|\big)$$
$$\leqslant \mathrm{O}(1/n)EY_{12}^2 \leqslant \mathrm{O}(1/n)E\big((L_h(X_1 - X_2))^2\big)$$
$$\leqslant \mathrm{O}(1/(nh))E|L_h(X_1 - X_2)| \leqslant \mathrm{O}(1/(nh))c_2 G(h).$$

In summary,

$$P(V_n(h) > \theta EL_n(X_1 - X_2)) \leqslant O(1/(nh))c_2/\big(\theta^2 c_1^2 G(h)\big),$$

which tends to zero when $h = 1/n^2$ and $\sqrt{h}\,G(h) \to \infty$ as $h \downarrow 0$. This concludes the verification of condition (ii). The proof of the theorem thus requires the verification of the existence of such a growth function $G$. This is accomplished in the next section. □

## 5. Some properties of the class of counterexamples

The existence of a suitable growth function $G$ for our class of counterexamples is established in a series of lemmas.

**Lemma 4.** *Assume the $X_1$ has a monotone $\downarrow$ density $f$ on $[0, \infty)$, and that $x^{3/4}f(x) \to \infty$ as $x \downarrow 0$. Then the density $g$ of $X_1 - X_2$ is symmetric, unimodal at the origin and $\sqrt{h}\,g(h) \to \infty$ as $h \to 0$.*

**Proof.** For $x > 0$, we have

$$g(x) = \int_0^\infty f(y)f(x+y)\,\mathrm{d}y,$$

which is obviously monotone $\downarrow$ in $x$. Furthermore,

$$\sqrt{x}\,g(x) = \int_0^\infty \sqrt{x} f(y)f(x+y)\,\mathrm{d}y \geqslant \int_x^{2x} \sqrt{x} f^2(2x)\,\mathrm{d}y \geqslant x^{3/2}f^2(2x) \to \infty$$

as $x \downarrow 0$. □

**Lemma 5.** *Let $K$ be a symmetric unimodal density, and define $L \triangleq 2K - K*K$, where "$*$" is the convolution operator. Then, for all $u > 0$,*

$$\int_0^u (2K(x) - K*K(x))\,\mathrm{d}x \geqslant \int_0^u K(x)\,\mathrm{d}x.$$

**Proof.** Let $X$, $X'$ be i.i.d. random variables with density $K$. Then, we need to show that for all $u > 0$, $\int_0^u K \geqslant \int_0^u K*K$, or, in other words, $P(0 \leqslant X \leqslant u) \geqslant P(0 \leqslant X + X' \leqslant u)$. This is easily seen to be the case by a geometric argument in the plane, by considering the joint (product) density of $(X, X')$. We have

$$\begin{aligned} P(X \in [0, u]) &= P((X, X') \in [0, u] \times [-u, u]) + \int_0^u \int_u^\infty K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x \\ &\quad + \int_0^u \int_{-\infty}^{-u} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x \\ &\geqslant \int_{-u}^u \int_{-y}^{u-y} K(x)\,\mathrm{d}x\,K(y)\,\mathrm{d}y + \int_u^\infty \int_{-y}^{u-y} K(x)\,\mathrm{d}x\,K(y)\,\mathrm{d}y \\ &\quad + \int_{-\infty}^{-u} \int_{-y}^{u-y} K(x)\,\mathrm{d}x\,K(y)\,\mathrm{d}y \\ &= P(X + X' \in [0, u]). \end{aligned}$$

