

## LECTURE 2

# Local Asymptotics

Historically, the greatest attention in the statistical literature has been paid to parametric problems. This is especially so in the case of independent identically distributed observations. In most of this development, a very detailed assumption is made about the common distribution of the observations, for example, a Gaussian distribution. The typical result is that under appropriate smoothness assumptions on the common distribution, reasonable estimates of the parameters converge at the rate  $n^{-1/2}$ . There are, of course, nonparametric problems of some vintage, for example, that are concerned with estimation of the common distribution function of the observations. In that context one still has convergence at the rate  $n^{-1/2}$ . In estimating the distribution function, one is estimating what one might call a global function. In recent years, there has been considerable interest in function estimation with a local character, for example, of a probability density, a regression function or a spectral density. There, typically the rates of convergence are slower. There is by now an enormous literature and we shall by no means try to cover it. Unfortunately we may not be fully accurate in attributing ideas or developments to those most responsible. Perhaps the best we can do is to follow a few suggestive ideas that touch on many developments and give some insight into typical results and directions. It should also be noted that when one tries to extend results for such local curve estimates to dependent observations with short-range dependence, many of the asymptotic results have the same character as in the case of independent observations. A basic motivation for investigations of this type is a skepticism or doubt relative to the usual assumptions in the classical finite parameter theory. The usual assumption of a specific finite parameter family of densities is regarded as unconvincing. The idea is that the data should be used to estimate or test the distributional or regression form. Practically all results have an asymptotic character and before applying such results one should try to get an idea of the extent to which such asymptotics provide useful finite sample approximations. We cannot pursue this difficult but important question here. Our exposition will mainly center on

the asymptotic results. An interesting discussion of the applicability of such results can be found in the book of Silverman (1986).

Let us first consider estimates of the density function. Let  $X_1, X_2, \dots, X_n$  be independent identically distributed random variables with density function  $f(x)$ . There are many possible types of estimates of  $f(x)$ , each based on a different representation of  $f$  as, for example, in terms of an orthogonal or Fourier representation. For the sake of simplicity, we consider a kernel estimate

$$f_n(x) = \frac{1}{nb(n)^d} \sum_{j=1}^n \omega\left(\frac{x - X_j}{b(n)}\right)$$

given in terms of a weight function  $\omega(u)$  integrable with

$$\int \omega(u) du = 1.$$

Here the random variables are assumed to be  $d$ -dimensional. The mean of the estimate

$$\begin{aligned} Ef_n(x) &= b(n)^{-d} \int \omega\left(\frac{x-u}{b(n)}\right) f(u) du \\ &= \int \omega(v) f(x - b(n)v) dv \end{aligned}$$

if  $\omega$  is also bounded and the variance

$$\begin{aligned} \sigma^2(f_n(x)) &= \frac{1}{nb(n)^{2d}} \left[ \int \omega\left(\frac{x-u}{b(n)}\right)^2 f(u) du - \left\{ \int \omega\left(\frac{x-u}{b(n)}\right) f(u) du \right\}^2 \right] \\ &= \frac{1}{nb(n)^d} \left[ \int \omega(v)^2 f(x - b(n)v) dv \right. \\ &\quad \left. - b(n)^d \left( \int \omega(v) f(x - b(n)v) dv \right)^2 \right]. \end{aligned}$$

If the density function  $f$  is bounded and is continuous at  $x$  and  $b(n) \rightarrow 0$  as  $n \rightarrow \infty$ , the bias

$$Ef_n(x) - f(x) = \int \omega(v) [f(x - b(n)v) - f(x)] dv = b_n(x)$$

tends to zero as  $n \rightarrow \infty$ . Later we shall see that in an essential way density estimates are biased and there is an interest in getting estimates on the rate at

which the bias tends to zero as  $n \rightarrow \infty$ . The variance  $\sigma_n^2(f_n(x))$  tends to zero as  $n \rightarrow \infty$  if  $f$  is bounded and  $nb(n)^d \rightarrow \infty$ ,  $b(n) \rightarrow 0$ . If  $f$  is continuous at  $x$  and  $f(x) > 0$

$$\sigma^2(f_n(x)) \simeq \frac{1}{nb(n)^d} f(x) \int \omega(v)^2 dv$$

as  $n \rightarrow \infty$ . The mean square error of the estimate at  $x$

$$E|f_n(x) - f(x)|^2 = b_n(x)^2 + \sigma^2(f_n(x)).$$

If  $\omega$  is bounded and  $nb(n)^d \rightarrow \infty$ ,

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

is asymptotically normal with mean zero and variance

$$f(x) \int \omega(v)^2 dv$$

if  $f$  is continuous at  $x$  [see Rosenblatt (1971)]. That implies that under these circumstances

$$E|U_n|^p \rightarrow 2^{p/2} \Gamma((p+1)/2) / \sqrt{\pi},$$

where  $U_n = \{nb(n)^d\}^{1/2} \{f(x) \int \omega(v)^2 dv\}^{-1/2} [f_n(x) - Ef_n(x)]$  if  $p > 0$ . If  $\omega$  has finite moments of second order

$$(2.1) \quad \int |\omega(u)| \prod_{j=1}^d |u_j^{n_j}| du < \infty, \quad \sum_{j=1}^d n_j = 2, \quad n_j \geq 0,$$

and zero first order moments

$$(2.2) \quad \int \omega(u) u_j du = 0, \quad j = 1, \dots, d,$$

with  $f$  continuously differentiable up to second order with bounded derivatives, one can show that the bias

$$b_n(x) = \frac{1}{2} b(n)^2 \sum_{|\alpha|=2} D^\alpha f(x) \int \omega(u) u^\alpha du + o(b(n)^2).$$

