Z. Wahrscheinlichkeitstheorie verw. Gebiete
57, 453-476 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

On the Histogram as a Density Estimator: L_2 Theory

David Freedman[1]* and Persi Diaconis[2]**

[1] Statistics Department, University of California, Berkeley, CA94720, USA
[2] Statistics Department, Stanford University, Stanford, CA94305, USA

1. Introduction

Let f be a probability density on an interval I, finite or infinite: I includes its finite endpoints, if any; and f vanishes outside of I. Let $X_1, \ldots, X_k$ be independent random variables, with common density f. The empirical histogram for the X's is often used to estimate f. To define this object, choose a reference point $x_0 \in I$ and a cell width h. Let N_j be the number of X's falling in the jth class interval:

$$[x_0 + j h, x_0 + (j+1) h).$$

On this interval the height of the histogram $H(x)$ is defined as

$$N_j / k h.$$

This definition forces the area under H to be 1. The dependence of H on k and h is suppressed in the notation.

On the average, how close does H come to f? A standard measure of discrepancy is the mean square difference:

$$\delta^2 = E\{\int_I [H(x) - f(x)]^2 \, dx\}. \tag{1.1}$$

This quantity is analyzed on the following assumptions:

(1.2) $f \in L_2$ and f is absolutely continuous on I, with a.e. derivate f'

(1.3) $f' \in L_2$ and f' is absolutely continuous on I, with a.e. derivative f''

(1.4) $f'' \in L_p$ for some p with $1 \leqq p \leqq 2$.

* Research partially supported by NSF Grant MCS-80-02535
** Research partially supported by NSF Grant MCS-80-24649

0044-3719/81/0057/0453/$04.80

Conditions (1.3) and (1.4) have the (non-obvious) consequence that f' is continuous and vanishes at ∞. In particular, f' is bounded; see (2.21) below. Also, f' is in fact the ordinary (everywhere) derivative of f. Likewise, f is continuous and vanishes at ∞. It will also be assumed that

(1.5) I is the union of class intervals.

For instance, if $I=[0,1]$ and $x_0=0$, condition (1.5) requires that $h=1/N$ for some positive integer N. By present conditions, if $I=[0,1]$, then f and f' are continuous on I, even at 0 and 1.

(1.6) **Theorem.** *Assume* (1.1–1.5). *Let*

$$\gamma=\int_I f'(x)^2\,dx>0$$

$$\beta=\tfrac{1}{4}\cdot 6^{2/3}\cdot\gamma^{1/3}$$

$$\alpha=6^{1/3}\,\gamma^{-1/3}.$$

Then, the cell width h which minimizes the δ^2 *of* (1.1) *is* $\alpha k^{-1/3}+O(k^{-1/2})$, *and at such h's,* $\delta^2=\beta k^{-2/3}+O(k^{-1})$.

The technique developed to prove (1.6) can be used to give a result under weaker conditions.

(1.7) **Theorem.** *Suppose* $f\in L_2$ *is absolutely continuous with a.e. derivative* $f'\in L_2$ *and* $\int f'(x)^2\,dx>0$. *Suppose* (1.5) *as well. Define* α *and* β *as in* (1.6). *Then the cell width which minimizes the* δ^2 *of* (1.1) *is* $\alpha k^{-1/3}+o(k^{-1/3})$ *and at such h's,* $\delta^2=\beta k^{-2/3}+o(k^{-2/3})$.

Such results suggest that the discrepancy δ^2 can be made small by choosing the cell width h as $\alpha k^{-1/3}$. Of course, this depends on γ, which will be unknown in general cases. In principle, γ can be estimated from the data, as in Woodroofe (1968). However, numerical computations, which will be reported elsewhere, suggest that the following simple, robust rule for choosing the cell width h often gives quite reasonable results.

(1.8) *Rule.* Choose the cell width as twice the interquartile range of the data, divided by the cube root of the sample size.

Rough versions of (1.6) and (1.7) seem part of the folklore. Two recent references providing formal computations are Tapia and Thompson (1978), and Scott (1979).

We hope to study the random variable $\Delta^2=\int[H(x)-f(x)]^2\,dx$ in a future paper. The standard deviation of Δ^2 is of smaller order than $E(\Delta^2)=\delta^2$. Thus, choosing h to minimize δ^2 is a sensible way to get a small Δ^2. To be a bit more precise, the standard deviation of Δ^2 is of order $k^{-1}h^{-1/2}\sim k^{-5/6}$ for the optimal $h\sim k^{-1/3}$. Using (1.6), the minimal discrepancy Δ^2 is of order $k^{-2/3}$ give or take a nearly normal random variable of the smaller order $k^{-5/6}$.

The histogram may be considered a very old-fashioned way of estimating densities. However, histograms are easy to draw; and, unlike kernel estimators, are very widely used in applied work. Mathematical aspects of density estimation are surveyed by Rosenblatt (1971), Cover (1972), Wegman (1972), Tarter and Kronmal (1976), Fryer (1977), Wertz and Schneider (1979), and references listed therein. These papers report a great deal of careful work on discrepancy at a point, and on global results for kernel estimates and other "generalized" histograms. The results show that the mean square error of kernel estimates tends to zero like a constant times $k^{-4/5}$, while (1.6) implies that the mean square error of histograms tends to zero like a constant times $k^{-2/3}$. Asymptotically, this rate is worse, a fact which seems to have stopped further work on the mathematics of histograms. However, for finite sample sizes, the constants determine everything. For example, take $k=500$: then $k^{-4/5} \doteq 0.007$ while $k^{-2/3} \doteq 0.016$. The asymptotic rate of $k^{-4/5}$ can be achieved using another old-fashioned object: the frequency polygon. This is provable with the techniques of this paper.

Before describing our results more carefully, it is helpful to separate the discrepancy (1.1) into sampling error and bias components. To this end, let

$$(1.9) \qquad f_h(x) = \frac{1}{h} \int_{x_0+nh}^{x_0+(n+1)h} f(u)\,du \qquad \text{for } x_0+nh \leqq x < x_0+(n+1)h.$$

(1.10) **Proposition.** *Suppose* $f \in L_2$, *and* (1.5). *Then*

$$E\{\int_I [H(x)-f(x)]^2\,dx\} = \frac{1}{kh} - \frac{1}{k}\int_I f_h(x)^2\,dx + \int_I [f_h(x)-f(x)]^2\,dx.$$

Proof. Suppose $x_0+nh \leqq x < x_0+(n+1)h$. Then $H(x)=N_n/kh$, and N_n is binomial with number of trials k and success probability $p_{nh}=hf_h(x)$. In particular,

$$E\{H(x)\} = f_h(x),$$

$$\mathrm{Var}\{H(x)\} = \frac{1}{kh} f_h(x)\,[1-hf_h(x)]$$

and

$$E\{[H(x)-f(x)]^2\} = \frac{1}{hk} f_h(x) - \frac{1}{k} f_h(x)^2 + [f_h(x)-f(x)]^2.$$

Now integrate in x over I. □

The term $\int (f_h-f)^2$ in (1.10) represents the bias in using discrete intervals of width h. Reducing h diminishes this bias, at the expense of increasing the sampling error term $1/kh$, for the number of observations per cell will decrease as h gets smaller. The tension between these two is resolved by (1.6) and (1.7).

Section 2 of this paper is about the bias term $\int (f_h-f)^2$; Sect. 3 gives examples to show what happens when the regularity conditions like (1.3) and (1.4) are relaxed. In particular, (1.7) fails for some beta and chi-squared densities. Section 4 gives the proof of (1.6) and (1.7). Clearly, the uniform density requires special treatment, since the optimal number of class intervals is one. This density is excluded by the condition that $\int f'^2 > 0$, which surfaces in Lemma (4.5) of Sect. 4.

2. The Bias Term

To begin with assume only that

(2.1) f is an L_2 function on the interval I.

Define f_h by (1.9). Let J be a union of class intervals. Clearly,

$$\int_J f_h(x)\,dx = \int_J f(x)\,dx \tag{2.2}$$

$$\int_J f_h(x)^2\,dx \leqq \int_J f(x)^2\,dx \tag{2.3}$$

(2.4) the f_h are square integrable uniformly in h.

