Oversmoothed Nonparametric Density Estimates

GEORGE R. TERRELL and DAVID W. SCOTT*

The optimal histogram for a sample of size n from a density defined on the entire real line requires at least $(2n)^{1/3}$ bins, under mild smoothness conditions. Similar bounds exist for the frequency polygon and kernel estimators. Values near these bounds give nearly optimal results for a variety of smooth densities, providing good, quick density estimates.

KEY WORDS: Histogram; Frequency polygon; Kernel density estimators; Smoothing.

1. INTRODUCTION

Although elementary methods of nonparametric density estimation have a long history, satisfactory studies of theory are relatively recent. Scott (1979) gave a theoretical treatment of the histogram (see also Freedman and Diaconis 1981), and Scott (1983) developed the theory of the frequency polygon from a similar point of view. The data-based choice of a bin width is a problem beset with serious practical difficulties, since higher derivatives of the density must be estimated. In this article, we find practical lower bounds on the number of bins required to cover the data under certain mild restrictions on the class of densities to be approximated. Quick density estimates based on these lower bounds are shown to give good results for normal data.

2. SOME ASYMPTOTICALLY OPTIMAL DENSITY ESTIMATES

The histogram of a simple random sample, $x_1, \ldots, x_n$, is defined on a mesh, $t_0 < t_1 < \cdots < t_p$, where $t_0 \leqq \min\{x_i\}$ and $\max\{x_i\} < t_p$. We will consider the special case $t_{i+1} - t_i = h \; \forall \; i$, where h is the bin width. Let $f_j = \#\{x_i \mid t_{j-1} \leqq x_i < t_j\}/nh$. Then $\hat{f}(x) = f_j$ for $x \in [t_{j-1}, t_j)$, $j = 1, \ldots, p$, and zero elsewhere. Scott (1979) showed that if the quality of the approximation is measured by the integrated mean squared error (IMSE),

$$IMSE = \int_{-\infty}^{\infty} E[\{\hat{f}(x) - f(x)\}^2],$$

then the IMSE is asymptotically minimized by choosing the bin width

$$h_0 = \left[6 \Big/ n \int_{-\infty}^{\infty} \{f'(x)\}^2 \, dx\right]^{1/3}, \tag{2.1}$$

where $\int_{-\infty}^{\infty} \{f'(x)\}^2 \, dx$ is a measure of the roughness of the underlying density. This depends on the existence of certain derivatives of the density (conditions A in Appendix A; see Freedman and Diaconis 1981). With this choice, the IMSE decreases to zero at a rate proportional to $n^{-2/3}$.

Scott (1983) showed that the IMSE of the frequency polygon is of order $n^{-4/5}$ and thus possesses the ease of use of a histogram and the convergence rate of a kernel density estimate. Let the bin centers $(t_j + t_{j-1})/2$ be denoted by m_j. Then the frequency polygon is

$$\hat{f}(x) = (1/h)(m_{j+1} - x)f_j + (1/h)(x - m_j)f_{j+1}$$

for $m_j \leqq x < m_{j+1}$. Thus $\hat{f}$ is a continuous, piecewise linear function connecting the bin centers of a histogram. In this case, the optimal bin width is

$$h_1 = 2\left[\frac{15}{49} \Big/ n \int_{-\infty}^{\infty} \{f''(x)\}^2 \, dx\right]^{1/5} \tag{2.2}$$

under conditions on the existence of derivatives (conditions B in Appendix A).

There remains the problem of estimating the optimal values of these bin widths from the data. Several authors—including Duin (1976), Habbema et al. (1974), Scott et al. (1977), Silverman (1978), Tarter and Kronmal (1976), and Wahba (1981)—have investigated the closely related problem of estimating smoothing parameters for kernel and series estimators from the data, with some success (Scott and Factor 1981). Each solution is computationally complex and sensitive to certain kinds of difficulties with the data.