Here  $D^\alpha$  represents the derivative with respect to  $x$  of order  $\alpha$ , that is, if  $\alpha = (\alpha_1, \dots, \alpha_d)$ ,  $|\alpha| = \alpha_1 + \dots + \alpha_d$ , is the  $d$ -vector with nonnegative integer entries,  $D^\alpha = (\partial/\partial x_1)^{\alpha_1} \dots (\partial/\partial x_d)^{\alpha_d}$ . Also  $u^\alpha = u_1^{\alpha_1} \dots u_d^{\alpha_d}$ . With finite moments of  $\omega$  up to  $k$ th order, moments of order up to  $k-1$  zero and bounded continuous differentiability of  $f$  up to order  $k$ , an appropriate use of Taylor's formula with remainder shows that

$$b_n(x) \simeq \frac{1}{k!} b(n)^k \sum_{|\alpha|=k} D^\alpha f(x) \int \omega(u) u^\alpha du + o(b(n)^k).$$

However, this condition with  $k > 2$  forces us to use negative weight functions. The best one can do with positive weight functions is  $O(b(n)^2)$  for the bias when one has sufficient smoothness (with bounds on the derivatives). A global

measure of approximation is given by the integrated mean square error

$$\int E|f_n(x) - f(x)|^2 dx = \int b_n(x)^2 dx + \int \sigma^2(f_n(x)) dx.$$

The asymptotic expression for local mean square error under the smoothness conditions assumed up to second order and the conditions (2.1) and (2.2) on the weight function is

$$E|f_n(x) - f(x)|^2 \simeq \frac{1}{4}b(n)^4 \left( \sum_{|\alpha|=2} D^\alpha f(x) \int \omega(u) u^\alpha du \right)^2 + \frac{1}{nb(n)^d} f(x) \int \omega(v)^2 dv + o\left(b(n)^4 + \frac{1}{nb(n)^d}\right).$$

It is clear that one gets the most rapid rate of decay to zero if

$$b(n) = Cn^{-1/(4+d)}$$

with

$$C = \left\{ \frac{Bd}{A} \right\}^{1/(4+d)},$$

where

$$A = \left( \sum_{|\alpha|=2} D^\alpha f(x) \int \omega(u) u^\alpha du \right)^2$$

and

$$B = f(x) \int \omega(v)^2 dv.$$

Then

$$(2.3) \quad E|f_n(x) - f(x)|^2 \simeq n^{-4/(4+d)} A^{d/(4+d)} B^{4/(4+d)} d^{4/(4+d)} \left(1 + \frac{4}{d}\right) \frac{1}{4} + o(n^{-4/(4+d)}).$$

Notice that already in the one-dimensional case the magnitude of the locally optimal  $b(n)$  is  $n^{-1/5}$ , a power that decreases rather slowly to zero as  $n \rightarrow \infty$ .

Let us now look at the integrated mean square error assuming that  $\omega, f$  are also in  $L^2$ . Then

$$\begin{aligned} \int E|f_n(x) - f(x)|^2 dx &= \int \left\{ \int \omega(v) [f(x - b(n)v) - f(x)] dv \right\}^2 dx \\ &+ \left[ \frac{1}{nb(n)^d} \int \omega^2(v) \int f(x - b(n)v) dx dv \right. \\ &\quad \left. - \frac{1}{n} \int \left( \int \omega(v) f(x - b(n)v) dv \right)^2 dx \right] \\ &= (1) + (2). \end{aligned}$$

Notice that

$$(1) = \int \int \omega(v)\omega(v') \int [f(x - b(n)v) - f(x)] [f(x - b(n)v') - f(x)] dx dv dv'.$$

The inner integral in  $x$  tends to zero as  $b(n) \downarrow 0$  and is bounded in absolute value by a function integrable with weight function  $|\omega(v)\omega(v')|$ .

The suggestion of Epanechnikov (1969) was that one ought to look for a weight function (in the case  $d = 1$ ) that minimizes

$$\int \omega^2(v) dv$$

subject to the restraints (i)  $\int \omega(v) dv = 1$ ; (ii)  $\omega(v) = \omega(-v)$ ; (iii)  $\int v^2 \omega(v) dv = 1$ . It is easy to see that the solution to this simple variational problem is the weight function

$$(2.4) \quad \omega_0(v) = \begin{cases} \frac{3}{4} 5^{-1/2} (1 - v^2/5) & \text{if } |v| \leq 5^{1/2}, \\ 0 & \text{otherwise.} \end{cases}$$

This variational problem was resolved by Hodges and Lehmann (1956) in a nonparametric context. It is interesting to compare this optimal weight function with others like the uniform and Gaussian weight functions (see Table 2.1) by computing the ratio  $r = \int \omega^2(u) du / \int \omega_0^2(u) du$ , where  $\omega_0$  is the weight function (2.4). It is clear that there is not much of a difference. Thus, the shape of the weight function does not make much of a difference on asymptotic grounds relative to the mean square error. However, in multidimensional problems, the shape of the weight function may be of greater importance. One should note, however, in the one-dimensional case, that from a visual perspective, continuity of the weight function is important. By the Lebesgue convergence theorem, it is clear that (1)  $\rightarrow 0$  as  $b(n) \rightarrow 0$ . Also

$$(2) \approx \frac{1}{nb(n)^d} \int \omega^2(v) dv$$

as  $n \rightarrow \infty$ ,  $b(n) \rightarrow 0$ . It is clear from these simple estimates that if  $\omega, f \in L, L^2$

TABLE 2.1  
A comparison of weight functions

$\omega$	$L = \int \omega^2(u) du$	$r$
$\omega_0$	$3 \cdot 5^{-3/2}$	1
$\begin{cases} 1/6^{1/2} -  y /6 & \text{if }  y  \leq 6^{1/2}, \\ 0 & \text{otherwise} \end{cases}$	$6^{1/2}/9$	1.015
$(2\pi)^{-1/2} e^{-y^2/2}$	$2^{-1} \pi^{-1/2}$	1.015
$\begin{cases} \frac{1}{2} 3^{-1/2} & \text{if }  y  \leq 3^{1/2}, \\ 0 & \text{otherwise} \end{cases}$	$\frac{1}{2} 3^{-1/2}$	1.076

and  $b(n) \downarrow 0$ ,  $nb(n)^d \rightarrow \infty$  as  $n \rightarrow \infty$ , then

$$(2.5) \quad E \int |f_n(x) - f(x)|^2 dx \rightarrow 0.$$

Other global measures of deviation such as

$$E \int |f_n(x) - f(x)| dx$$

that may under some circumstances appear to be more natural [see Devroye and Györfi (1985)] have been proposed. However, detailed estimates are usually easier to obtain for (2.5) and in part for that reason much research has focused on it.

