ON THE HISTOGRAM AS A DENSITY ESTIMATOR

BY

DAVID FREEDMAN and PERSI DIACONIS

TECHNICAL REPORT NO. 159
JULY 1980

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE HISTOGRAM AS A DENSITY ESTIMATOR

by

DAVID FREEDMAN AND PERSI DIACONIS

TECHNICAL REPORT NO. 159
JULY 1980

Prepared Under the Auspices
of
National Science Foundation Grant MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
On the Histogram as a Density Estimator

by

David Freedman
Statistics Department
University of California, Berkeley

and

Persi Diaconis
Statistics Department
Stanford University

Abstract

We derive approximations to the mean square difference between a histogram and the underlying density $f$. With sample size $k$, the optimal class width $h$ satisfies $h = \alpha k^{-1/3} + O(k^{-1/2})$ with $\alpha = 6^{1/3} \gamma^{-1/3}$ and $\gamma = \int (r')^2$. With this choice of $h$ the mean square difference is $\beta k^{-2/3} + O(k^{-1})$ with $\beta = \frac{1}{4} \times 6^{2/3} \gamma^{1/3}$. These results are proved under mild regularity conditions. Examples are given to show that these conditions cannot be weakened much.

Key words: Histogram, density estimation, $L_2$ approximation.
On the Histogram as a Density Estimator

by

David Freedman
Statistics Department
University of California, Berkeley

and

Persi Diaconis
Statistics Department
Stanford University

1. Introduction

Let \( f \) be a probability density on an interval \( I \), finite or infinite: we require \( I \) to include its finite endpoints, if any; and \( f \) to vanish 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 \( j \)th class interval:

\[
[x_0 + jh, x_0 + (j+1)h).
\]

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

\[
N_j / kh.
\]

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:
\begin{align}
\delta^2 &= \mathbb{E}\left\{ \int_I [H(x) - f(x)]^2 \, dx \right\} .
\end{align}

We will analyze this quantity on the following assumptions:

\begin{align}
\text{(1.2)} & \quad f \in L_2 \text{ and } f \text{ is absolutely continuous,} \\
& \quad \text{with a.e. derivative } f',
\end{align}

\begin{align}
\text{(1.3)} & \quad f' \in L_2 \text{ and } f' \text{ is absolutely continuous,} \\
& \quad \text{with a.e. derivative } f'' \text{; and } \int (f')^2 > 0
\end{align}

\begin{align}
\text{(1.4)} & \quad f'' \in L_p \text{ for some } p \text{ with } 1 \leq p \leq 2 .
\end{align}

Conditions (1.3) and (1.4) imply that \( f' \) is continuous and vanishes at \( \infty \). In particular, \( f' \) is bounded; compare (2.21) below. Likewise, \( f \) is continuous and vanishes at \( \infty \). Also, \( f' \) is in fact the ordinary (everywhere) derivative of \( f \). We also assume

\begin{align}
\text{(1.5)} & \quad I \text{ is the union of class intervals.}
\end{align}

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

Under conditions (1.1-1.5) we will show that the cell width \( h \) which minimizes \( \delta^2 \) is

\begin{align}
\text{(1.6)} & \quad \alpha k^{-1/3} + O(k^{-1/2}) ,
\end{align}
and at such $h$'s the quantity $\delta^2$ of (1.1) is

\begin{equation}
\beta k^{-2/3} + o(k^{-1}),
\end{equation}

where

\begin{align}
\alpha &= 6^{1/3} \gamma^{-1/3} \\
\beta &= \frac{1}{4} \cdot 6^{2/3} \cdot \gamma^{1/3} \\
\gamma &= \int_I f'(x)^2 \, dx > 0.
\end{align}

Result (1.6) suggests that the discrepancy $\delta^2$ can be made small by choosing the cell width $h$ as $\alpha k^{-1/3}$. Of course, this depends on $\gamma$ which will be unknown in general cases. In principle, $\gamma$ can be estimated from the data as in Woodroffe (1968). However, numerical computations which will be reported elsewhere suggest that the following simple, robust rule for choosing the cell width $h$ provides an adequate approximation.

\begin{equation}
(1.9) \textbf{Rule:} \text{ Choose the cell width as twice the interquantile range of the data divided by the cube root of the sample size.}
\end{equation}

We plan to study the random variable $\Delta^2 = \int [H(x) - f(x)]^2 \, dx$ in a future paper. Here we note only that the standard deviation of $\Delta^2$ is of smaller order than $E(\Delta^2) = \delta^2$. Thus, choosing $h$ to minimize $\delta^2$ is a sensible way to get a small $\Delta^2$. To be a bit more precise, the standard deviation of $\Delta^2$ is of order $k^{-1} h^{-1/2} \sim k^{-5/6}$ for the optimal
h \sim k^{-1/3}. Using (1.7), the minimal discrepancy $\Delta^2$ is of order $k^{-2/3}$
give or take a nearly normal random variable of the smaller order $k^{-5/6}$.

We are aware that the histogram may be considered a very old-
fashioned way of estimating densities. However, histograms are easy
to draw — and very widely used. Thus, a detailed analysis may be worth-
while. Rough versions of (1.6) and (1.7) seem part of the folklore, but
we are unable to give any references. To be specific, we consulted the
survey papers by Rosenblatt (1971), Cover (1972), Wegman (1972), Tarter
and Kronmal (1976), Fryer (1977), 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 $L_2$ error of kernel estimates
tends to zero like a constant times $k^{-2/5}$, while (1.7) implies that
the $L_2$ error of histograms tends to zero like a constant times $k^{-1/3}$.
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 $k^{-2/5} \approx 1/12$ while $k^{-1/3} \approx 1/8$. The asymptotic rate of $k^{-2/5}$
can be achieved using another old fashioned object: the frequency
polygon. This can be analyzed using 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

\begin{equation}
(1.10) \quad f_h(x) = \frac{1}{h} \int_{x_0+nh}^{x_0+(n+1)h} f(u) \, du
\end{equation}

for $x_0+nh \leq x < x_0+(n+1)h$. 