3. BOUNDS ON NUMBERS OF BINS

Sturges (1926) suggested that the number of bins in a histogram should be approximately $1 + \log_2 n$, where n is the sample size. More recently, Cencov (1962) noted that the number of bins should be proportional to the cube root of the sample size. Note that these results both depend on the fact that the range of a sample is necessarily finite, so only a finite number of bins is necessary to cover the data.

We begin by assuming that the sample comes from a density f defined on the entire real line but that f is zero outside a finite interval. In Section 5 we show how these results extend to densities supported on an infinite interval. We have the following theorem.

Theorem 1. For a density defined on the real line meeting conditions A, at least $(2n)^{1/3}$ bins are required for the asymptotically optimal histogram.

The derivative conditions require that the analyst believe that the data come from a reasonably smooth underlying distribution. Examples suggest, however, that the inequality is still sharp for some classes of densities that are no more than absolutely continuous. Further, the lower bound still holds true even for discontinuous piecewise-constant densities, since it can then be shown that the optimal number of bins is proportional to the square root of n.

Similarly, we have the next theorem.

Theorem 2. For a density on the real line meeting conditions B, at least $(147n/2)^{1/5}$ bins are required for the asymptotically optimal frequency polygon.

* George R. Terrell is Visiting Assistant Professor and David W. Scott is Associate Professor of Mathematical Sciences, both at Rice University, Houston, TX 77251-1892. This research was supported by ARO Grant DAAG-29-82-K-0014 and NASA/JSC Grant NCC9-10.

Journal of the American Statistical Association
March 1985, Vol. 80, No. 389, Theory and Methods

Again, more bins are required for densities not satisfying conditions B, so the inequality still holds.

4. DENSITIES WITH COMPACT SUPPORT

The following lemma identifies the class of smoothest densities we need.

Lemma 1. Among all densities $g(x)$ defined on the entire real line such that $g(x) = 0$ for $|x| > \frac{1}{2}$, the density

$$f_k(x) = \frac{(2k+1)!}{2^{2k}(k!)^2}(1-4x^2)^k, \qquad |x| \leqq \tfrac{1}{2},$$
$$= 0, \qquad |x| > \tfrac{1}{2}, \tag{4.1}$$

minimizes

$$\int_{-\infty}^{\infty} \{g^{(k)}(x)\}^2\,dx \tag{4.2}$$

for $k = 1, 2, \ldots$.

Proof. It is sufficient to consider only densities $g(x)$ for which $g^{(k-1)}(x)$ is absolutely continuous on the real line; otherwise (4.2) is infinite. Let $e(x) = g(x) - f(x)$, where $f = f_k$ is given by (4.1). It follows from the absolute continuity of $g^{(k-1)}$ and $f^{(k-1)}$ that $\int_{-\infty}^{\infty} e(x)\,dx = 0$ and $e^{(i)}(\pm\frac{1}{2}) = 0$ for $i = 0, \ldots, k-1$. Clearly

$$\int \{g^{(k)}(x)\}^2\,dx = \int \{f^{(k)}(x)\}^2\,dx + \int \{e^{(k)}(x)\}^2\,dx + 2\int f^{(k)}(x)e^{(k)}(x)\,dx.$$

Now

$$\int_{-1/2}^{1/2} f^{(k)}(x)e^{(k)}(x)\,dx = (-1)^k \int_{-1/2}^{1/2} f^{(2k)}(x)e(x)\,dx$$

from integration by parts and from invoking the boundary conditions on $e(x)$. But $f^{(2k)}$ is constant, so by the integral constraint on $e(x)$,

$$\int_{-1/2}^{1/2} f^{(k)}(x)e^{(k)}(x)\,dx = 0.$$

Therefore $\int_{-\infty}^{\infty} \{g^{(k)}(x)\}^2\,dx \geqq \int_{-\infty}^{\infty} \{f^{(k)}(x)\}^2\,dx$.