In fact, if  $\omega$  is a bounded symmetric integrable weight function

$$\omega(u) = \omega(-u)$$

with

$$\int |\omega(u)| |u|^2 du < \infty$$

and  $f$  is a bounded density function with continuous partial derivatives up to second order and second order partials in  $L^2$ , one can show that

$$\begin{aligned} E \int |f_n(x) - f(x)|^2 dx &= \frac{1}{4} b(n)^4 \int \left| \sum_{|\alpha|=2} D^\alpha f(x) \int \omega(u) u^\alpha du \right|^2 dx \\ &\quad + \frac{1}{nb(n)^d} \cdot \int \omega(u)^2 du + o\left(b(n)^4 + \frac{1}{nb(n)^d}\right) \end{aligned}$$

as  $n \rightarrow \infty$ ,  $b(n) \rightarrow 0$ ,  $nb(n)^d \rightarrow \infty$ . This can readily be seen by making use of the fact that

$$\begin{aligned} f(x - b(n)u) - f(x) &= -b(n) \sum_j u_j D_j f(x) \\ &\quad + \frac{1}{2} b(n)^2 \sum_{j,k} u_j u_k D_j D_k f(x - \theta b(n)u) \end{aligned}$$

for some  $\theta$ ,  $|\theta| < 1$ , with  $D_j = (\partial/\partial x_j)$ .

One can construct estimates of a regression function patterned on the kernel estimate of a density function. Assume that one has independent identically distributed random variables  $(X_j, Y_j)$ ,  $j = 1, \dots, n$ , with  $X$ 's  $d$ -dimensional and the  $Y$ 's one-dimensional. Let  $g(y, x)$  be the joint density of  $Y$  and  $X$  with  $f(x)$  the marginal density of  $X$ . Let  $r(x)$  be the regression function of  $Y$  on  $X$ ,

$$r(x) = E(Y|X = x).$$

As before, let  $\omega(u)$  be a weight function on  $R^d$  with  $\int \omega(u) du = 1$ . Then a plausible kernel estimate of the regression function  $r(x)$  would be given by

$$r_n(x) = \frac{1}{nb(n)^d} \sum_{j=1}^n Y_j \omega\left(\frac{x - X_j}{b(n)}\right) / f_n(x) = \frac{a_n(x)}{f_n(x)},$$

where  $f_n(x)$  is the kernel estimate of the density function  $f(x)$  based on the weight function  $\omega$  [see Nadaraya (1964) and Watson (1964)]. Here it is not so convenient to directly talk about the mean and variance of  $r_n(x)$ . We shall consider very good approximations to  $r_n(x)$  and determine estimates for the bias and variance of these approximations. Let us first note that

$$\begin{aligned} r_n(x) &= [Ea_n(x) + \{a_n(x) - Ea_n(x)\}]\{Ef_n(x)\}^{-1} \\ &\quad \times \left\{1 + [f_n(x) - Ef_n(x)](Ef_n(x))^{-1}\right\}^{-1} \\ (2.6) \quad &= Ea_n(x)\{Ef_n(x)\}^{-1} + \{a_n(x) - Ea_n(x)\}\{Ef_n(x)\}^{-1} \\ &\quad - \{f_n(x) - Ef_n(x)\}Ea_n(x)\{Ef_n(x)\}^{-2} \\ &\quad + O\left[(a_n(x) - Ea_n(x))^2 + (f_n(x) - Ef_n(x))^2\right]. \end{aligned}$$

It is the expression on the right side of (2.6) aside from the order of magnitude term that is the approximation to  $r_n(x)$  that we shall consider in some detail. Now

$$\begin{aligned} \frac{Ea_n(x)}{Ef_n(x)} &= (r(x)f(x) + \{Ea_n(x) - r(x)f(x)\})\frac{1}{f(x)} \\ (2.7) \quad &\quad \times \left(1 - \frac{Ef_n(x) - f(x)}{f(x)} + O\{Ef_n(x) - f(x)\}^2\right) \\ &= r(x) + \{Ea_n(x) - r(x)f(x)\}f(x)^{-1} - \frac{r(x)}{f(x)}(Ef_n(x) - f(x)) \\ &\quad + O(\{Ef_n(x) - f(x)\}^2 + \{Ea_n(x) - f(x)\}^2). \end{aligned}$$

The second and third terms on the right of (2.7) are analogous to the bias. If the weight function  $\omega(u)$  is symmetric and of compact support with  $f(x)$  and  $r(x)$  continuously differentiable up to second order and  $f(x)$  positive,

$$\begin{aligned} &\{Ea_n(x) - r(x)f(x)\}f(x)^{-1} - \frac{r(x)}{f(x)}(Ef_n(x) - f(x)) \\ (2.8) \quad &= \frac{1}{2} \frac{b(n)^2}{f(x)} \left[ \sum_{|\alpha|=2} \{D^\alpha(r(x)f(x)) - r(x)D^\alpha f(x)\} \int v^\alpha \omega(v) dv \right] \\ &\quad + o(b(n)^2) \end{aligned}$$

as  $b(n) \rightarrow 0$ . Let us assume that  $EY^2 < \infty$  and set

$$r^{(2)}(x) = E(Y^2|X = x).$$

If  $\omega(u)$  is bounded and has bounded support,  $r^{(2)}(x)$ ,  $r(x)$ ,  $f(x)$  are continuous and  $nb(n)^d \rightarrow \infty$  as  $n \rightarrow \infty$ ,  $b(n) \rightarrow 0$ , the expression