(1.11) Proposition. Suppose $f \in L^2$, and (1.5). Then

$$E \left\{ \int_I [H(x) - f(x)]^2 \, dx \right\}$$

equals

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

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

$$E[H(x)] = f_h(x)$$

$$\text{Var}[H(x)] = \frac{1}{kh} f_h(x) [1 - h f_h(x)]$$

and

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

Now integrate in $x$ over $I$. \qed

The term $\int (f_h - f)^2$ in (1.11) 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 (1.9).

In Section 2 of this paper we will study the bias term $\int (f_h - f)^2$; Section 3 gives examples to show what happens when the regularity
conditions like (1.3) and (1.4) are relaxed. In particular, (1.6) and (1.7) fail for beta and chi-squared densities. In Section 4, we prove (1.6) and (1.7) using our regularity conditions. Clearly, the uniform density requires special treatment since only one class interval is needed. This density is excluded by condition (1.3) which surfaces in section (1.4).

2. The bias term

We begin by assuming only that

\begin{equation}
(2.1) \quad f \text{ is an } L_2 \text{ function on the interval } I.
\end{equation}

We define \( f_h \) by (1.10). Let \( J \) be a union of class intervals.

Clearly,

\begin{equation}
(2.2) \quad \int_J f_h(x) \, dx = \int_J f(x) \, dx
\end{equation}

\begin{equation}
(2.3) \quad \int_J f_h(x)^2 \, dx \leq \int_J f(x)^2 \, dx
\end{equation}

\begin{equation}
(2.4) \quad \text{the } f_h \text{ are square integrable uniformly in } h.
\end{equation}

Also

\begin{equation}
(2.5) \quad \int_I (f_h - f)^2 \to 0 \text{ as } h \to 0.
\end{equation}

For the proof of (2.5), approximate \( f \) in \( L_2 \) by a continuous function with compact support. We will need estimates on the rate of convergence in (2.5), and for this, additional assumptions are needed. One such assumption is
(2.6) \( f \) is an \( L_2 \) function on the interval \( I \), which is absolutely continuous with a.e. derivative \( f' \), and \( f' \in L_2 \).

Under (2.6) the bias term tends to zero like \( h^2 \). More precisely;

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

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

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,

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

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

7
\[ \int_{0}^{h} f^2 = \int_{0}^{h} \int_{0}^{x} g(u) g(v) \, du \, dv \, dx \]

\[ = \int_{0}^{h} \int_{0}^{h} \int_{u,v} g(u) g(v) \, dx \, du \, dv \]

\[ = \int_{0}^{h} \int_{0}^{h} (h-u,v) g(u) g(v) \, du \, dv . \]

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

\[ \phi_h(u,v) = (h-u,v) - \frac{1}{h} (h-u)(h-v) \]

\[ = (u+v) - (uv) - \frac{1}{h} uv \]

\[ = u+v - \frac{1}{h} uv . \]

This defines \( \phi_h \) as a function from \( 0 \leq u,v \leq h \). Note that \( \phi(u,0) = \phi(u,h) = \phi(0,v) = \phi(h,v) = 0 \). Define \( \phi \) on the whole plane by periodic continuation.
Let
\[ \delta_{nh}(g) = \frac{1}{h^2} \int_{nh}^{(n+1)h} (f - f_h)^2 \, g \, du \, dv - \frac{1}{12} \int_{nh}^{(n+1)h} g^2 \, du \, dv . \]

We have proved that
\[ \delta_{nh}(g) = \frac{1}{h^2} \int_{nh}^{(n+1)h} \int_{nh}^{(n+1)h} \phi_h(u,v) g(u) g(v) \, du \, dv - \frac{1}{12} \int_{nh}^{(n+1)h} g^2 \, du \, dv . \]

We must show 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 \( g \). We need only show that
\[ \Sigma_n \delta_{nh}(g) - \Sigma_n \delta_{nh}(g_0) \]
is uniformly small as \( h \to 0 \). Of course,
\[ \left| \left( \int \frac{1}{2} g^2 \right) - \left( \int \frac{1}{2} g_0^2 \right) \right| \leq \| g - g_0 \| \]
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 \( |\phi_h| \leq h \), and
\[ |g(u)g(v) - g_0(u)g_0(v)| \leq |g(u) - g_0(u)| \cdot |g(v)| + |g(v) - g_0(v)| \cdot |g_0(u)| \]

so \( h|\Delta_{nh}| \leq \alpha_{nh} + \beta_{nh} \), where

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

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

Using the Cauchy-Schwarz inequality,

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

\[
\leq h^2 \int_I (g - g_0)^2 \cdot \int_I g^2.
\]

Likewise,

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

So

\[
(\Sigma_n |\Delta_{nh}|)^2 \leq 2 h^{-2} (\Sigma_n \alpha_{nh}^2 + \Sigma_n \beta_{nh}^2)
\]

\[
\leq 2 \int_I (g - g_0)^2 \cdot \int_I (g^2 + g_0^2).
\]

is small. \( \square \)
(i) If \( f' \notin L_2 \), then \( f_h - f \) 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 \( \| (f_h - f)/h \|^2 + 1/12 \| f' \|^2 > 0 \); this rules out strong convergence to 0. To argue weak convergence to 0, let \( \psi \in L_2 \). Then

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

where now, writing \( I\{ \} \) for the function which is 1 if the statement in braces is true, and 0 otherwise,

\[
\phi_h(u,v) = (1 - h^{-1} u) - I\{u \leq v\}.
\]

As before, (2.8) vanishes if \( \psi \) is constant on \([0,h]\), and \( |\phi_h| \leq 1\), so \( \psi \) can be replaced by a function constant over the class intervals, without disturbing \( 1/n \int_1^f (f - f_h)\psi \) very much.