Proofs of Theorems. By the lemma, among all densities on the real line with support $[-\frac{1}{2}, \frac{1}{2}]$, $f_1(x) = \frac{3}{2}(1-4x^2)$ minimizes $\int_{-\infty}^{\infty} \{g'(x)\}^2\,dx$. We need two further facts about smoothness:

1. $\int \{g'(x+\theta)\}^2\,dx = \int \{g'(x)\}^2\,dx$
2. $\int [\{rg(rx)\}']^2\,dx = r^3 \int \{g'(x)\}^2\,dx$

for any density g for which $\int \{g'(x)\}^2\,dx$ exists. Further, $\int \{f_1'(x)\}^2\,dx = 12$. Let D be the length of the interval of support of the density g. Then by using the preceding facts, the result of the lemma may be restated as $\int \{g'(x)\}^2\,dx \geqq 12/D^3$. Now the class of densities that meet conditions A and have an interval of support of length D is a subset of the class of densities in the lemma with $k = 1$, so the inequality is still true for the smaller function space. From (2.1), it follows that

$$h_0 \leqq [D^3/2n]^{1/3}, \tag{4.3}$$

so the number of bins $\geqq D/h_0 \geqq (2n)^{1/3}$, which is Theorem 1.

The proof of Theorem 2 is precisely parallel. Again by the lemma, among all densities on the real line that have support $[-\frac{1}{2}, \frac{1}{2}]$, $f_2(x) = \frac{15}{8}(1-4x^2)^2$ minimizes $\int \{g''(x)\}^2\,dx$. We have

1. $\int \{g''(x+\theta)\}^2\,dx = \int \{g''(x)\}^2\,dx$
2. $\int [\{rg(rx)\}'']^2\,dx = r^5 \int \{g''(x)\}^2\,dx$

for any density for which the expressions make sense. Further, $\int \{f_2''(x)\}^2\,dx = 720$. Thus the lemma can be written as

$$\int \{g''(x)\}^2\,dx \geqq 720/D^5. \tag{4.4}$$

Again, densities that meet conditions B and have an interval of support of length D constitute a subset of the densities described in the case $k = 2$ of the lemma, so this bound certainly applies to the smaller set. From (2.2), we get

$$h_1 \leqq (2D^5/147n)^{1/5} \tag{4.5}$$

and thus the number of bins $\geqq D/h_1 \geqq (147n/2)^{1/5}$.

5. DENSITIES WITH INFINITE SUPPORT

Theorems 1 and 2 have nontrivial formal content only for underlying densities with a finite interval of support. In this section, we show how the bounds may be applied to infinite-support densities by constructing bins over an interval slightly larger than the sample range. In Section 6, we perform a Monte Carlo experiment to evaluate the loss of efficiency for Gaussian data.

Let F be the continuous cdf of f. Then a well-known result in order statistics is

$$E[F(x_{(1)})] = E[1 - F(x_{(n)})] = 1/(n+1). \tag{5.1}$$

Hence for an interval of support $[a, b] \supseteq [x_{(1)}, x_{(n)}]$ for our histogram or frequency polygon, we may assume that at most a small area of order n^{-1} has been excluded. Then the following result generalizes Lemma 1.

Table 1. Estimated IMSE of Histogram and Frequency Polygon for N(0, 1) Data

	Histogram				Frequency polygon			
Sample size	No. bins	Oversmoothed	Optimal	Efficiency	No. bins	Oversmoothed	Optimal	Efficiency
108	6	.0170	.0162	.952	6	.00691	.00623	.902
864	12	.00516	.00442	.857	9	.00163	.00141	.865
2,916	18	.00258	.00201	.785	12	.000709	.000561	.792
6,912	24	.00157	.00114	.737	14	.000429	.000285	.665

Table 2. Pregnancy Difficulties in 283 Cases

Time of onset (week)	4	5	6	7	8	9	10	11	12	13	14	15	16
No. of cases	3	7	10	13	14	29	22	21	18	28	16	19	10
Time of onset (week)	17	18	19	20	21	22	23	24	25	26	27	28	
No. of cases	13	14	8	4	2	10	4	4	3	4	6	1	

Source: Pearce data (given in Kendall and Stuart 1969).