$$(2.9) \quad \{a_n(x) - Ea_n(x)\}\{Ef_n(x)\}^{-1} - \{f_n(x) - Ef_n(x)\}Ea_n(x)\{Ef_n(x)\}^{-2}$$

can be shown to be asymptotically normal with mean zero and variance

$$(2.10) \quad \frac{1}{nb(n)^d} \{r^{(2)}(x) - r^2(x)\}f(x)^{-1} \int \omega^2(v) dv$$

to the first order. Expressions (2.8) and (2.10) are corrections of the formulas (3.2) and (3.3) in Rosenblatt (1969). Formula (2.10) follows directly from the fact that

$$\begin{aligned} EY\omega\left(\frac{x-X}{b(n)}\right) &= \int y\omega\left(\frac{x-u}{b(n)}\right)g(y,u) du dy \\ &= \int y\omega(v)g(y, x-vb(n)) dv dy b(n)^d \\ &= \int r(x-vb(n))f(x-vb(n))\omega(v) dv b(n)^d, \end{aligned}$$

where  $g(y, x)$  is the joint density of  $Y, X$ . The formula is obtained by using a Taylor expansion with error for  $r(x)f(x)$ . Given finite moments of  $\omega$  up to  $k$ th order with moments up to order  $k-1$  zero and bounded continuous differentiability of  $r$  and  $f$  up to order  $k$ , the formula (2.8) is replaced by

$$\begin{aligned} &\{Ea_n(x) - r(x)f(x)\}f(x)^{-1} - \frac{r(x)}{f(x)}(Ef_n(x) - f(x)) \\ &= \frac{1}{k!} \frac{b(n)^k}{f(x)} \left[ \sum_{|\alpha|=k} \{D^\alpha(r(x)f(x)) - r(x)D^\alpha f(x)\} \int v^\alpha \omega(v) dv \right] \\ &\quad + o(b(n)^k). \end{aligned}$$

The estimate (2.3) obtained for kernel density estimates is concerned with local convergence at a point. If we look at the global measure when  $d = 1$  we have

$$\begin{aligned} E \int |f_n(x) - f(x)|^2 dx &= \frac{1}{4} b(n) \int \{f''(x)\}^2 dx \left( \int \omega(u) u^2 du \right)^2 \\ &\quad + \frac{1}{nb(n)^d} \int \omega(u)^2 du + o\left(b(n) + \frac{1}{nb(n)^d}\right). \end{aligned}$$



Here again one would have an optimum rate if  $b(n) = cn^{-1/5}$  with  $c = (B/A)^{1/5}$  but now  $A = \int \{f''(x)\}^2 dx (\int \omega(u) u^2 du)^2$ ,  $B = \int \omega(v)^2 dv$ . Then

$$E \int |f_n(x) - f(x)|^2 dx \approx n^{-4/5} A^{1/5} B^{4/5} (5/4).$$

If we could locally optimize  $b(n)$  (so it depends on  $x$ ) and then integrate we would get, conjecturally,

$$\begin{aligned} & \int E |f_n(x) - f(x)|^2 dx \\ &= n^{-4/5} \int |f''(x)|^{2/5} f(x)^{4/5} dx \left( \int \omega(u) u^2 du \right)^{2/5} \left( \int \omega(v)^2 dv \right)^{4/5} 5/4 \\ & \quad + o(n^{-4/5}), \end{aligned}$$

while if we use the optimal  $b(n)$  independent of  $x$  we obtain

$$= n^{-4/5} \left( \int \{f''(x)\}^2 dx \right)^{1/5} \left( \int \omega(u) u^2 du \right)^{2/5} \left( \int \omega(v)^2 dv \right)^{4/5} 5/4 + o(n^{-4/5}).$$

The inequality

$$\int |f''(x)|^{2/5} f(x)^{4/5} dx \leq \left( \int |f''(x)|^2 dx \right)^{1/5}$$

is a special case of the Hölder inequality. The contrast might be large when there are ranges with  $f''$  large but  $f$  small.

In our discussion of kernel estimates, we have assumed implicitly either that one is interested in estimating over all of  $R$ , or if the density function is defined only on interval  $[a, c]$ , that estimation is carried out at a point well away from the boundary points. If the point of estimation is at the boundary or too close to it, one can no longer make use of a weight function symmetric about zero. This suggests that one ought to investigate the existence of nonsymmetric weight functions that will allow one to obtain the same asymptotic properties as in the case of symmetric weight functions. Such weight functions will be useful in estimation at a boundary or close to it. We construct a simple family of weight functions that are useful in the estimation of density functions continuously differentiable up to second order. Our object is to consider a weight function

$$p(t; x) = \begin{cases} [1 - (t - x)^2](\alpha + \beta\{t - x\}) & \text{if } -1 + x \leq t \leq 1 + x, \\ 0 & \text{otherwise,} \end{cases}$$

with the property that

$$\int_{-1+x}^{1+x} p(t; x) dt = 1, \quad \int_{-1+x}^{1+x} tp(t; x) dt = 0.$$

Now

$$\begin{aligned} & \int_{-1+x}^{1+x} p(t; x) dt \\ &= \alpha(t-x) - \alpha \frac{(t-x)^3}{3} + \beta \frac{(t-x)^2}{2} - \beta \frac{(t-x)^4}{4} \Big|_{-1+x}^{1+x} \\ &= \frac{4}{3} \alpha = 1 \end{aligned}$$

and

$$\begin{aligned} \int_{-1+x}^{1+x} tp(t; x) dt &= \int_{-1+x}^{1+x} [(t-x)p(t; x) + xp(t; x)] dt \\ &= \beta \frac{(t-x)^3}{3} - \beta \frac{(t-x)^5}{5} \Big|_{-1+x}^{1+x} + x \\ &= 2\beta \left( \frac{1}{3} - \frac{1}{5} \right) + x = 0 \end{aligned}$$

or  $\frac{4}{15}\beta = -x$ ,  $\beta = -\frac{15}{4}x$ . Therefore on the interval  $[-1+x, 1+x]$

$$p(t; x) = \left[ 1 - (t-x)^2 \right] \left( \frac{3}{4} - \frac{15}{4}x(t-x) \right).$$

The use of this type of weight function is suggested in Müller's [(1988), page 73] study when too close to the boundary. If estimating at  $y$  with bandwidth  $b_n$ , when  $a + b_n \leq y \leq c - b_n$ , one can just make use of

$$\frac{1}{b_n} p\left(\frac{t-y}{b_n}, 0\right)$$

as weight function. However, if  $a \leq y < a + b_n \leq c - b_n$ , the suggestion is that one use weight function

$$\frac{1}{b_n} p\left(\frac{t-y}{b_n}, \frac{a-y+b_n}{b_n}\right).$$

The adjustment to interior behavior is smooth. However, it should be noted that if  $x$  is too large (greater than  $\frac{1}{5}$ ), the weight function will take on negative values.