Also, f_h converges to f in L_2:

$$\int_I (f_h - f)^2 \to 0 \quad \text{as } h \to 0. \tag{2.5}$$

For the proof of (2.5), approximate f in L_2 by a continuous function with compact support. Estimates on the rate of convergence in (2.5) will be helpful. For this, additional assumptions are needed. One such is:

(2.6) f is an L_2 function on the interval I, and f is absolutely continuous with a.e. derivative f', and $f' \in L_2$.

Under (2.6), the bias term on the left of (2.5) tends to zero like h^2. More precisely;

(2.7) **Proposition.** *Suppose* (2.6) *and* (1.5). *Let*

$$r(h) = \int_I [f_h(x) - f(x)]^2\,dx - \frac{1}{12} h^2 \int_I f'(x)^2\,dx.$$

Then $r(h) = o(h^2)$.

Proof. To ease the notation, write g for f', and set $x_0 = 0$. Focus on a specific class interval, for instance, $[0, h]$. Clearly,

$$f(x) = a + \int_0^x g(u)\,du$$

where $a = f(0)$. In computing $\int (f_h - f)^2$, the constant a will cancel, so it is harmless to set $a = 0$. Of course,

$$\int_0^h (f_h - f)^2 = \int_0^h f^2 - h f_h^2.$$

In what follows, $u \vee v = \max(u, v)$ and $u \wedge v = \min(u, v)$. Because $a = 0$,

$$\begin{aligned}\int_0^h f^2 &= \int_0^h \int_0^x \int_0^x g(u)\, g(v)\, du\, dv\, dx \\ &= \int_0^h \int_0^h \int_{u \vee v}^h g(u)\, g(v)\, dx\, du\, dv \\ &= \int_0^h \int_0^h (h - u \vee v)\, g(u)\, g(v)\, du\, dv.\end{aligned}$$

Likewise,

$$f_h = \frac{1}{h} \int_0^h (h-u)\, g(u)\, du$$

so

$$h f_h^2 = \frac{1}{h} \int_0^h \int_0^h (h-u)(h-v)\, g(u)\, g(v)\, du\, dv$$

and

$$\int_0^h (f_h - f)^2 = \int_0^h \int_0^h \phi_h(u, v)\, g(u)\, g(v)\, du\, dv$$

where

$$\begin{aligned}\phi_h(u, v) &= (h - u \vee v) - \frac{1}{h}(h-u)(h-v) \\ &= (u+v) - (u \vee v) - \frac{1}{h} uv \\ &= u \wedge v - \frac{1}{h} uv.\end{aligned}$$

This defines ϕ_h as a function from $0 \leqq u$, $v \leqq h$. Note that $\phi(u, 0) = \phi(u, h) = \phi(0, v) = \phi(h, v) = 0$. Define ϕ on the whole plane by periodic continuation.

Let

$$\delta_{nh}(g) = \frac{1}{h^2} \int_{nh}^{(n+1)h} (f - f_h)^2 - \frac{1}{12} \int_{nh}^{(n+1)h} g^2.$$

The argument thus far shows that

$$\delta_{nh}(g) = \frac{1}{h^2} \int_{nh}^{(n+1)h} \int_{nh}^{(n+1)h} \phi_h(u, v)\, g(u)\, g(v)\, du\, dv - \frac{1}{12} \int_{nh}^{(n+1)h} g^2.$$

It will now be shown that $\Sigma_n \delta_{nh}(g) \to 0$ as $h \to 0$.

If g is constant on $[nh, (n+1)h]$, a direct computation shows that $\delta_{nh}(g) = 0$. But g may be approximated closely in L_2 by a function g_0 which is constant on each class interval: for instance, apply (2.5) to g. It remains to show that

$$\Sigma_n \delta_{nh}(g) - \Sigma_n \delta_{nh}(g_0)$$

is uniformly small as $h \to 0$. Of course,

$$|(\int g^2)^{\frac{1}{2}} - (\int g_0^2)^{\frac{1}{2}}| \leqq \|g - g_0\|$$

is small, so it remains only to show that $\Sigma_n \Delta_{nh}$ is small, where

$$\Delta_{nh} = \frac{1}{h^2} \int_{nh}^{(n+1)h} \int_{nh}^{(n+1)h} \phi_h(u,v)\,[g(u)\,g(v) - g_0(u)\,g_0(v)]\,du\,dv.$$

Now $|\phi_h| \leqq h$, and

$$|g(u)\,g(v) - g_0(u)\,g_0(v)| \leqq |g(u) - g_0(u)| \cdot |g(v)| + |g(v) - g_0(v)| \cdot |g_0(u)|$$

so $h|\Delta_{nh}| \leqq \alpha_{nh} + \beta_{nh}$, where

$$\alpha_{nh} = \int_{nh}^{(n+1)h} |g(v) - g_0(v)|\,dv \cdot \int_{nh}^{(n+1)h} |g(v)|\,dv,$$

$$\beta_{nh} = \int_{nh}^{(n+1)h} |g(v) - g_0(v)|\,dv \cdot \int_{nh}^{(n+1)h} |g_0(v)|\,dv.$$

Using the Cauchy-Schwarz inequality,

$$[\Sigma_n \alpha_{nh}]^2 \leqq \Sigma_n \left(\int_{nh}^{(n+1)h} |g(u) - g_0(u)|\,du\right)^2 \cdot \Sigma_n \left(\int_{nh}^{(n+1)h} |g(v)|\,dv\right)^2$$
$$\leqq h^2 \int_I (g - g_0)^2 \cdot \int_I g^2.$$

Likewise,

$$[\Sigma_n \beta_{nh}]^2 \leqq h^2 \int_I (g - g_0)^2 \cdot \int_I g_0^2.$$

So

$$(\Sigma_n |\Delta_{nh}|)^2 \leqq 2h^{-2}[(\Sigma_n \alpha_{nh})^2 + (\Sigma_n \beta_{nh})^2]$$
$$\leqq 2 \int_I (g - g_0)^2 \cdot \int_I (g^2 + g_0^2)$$

is small. □

Notes. (i) If $f' \notin L_2$, then $\int (f_h - f)^2$ need not be of order h^2: see example (3.1).

(ii) If (2.6) holds and $f' \neq 0$, then $(f_h - f)/h$ converges weakly in L_2 to 0, but not strongly (in L_2 norm). Indeed, the proposition shows that $\|(f_h - f)/h\|^2 \to 1/12\, \|f'\|^2 > 0$; this rules out strong convergence to 0. To argue weak convergence to 0, let $\psi \in L_2$. Write $I\{\ \}$ for the function which is 1 if the statement in braces is true, and 0 otherwise, and now let

$$\phi_h(u,v) = (1 - h^{-1}u) - I\{u \leqq v\}.$$

Then

(2.8)
$$\frac{1}{h} \int_0^h (f_h - f)\,\psi = \frac{1}{h} \int_0^h \int_0^h \phi_h(u,v)\,g(u)\,\psi(v)\,du\,dv.$$

As before, (2.8) vanishes if ψ is constant on $[0,h]$, and $|\phi_h|\leqq 1$, so ψ can be replaced by a function constant over the class intervals, without disturbing $1/h\int_I (f_h-f)\psi$ very much.

For later use, it will be helpful to improve the $o(h^2)$ error term in (2.7) to $o(h^3)$. To accomplish this, an additional regularity condition like (1.4) is needed. Indeed, example (3.11) below constructs a nonnegative $f\in L_1\cap L_2$, such that $f'\in L_2$ and $f''\in L_\infty$; but $r(h)$ is only of order $h^2\Big/\left(\log\frac{1}{h}\right)^3$ as $h\to 0$.

As a preliminary,

(2.9) Let $\theta(u)=10u(1-u)(1-2u)$ for $0\leqq u\leqq 1$, and be continued periodically over the line.

The function $\theta(u)$ is a constant multiple of the third Bernoulli polynomial: see Sect. 1.2., 11.2 of Knuth (1973).

(2.10) **Lemma.** *$\theta(u)$ vanishes at 0, $\frac{1}{2}$ and 1. It is positive on $(0,\frac{1}{2})$ and antisymmetric about $\frac{1}{2}$, so $\int_0^1 \theta(u)\,du=0$. Furthermore, $|\theta|\leqq 1$.*

(2.11) **Lemma.** *Let $\psi\in L_1$. Then $\int_I \theta(u/h)\,\psi(u)\,du\to 0$ as $h\to 0$.*

Proof. This is a variation on the Riemann-Lebesgue lemma. To prove it replace ψ by a nearby function in L_1 constant on each class interval. □

The form of the next theorem may seem curious, but it gives sharp estimates for $\int(f_h-f)^2$.