Lemma 2. Consider the class of all real densities with $F(a) = e_1$ and $1 - F(b) = e_2$, where $e_1, e_2 > 0$ and $e_1 + e_2 < 1$. Then (a) the member of that class with least $\int_{-\infty}^{\infty} f'(x)^2\, dx$ is the unique continuously differentiable piecewise quadratic density with knots at $\alpha < a < b < \beta$, where α, β are determined by a, b, e_1, and e_2 and $f(x) = 0$ for $x > \beta$ and $x < \alpha$; and (b) the member of that class with least $\int_{-\infty}^{\infty} f''(x)^2\, dx$ is the unique piecewise quartic density with knots at $\alpha < a < b < \beta$, which is C^2 on the real line and thrice differentiable at a and b and for which $f(x) = 0$ for $x < \alpha$ and $x > \beta$.

A proof is given in Appendix B. Consider the behavior of the function in (a) for the special case in which $e_1 = e_2 = e$. Without loss of generality, we may choose $[a, b] = [-\frac{1}{2}, \frac{1}{2}]$ so that $f(x)$ will be symmetric about zero and $\alpha = -\beta$. Then the optimal density will have the form

$$\begin{aligned} f(x) &= c_1(x + \beta)^2, && -\beta \leqq x \leqq -\tfrac{1}{2} \\ &= c_2 - c_3x^2, && -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2} \\ &= c_1(x - \beta)^2, && \tfrac{1}{2} \leqq x \leqq \beta. \end{aligned}$$

The four unknowns—β, c_1, c_2, and c_3—are determined by the following conditions: (a) $\int_{1/2}^{\beta} f(x)\, dx = e$, (b) $\int_{-1/2}^{1/2} f(x)\, dx = 1 - 2e$, (c) $f(\frac{1}{2}+) = f(\frac{1}{2}-)$, and (d) $f'(\frac{1}{2}+) = f'(\frac{1}{2}-)$. We find that

$$\begin{aligned} c_3 &= 3[2 + 5e - 3\sqrt{e(e + 4)}] \\ c_2 &= \tfrac{3}{4}[2 - e - \sqrt{e(e + 4)}] \\ \beta &= 2c_2/c_3 = \tfrac{1}{2} + \sqrt{e} \quad \text{for } e \text{ small.} \end{aligned}$$

Thus for the choice $[a, b] = [x_{(1)}, x_{(n)}]$, we expect the interval $[-\beta, \beta]$ to be wider by the fraction $2/(n + 1)^{1/2}$. Let $e \to 0$ in our computation above. Then

$$\begin{aligned} \lim_{e \to 0} f(x) &= \tfrac{3}{2}(1 - 4x^2), && |x| \leqq \tfrac{1}{2}, \\ &= 0, && |x| > \tfrac{1}{2}, \end{aligned}$$

the density of Lemma 1. Further, $\lim_{e \to 0} \int_{-\infty}^{\infty} f'(x)^2 = 12$, as required.

Thus the oversmoothed window width h approaches the width that will require $(2n)^{1/3}$ bins to cover the original histogram support $[a, b]$. We have now verified that the bound given in Theorem 1 is an asymptotic lower bound for a class of densities that is not assumed to have compact support. In a similar manner, we may use (b) to show that the bounds in Theorem 2 are asymptotically valid for this class of densities.

6. INEFFICIENCY DUE TO OVERSMOOTHING

In this section, we study the increase in integrated mean squared error that results from using approximately the numbers of bins in the lower bounds over the sample range of normal data. Histograms and frequency polygons of 2,000 samples of four different sizes were generated by the $N(0, 1)$ pseudo-random-number generator GGNPM in the IMSL library. They were constructed in two ways: first by using the asymptotic optima of (2.1) and (2.2) with bin origin selected randomly from the interval $(0, h)$, and second, by using bin widths h given by the lower bounds $(2n)^{1/3}$ and $(147n/2)^{1/5}$ bins, respectively, divided into the sample range. The IMSE