We now wish 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. See Section 3 below for examples showing that this condition cannot be weakened much. As a preliminary,

\[
(2.9) \quad \text{Let } \theta_h(u) = 10\left(\frac{u}{h}\right)(1 - \frac{u}{h})(1 - \frac{2u}{h}) \text{ for } 0 \leq u \leq h, \text{ and be continued periodically over the line.}
\]
The function $\theta_h(u)$ is a constant multiple of the third Bernoulli polynomial, see section 1.2, 11.2 of Knuth (1973).

(2.10) **Lemma.** $\theta_h$ vanishes at $0, \frac{h}{2}$ and $h$. It is positive on $(0, \frac{h}{2})$ and anti-symmetric about $\frac{h}{2}$, so $\int_0^h \theta_h(u) du = 0$. Furthermore, $|\theta_h| \leq 1$. \[\square\]

(2.11) **Lemma.** Let $\psi \in L_1$. Then $\int I \theta_h \psi \to 0$ as $h \to 0$.

**Proof.** This is a variation on the Riemann-Lebesgue lemma. To prove it replace $\psi$ by a nearby function in $L_1$ constant on each class interval. \[\square\]

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 $\mu$. Let $\mu^+$ and $\mu^-$ be the positive and negative parts of $\mu$, $|\mu| = \mu^+ + \mu^-$, and

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

Assume

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

Then $f' \in L_1(\mu)$. Define $r(h)$ as in (2.7). Then
\[ |r(h) - \frac{1}{60} h^3 \int_1^h \theta_h(x-x_0) f'(x) \mu(dx)| \leq \frac{3}{2} h^3 D_h. \]

**Proof.** Without loss of generality, set \( x_0 = 0 \). To show that \( f' \in L_1(\mu) \), we claim that for any \( \xi \in [0,h] \),

\[(2.14) \quad \int_0^h |f'| \, |d\mu| \leq |f'(\xi)| |d_{oh}| + d_{oh}^2. \]

In (2.14) and below we have written \( |d\mu| \) to indicate integration with \( |\mu| \). To verify (2.14), split the interval of integration at \( \xi \). Now

\[ \int_\xi^h |f'| \, |d\mu| = \int_\xi^h |f'(\xi) + \int_\xi^\nu \, d\mu| \, |\mu(\nu)| \]
\[ \leq |f'(\xi)| \cdot |\mu|((\xi,h]) \]
\[ + \int_\xi^h \int_\xi^h |d\mu| \, |d\mu| \]
\[ \leq |f'(\xi)| \cdot |\mu|((\xi,h]) \]
\[ + |\mu|((\xi,h])^2. \]

Likewise, for the integral from 0 to \( \xi \). Finally,

\[ (|\mu|[0,\xi])^2 + (|\mu|[\xi,h])^2 \leq (|\mu|[0,h])^2. \]

This completes the proof of (2.14).

Now for any \( \xi_n \in [nh,(n+1)h] \),
\[\int_{n\h}^{(n+1)\h} |f'| \, d\mu \leq |f'(\xi_n)| \, d_{n\h} + d_{n\h}^2.\]

Sum, and use the Cauchy-Schwarz inequality:

\[\int |f'| \, d\mu \leq \left[ n \cdot f'(\xi_n)^2 \cdot D_h \right] \frac{1}{2} + D_h\]

\[\leq \left[ D_h \cdot \frac{1}{h} \int_I (f'(x))^2 \, dx \right] \frac{1}{2} + D_h\]

with suitably chosen \(\xi_n\).

This completes the proof that \(f' \in L_1(\mu)\). Since \(\theta_h\) is bounded, \(\theta_h \cdot f' \in L_1(\mu)\) as well. We turn now to the main inequality. Clearly, it is enough to prove that

\[\int_0^h (f_h - f)^2 = \frac{1}{12} \h^2 \int_0^h (f')^2 - \frac{1}{60} \h^3 \int_0^h \theta_h \cdot f' \, d\mu \leq \frac{3}{2} \h^3 d_{oh}^2.\]

Now

\[f'(x) = b + \int_0^x \mu(du)\]

\[f(x) = a + bx + \int_0^x (x-u) \, \mu(du)\]

The constant \(a\) cancels in \(f_h - f\), so it is harmless to take \(a = 0\).