In the paper of Boneva, Kendall and Stefanov (1971), they generated what they called a histospline estimate of a density function. Basically they did this by considering the histogram corresponding to the bins and carrying out an area matching procedure relative to the histogram. The partial sums of the histogram were fitted by a cubic spline with appropriate boundary conditions. The fitted cubic spline was then differentiated and the derivative was taken as an estimate of the underlying density. We consider the procedure in the formulation of Lii and Rosenblatt (1975) where certain asymptotic local properties were obtained in the interior of the interval considered.

Assume  $f$  is a continuous density on  $[0, 1]$ .  $X_1, \dots, X_n$  are assumed to be independent identically distributed with density  $f$ . For convenience, we assume uniform bin size. Let

$$y_k = F_n\left(\frac{k}{N}\right), \quad k = 0, 1, \dots, N = \frac{1}{h},$$

where  $F_n(x)$  is the sample distribution function and  $h = 1/N$  the bin size. Consider  $s_n(x)$  the cubic spline interpolator of  $F_n$  with knots at  $x_j = j/N$ ,  $j = 0, 1, \dots, N$ , with boundary conditions  $f(0) = s_n'(0) = y_0'$ ,  $f(1) = s_n'(1) = y_n'$ . The results mentioned below are still valid for other conventional boundary conditions for cubic splines such as

$$s_n''(0) = 0 = s_n''(1)$$

or with periodic boundary conditions. The estimate of the density  $f(x)$  is  $s_n'(x)$ .

PROPOSITION. *If  $f \in C^3[0, 1]$ , the bias if  $h \downarrow 0$  is*

$$b_n(x) = Es_n'(x) - f(x) = \frac{f^{(3)}(x)}{4!} h^3 \{(1-r)^4 - r^4 - (1-r)^2 + r^2 + o(1)\},$$

where  $0 < x < 1$ ,  $x \in [x_{i-1}, x_i]$  with  $x_{i-1} = [Nx]/N$  and  $r = (1/h)(x - x_{i-1})$ . Here  $[y]$  is the greatest integer less than or equal to  $y$ .

Let  $\sigma = \sqrt{3} - 2$ . One can then give a comparable estimate for the variance.

PROPOSITION. *Let  $f \in C[0, 1]$ . The variance of the estimate  $s_n'(x)$  of  $f(x)$  is*

$$\frac{f(x)}{nh} A(r) + O\left(\frac{h}{n}\right)$$

if  $0 < x < 1$  is fixed and  $nh \rightarrow \infty$ ,  $h \rightarrow 0$ , where

$$\begin{aligned} A(r) = & 1 - \frac{3(1-\sigma)}{2+\sigma} \left(2r^3 - 2r + \frac{1}{3}\right) + \frac{9}{4} \left(\frac{1-\sigma}{2+\sigma}\right)^2 \\ & \times \left[ \left(2r^2 - 2r + \frac{1}{3}\right)^2 + \left\{ \left(r^2 - \frac{1}{3}\right) + \sigma \left(\frac{1}{3} - (1-r)^2\right) \right\}^2 \frac{1}{1-\sigma^2} \right. \\ & \left. + \left\{ \left(r^2 - \frac{1}{3}\right) + \frac{1}{\sigma} \left(\frac{1}{3} - (1-r)^2\right) \right\}^2 \frac{\sigma^2}{1-\sigma^2} \right]. \end{aligned}$$

Clearly  $h$  plays the role of a bandwidth. From these propositions it seems clear that the histospline has a local oscillation in bias and variance in the interior that is a reflection of the binning. An unpleasant feature of the histospline as an estimate of the density is the possibility, unlikely though it is, that the estimate might be negative. A density estimate that does not have this

feature is the version of a maximum penalized likelihood estimate considered by Silverman (1982).

We exhibit estimates of a density function made in analyzing readings of the turbulent wind velocity in a fixed direction obtained in Kansas and supplied by Dr. Wyngaard. The data were sampled 3200 times per second for about an hour. The data have been discussed in an article of Tennekes and Wyngaard (1972). Figures 2.1, 2.2 and 2.3 correspond to different sections of the left tail. These figures are from Lii and Rosenblatt (1975). (Thanks are due the editor of *Computing and Mathematics with Applications* in which they appeared for giving permission for reproduction of the figures here.) The data were already binned. Estimates are graphed for the left tail of the density of  $-(\partial u/\partial t)$  (with  $u$  velocity and  $t$  time). Three figures cover adjoining sections of the tail. The vertical scale is in terms of a logarithmic transform of the data. The kernel type of fit is in terms of a piecewise linear curve. The weight function used is

$$\omega(x) = \begin{cases} \frac{1}{2} & \text{if } |x| \leq \frac{1}{2}, \\ \frac{1}{4} & \text{if } \frac{1}{2} \leq |x| \leq \frac{3}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

The bandwidth is three bins. There are two spline fits, one with a cell width of one bin and the other with a cell width of two bins. It is already clear that determination of the bandwidth is an important issue. But here it seems to be clear that the range is so wide that different values of the bandwidth should be taken over different sections. The boundary conditions used for the histosplines is  $y_0'' = y_n''$ . The data were also fitted with an exponential  $f(x) = A \exp(-B|x|^C)$  using least squares from -32nd bin to -192nd bin. The estimated values of  $A$ ,  $B$  and  $C$  were  $A = 0.74$ ,  $B = 4.2$ ,  $C = 0.41$ . In the past the proposal was made that the rate of energy dissipation in high Reynold's number turbulence should have a lognormal distribution. The experimental data fitted here suggest a slower rate of decay in the tail. One should note here that the data analyzed are dependent and this in a particular example provides a motivation for looking at properties of density estimates when one has dependence.