(2.12) **Theorem.** *Suppose* (1.5) *and* (2.6). *Suppose f' is locally of bounded variation, determining the signed measure μ. Let μ^+ and μ^- be the positive and negative parts of μ, $|\mu|=\mu^++\mu^-$, and*

$$d_{nh}=|\mu|\{[x_0+nh,x_0+(n+1)h]\}.$$

Assume

(2.13) $$D_h=\Sigma_n d_{nh}^2<\infty.$$

Then $f'\in L_1(\mu)$. Define r(h) as in (2.7). Then

$$\left|r(h)-\frac{1}{60}h^3\int_I \theta[(x-x_0)/h]f'(x)\,\mu(dx)\right|\leqq\tfrac{3}{2}h^3 D_h.$$

Proof. Without loss of generality, set $x_0=0$. The first step is to show that $f'\in L_1(\mu)$. First, it will be shown that for any $\xi\in[0,h]$,

(2.14) $$\int_0^h |f'|\,|d\mu|\leqq|f'(\xi)|\,d_{oh}+d_{oh}^2.$$

In (2.14) and below, $|d\mu|$ indicates integration with $|\mu|$. To verify (2.14), split the interval of integration at ξ. Now

$$\begin{aligned}\int_0^\xi |f'|\,|d\mu| &= \int_0^\xi \left|f'(\xi)-\int_v^\xi d\mu\right| |\mu(dv)| \\ &\leqq |f'(\xi)|\cdot|\mu|\{[0,\xi]\}+\int_0^\xi\int_0^\xi |d\mu|\,|d\mu| \\ &= |f'(\xi)|\cdot|\mu|\{[0,\xi]\}+|\mu|\{[0,\xi]\}^2.\end{aligned}$$

Likewise, for the integral from ξ to h. Finally,

$$|\mu|\{[0,\xi]\}^2+|\mu|\{(\xi,h]\}^2 \leqq |\mu|\{[0,h]\}^2.$$

This completes the proof of (2.14).

Now for any $\xi_n \in [nh,(n+1)h]$,

$$\int_{nh}^{(n+1)h} |f'|\,|d\mu| \leqq |f'(\xi_n)|\,d_{nh}+d_{nh}^2.$$

Sum, and use the Cauchy-Schwarz inequality:

$$\begin{aligned}\int |f'|\,|d\mu| &\leqq [D_h \Sigma_n f'(\xi_n)^2]^{1/2}+D_h \\ &\leqq \left[D_h\cdot\frac{1}{h}\int_I f'(x)^2\,dx\right]^{1/2}+D_h\end{aligned}$$

with suitably chosen ξ_n. This completes the proof that $f'\in L_1(\mu)$.

Write $\theta_h(u)=\theta(u/h)$. Since θ_h is bounded, $\theta_h f'\in L_1(\mu)$ as well. Turn now to the main inequality. Clearly, it is enough to prove that

$$\left|\int_0^h (f_h-f)^2-\frac{1}{12}h^2\int_0^h (f')^2-\frac{1}{60}h^3\int_0^h \theta_h\, f'\,d\mu\right| \leqq \tfrac{3}{2}h^3 d_{oh}^2. \tag{2.15}$$

Now

$$f'(x)=b+\int_0^x \mu(dv)$$

$$f(x)=a+bx+\int_0^x (x-v)\,\mu(dv).$$

The constant a cancels in f_h-f, so it is harmless to take $a=0$. Then

$$\begin{aligned}f_h &= \tfrac{1}{2}bh+\frac{1}{h}\int_0^h\int_0^x (x-v)\,\mu(dv)\,dx \\ &= \tfrac{1}{2}bh+\frac{1}{h}\int_0^h\int_v^h (x-v)\,dx\,\mu(dv) \\ &= \tfrac{1}{2}bh+\frac{1}{2h}\int_0^h (h-v)^2\,\mu(dv).\end{aligned}$$

Thus,

(2.16) $$h f_h^2 = \tfrac{1}{4} b^2 h^3 + \tfrac{1}{2} b h \int_0^h (h-v)^2 \mu(dv) + \varepsilon_1$$

where

$$\varepsilon_1 = \frac{1}{4h}\left[\int_0^h (h-v)^2 \mu(dv)\right]^2 \leqq \tfrac{1}{4} h^3 d_{oh}^2.$$

Likewise,

(2.17) $$\frac{1}{12} h^2 \int_0^h (f')^2 = \frac{1}{12} b^2 h^3 + \tfrac{1}{6} b h^2 \int_0^h (h-v)\,\mu(dv) + \varepsilon_2$$

where

$$\varepsilon_2 = \frac{1}{12} h^2 \int_0^h \left[\int_0^x \mu(dv)\right]^2 dx \leqq \frac{1}{12} h^3 d_{oh}^2.$$

And

(2.18) $$\int_0^h f^2 = \tfrac{1}{3} b^2 h^3 + b \int_0^h [\tfrac{2}{3}(h^3 - v^3) - v(h^2 - v^2)]\,\mu(dv) + \varepsilon_3$$

where

$$\varepsilon_3 = \int_0^h \left[\int_0^x (x-v)\,\mu(dv)\right]^2 dx \leqq h^3 d_{oh}^2$$

because $\left|\int_0^x (x-v)\,\mu(dv)\right| \leqq h\, d_{oh}$.

Combining (2.16–2.18) gives that

(2.19) $$\left|\int_0^h (f_h - f)^2 - \frac{1}{12} h^2 \int_0^h (f')^2 - b \int_0^h \psi\, d\mu\right| \leqq \tfrac{4}{3} h^3 d_{oh}^2$$

where

$$\psi(u) = \tfrac{2}{3}(h^3 - u^3) - u(h^2 - u^2) - \tfrac{1}{6} h^2 (h-u) - \tfrac{1}{2} h (h-u)^2 = \frac{1}{60} h^3\, \theta(u/h).$$

It remains to estimate

$$\varepsilon_4 = \frac{1}{60} h^3 \int_0^h \theta(v/h)\,(f'(v) - b)\,\mu(dv) = \frac{1}{60} h^3 \int_0^h \int_0^v \theta(v/h)\,\mu(du)\,\mu(dv).$$

Since $|\theta| \leqq 1$,

$$|\varepsilon_4| \leqq \frac{1}{60} h^3 d_{oh}^2. \quad \square$$

(2.20) **Corollary.** *Suppose* (1.2–1.5). *Define* $r(h)$ *as in* (2.7). *Then* $r(h)=o(h^3)$.

Proof. Assume without loss of generality that $x_0=0$. The idea is to use (2.12). To estimate D_h, choose q so that $\frac{1}{p}+\frac{1}{q}=1$, where p appears in (1.4) and $1 \leqq p \leqq 2$, so $\frac{1}{2} \leqq q \leqq \infty$. Now use Holder's inequality:

$$d_{nh} = \int_{nh}^{(n+1)h} 1 \cdot |f''(x)|\, dx \leqq h^{\frac{1}{q}} \left[\int_{nh}^{(n+1)h} |f''|^p \right)^{\frac{1}{p}}.$$

So

$$D_h \leqq h^{\frac{2}{q}} \Sigma_n \left(\int_{nh}^{(n+1)h} |f''|^p \right)^{\frac{2}{p}} \leqq h^{2-\frac{2}{p}} \beta(h) \int_I |f''|^p$$

where

$$\beta(h) = \sup_n \left(\int_{nh}^{(n+1)h} |f''|^p \right)^{\frac{2}{p}-1}.$$

If $p=2$, then $\beta(h)=1$ for all h, and $D_h=0(h)=o(1)$. If $p=1$, then $h^{2-2/p}=1$ for all h, and $\beta(h)\to 0$ as $h\to 0$, so $D_h=o(1)$. Likewise, if $1<p<2$, then $D_h = o(h^{2-2/p})=o(1)$. As (2.12) shows, $f'f'' \in L_1$ and

$$|r(h)| \leqq \frac{1}{60} h^3 |\alpha(h)| + \tfrac{3}{2} h^3 D_h$$

where

$$\alpha(h) = \int_I \theta(x/h) f'(x) f''(x)\, dx.$$

Now $\alpha(h)\to 0$ as $h \to 0$, by (2.11). $\square$

Notes. (i) With the assumptions and notation of (2.20), not only is $f' \cdot f'' \in L_1$, but $f' \in L_q$. This is so by assumption for $p=2$. If $p<2$, then $q>2$, and

$$|f'|^q = |f'|^{q-2} \cdot |f'|^2.$$

But f' is bounded by (2.21) below.