Then

\[f_h = \frac{1}{2} bh + \frac{1}{h} \int_0^h \int_0^x (x-u) \, \mu(du) \, dx\]

\[= \frac{1}{2} bh + \frac{1}{h} \int_0^h \int_u^h (x-u) \, dx \, \mu(du)\]

\[= \frac{1}{2} bh + \frac{1}{2h} \int_0^h (h-u)^2 \, \mu(du)\]
Thus,

\[(2.16) \quad h f_h^2 = \frac{1}{4} b^2 h^3 + \frac{1}{2} bh \int_0^h (h-u)^2 \mu(du) + \epsilon_1 \]

where

\[\epsilon_1 = \frac{1}{4h} \left[ \int_0^h (h-u)^2 \mu(du) \right]^2 \]

\[= \frac{1}{4} h^3 d_{oh}^2.\]

Likewise,

\[(2.17) \quad \frac{1}{12} \int_0^h (f')^2 = \frac{1}{12} b^2 h^3 + \frac{1}{6} bh^2 \int_0^h (h-u) \mu(du) + \epsilon_2 \]

where

\[\epsilon_2 = \frac{1}{12} \int_0^h \left[ \int_0^x \mu(du) \right]^2 dx \]

\[\leq \frac{1}{12} h^3 d_{oh}^2.\]

And

\[(2.18) \quad \int_0^h f^2 = \frac{1}{3} b^2 h^3 + b \int_0^h [2(h^3-u^3) - u(h^2-u^2)] \mu(du) + \epsilon_3 \]

where
\[ \varepsilon_3 = \int_0^h \int_0^x (x-u) \mu(du) \, dx \]

\[ \leq h^3 d^2_{oh} \]

because \( |\int_0^x (x-u) \mu(du)| \leq h d_{oh} \).

Combining (2.16-2.18) gives that

\[ (2.19) \quad \left| \int_0^h (f - f)^2 - \frac{h^2}{12} \int_0^h (f')^2 - b \int_0^h \psi \, d\mu \right| \leq \frac{4}{3} h^3 d_{oh} \]

where

\[ \psi(u) = \frac{2}{3} (h^3 - u^3) - u(h^2 - u^2) - \frac{h^2}{6} (h-u) - \frac{1}{2} h(h-u)^2 \]

\[ = \frac{1}{60} h^3 \theta_h(u). \]

We now have to estimate

\[ \varepsilon_4 = \frac{1}{60} h^3 \int_0^h \theta_h(f' - b) \, d\mu \]

\[ = \frac{1}{60} h^3 \int_0^u \int_0^h \theta_h(u) \, \mu(dv) \, \mu(du). \]

Since \( |\theta_h| \leq 1 \),

\[ |\varepsilon_4| \leq \frac{1}{60} h^3 d^2_{oh}. \]

\[ (2.20) \textbf{Corollary.} \text{ Suppose (1.2-1.5). Then } f' \cdot f'' \in L_1. \text{ Define } r(h) \text{ as in (2.7)}: \]
\[ r(h) = \int_I (f_h - f)^2 - \frac{1}{12} h^2 \int_I (f')^2. \]

Then \( r(h) = O(h^3) \). Choose \( q \) so that \( \frac{1}{p} + \frac{1}{q} = 1 \), where \( p \) appears in (1.4) and \( 1 \leq p \leq 2 \). Then, for \( h \leq 1 \),

\[ |r(h)| \leq \frac{1}{60} h^3 \int_I |f''| + \frac{3}{2} h^5 - \frac{2}{p} \cdot \alpha_h \cdot \int_I |f''|^p \]

where

\[ \alpha_h = \sup_n \left( \frac{x_{0+\frac{(n+1)}{h}}}{x_{0+nh}} \right)^{\frac{2}{p} - 1}. \]

**Proof.** As usual, set \( x_0 = 0 \) and \( g = f'' \). To estimate \( D_h \), use Hölder's inequality:

\[ d_{nh} = \int_{nh}^{(n+1)h} 1 \cdot |g(x)| \, dx \]
\[ \leq \frac{1}{q} \left( \int_{nh}^{(n+1)h} |g| \right)^{\frac{1}{p}} \]

so

\[ D_h \leq \frac{2}{q} \sum_{n} \left( \int_{nh}^{(n+1)h} |g| \right)^{\frac{2}{p}} \]
\[ \leq h^{2 - \frac{2}{p}} \alpha_h \int_I |g|^p. \]

In particular, \( D_h < \infty \), proving indirectly that \( f' \cdot f'' \in L_1 \). Consequently, \( \int \theta_h f'f'' \to 0 \) by (2.11). Also, \( D_h \to 0 \) as \( h \to 0 \): if
p = 1, this is because \( \alpha_h \to 0 \), and if \( p > 1 \), this is because \( 2/p < 2 \). Thus \( r(h) = O(h^3) \).

Notes.

(i) Not only is \( f' \cdot f'' \in L_1 \), but \( f' \in L_q \). This is as 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 \( \int_0^h f'f'' \) is of order \( h \), as is \( D_h \), so \( r(h) \) is of order \( h^4 \).

(iii) However, example (3.2) below constructs an \( f \) with \( f'' \in C[0,1] \), yet \( \int_0^h f'f'' \) 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 Sections 2 and 3 of Chapter 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 \( \psi \) be absolutely continuous, with a.e. derivative \( \psi' \in L_\beta \) for some \( \beta \geq 1 \). Then \( \psi \) vanishes at \( \infty \).

Proof. Suppose, e.g., \( \lim \sup_{x \to \infty} \psi(x) > 0 \). Then we can construct a sequence of numbers

\[
a_1 < b_1 < a_2 < b_2 < \ldots
\]

with \( a_n \to \infty \) and \( \psi(a_n) = \epsilon > 0 \) and \( \psi(b_n) = \frac{1}{2} \epsilon \) and \( \psi(x) \geq \frac{1}{2} \epsilon \).
on \([a_i, b_i]\). In particular, the sum \(\sum (b_i - a_i) < \infty\). However,

\[
\int_{a_i}^{b_i} \psi' = -\frac{1}{2} \varepsilon
\]

so

\[
\int_{a_i}^{b_i} |\psi'|^\beta \geq \left(\frac{1}{2} \varepsilon\right)^\beta / (b_i - a_i)^{\beta-1}
\]

and the sum is infinite.