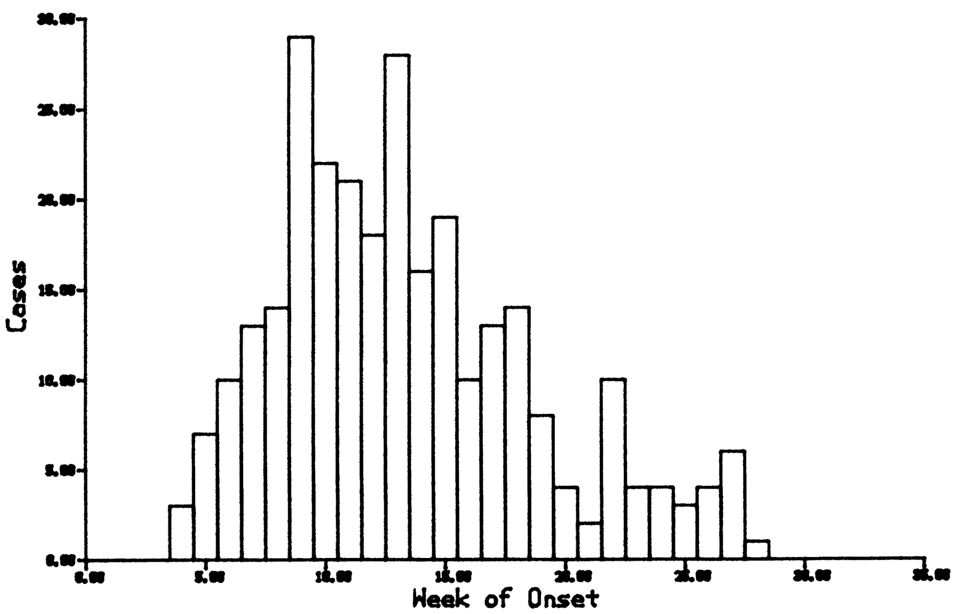

Figure 1. Histogram of Raw Data in Table 1. The bin width is one week.

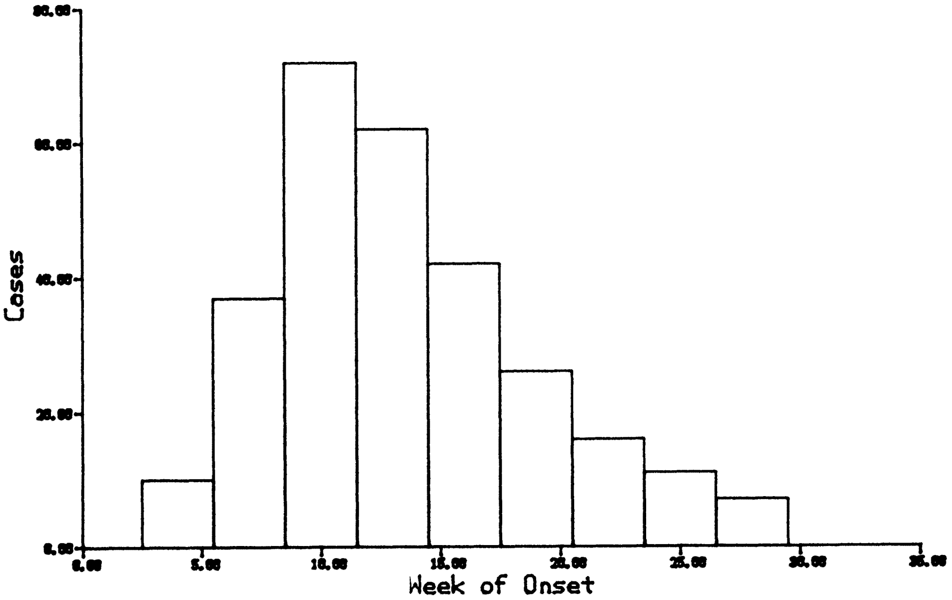

Figure 2. Nearly Optimal Histogram of Table 1 Data. The bin width is three weeks.

was estimated by averaging the 2,000 integrated squared errors, where the histogram bins of width h were constructed to cover at least the interval $(-8, 8)$. The standard error of each average suggests that the digits shown are correct (see Table 1). We conclude that for moderate sample sizes the inefficiency is modest, considering the simplicity of the procedure.