In Lecture 1, smoothing splines were briefly mentioned. We shall at this point discuss some results that have been obtained when the errors  $\varepsilon_i$  in (1.1) are orthogonal random variables with

$$E\varepsilon_i \equiv 0, \quad E\varepsilon_i^2 = \sigma^2 > 0.$$

In Rice and Rosenblatt (1983), a measure closely related to (1.2),

$$\frac{1}{n} \left[ \frac{1}{4} \{y_0 + y_n - f(0) - f(1)\}^2 + \sum_{k=1}^{n-1} \{y_k - f(x_k)\}^2 \right] + \lambda \int_0^1 [f^{(2)}(t)]^2 dt,$$

was considered for the case  $m = 2$  where the points  $x_i = i/n$ ,  $i = 0, 1, \dots, n$ . The function  $f(x) = f_n(x, \lambda)$  minimizing this functional is taken as an estimate of the unknown smooth regression function  $g(x)$ . It is clear that as the sample size  $n \rightarrow \infty$ , one should let  $\lambda = \lambda(n) \rightarrow 0$  at an appropriate rate. The

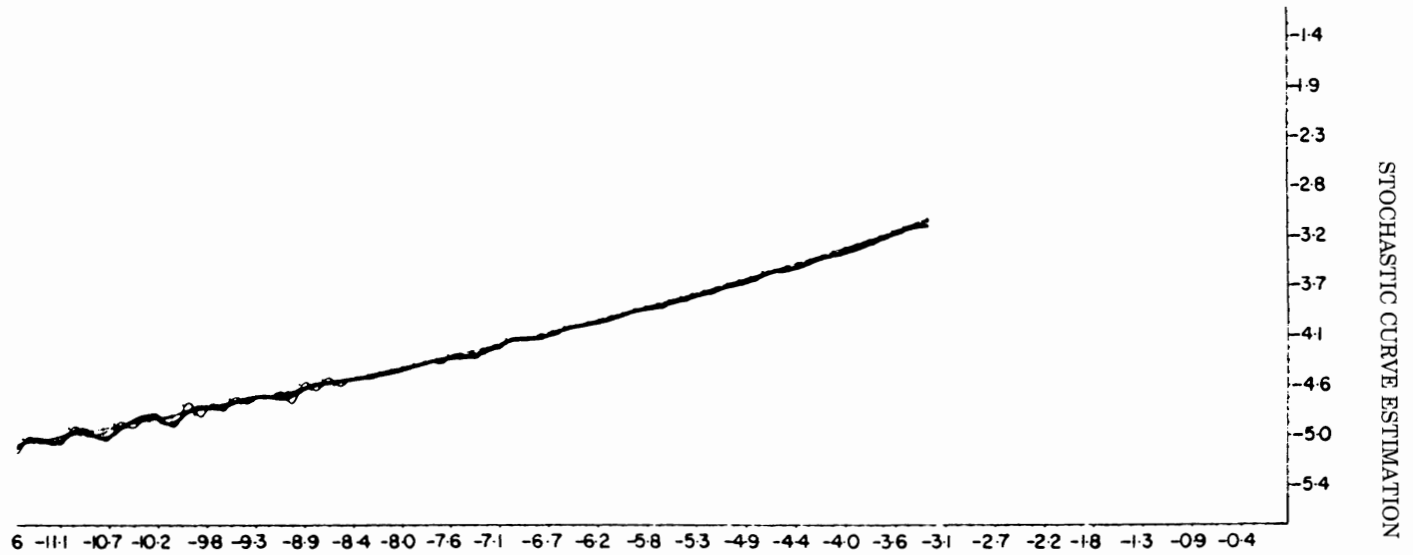


FIG. 2.1. Estimation of left tail of the probability density of turbulent wind velocity derivative. Turbulent Reynold's number  $\approx 8000$ . Histogram, kernel and two spline estimates. Reproduced by permission from Computing and Mathematics with Applications.