In the course of proving these results, we discovered an interesting variation on Cauchy-Riemann sums.

(2.22) Lemma. Suppose \(\phi\) is \(L_1\) on the finite interval \([a, b]\), and is absolutely continuous, with a.e. derivative \(\phi'\). Let \(\xi \in [a, b]\).

Then

\[
\int_a^b |\phi(x) - \phi(\xi)| \, dx \leq (b-a) \cdot \int_a^b |\phi'(x)| \, dx.
\]

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

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

so

19
\[
\int_a^b |\phi(x) - \phi(a)| \leq \int_a^b \int_a^x |\phi'(u)| du \, dx
\]

\[
= \int_a^b \int_u^b |\phi'(u)| dx \, du
\]

\[
= \int_a^b (b-u) |\phi'(u)| du
\]

\[
\leq (b-a) \cdot \int_a^b |\phi'(u)| du.
\]

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

\[
\phi(x) = nx \quad \text{for} \quad 0 \leq x \leq 1/n
\]

\[
= 1 \quad \text{for} \quad 1/n \leq x \leq 1.
\]

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 \(\phi\) is \(L_1\) on \((-\infty, \infty)\), and is absolutely continuous with a.e. derivative \(\phi'\). 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 \( \xi_n \) arbitrarily in \( [a_n, a_{n+1}] \) and let

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

Then

\[
\left| \int_{-\infty}^{\infty} \phi(x) dx - \sum_n \phi(\xi_n)(a_{n+1} - a_n) \right| \leq h \int_{-\infty}^{\infty} |\phi'(x)| dx.
\]

**Proof.** The left hand side is at most

\[
\sum_n \int_{a_n}^{a_{n+1}} |\phi(x) - \phi(\xi_n)| dx.
\]

**Remarks:** The arguments for (2.22) and (2.24) work, in exactly the same way, when \( \phi \) is merely locally of bounded variation, determining the signed measure \( \mu \) and 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 \( \xi_n \). Of course, the right hand side of the bound is infinite. For related material, see the discussion of direct Riemann integrability in Section 11.1 of Feller (1971).
3. Examples

(3.1) Example. Suppose \( f \in L_2 \) 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 \( .5 < \alpha < 1.5 \). Then \( \int_0^1 (f_h - f)^2 \) is of order \( h^{2\alpha-1} \).

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

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

and

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

so

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

where

\[
q_n = \frac{\alpha^2}{2\alpha-1} \left( (n+1)^{2\alpha-1} - n^{2\alpha-1} \right) - ((n+1)^\alpha - n^\alpha)^2.
\]

Thus, \( q_0 = \frac{\alpha^2}{2\alpha-1} \), and for \( n \geq 1 \) \( q_n \) equals

\[
\frac{1}{12} \alpha^2 (\alpha-1)^2 n^{2\alpha-4} + o(n^{2\alpha-5})
\]

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

22
\[ \int_0^1 (f_h - f)^2 = \left( \sum_{n=0}^{N-1} q_n \right) h^{2\alpha-1} = q h^{2\alpha-1} + O(h^2). \]

**Note.**

If \( \alpha = 1.5 \), then \( 2\alpha - 1 = 2 \), but the argument breaks down because \( \sum_n n^{2\alpha-4} \) 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 \( 0 \).

**Construction.** Refer back to (2.9-2.13). Let \( h_j = 1/2^j \) and

\[ g = \sum_{j=1}^{\infty} \theta_j h_j^2 \] on \([0,1]\).

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

\[ f'(x) = b + \int_0^x g(u)du \text{ and } f(x) = \int_0^x f'(u)du. \]
Choose \( b \) so \( f' \geq 0 \) on \([0,1]\). Clearly, \( d_{nh} \leq h \cdot \max |g| \), so

\[
D_h \leq \frac{1}{h} \cdot h^2 \cdot (\max |g|)^2 = 0(h).
\]

To evaluate \( \int_0^1 \theta_{h_j} f' f'' \), notice that \( h_k \) divides \( h_j \) evenly when \( k > j \), so \( \theta_{h_j} \) is orthogonal to \( \theta_{h_j} \) if \( k \neq j \). Thus

\[
\int_0^1 \theta_{h_j} f' f'' = \int_0^1 f' (\theta_{h_j})^2 / j^2.
\]

As is easily verified,

\[
\int_0^1 f' \theta_{h_j}^2 + \alpha \int_0^1 f' \quad \text{as } h \to 0,
\]

where \( \alpha = \int_0^1 \theta_1^2 > 0 \). Finally, \( \int_0^1 \theta_{h_j} f' f'' \) is of order \( 1/j^2 \), namely, \( 1/\log \frac{1}{h_j} \).

Condition (1.4) constrains \( f'' \) to lie in \( L_p \) for some \( p \) with \( 1 \leq p \leq 2 \). 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.