(ii) If f is smooth, then $\int \theta_h f' f''$ is of order h, as is D_h, so $r(h)$ is of order h^4.

(iii) However, example (3.3) below constructs an f with $f'' \in C[0,1]$, yet $\int \theta_h f' f''$ is only of order $1/\log \frac{1}{h}$. Now D_h is of order h, so $r(h)$ is of order $h^3/\log \frac{1}{h}$.

The following result has been used several times above. Similar results appear in Sect. 2 and 3 of Chap. 5 of Beckenbach and Bellman (1965).

(2.21) **Lemma.** *Suppose $I=(-\infty,\infty)$. Let $\psi\in L_\alpha$ on I for $0<\alpha<\infty$ and let ψ be absolutely continuous, with a.e. derivative $\psi'\in L_\beta$ for some $\beta\geqq 1$. Then ψ vanishes at ∞.*

Proof. Suppose, e.g., $\limsup_{x\to\infty}\psi(x)>0$. There is a sequence of numbers

$$a_1<b_1<a_2<b_2<\ldots$$

with $a_n\to\infty$ and $\psi(a_i)=\varepsilon>0$ and $\psi(b_i)=\frac{1}{2}\varepsilon$ and $\psi(x)\geqq\frac{1}{2}\varepsilon$ on $[a_i,b_i]$. In particular, $\Sigma(b_i-a_i)<\infty$. However,

$$\int_{a_i}^{b_i}\psi'=-\tfrac{1}{2}\varepsilon$$

so

$$\int_{a_i}^{b_i}|\psi'|^\beta\geqq(\tfrac{1}{2}\varepsilon)^\beta/(b_i-a_i)^{\beta-1}$$

and the sum is infinite. □

While thinking about these results we discovered an interesting variation on Cauchy-Riemann sums.

(2.22) **Lemma.** *Suppose ϕ is absolutely continuous on the finite interval $[a,b]$. Let $\xi\in[a,b]$. Then*

$$\int_a^b|\phi(x)-\phi(\xi)|\,dx\leqq(b-a)\cdot\int_a^b|\phi'(x)|\,dx.$$

Proof. Assume without loss of generality that $\xi=a$: if not, just split $[a,b]$ at ξ. Now

$$\phi(x)-\phi(a)=\int_a^x\phi'(u)\,du$$

so

$$\begin{aligned}\int_a^b|\phi(x)-\phi(a)|&\leqq\int_a^b\int_a^x|\phi'(u)|du\,dx\\&=\int_a^b\int_u^b|\phi'(u)|\,dx\,du\\&=\int_a^b(b-u)\,|\phi'(u)|\,du\\&\leqq(b-a)\cdot\int_a^b|\phi'(u)|\,du.\quad\square\end{aligned}$$

(2.23) *Example.* Let $a=\xi=0$ and $b=1$. Let n be a positive integer, let

$$\begin{aligned}\phi(x)&=nx\quad\text{for}\quad 0\leqq x\leqq 1/n\\&=1\quad\text{for}\quad 1/n\leqq x\leqq 1.\end{aligned}$$

Then

$$\int_0^1 \phi(x)\,dx = 1 - \frac{1}{2n}$$

and

$$\int_0^1 \phi'(x)\,dx = 1$$

so the ratio of the two integrals is arbitrarily close to 1. □

(2.24) **Corollary.** *Suppose ϕ is L_1 and absolutely continuous on $(-\infty, \infty)$. Let a_n be a monotone bilateral sequence of real numbers, with $a_n \to -\infty$ as $n \to -\infty$ and $a_n \to +\infty$ as $n \to +\infty$. Choose ξ_n arbitrarily in $[a_n, a_{n+1}]$ and let*

$$h = \sup_n (a_{n+1} - a_n).$$

Then

$$\left| \int_{-\infty}^{\infty} \phi(x)\,dx - \Sigma_n \phi(\xi_n)(a_{n+1} - a_n) \right| \leqq h \int_{-\infty}^{\infty} |\phi'(x)|\,dx.$$

Proof. The left hand side is at most

$$\Sigma_n \int_{a_n}^{a_{n+1}} |\phi(x) - \phi(\xi_n)|\,dx. \quad \square$$

Remarks. The arguments for (2.22) and (2.24) work, in exactly the same way, when ϕ is only assumed to be locally of bounded variation, determining the signed measure μ with variation $|\mu|$. The integrals on the right hand side of the inequalities are replaced by $|\mu|[a,b]$ and $|\mu|(-\infty,\infty)$ respectively. This includes (2.22) and (2.24) since $|\mu|[a,b] = \int_a^b |\phi'|$. It is easy to construct examples where the Riemann sum is not a good approximation to a smooth L_1 function. Take triangles of height 1, centered at the positive integers, the j-th triangle having base $1/j^2$. Smooth the triangles, and define the function to be zero elsewhere. This function has positive, finite integral, but the Riemann sum approximation can be zero or infinite depending on the choice of a_n and ξ_n. Of course, the right hand side of the bound is infinite. For related material, see the discussion of direct Riemann integrability in Sect. 11.1 of Feller (1971).

3. Examples

(3.1) *Example.* Suppose f is L_2 on $[0,1]$, and is absolutely continuous, but $f' \notin L_2$. Then $\int_0^1 (f_h - f)^2$ need not be of order h^2. Consider the beta distribution: $F(x) = x^\alpha$, so $f(x) = \alpha x^{\alpha-1}$ and $f'(x) = \alpha(\alpha-1)x^{\alpha-2}$. Choose $\alpha \neq 1$ with $0.5 < \alpha < 1.5$. Then $\int_0^1 (f_h - f)^2$ is of order $h^{2\alpha-1}$.

Proof. Let $h=1/N$. On $[nh,(n+1)h]$,

$$hf_h=\int_{nh}^{(n+1)h} f=[(n+1)^\alpha-n^\alpha]\,h^\alpha$$

and

$$hf_h^2=[(n+1)^\alpha-n^\alpha]^2\,h^{2\alpha-1}$$

so

$$h^{1-2\alpha}\int_{nh}^{(n+1)h}(f_h-f)^2=q_n$$

where

$$q_n=\frac{\alpha^2}{2\alpha-1}[(n+1)^{2\alpha-1}-n^{2\alpha-1}]-[(n+1)^\alpha-n^\alpha]^2. \tag{3.2}$$

Thus, $q_0=(\alpha-1)^2/2\alpha-1$, and for $n\geqq 1$,

$$q_n=\frac{1}{12}\alpha^2(\alpha-1)^2\,n^{2\alpha-4}+O(n^{2\alpha-5}).$$

Now $2\alpha-4<-1$ so $\sum_{n=0}^{\infty} q_n=q<\infty$. Also, $q_n>0$ by (3.2), so $q>0$, and

$$\int_0^1 (f_h-f)^2=\left(\sum_{n=0}^{N-1} q_n\right)h^{2\alpha-1}=qh^{2\alpha-1}+O(h^2). \quad \square$$

Note. If $\alpha=1.5$, then $2\alpha-1=2$, but the argument breaks down because $\Sigma_n n^{2\alpha-4}$ diverges. Then $\int(f_h-f)^2$ is of order $h^2\log\frac{1}{h}$. When $\alpha=1$, the argument applies, but $q=0$ because each $q_n=0$. When $\alpha=1/2$, the density f is not in L_2.

(3.3) *Example.* Suppose f satisfies (1.2-1.4) on $I=[0,1]$ and $f''=g$ is continuous on $[0,1]$. Still

$$r(h)=\int_0^1 (f_h-f)^2-\frac{1}{12}h^2\int_0^1 (f')^2$$

can be of order $h^3/\log\frac{1}{h}$, rather than of order h^4, along a sequence of h's tending to 0. See the notes following Corollary (2.20).