The minorization is done on a termwise basis. The second and third terms are minorized after observing that for fixed $|y| \geqslant u$, $\int_0^u K(x)\,\mathrm{d}x \geqslant \int_{-y}^{u-y} K(x)\,\mathrm{d}x$, with equality occurring if and only if $y = u$. The first term is minorized by noting that

$$\begin{aligned} &\int_0^u \int_{-u}^u K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x \\ &= \int_0^u \int_{-x}^{u-x} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x + \int_0^u \int_{-u}^{-x} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x + \int_0^u \int_{u-x}^u K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x \\ &= \int_0^u \int_{-x}^{u-x} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x + \int_{-u}^0 \int_{-x}^u K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x + \int_0^u \int_{-u}^{x-u} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x \\ &\geqslant \int_0^u \int_{-x}^{u-x} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x + \int_{-u}^0 \int_{-x}^u K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x + \int_u^{2u} \int_{-u}^{u-x} K(y)\,\mathrm{d}y\,K(x)\,\mathrm{d}x \\ &= \int_{-u}^u \int_{-y}^{u-y} K(x)\,\mathrm{d}x\,K(y)\,\mathrm{d}y. \qquad \square \end{aligned}$$

**Lemma 6.** *Assume that $X_1$ has a monotone $\downarrow$ density $f$ on $[0, \infty)$, and that $x^{3/4}f(x) \to \infty$ as $x \downarrow 0$. Let $K$ be a bounded symmetric unimodal density, and define the function $L = 2K - K*K$. Then there exists a function $G$ such that $\sqrt{u}\,G(u) \to \infty$ as $u \downarrow 0$, and there are finite positive constants $c_1$ and $c_2$ such that*

$$c_1 G(h) \leqslant EL_h(X_1 - X_2) \leqslant E|L_h(X_1 - x_2)| \leqslant c_2 G(h), \quad \text{all } h \in (0, 1].$$

*It suffices to take $G(u) \triangleq (1/u)\int_0^u g(x)\,\mathrm{d}x$, where $g$ is the density of $X_1 - X_2$.*

**Proof.** From Lemma 4, we recall that the density $g$ of $X_1 - X_2$ is symmetric, unimodal at the origin, and that $\sqrt{h}\,g(h) \to \infty$ as $h \to 0$. By monotonicity, we have $G(u) \geqslant g(u)$ and $G(u) \uparrow \infty$ as $u \downarrow 0$.

The Lemma follows from these facts and the following inequalities that will be obtained below:

$$\begin{aligned} 2bK(b)G(bh) &\leqslant \int_{-\infty}^\infty K_h(x)g(x)\,\mathrm{d}x \leqslant \int_{-\infty}^\infty L_h(x)g(x)\,\mathrm{d}x \\ &\leqslant \int_{-\infty}^\infty |L_h(x)|g(x)\,\mathrm{d}x \leqslant 2\left(b\|L\|_\infty + \int_b^\infty |L|\right)G(bh), \end{aligned}$$

where $b$ is selected such that $K(b) > 0$. The four inequalities in this chain will be called I, II, III and IV. Of these, III is trivial. Inequality I is seen to hold as follows:

$$\int_{-\infty}^\infty K_h(x)g(x)\,\mathrm{d}x \geqslant K_h(bh)\int_{-bh}^{bh} g(x)\,\mathrm{d}x = 2K_h(bh)bhG(bh) = 2K(b)bG(bh).$$

Inequality IV is obtained easily:

$$\int_{-\infty}^{\infty} |L_h(x)| g(x)\, dx \leqslant \int_{-bh}^{bh} \frac{\|L\|_\infty}{h} g(x)\, dx + 2g(bh)\int_{bh}^{\infty} |L_h(x)|\, dx$$

$$= \frac{2bhG(bh)\|L\|_\infty}{h} + 2g(bh)\int_b^\infty |L| \leqslant 2G(bh)\Big(b\|L\|_\infty + \int_b^\infty |L|\Big).$$

To obtain inequality II, we use Lemma 5. $2g$ is the density of $|X_1 - X_2|$ on $[0, \infty)$. By Khinchine's theorem for unimodal densities, $2g$ is also the density of a random variable $UZ$, where $U$ has the uniform [0, 1] density, and $Z \geqslant 0$ is independent of $U$. Exploiting the symmetry and unimodality, we thus have

$$2\int_0^\infty L_h(x) g(x)\, dx = 2\int_0^\infty L(z) g(zh)\, dz = \frac{1}{h}\int_0^\infty L(z)(2hg(zh))\, dz$$

$$= E\Big(\frac{1}{h}L(UZ)\Big) = E\frac{1}{h}\int_0^1 L(uZ)\, du + E\frac{1}{h}\int_0^Z L(v)\, dv$$

$$\geqslant E\frac{1}{h}\int_0^Z K(v)\, dv = 2\int_0^\infty K_h(x) g(x)\, dx,$$

which was to be shown. □

## References

Bowman, A.W. (1982), A comparative study of some kernel-based non-parametric density estimators, *J. Statist. Comput. Simulation* **21**, 313–327.