(3.4) Lemma. Suppose \( f \) is quadratic on \([d,d+h]\). Then

\[
\int_d^{d+h} (f_h - f)^2 - \frac{1}{12} h^2 \int_d^{d+h} f''^2 = \frac{1}{12} f'(d) f''(d) h^4 - \frac{1}{45} f''(d)^2 h^5. \]

We now define a "bump" of height parameter \( b \), width parameter \( \epsilon \), and starting point \( a \). This function on \([a,a+4\epsilon]\) is characterized by the requirements
\[ f''(x) = b \text{ for } a \leq x < a + \varepsilon, \]
\[ = -b \text{ for } a + \varepsilon \leq x < a + 3\varepsilon, \]
\[ = b \text{ for } a + 3\varepsilon \leq x < a + 4\varepsilon, \]
\[ f'(a) = 0, \]
\[ f(a) = 0. \]

(3.5) **Lemma.** Let \( f \) be a bump of height parameter \( b \), width parameter \( \varepsilon \), 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) \[ f(a + 4\varepsilon) = \int_0^{a + 4\varepsilon} f' = 0 \]

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

(v) \[ \int_a^{a + 4\varepsilon} (f')^2 = A b^2 \varepsilon^3 \]

(vi) \[ \int_a^{a + 4\varepsilon} f^2 = B b^2 \varepsilon^5 \]

(vii) \[ \int_a^{a + 4\varepsilon} f = C b \varepsilon^3. \]

Here, \( A, B, C \) are positive, finite constants, where exact value is immaterial. \[ \square \]
We now make a "bump function" \( f \) on \([0, \infty)\) as follows. Choose a sequence of height parameters \( b_j \), width parameters \( \varepsilon_j \), and multiplicities \( n_j \). We require \( f \) to have bumps starting at \( 0, 1, 2, \ldots \). The first \( n_1 \) bumps all have height parameters \( b_1 \) and width parameters \( \varepsilon_1 \). The next \( n_2 \) bumps all have height parameters \( b_2 \) and width parameters \( \varepsilon_2 \); and so on. We require \( b_j > 0 \), and \( \varepsilon_j = 1/4^j \) for some positive integer \( \gamma \), and \( H_j \) to be a positive integer. As usual,

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

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 \( \varepsilon_{j+1} \) to divide \( \varepsilon_j \) evenly. As a result, the early bump error is easily estimated from (3.4). Indeed, fix \( h = \varepsilon_j \) and consider the bump on \( J = [a, a+4\varepsilon_j] \) where \( i \leq j \). Let \( M = \varepsilon_i / \varepsilon_j = 4^{j-i} \). There are \( M \) class intervals which evenly cover \( [a, a+\varepsilon_i] \); another \( M \) which covers \( [a+\varepsilon_i, a+2\varepsilon_i] \); etc. On each such class interval the bump is quadratic. The \( h^4 \)-term in (3.4) cancels when summed over the whole bump, by symmetry. Indeed, consider

\[
\sum_{m=0}^{4M-1} f'(a+mh) f''(a+mh).
\]

The term in \( m = 0 \) vanishes because \( f'(a) = 0 \). For \( 0 < m < 4M \), the term in \( m \) cancels the term in \( 4M-m \). This completes our discussion of the \( h^4 \)-error, and the \( h^5 \)-error is easy to handle because \( (f'')^2 \) is constant over the bump. We have proved:

(3.6) The early-bump error is

\[-\frac{4}{45} \varepsilon_j \sum_{i=1}^4 n_i b_i^2 \varepsilon_i.\]

As (3.5v) shows,

(3.7) The incomplete-f' error is
\[- \frac{1}{12} A \varepsilon_j^2 \sum_{i=j+1}^{\infty} n_i b_i^2 \varepsilon_i^3.\]

We have required \(\varepsilon_j \geq 4\varepsilon_{j+1}\). As a result, (3.5vi) implies

(3.8) The incomplete-\(f\) error is

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

(3.9) Example. There is an \(f \geq 0\) on \([0, \infty)\) which is \(L_1\), absolutely continuous, and vanishing at \(\infty\); further \(f' \in L_2\) is absolutely continuous, and \(f'' \in L_p\) for all \(p \geq 4\). However,

\[r(h) = \left( (f_h - f)^2 - \frac{1}{12} h^2 \right) f'^2\]

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

Construction. We choose \(b_i = 1/(i\cdot 4^i)\), and \(\varepsilon_i = 1/4^i\), and \(n_i = 4^5 i\). The early-bump error is of order \(\varepsilon_j^2/J^4\), as is the incomplete-\(f\) error. The incomplete-\(f'\) error is dominant, being of order \(\varepsilon_j^2/J^3\).

(3.10) Example. Fix \(p\) with \(2 < p < 4\). There is an \(f \geq 0\) on \([0, \infty)\) which is \(L_2\), absolutely continuous, and vanishing at \(\infty\); further, \(f' \in L_2\) is absolutely continuous, and \(f'' \in L_p\). However,

\[r(h) = \left( (f_h - f)^2 - \frac{1}{12} h^2 \right) f'^2\]
is only of order \( h^2/\log \frac{1}{h} \), rather than \( h^3 \), at least on a sequence \( h = 4^{-j} \to 0 \). This \( f \) is not \( L_1 \).

**Construction.** Choose \( c \geq 2/(p-2) \) such that \( 2c \) is an integer. Set \( d = 3+2c \). Then \( b_1 = 1/(4^{c_1}) \), and \( \varepsilon_1 = 1/4^{i_1} \), and \( n_i = 4^{d_i} \).

(3.11) **Example.** Fix \( p \) with \( 0 < p < 2/3 \). There is an \( f \geq 0 \) on \([0, \infty)\) which is \( L_2 \), absolutely continuous, and vanishing at \( \infty \); further \( f' \in L_2 \) is absolutely continuous, and \( f'' \in L_p \). However,

\[
r(h) = \int (f_h - f)^2 - \frac{1}{12} h^2 \int f'^2
\]

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

**Construction.** Let \( c = 2/(2-p) \) and \( d = 3-2c > 0 \). Typically, \( d \) is not an integer. Let \( b_1 = 4^{c_1} \), \( \varepsilon_1 = 1/4^{i_1} \), and let \( n_2 \) be the integer part of \( 4^{d_i}/\varepsilon^2 \).