The construction uses notation defined in (2.9-2.13). A preliminary lemma is needed.

(3.4) **Lemma.** *Let θ be defined by* (2.9). *Let ψ be absolutely continuous on* $[0,1]$ *with a.e. derivative ψ'. Let m and n be nonnegative integers. There are finite, positive constants A and B, which do not depend on ψ, m, or n, such that:*

$$\int_0^1 \theta(nu)\,\psi(u)\,du\leqq\frac{1}{n}A\int_0^1|\psi'(u)|\,du \tag{3.4a}$$

$$\int_0^1 \theta(mu)\,\theta(nu)\,\psi(u)\,du\leqq\frac{m\wedge n}{m\vee n}B\int_0^1[|\psi(u)|+|\psi'(u)|]\,du. \tag{3.4b}$$

Proof. Claim (a). Let $\tilde{\theta}(u)=\int_0^u \theta(v)\,dv$. The periodicity of θ implies that $\tilde{\theta}\geqq 0$. Likewise, $\tilde{\theta}$ has period 1 and vanishes at all the integers. Let $A=\max\tilde{\theta}$. Integrate by parts:

$$\int_0^1 \theta(nu)\,\psi(u)\,du=-\frac{1}{n}\int_0^1 \tilde{\theta}(nu)\,\psi'|u|\,du.$$

Claim (b). Suppose $m\leqq n$. Apply claim *(a)* to the function $\theta(mu)\,\psi(u)$. □

Construction. Let $h_j=1/2^{j^2}$ and define

$$g(u)=\sum_{j=1}^{\infty}\theta(u/h_j)/j^2 \quad \text{on} \quad [0,1].$$

Clearly, g is continuous (but not much more). Let

$$f'(x)=b+\int_0^x g(u)\,du \quad \text{and} \quad f(x)=\int_0^x f'(u)\,du.$$

Choose b so $f'\geqq 0$ on $[0,1]$.

Now $r(h)$ can be estimated using Theorem (2.12). In the notation of that theorem, the measure μ is absolutely continuous with density g. Clearly, $d_{nh}\leqq h\cdot\max|g|$, so

$$D_h\leqq\frac{1}{h}\cdot h^2\cdot(\max|g|)^2=O(h).$$

What is left is to estimate $h^3\int\theta(x/h)f'(x)f''(x)\,dx$.

Recall that $f''(u)=g(u)=\Sigma\,\theta(u/h_k)\,k^2$, so

$$\int_0^1 \theta(u/h_j)f'(u)f''(u)\,du=\sum_{k=1}^{j-1}\frac{1}{k^2}\int_0^1\theta(u/h_k)\,\theta(u/h_j)f'(u)\,du+\frac{1}{j^2}\int_0^1\theta^2(u/h_j)f'(u)\,du+\sum_{k=j+1}^{\infty}\frac{1}{k^2}\int_0^1\theta(u/h_k)\,\theta(u/h_j)f'(u)\,du. \tag{3.5}$$

The middle term on the right side of (3.5) is the dominant one, for

$$\int_0^1\theta^2(u/h)f'(u)\,du\to\alpha\int_0^1 f'(u)\,du \quad \text{as} \quad h\to 0,$$

where $\alpha=\int_0^1\theta^2(u)\,du>0$. Thus,

$$\frac{1}{j^2}\int_0^1\theta(u/h_j)f'(u)\,du$$

is of order $1/j^2$, namely, $1/\log\frac{1}{h_j}$, as $j\to\infty$.

It will now be shown that the two sums on the right in (3.5) are negligible. Of course, $\theta(u/h_k)=\theta(2^{k^2}u)$. Use (3.4b) on the first sum, with f' for ψ: when $k<j$,

$$\int_0^1 \theta(u/h_k)\,\theta(u/h_j)\,f'(u)\,du \leqq 2^{k^2-j^2}B_1$$

where B is from (3.4b) and

$$B_1 = B\cdot\int_0^1 (|f'|+|f''|).$$

The first sum is at most

$$B_1\cdot\sum_{k=1}^{j-1}\frac{1}{k^2}2^{k^2-j^2}\leqq B_2/2^{2j}$$

where $B_2=2B_1\cdot\sum_{k=1}^{\infty}1/k^2$, because $k^2-j^2\leqq(j-1)^2-j^2=-2j+1$.

Similarly, use (3.4b) on the second sum, with f' for ψ: when $k>j$,

$$\int_0^1 \theta(u/h_k)\,\theta(u/h_j)\,f'(u)\,du \leqq 2^{j^2-k^2}B_1.$$

The second sum is at most

$$B_1\cdot\sum_{k=j+1}^{\infty}\frac{1}{k^2}2^{j^2-k^2}\leqq B_3/2^{2j}$$

where $B_3=\frac{1}{2}B_1\cdot\sum_{k=1}^{\infty}1/k^2$, because $j^2-k^2\leqq j^2-(j+1)^2=-2j-1$. □

Condition (1.4) constrains f'' to lie in L_p for some p with $1\leqq p\leqq 2$. This guarantees that $r(h)=o(h^3)$ by (2.20). Other values of p will not do, as the next sequence of examples shows. The densities are made up of an infinite sequence of quadratic "bumps". The conditions for (2.20) demand $f\in L_2$. In the examples, usually $f\notin L_1$.

(3.6) **Lemma.** *Suppose f is quadratic on $[d, d+h]$. Then*

$$\int_d^{d+h}(f_h-f)^2-\frac{1}{12}h^2\int_d^{d+h}f'^2=-\frac{1}{180}f''(d)^2h^5.$$

Now define a "bump" of height parameter b, width parameter ε, and starting point a. This function f on $[a, a+4\varepsilon]$ is characterized by the requirements

$$\begin{aligned} f''(x)&=b &&\text{for } a\leqq x<a+\varepsilon,\\ &=-b &&\text{for } a+\varepsilon\leqq x<a+3\varepsilon,\\ &=b &&\text{for } a+3\varepsilon\leqq x<a+4\varepsilon,\\ f'(a)&=0, &&\\ f(a)&=0. && \end{aligned}$$

(3.7) **Lemma.** Let f be a bump of height parameter b, width parameter ε, and starting point a. Then

(i) $f'(a+4\varepsilon)=\int_a^{a+4\varepsilon} f''=0,$

(ii) $\max f'=b\varepsilon$ and $\min f'=-b\varepsilon,$

(iii) $f(a+4\varepsilon)=\int_a^{a+4\varepsilon} f'=0,$

(iv) $\max f=b\varepsilon^2$ and $\min f=0,$

(v) $\int_a^{a+4\varepsilon} (f')^2=Ab^2\varepsilon^3,$

(vi) $\int_a^{a+4\varepsilon} f^2=Bb^2\varepsilon^5,$

(vii) $\int_a^{a+4\varepsilon} f=Cb\varepsilon^3.$

Here, A, B, C are positive, finite constants, whose exact value is immaterial.

Now make a "bump function" f on $[0,\infty)$ as follows. Choose a sequence of height parameters b_j, width parameters ε_j, and multiplicities n_j. The function f will have bumps starting at 0, 1, 2, ... The first n_1 bumps all have height parameters b_1 and width parameters ε_1. The next n_2 bumps all have height parameters b_2 and width parameters ε_2; and so on. Here $b_j>0$, $\varepsilon_j=1/4^{\gamma j}$ for some positive integer γ, and n_j is a positive integer. The remainder

$$r(h)=\int_0^\infty (f_h-f)^2-\frac{1}{12}h^2\int_0^\infty (f')^2$$

is to be estimated for $h=\varepsilon_j$ and $x_0=0$. Let $n=n_1+\ldots+n_j$. Now

$$r(h)=r_1(h)+r_2(h)+r_3(h).$$

Here

$$r_1(h)=\int_0^n (f_h-f)^2-\frac{1}{12}h^2\int_0^n f'^2$$

will be called the "early bump error". It depends only on the first n bumps. Next,

$$r_2(h)=-\frac{1}{12}h^2\int_n^\infty (f')^2$$

is the "incomplete-f' error", and depends only on bumps $n+1$, $n+2,\ldots$. Finally,

$$r_3(h)=\int_n^\infty (f_h-f)^2$$

is the "incomplete-f error", and it too depends only on bumps $n+1$, $n+2,\ldots$.