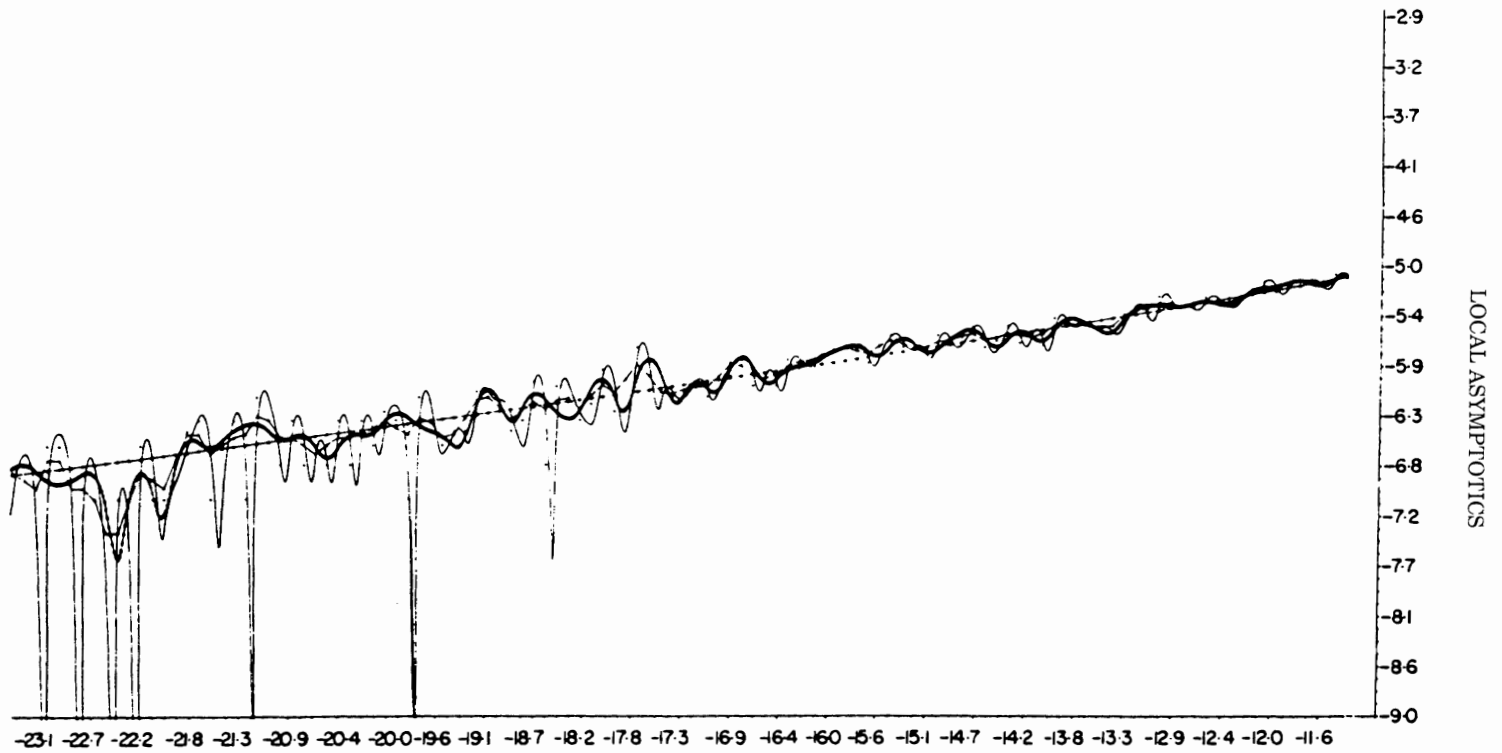


FIG. 2.2. Estimation of left tail of the probability density of turbulent wind velocity derivative. Turbulent Reynold's number  $\approx 8000$ . Histogram, kernel and two spline estimates. Reproduced by permission from Computing and Mathematics with Applications.

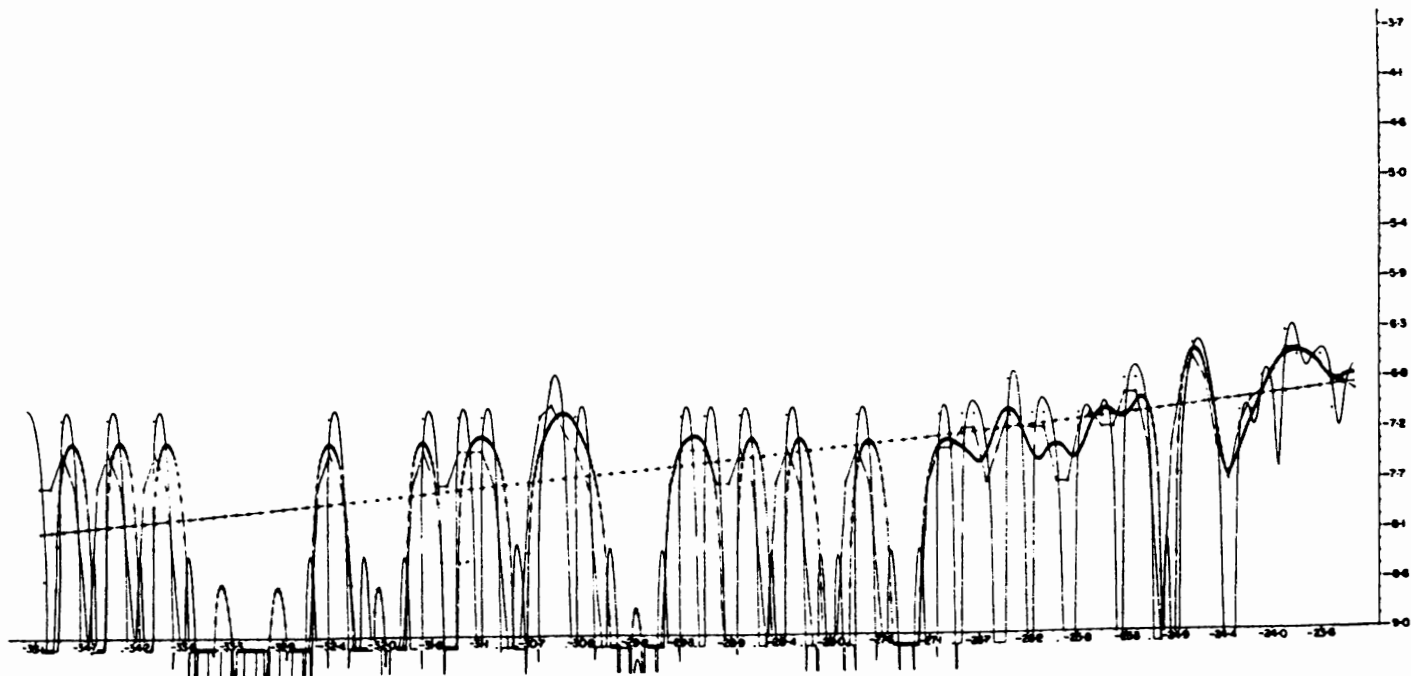


FIG. 2.3. Estimation of left tail of the probability density of turbulent wind velocity derivative. Turbulent Reynold's number  $\approx 8000$ . Histogram, kernel and two spline estimates. Reproduced by permission from Computing and Mathematics with Applications.