(3.12) **Example.** Fix \( p \) with \( 2/3 \leq p < 1 \), and \( \theta \) with \( p < \theta < 1 \). There is an \( f \geq 0 \) on \([0, \infty)\) which is \( L_2 \), absolutely continuous, and vanishing at \( \infty \); further \( f' \in L_2 \) is absolutely continuous, and \( f'' \in L_p \). However,

\[
r(h) = \int (f_h - f)^2 - \frac{1}{12} h^2 \int f'^2
\]

is only of order \( h^{5-(2/\theta)} \), rather than \( h^3 \), along a sequence of \( h \)'s tending to \( 0 \). This \( f \) is not \( L_1 \).
Construction. Let \( \gamma \) be a large positive integer, to be chosen later. Let \( b_i = 4^{\gamma_1/\theta} \) and \( \varepsilon_i = 4^{-\gamma_1} \). Here, the three errors are of the same order of magnitude, viz. \( \varepsilon_i^{5-(2/\theta)} \). However, the constants depend on \( \gamma \), and for large \( \gamma \), the incomplete-f' error dominates.

Note.

Similar examples (with \( \rho < 1 \)) 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

In this section we prove (1.6) and (1.7). These results give an approximation to the cell width \( h^* \) which minimizes the expected \( L_2 \) error, and the size of this error at \( h^* \). The idea of the argument is to approximate the expected \( L_2 \) error by a simpler function of \( h \) involving the decomposition (1.10) into bias and sampling error, and the expansion for the bias given in Section 2. The minimizing \( h^* \) for the simple approximation is easy to compute and will be shown to be suitably close to \( h^* \).

Let

\[
\psi_k(h) = \text{E}\left\{ \left[ h(x) - f(x) \right]^2 dx \right\}
\]

(4.1a)

\[
\phi_k(h) = \frac{1}{kh} + bh^2
\]

(4.1b)

\[
b = \frac{1}{12} \int f'(x)^2 \, dx
\]

(4.1c)

\[
d = \int f(x)^2 \, dx .
\]

(4.1d)
Suppose (1.2-3-4). Then, there is a positive constant $c$, depending on $f$, such that for $0 \leq h \leq 1$,

\begin{equation}
\phi_k(h) - \frac{d}{k} \leq \psi_k(h) \leq \phi_k(h) + ch^3.
\end{equation}

This follows from (1.11): relation (2.3) shows $\int f_h^2 \leq \int f^2$, and the bias term is estimated by (2.20). If, for definiteness, $p = 1$ in (1.4), i.e., $f'' \in L_1$, then $c$ can be chosen as

\[
\frac{1}{60} \int_I |f'f''| + \frac{3}{2} \left[ \int_I |f''|^2 \right]^2.
\]

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

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

(4.4) **Lemma.** $\psi_k(h)$ is a continuous function of $h$.

**Proof.** The $(f_h - f)^2$ 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 (1.11).

Clearly, $\psi_k(h)$ tends to infinity as either $h$ tends to zero or infinity. It follows that there is a minimizing value of $h$.

(4.5) **Theorem.** Suppose (1.2-3-4). Let $h_k^*$ be a value of $h$ which minimizes $\psi_k(h) = \mathbb{E}(H-f)^2$. Then
\[ |h_k^* - h_k| < A/k^{1/2} \]

\[ |\psi_k(h_k^*) - \phi_k(h_k)\psi_k| < B/k. \]

These bounds are useful, because \( h_k \) is of order \( 1/k^{1/3} >> 1/k^{1/2} \), and \( \phi_k(h_k) \) is of order \( 1/k^{2/3} >> 1/k \). Numerical values for \( A \) and \( B \) will be given below. Before proving (4.5), it will be helpful to do some preliminary analysis.

(4.6) Lemma. (a) \( \phi_k(h) \geq \phi_k(h_k) + b(h-h_k)^2 \)

(b) \( \phi_k(h) \leq \phi_k(h_k) + 3b(h-h_k)^2 \) if \( h > h_k \)

(c) \( \phi_k(h) \leq \phi_k(h_k) + 3b(h-h_k)^2 + |h-h_k|^3/k^2 \) 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,

\[ \phi_k(h) = \phi_k(h_k) + (h-h_k) \phi_k'(h_k) + \frac{1}{2} (h-h_k)^2 \phi_k''(h_k) + \frac{1}{6} (h-h_k)^3 \phi_k^{(3)}(\xi), \]

with \( h_k < \xi < h \). Of course, \( \phi_k'(h_k) = 0 \), and \( \phi_k''(h_k) = b \), 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 $\phi_k$ is changing shape as $k$ grows.

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

**Proof.** $\min \psi_k(\cdot) \leq \psi_k(h_k) \leq \phi_k(h_k) + ch_k^3$, by (4.2).

If $h$ is more than $A/k^{1/2}$ away from $h_k$, then $\psi_k(\cdot)$ is bigger than the approximate minimum. We first prove this on the side condition that $h$ is of the same order as $h_k$, namely, $1/k^{1/3}$.

(4.8) **Lemma.** Choose $A$ so large that

$$6A^2 > 9 \cdot \frac{b}{2b} + d$$

If $h \leq 2h_k$, but $|h-h_k| \geq A/k^{1/2}$, then

$$\psi_k(h) > \phi_k(h_k) + \frac{c}{2b} \cdot \frac{1}{k}.$$  

In particular, the minimum of $\psi_k(\cdot)$ cannot be found in this range of $h$'s, by (4.7).