We have required ε_{j+1} to divide ε_j evenly. As a result, the early bump error is easily estimated from (3.6). Indeed, fix $h=\varepsilon_j$ and consider the bump on $J=[a, a+4\varepsilon_i]$ where $i \leqq j$. Let $M=\varepsilon_i/\varepsilon_j=4^{\gamma(j-i)}$. There are M class intervals which evenly cover $[a, a+\varepsilon_i]$; another M which cover $[a+\varepsilon_i, a+2\varepsilon_i]$; etc. On each such class interval the bump is quadratic. This proves:

(3.8) The early-bump error is

$$-\frac{4}{180}\varepsilon_j^4 \sum_{i=1}^{j} n_i b_i^2 \varepsilon_i.$$

As (3.7v) shows,

(3.9) The incomplete-f' error is

$$-\frac{1}{12}A\varepsilon_j^2 \sum_{i=j+1}^{\infty} n_i b_i^2 \varepsilon_i^3.$$

Now $\varepsilon_j \geqq 4\varepsilon_{j+1}$; as a result, (3.7vi–vii) imply

(3.10) The incomplete-f error is

$$B\sum_{i=j+1}^{\infty} n_i b_i^2 \varepsilon_i^5 - C^2\frac{1}{\varepsilon_j}\sum_{i=j+1}^{\infty} n_i b_i^2 \varepsilon_i^6.$$

(3.11) *Example.* There is an $f \geqq 0$ on $[0, \infty)$ which is L_1 and L_2 and absolutely continuous; furthermore, $f' \in L_2$ is absolutely continuous; and $f'' \in L_\infty$ vanishes at ∞. However,

$$r(h)=\int_0^\infty (f_h-f)^2 - \frac{1}{12}h^2\int_0^\infty f'^2$$

is only of order $h^2\Big/\left(\log\frac{1}{h}\right)^3$ rather than $o(h^3)$, at least on a sequence $h_j=4^{-j}\to 0$.

Construction. Choose $b_j=1/j^2$, $\varepsilon_j=4^{-j}$, and $n_j=4^{3j}$. In view of (3.7),

$f \in L_1$ because $\Sigma n_j b_j \varepsilon_j^3 < \infty$,
$f \in L_2$ because $\Sigma n_j b_j^2 \varepsilon_j^5 < \infty$,
$f' \in L_2$ because $\Sigma n_j b_j^2 \varepsilon_j^3 < \infty$,
$f'' \in L_\infty$ vanishes at ∞ because $b_j \to 0$.

Also, $r(h)$ can be estimated using (3.8–9–10). The early-bump error is of order ε_j^2/j^4, as is the incomplete-f error. The incomplete-f' error is dominant, being of order ε_j^2/j^3. □

(3.12) *Example.* There is an $f \geqq 0$ on $[0, \infty)$ which is L_2 and absolutely continuous; furthermore, $f' \in L_2$ is absolutely continuous, and $f'' \in L_p$ for all $p \geqq 4$.

However, $r(h)$ is only of order $h^2 \Big/ \left(\log \frac{1}{h}\right)^3$ rather than $o(h^3)$, at least on a sequence $h_j = 4^{-j} \to 0$. This f is not L_1.

Construction. Choose $b_i = 1/(i^2 4^i)$, $\varepsilon_i = 1/4^i$, and $n_i = 4^{5i}$. □

(3.13) *Example.* Fix p with $2 < p < 4$. There is an $f \geqq 0$ on $[0, \infty)$ which is L_2 and absolutely continuous; furthermore, $f' \in L_2$ is absolutely continuous, and $f'' \in L_p$. However, $r(h)$ is only of order $h^2/\log \frac{1}{h}$, rather than $o(h^3)$, at least on a sequence $h_j = 4^{-j} \to 0$. This f is not L_1.

Construction. Choose $c \geqq 2/(p-2)$ such that $2c$ is an integer. Set $d = 3 + 2c$. Then $b_i = 1/(i 4^{ci})$, and $\varepsilon_i = 1/4^i$, and $n_i = 4^{di}$. □

(3.14) *Example.* Fix p with $0 < p < 2/3$. There is an $f \geqq 0$ on $[0, \infty)$ which is L_2 and absolutely continuous; furthermore, $f' \in L_2$ is absolutely continuous, and $f'' \in L_p$. However, $r(h)$ is only of order $h^2/\log \frac{1}{h}$, rather than $o(h^3)$, at least on a sequence $h_j = 4^{-j} \to 0$. This f is not L_1.

Construction. Let $c = 2/(2-p)$ and $d = 3 - 2c > 0$. Typically, d is not an integer. Let $b_i = 4^{ci}$, $\varepsilon_i = 1/4^i$, and let n_i be the integer part of $4^{di}/i^2$. □

(3.15) *Example.* Fix p with $2/3 \leqq p < 1$, and θ with $p < \theta < 1$. There is an $f \geqq 0$ on $[0, \infty)$ which is L_2 and absolutely continuous; furthermore, $f' \in L_2$ is absolutely continuous, and $f'' \in L_p$. However, $r(h)$ is only of order $h^{5-(2/\theta)}$, rather than $o(h^3)$, along a sequence of h's tending to 0. This f is not L_1.

Construction. Let $n_i = 1$. Let γ be a large positive integer, to be chosen later. Let $b_i = 4^{\gamma i/\theta}$ and $\varepsilon_i = 4^{-\gamma i}$. Here, the three errors are of the same order of magnitude, viz. $\varepsilon_j^{5-(2/\theta)}$. However, for large γ, the incomplete-f' error dominates. □

Note. Similar examples (with $p < 1$) may be constructed starting with the function $f(x) = \alpha x^{\alpha - 1}$ for $1.5 < \alpha < 2$. However, the calculations are quite tedious.

4. The Optimization

Theorems (1.6) and (1.7) are proved in this section. The following notation will be used throughout: Let

(4.1 a) $$\psi_k(h) = E\{\int_I [H(x) - f(x)]^2 \, dx\},$$

(4.1 b) $$\phi_k(h) = \frac{1}{kh} + b h^2,$$

(4.1 c) $$b = \frac{1}{12} \int f'(x)^2 \, dx,$$

(4.1 d) $$d = \int_I f(x)^2 \, dx.$$

Both theorems give an approximation to the cell width h^* which minimizes the expected L_2 error $\psi_k(h)$, and the size of this error at h^*. The argument will show that $\psi_k(h)$ is a continuous function of h on $(0, \infty)$, tending to ∞ as h tends to 0, and tending to some positive limit as h tends to infinity. The latter limit is bounded away from 0, as k tends to ∞. Further, $\inf_h \psi_k(h)$ is of order $k^{-2/3} \to 0$. As a result, $\inf_h \psi_k(h)$ is attained, say at h_k^*. To begin, it is useful to introduce an approximation to $\psi_k(h)$; this is $\phi_k(h)$ defined in (4.1b). The first lemma shows that $\phi_k(h)$ achieves its minimum at $\alpha k^{-1/3}$ and at this minimum is of size $\beta k^{-2/3}$. These are the lead terms of (1.6) and (1.7). All preliminary lemmas are proved under the assumptions of (1.7).

(4.2) **Lemma.** *$\phi_k(\cdot)$ is minimized at $h_k = (2bk)^{-1/3} = \alpha k^{-1/3}$, and*

$$\phi_k(h_k) = 3 \cdot 2^{-2/3} \cdot b^{1/3} \cdot k^{-2/3} = \beta k^{-2/3}.$$

(4.3) **Lemma.** (a) $\phi_k(h) \geqq \phi_k(h_k) + b(h - h_k)^2$,

(b) $\phi_k(h) \leqq \phi_k(h_k) + 3b(h - h_k)^2$ *if* $h > h_k$,

(c) $\phi_k(h) \leqq \phi_k(h_k) + 3b(h - h_k)^2 + |h - h_k|^3/kh^4$ *if* $h < h_k$.

Proof. Claim (a). Consider the difference between the left side and the right. The derivative turns out to be positive to the right of h_k, and negative to the left. Clearly, the difference is 0 at h_k, completing the argument.