paper was concerned with the estimation of

$$\begin{aligned} E \int_0^1 [f_n(x; \lambda(n)) - g(x)]^2 dx \\ = \int_0^1 \sigma^2(f_n(x; \lambda(n))) dx + \int_0^1 \{E f_n(x; \lambda(n)) - g(x)\}^2 dx \end{aligned}$$

as  $n \rightarrow \infty$ . If  $g \in C^2$  and  $\lambda^3(n)n^8 \rightarrow \infty$ ,  $\lambda(n) \rightarrow 0$  as  $n \rightarrow \infty$  it was shown that

$$(2.11) \quad \int_0^1 \sigma^2(f_n(t, \lambda(n))) dt \approx \frac{\sigma^2 \lambda^{-1/4}}{n} 3 \cdot 2^{-7/2}.$$

If one makes the stronger assumption that  $f \in C^4$  and  $f^{(2)}(0)$  or  $f^{(2)}(1) \neq 0$

$$\int [E f_n(x) - g(x)]^2 dx \approx [\{f^{(2)}(0)\}^2 + \{f^{(2)}(1)\}^2] \lambda^{5/4} 2^{-3/2}$$

and if  $f^{(2)}(0) = f^{(2)}(1) = 0$ , but  $f^{(3)}(0)$  or  $f^{(3)}(1) \neq 0$ , then

$$\int [E f_n(x) - g(x)]^2 dx \approx [\{f^{(3)}(0)\}^2 + \{f^{(3)}(1)\}^2] \lambda^{7/4} 2^{-3/2}.$$

This shows that stronger smoothness assumptions than  $f \in C^2$  can lead to bias effects dominated by the boundary behavior at 0 and 1. These results were obtained by detailed estimation using Fourier analysis.

A sketch of an argument given by Speckman (1981) as it relates to (1.2) for positive integral  $m$  is now given. Let

$$W_2^{(m)} = \{f \in L^2[0, 1]; D^{(m)}f \in L^2([0, 0])\}.$$

The points  $x_i = x_{i,n}$ ,  $i = 1, \dots, n$ , are assumed to satisfy the following condition. Let  $p(t)$  be a continuous function bounded away from zero on  $[0, 1]$  and such that

$$\int_0^1 p(t) dt = 1.$$

Assume that

$$\frac{2i-1}{2n} = \int_0^{x_{i,n}} p(t) dt.$$

It is then suggested that for  $f, g \in W_2^{(n)}$ ,

$$\frac{1}{n} \sum_{i=1}^n f(x_i) g(x_i) \rightarrow \int_0^1 f(t) g(t) p(t) dt$$

as  $n \rightarrow \infty$  and so it is reasonable to approximate

$$(2.12) \quad \min_h \left[ \frac{1}{n} \sum_{i=1}^n (h(x_i) - g(x_i))^2 + \lambda \int_0^1 (h^{(m)}(x))^2 dx \right]$$



by

$$(2.13) \quad \min_h \left[ \int_0^1 (h(x) - g(x))^2 p(x) dx + \int_0^1 (h^{(m)}(x))^2 dx \right].$$

We consider the case of just uniform density  $p(x) \equiv 1$  of knots. One actually wants to obtain the random function  $f_n(x; \lambda(n))$  that minimizes (1.2). But

$$h_{n,\lambda}(x) = Ef_n(x; \lambda(n))$$

can be shown to be the solution of (2.13). On considering local variations of the functional in (2.13), one finds that the minimizing function is characterized by the following equation and boundary conditions (the Euler equations for the problem)

$$(2.14) \quad \begin{aligned} (-1)^m \lambda h^{(2m)}(t) + h(t) &= g(t), & t \in [0, 1], \\ h^{(j)}(0) &= h^{(j)}(1), & j = m, m+1, \dots, 2m-1. \end{aligned}$$

The integrated squared bias

$$b_n^2(\lambda) = \int_0^1 (Ef_n(x; \lambda(n)) - g(x))^2 dx$$

for  $m \geq 2$  can be shown to be well approximated by

$$b^2(\lambda) = \int_0^1 (h_\lambda(x) - g(x))^2 dx$$

as  $\lambda(n) \downarrow 0$ ,  $n \rightarrow \infty$ ,  $n^2 \lambda(n) \rightarrow \infty$  with  $h_\lambda$  the solution of the system (2.14). The solution  $h_\lambda$  of (2.14) is given in terms of the Green's function  $G_\lambda(t, s)$  of the system by

$$h_\lambda(t) = \int_0^1 G_\lambda(t, s) g(s) dx.$$

Let

$$D_t^i D_s^j G_\lambda(t, s) = G_\lambda^{i,j}(t, s).$$

$G_\lambda$  is determined by the conditions [refer to Coddington and Levinson (1955) for a discussion]:

(i)  $G_\lambda(t, s)$  is a symmetric function of  $(t, s)$  and satisfies the differential equation of (2.14) for  $t \neq s$  in  $[0, 1]$ .

(ii)  $G_\lambda^{j,0}(0, s) = G_\lambda^{j,0}(1, s) = 0$  for  $j = m, m+1, \dots, 2m-1$  and  $s \in (0, 1)$ .

(iii)  $G_\lambda(\cdot, s) \in C^{(2m-2)}[0, 1]$  if  $s \in [0, 1]$ .

(iv)  $G_\lambda^{2m-1,0}(s+, s) - G_\lambda^{2m-1,0}(s-, s) = (-1)^m \lambda^{-1}$

for  $s \in (0, 1)$  with appropriate one-sided derivatives understood here. Using

the properties just for the Green's function, integration by parts with  $g \in W_2^{(m)}$  leads to

$$\begin{aligned} h_\lambda(t) &= g(t) + \lambda(-1)^m \int_0^1 G_\lambda^{0,2m-1}(t,s) g'(s) ds \\ &= g(t) - \lambda \int_0^1 G_\lambda^{0,m}(t,s) g^{(m)}(s) ds. \end{aligned}$$

Remarks of Messer and Goldstein (1989) on an approximation to the Green's function for  $m = 2, 3$  enable one to get good estimates of the integrated squared bias.

One can show that the estimate for the integrated variance as given in (2.11) for  $m = 2$  is still valid in this formulation.