Bowman, A.W. (1984), An alternative method of cross-validation for the smoothing of density estimates, *Biometrika* **71**, 353–360.

Bretagnolle, J. and C. Huber (1979), Estimation des densités: risque minimax, *Z. Wahrsch. Verw. Gebiete* **47**, 119–137.

Broniatowski, P. Deheuvels and L. Devroye (1989), On the relationship between stability of extreme order statistics and convergence of the maximum likelihood kernel density estimate, *Ann. Statist.* **17**.

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

Devroye, L. (1987), An $L_1$ asymptotically optimal kernel estimate, to appear.

Devroye, L. (1987), *A Course in Density Estimation* (Birkhauser, Boston, MA).

Devroye, L. and L. Györfi (1985), *Nonparametric Density Estimation: The $L_1$ View* (Wiley, New York).

Duin, R.P.W. (1976) On the choice of smoothing parameters for Parzen estimators of probability density functions, *IEEE Trans. Comput.* **C-25**, 1175–1179.

Habbema, J.D.F., J. Hermans and K. Vandenbroek (1974), A stepwise discriminant analysis program using density estimation, in: G. Bruckmann, ed., *Compstat 1974* (Physica Verlag, Wien) pp. 101–110.

Hall, P. (1983), Large-sample optimality of least squares cross-validation in density estimation, *Ann. Statist.* **11**, 1156–1174.

Hall, P. and J.S. Marron (1987), Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation, *Probab. Theory Rel. Fields* **74**, 567–581.

Hall, P. and M.P. Wand (1988), Minimizing $L_1$ distance in nonparametric density estimation, *J. Multivariate Anal.*

Marron, J.S. (1987), A comparison of cross-validation techniques in density estimation, *Ann. Statist.* **15**, 152–162.

Nadaraya, E.A. (1974), On the integral mean square error of some nonparametric estimates for the density function, *Theory Probab. Appl.* **19**, 133–141.

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

Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* **27**, 832–837.

Rudemo, M. (1982), Empirical choice of histograms and kernel density estimators, *Scand. J. Statist.* **9**, 65–78.

Schuster, E.F. and G.G. Gregory (1981), On the nonconsistency of maximum likelihood nonparametric density estimators, in: W.F. Eddy, ed., *Computer Science and Statistics: Proc. 13th Symp. Interface* (Springer, New York) pp. 295–298.

Scott, D.W. and L.E. Factor (1981), Monte Carlo study of three data-based nonparametric probability density estimators, *J. Amer. Statist. Assoc.* **76**, 9–15.

Scott, D.W. and G.R. Terrell (1986), Biased and unbiased cross-validation in density estimation, Tech. Rept. 23, Department of Statistics, Stanford University (Stanford, CA).

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London).

Stone, C.J. (1984), An asymptotically optimal window selection rule for kernel density estimates, *Ann. Statist.* **12**, 1285–1297.

Stuetzle, W. and Y. Mittal (1979), Some comments on the asymptotic behavior of robust smoothers, in: T. Gasser and M. Rosenblatt, eds., *Proc. Heidelberg Workshop* (Springer, Heidelberg) pp. 191–195.

Watson, G.S. and M.R. Leadbetter (1963), On the estimation of the probability density, *Ann. Math. Statist.* **34**, 480–491.

Woodroofe, M. (1970), On choosing a delta sequence, *Ann. Math. Statist.* **41**, 1665–1671.