Claim (b). By Taylor's theorem,

$$\phi_k(h) = \phi_k(h_k) + (h - h_k)\, \phi_k'(h_k) + \tfrac{1}{2}(h - h_k)^2\, \phi_k''(h_k) + \tfrac{1}{6}(h - h_k)^3\, \phi_k^{(3)}(\xi),$$

with $h_k < \xi < h$. Of course, $\phi_k'(h_k) = 0$, and $\phi_k''(h_k) = 6b$, and $\phi_k^{(3)}(h) = -6/kh^4 < 0$.

Claim (c). This is like (b). □

Note. The bounds in (4.6a–b) are a bit surprising because the coefficient b does not depend on k. At h_k, of course, $\phi_k^{(3)}$ is of order $-k^{1/3}$, so the function ϕ_k is changing shape as k grows.

(4.4) **Lemma.** (a) *$\psi_k(h)$ is a continuous function of h for $0 < h < \infty$.*

(b) $\lim_{h \to 0} \psi_k(h) = \infty$.

Proof. Claim (a). The $(f_h - f)$ and f_h are uniformly square integrable by (2.4); as $h_n \to h$, clearly $f_{h_n} \to f_h$ a.e. So $f_{h_n} \to f_h$ in L_2. Now use (1.10).

Claim (b). Use (1.10). □

The next job is to estimate $\inf_h \psi_k(h)$ carefully, and show that unless h is rather close to the h_k of (4.2), $\psi_k(h)$ is too large to be the inf. It is convenient to estimate $\psi_k(h)$ separately in three zones: $0 < h < \delta$, and $\delta \leqq h \leqq L$, and $L \leqq h < \infty$. Only the first zone will matter.

(4.5) **Lemma.** *For any $\delta>0$ and $L>\delta$ there are positive numbers $\theta_{\delta L}$ and $k_{\delta L}$ such that $k>k_{\delta L}$ implies $\min_h \{\psi_k(h): \delta\leqq h\leqq L\}\geqq\theta$.*

Proof. In view of (1.10) and (2.3)

$$\psi_k(h)\geqq\int_I (f_h-f)^2-\frac{1}{k}\int_I f^2.$$

The first term on the right is a continuous function of h, as in (4.4). It cannot vanish: if it did, $f\equiv f_h$; either f is discontinuous, or $f'\equiv 0$; both possibilities are ruled out by hypothesis. At this point we use the condition $\int f'^2>0$ to exclude the possibility that f is, e.g., uniform over $[0,1]$, in which case $h=1$ is optimal. Let θ_0 be the minimum over h with $\delta\leqq h\leqq L$ of

$$\int_I (f_h-f)^2.$$

So $\theta_0>0$. For k large, $\frac{1}{k}\int_I f^2<\frac{1}{2}\theta_0$. □

(4.6) **Lemma.** *For any $\delta>0$ there are positive numbers θ_δ and k_δ such that $\psi_k(h)\geqq\theta_\delta$ for all $h\geqq\delta$ and $k\geqq k_\delta$.*

Proof. As $h\to\infty$, it is clear that $f_h\to 0$ pointwise. The convergence is L_2 by uniform integrability (2.4). So $\int(f_h-f)^2\to\int f^2$. Choose L so large that $h>L$ entails $\int(f_h-f)^2>\frac{1}{2}\int f^2$. Then use (4.5). □

The argument for (1.7) is easier than the argument for (1.6), and will be presented first.

Proof of Theorem (1.7). Fix ε with $0<\varepsilon<b$: (see 4.1c). Use (2.7) to choose $\delta>0$ so small that $|r(h)|\leqq\varepsilon h^2$ for $0<h\leqq\delta$. Now use (1.10) and (2.3):

$$\phi_k(h)-\varepsilon h^2-\frac{d}{k}\leqq\psi_k(h)\leqq\phi_k(h)+\varepsilon h^2 \quad \text{for} \quad 0\leqq h\leqq\delta. \tag{4.7}$$

In particular, the infinimum of $\psi_k(h)$ over h with $0<h\leqq\delta$ is smaller than

$$\min_h\,[\phi_k(h)+\varepsilon h^2]=3\cdot 2^{-2/3}\cdot(b+\varepsilon)^{1/3}\cdot k^{-2/3}$$

and larger than

$$-\frac{d}{k}+[\min_h \phi_k(h)-\varepsilon h^2]=-\frac{d}{k}+3\cdot 2^{-2/3}\cdot(b-\varepsilon)^{1/3}\cdot k^{-2/3}.$$

Here, (4.2) has been used with $b\pm\varepsilon$ in place of b; and k is so large that $[2(b-\varepsilon)\,k]^{-1/3}<\delta$. Because ε was arbitrary, the infimum of $\psi_k(h)$ over h with $0<h\leqq\delta$ is

$$3\cdot 2^{-2/3}\cdot b^{1/3}\cdot k^{-2/3}+o(k^{-2/3}). \tag{4.8}$$

Now (4.4-6) show that $\psi_k(\cdot)$ has a global minimum, say at h_k^*, any such h_k^* tends to 0 as $k\to\infty$, and $\psi_k(h_k^*)=\phi_k(h_k)+O(k^{-2/3})$.

To bound the location of h_k^*, apply (4.3) with $b-\varepsilon$ in place of b, and use (4.7) again. For $0<h\leqq\delta$,

$$\psi_k(h)\geqq\beta_\varepsilon k^{-2/3}-\frac{d}{k}+(b-\varepsilon)(h-h_k)^2,$$

where

$$\beta_\varepsilon=3\cdot 2^{-2/3}\cdot(b-\varepsilon)^{1/3}.$$

If $|h-h_k|\geqq\eta k^{-1/3}$, and ε is small, and k is large, then

$$\psi_k(h)\geqq\beta_\varepsilon k^{-2/3}-\frac{d}{k}+(b-\varepsilon)\eta^2 k^{-2/3}>\min\psi_k.$$

In particular, any h_k^* must be within $\eta k^{-1/3}$ of h_k, for k large.

Theorem (1.7) asserts a bit more than has been proved so far: that for any h suitably close to h_k, $\psi_k(h)$ is close to its minimum. Thus, suppose $|h-h_k|\leqq\eta k^{-1/3}$, where η is small. To finish the proof, $\psi_k(h)$ will be estimated above and below. First, if $\eta\leqq\frac{1}{2}(2b)^{-1/3}$, then

$$\tfrac{1}{2}h_k\leqq h\leqq 2h_k.$$

Now $\psi_k(h)$ can be estimated from below using (4.7) and (4.2):

$$\begin{aligned}\psi_k(h)&\geqq\phi_k(h)-\varepsilon h^2-\frac{d}{k}\\&\geqq\phi_k(h_k)-\varepsilon 4h_k^2-\frac{d}{k}.\end{aligned}$$

Since ε is arbitrary, and h_k is of order $k^{-1/3}$,

$$\psi_k(h)\geqq\phi_k(h_k)+o(k^{-2/3}).$$

The estimate for $\psi_k(h)$ from above is very similar when $h>h_k$; see (4.3b). So, suppose

$$\tfrac{1}{2}h_k\leqq h_k-\eta k^{-1/3}\leqq h\leqq h_k.$$

Now use (4.7) and (4.3c):

$$\psi_k(h)\leqq\phi_k(h_k)+\varepsilon h_k^2+3b\eta^2 k^{-2/3}+T$$

where

$$\begin{aligned}T&=|h-h_k|^3/kh^4\leqq(\eta k^{-1/3})^3/k(\tfrac{1}{2}h_k)^4\\&\leqq 2^4\cdot\eta^3\cdot(2b)^{4/3}\cdot k^{-2/3}.\end{aligned}$$

Again, ε is arbitrary and h_k is of order $k^{-1/3}$. Also η is arbitrary; so if $|h-h_k|=o(k^{-1/3})$,

$$|\psi_k(h)-\phi_k(h_k)|=o(k^{-2/3}),$$

as desired. □

Note. We guess that h_k^* is unique, but cannot prove this without additional conditions.

Turn now to the proof of Theorem (1.6). Assume (1.1–1.5). This is stronger than the assumptions for (1.7), so for any $\delta>0$, the infimum over all h of $\psi_k(\cdot)$ is achieved in $0<h<\delta$ and tends to 0 as k tends to ∞. The region $0<h\leqq\delta$ will be split into the following zones, defined in terms of h_k from (4.2) and a constant A to be chosen later:

- $|h-h_k|\leqq A/k^{1/2}$
- $|h-h_k|>A/k^{1/2}$ but $h<2h_k$
- $2h_k\leqq h\leqq\delta$.