**Proof.** From (4.6a),

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

From (4.2), then,
\[ \psi_k(h) \geq \phi_k(h) - \frac{d}{k} - c h^3 \]
\[ \geq \phi_k(h) - \frac{d}{k} - 8 c h^3_k \]
\[ = \phi_k(h_k) - (8 \frac{c}{2b} + d) \cdot \frac{1}{k} \]
\[ \geq \phi_k(h_k) + (b A^2 - 8 \frac{c}{2b} - d) \cdot \frac{1}{k} \]
\[ > \phi_k(h_k) + \frac{c}{2b} \cdot \frac{1}{k} . \]

Next, we show \( \psi_k(h) = \phi_k(h_k) + O(h) \) for \( h \)'s within \( A/k^{1/2} \) of \( h_k \).

(4.9) **Lemma.** If \( \frac{1}{2} h_k \leq h \leq 2h_k \) and \( |h-h_k| \leq A/k^{1/2} \), then

(a) \( \psi_k(h) \geq \phi_k(h_k) - (8 \frac{c}{2b} + d) \cdot \frac{1}{k} \)

(b) \( \psi_k(h) \leq \phi_k(h_k) + (3b A + 8 \frac{c}{2b}) \cdot \frac{1}{k} + (16b)^{4/3} A^3 \frac{1}{k^{7/6}} \).

**Proof.** **Claim** (a). As before, (4.2) implies

\[ \psi_k(h) \geq \phi_k(h) - (8 \frac{c}{2b} + d) \frac{1}{k} \]

and \( \phi_k(h) \geq \phi_k(h_k) \) by (4.3).

**Claim** (b). First, suppose \( h > h_k \). By (4.2) and (4.6b),

34
\[ \psi_k(h) \leq \phi_k(h) + 8 \frac{c}{2b} \cdot \frac{1}{k} \]

\[ \leq \phi_k(h_k) + (3b^2 A + 8 \frac{c}{2b}) \cdot \frac{1}{k}. \]

Second, suppose \( h < h_k \). Then we must add an extra term \( T \) to the upper bound:

\[ T = \frac{|h-h_k|^3}{kh^4} \]

\[ \leq \frac{A^3}{k^{5/2}} (2h_k)^4 \]

\[ \leq (16b)^4/3 A^3/k^{7/6}. \]

\[ \square \]

**Note.**

For sufficiently large \( k \), if \( |h-h_k| \leq A/k^{1/2} \), then

\[ \frac{1}{2} h_k \leq h \leq 2h_k \]

eventually.

Next, we will choose a small, positive \( \delta \) and show that \( \min \psi_k(\cdot) \) cannot be found among \( h \)'s in the range

(4.10) \[ 2h_k \leq h \leq \delta. \]

(4.11) **Lemma.** Choose \( \delta \) positive, but so small that

\[ 2b > 3c \delta + \frac{1}{8} \cdot 2b. \]

Then \( \phi_k(h) - ch^3 \) is a monotone increasing function of \( h \) for \( h \) in the interval (4.10).
Corollary. For $h$ in the interval (4.10), and $k > k_0$, 

$$
\psi_k(h) > \phi_k(h_k) + \frac{c}{2b} \cdot \frac{1}{k}.
$$

In particular, the minimum of $\psi_k(h)$ cannot be found among these $h$'s, by (4.17).

Proof. We estimate as follows.

\[
\psi_k(h) \geq \phi_k(h) - ch^3 - \frac{d}{k} \quad \text{by (4.2)},
\]

\[
\geq \phi_k(2h_k) - 8h_k^3 - \frac{d}{k} \quad \text{by (4.11)},
\]

\[
\geq \phi_k(h_k) + b h_k^2 - 8 h_k^3 - \frac{d}{k} \quad \text{by (4.6a)},
\]

\[
= \phi_k(h_k) + \frac{c}{2b} \cdot \frac{1}{k} + \tau_k,
\]

where

\[
\tau_k = b h_k^2 - 8 h_k^3 - \frac{d}{k} - \frac{c}{2b} \frac{1}{k}
\]

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

Lemma. Let $L$ be large but finite. Then

$$
\min_h \{\psi_k(h) : \delta \leq h \leq L\} \geq \theta, \quad \text{for } k > k_0 \delta L,
$$

where $\theta$ is a positive number which does not depend on $k$. 

36
Since the approximate minimum for $\psi_k$ found in (4.7) goes to zero as $k$ goes to infinity, for large $k$ the minimum of $\psi_k$ is not to be found in $[\delta, L]$.

**Proof.** In view of (1.11) and (2.3)

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

The first term on the right is a continuous function of $h$, by (4.4). It cannot vanish: if it did, $f \equiv f_h$; either $f$ is discontinuous, violating (1.2); or $f' \equiv 0$, violating (1.3). It is at this point that we use the conditions to exclude the possibility that $f$ is, e.g., uniform over $[0,1]$, in which case $h = 1$ is optimal. Let

$$\theta_0 = \min_h \int (f_h - f)^2 : \delta \leq h \leq L.$$  

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

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 $\int (f_h - f)^2 > \frac{1}{2} \int f^2$. Clearly,

(4.14) **Lemma.** $\lim \inf_{k \to \infty} \inf_h \{\psi_k(h) : h > L\} > \frac{1}{2} \int f^2 > 0$.

Again, the minimum of $\psi_k(\cdot)$ cannot be found among the $h$'s with $h > L$. We are forced to the conclusion that for large $k$, the only $h$'s minimizing $\psi_k(h)$ are to be found in the interval $h_k \pm A/\sqrt{k}$; and on that whole interval, $\psi_k(h) = \phi_k(h_k) + O(1/k)$. This proves (4.5). □

37
References