For any small positive constant c there is a δ_0 such that for $0<h\leqq\delta_0$

$$\phi_k(h)-\frac{d}{k}-ch^3\leqq\psi_k(h)\leqq\phi_k(h)+ch^3. \tag{4.9}$$

This follows from (1.10): relation (2.3) shows $\int f_h^2\leqq\int f^2$ and the bias term is estimated by (2.20).

(4.10) **Lemma.** *Choose c and δ_0 as in (4.9). Let k be so large that $2h_k\leqq\delta_0$. Fix A finite and positive. If $\frac{1}{2}h_k\leqq h\leqq 2h_k$, and $|h-h_k|\leqq A/k^{1/2}$, then*

(a) $\psi_k(h)\geqq\phi_k(h_k)-\left(4\cdot\frac{c}{b}+d\right)\cdot\frac{1}{k}$,

(b) $\psi_k(h)\leqq\phi_k(h_k)+\left(3b^2A+4\frac{c}{b}\right)\cdot\frac{1}{k}+(16b)^{4/3}A^3\frac{1}{k^{7/6}}$.

Proof. Claim (a). Since $h^3\leqq 8h_k^3=4/b$, relation (4.9) implies

$$\psi_k(h)\geqq\phi_k(h)-\left(4\cdot\frac{c}{b}+d\right)\frac{1}{k}$$

and $\phi_k(h)\geqq\phi_k(h_k)$ by (4.3).

Claim (b). First, suppose $h>h_k$. By (4.9) and (4.3b),

$$\begin{aligned}\psi_k(h)&\leqq\phi_k(h)+4\cdot\frac{c}{b}\cdot\frac{1}{k}\\&\leqq\phi_k(h_k)+\left(3bA^2+4\cdot\frac{c}{b}\right)\cdot\frac{1}{k}.\end{aligned}$$

Second, suppose $h<h_k$. Then an extra term T must be added to the upper bound:

$$\begin{aligned}T&=|h-h_k|^3/kh^4\leqq A^3/[k^{5/2}(h_k/2)^4]\\&\leqq(16b)^{4/3}A^3/k^{7/6}.\quad\square\end{aligned}$$

Note. For sufficiently large k, if $|h-h_k|\leqq A/k^{1/2}$, then $\frac{1}{2}h_k\leqq h\leqq 2h_k$ eventually.

The next lemma gives a careful upper bound for $\min\psi_k$.

(4.11) **Lemma.** *The minimum of* $\psi_k(\cdot)$ *is at most* $\phi_k(h_k)+\frac{c}{2b}\cdot\frac{1}{k}$.

Proof. $\min\psi_k(h)\leqq\psi_k(h_k)\leqq\phi_k(h_k)+ch_k^3$ by (4.9). □

If h is more than $A/k^{1/2}$ away from h_k, then $\psi_k(\cdot)$ is larger than the upper bound of (4.11). Consider first $h\leqq 2h_k$.

(4.12) **Lemma.** *Choose* A *so large that*

$$bA^2>5\cdot\frac{c}{b}+d.$$

If $h\leqq 2h_k$, *but* $|h-h_k|\geqq A/k^{1/2}$, *then*

$$\psi_k(h)>\phi_k(h_k)+\frac{c}{b}\cdot\frac{1}{k}.$$

In particular, the minimum of $\psi_k(\cdot)$ *cannot be found in this range of* h's, *by* (4.11).

Proof. From (4.3a),

$$\phi_k(h)\geqq\phi_k(h_k)+bA^2\cdot\frac{1}{k}. \tag{4.13}$$

Now

$$\begin{aligned}\psi_k(h)&\geqq\phi_k(h)-\frac{d}{k}-ch^3 \quad \text{by (4.9)},\\ &\geqq\phi_k(h)-\frac{d}{k}-8ch_k^3 \quad \text{because } h\leqq 2h_k,\\ &=\phi_k(h)-\left(4\cdot\frac{c}{b}+d\right)\cdot\frac{1}{k} \quad \text{because } h_k=(2bk)^{-1/3},\\ &\geqq\phi_k(h_k)+\left(bA^2-4\cdot\frac{c}{b}-d\right)\cdot\frac{1}{k} \quad \text{by (4.13)},\\ &>\phi_k(h_k)+\frac{c}{b}\cdot\frac{1}{k}.\ \square\end{aligned}$$

Finally, consider h's in the zone

$$2h_k\leqq h\leqq\delta. \tag{4.14}$$

(4.15) **Lemma.** *Choose* δ *positive, but smaller than* $\min\{\delta_0, b/3c\}$, *where* c *and* δ_0 *are as in* (4.9), *Then* $\phi_k(h)-ch^3$ *is a monotone increasing function of* h *in the interval* (4.14).

Proof. Clearly,

$$h^2\frac{\partial}{\partial h}[\phi_k(h)-ch^3]=2bh^3-\left[\frac{1}{k}+3ch^4\right].$$

If $h \geqq 2h_k$, then $bh^3 \geqq 8bh_k^3 = \frac{4}{k} > \frac{1}{k}$. On the other hand, if $h \leqq \delta$, then $bh^3 \geqq 3ch^4$. □

(4.16) **Corollary.** *Choose* δ *as in* (4.15). *For* $2h_k \leqq h \leqq \delta$, *and* $k > k_0$, $\psi_k(h) > \phi_k(h_k) + \frac{1}{2} b h_k^2$.

In particular, the minimum of $\psi_k(h)$ *cannot be found among these* h's, *by* (4.11).

Proof. Estimate as follows.

$$\begin{aligned}\psi_k(h) &\geqq \phi_k(h) - ch^3 - \frac{d}{k} \quad \text{by} \quad (4.9)\\ &\geqq \phi_k(2h_k) - c8h_k^3 - \frac{d}{k} \quad \text{by} \quad (4.15)\\ &\geqq \phi_k(h_k) + bh_k^2 - c8h_k^3 - \frac{d}{k} \quad \text{by} \quad (4.3\text{a})\\ &= \phi_k(h_k) + \tfrac{1}{2} bh_k^2 + \tau_k,\end{aligned}$$

where

$$\tau_k = \tfrac{1}{2} bh_k^2 - c8h_k^3 - \frac{d}{k}$$

is positive for sufficiently large k, because h_k is of order $1/k^{1/3}$. □

These bounds force the following conclusions: for large k the h's minimizing $\psi_k(\cdot)$ are to be found in the interval $h_k \pm A/k^{1/2}$; on that whole interval $\psi_k(h) = \phi_k(h_k) + O(1/k)$. This completes the proof of Theorem (1.6).

References

Bechenbach, E.F., Bellman, R.: Inequalities, 2nd ed. Berlin-Heidelberg-New York: Springer 1965

Cover, T.: A hierarchy of probability density function estimates. Frontiers of Pattern Recognition. New York: Academic Press 1972

Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd ed. New York: Wiley 1971

Fryer, M.J.: A review of some non-parametric methods of density estimation. J. Inst. Math. Appl. **20**, 335–354 (1977)

Knuth, D.: The Art of Computer Programming, Vol. 1, 2nd ed. Reading, Mass.: Addison-Wesley 1973

Rosenblatt, M.: Curve estimates. Ann. Statist. **2**, 1815–1852 (1971)

Scott, D.: On optimal and data-based histograms. Biometrika **66**, 605–610 (1979)

Tapia, R.A., Thompson, J.R.: Nonparametric Probability Density Estimation. Baltimore: Johns Hopkins 1978

Tarter, M.E., Kronmal, R.S.: An introduction to the implementation and theory of nonparametric density estimation Amer. Statist. **30**, 105–111 (1976)

Wegman, E.J.: Nonparametric probability density estimation I. A summary of available methods. Technometrics **14**, 533–546 (1972)

Wertz, W., Schneider, B.: Statistical density estimation: A bibliography. Internat. Statist Rev. **47**, 155–175 (1979)

Woodroofe, M.: On choosing a delta sequence. Ann. Math. Statist. **41**, 1665–1671 (1968)

Received September 8, 1980