7. APPLICATIONS

The foregoing considerations lead us to propose the following rule of thumb: For a first histogram of data from a density believed to be moderately smooth on the real line, use $(2n)^{1/3}$ equal-width bins or a convenient slightly larger number over the sample range. One of us has successfully taught histogram construction to an introductory applied statistics class, using this as the fundamental principle. A similar rule applies to the frequency polygon: For data presumed to be reasonably smooth, use $(147n/2)^{1/5}$ bins or a convenient slightly larger number for data on the real line.

Kendall and Stuart (1969, p. 27) gave the week of spontaneous miscarriage of 283 pregnancies (reported by T. V. Pearce in 1930; see Table 2). Notice that the manner of reporting these data results in a histogram with a bin width of one week. This histogram is too rough (see Fig. 1). Computing $(2 \times 283)^{1/3} = 8.3$ suggests 9 bins. As we see from Figure 2, the appearance of the resulting histogram suggests skewness to the left, but no other features.

The frequency polygon constructed in like manner must have at least $(147 \times 283/2)^{1/5} = 7.3$ bins. Unfortunately, 8 bins is not compatible with the nearest week prebinning, so once again we use 9 bins (see Fig. 3).

The application of our rule of thumb of course requires us to use the sample range to choose a range for the estimated

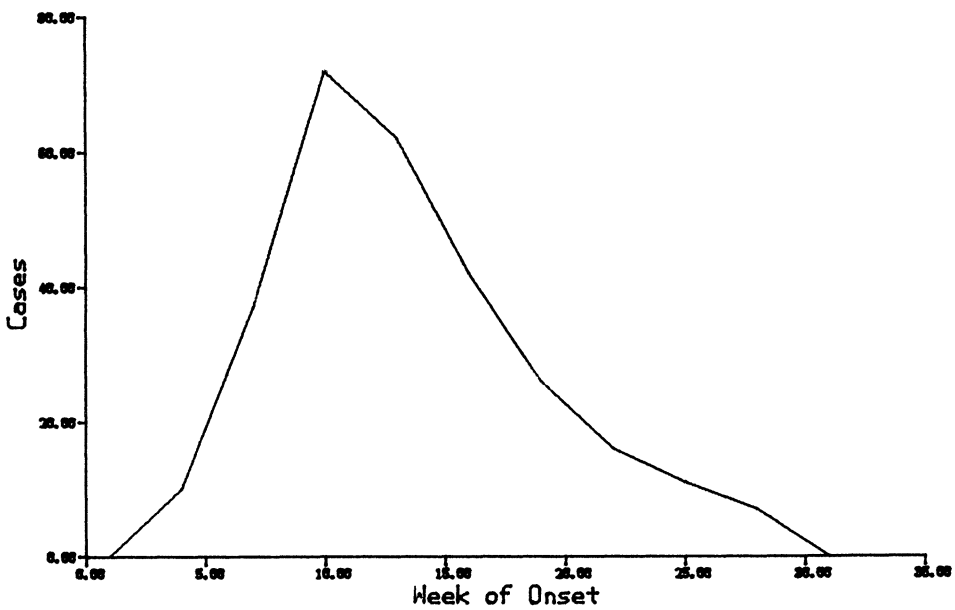

Figure 3. Nearly Optimal Frequency Polygon of Table 1 Data. The bin width is also three weeks.

Table 3. Bin Width Bounds

Density estimator	Bin width bounds	Optimal bin widths for Gaussian data with variance σ^2
Histogram	$h_0 \leqq 3.55\sigma n^{-1/3}$	$h_0 = 3.49\sigma n^{-1/3}$
Frequency polygon	$h_1 \leqq 2.24\sigma n^{-1/5}$	$h_1 = 2.15\sigma n^{-1/5}$
Kernel method (Epanechnikov kernel)	$h_2 \leqq 2.44\sigma n^{-1/5}$	$h_2 = 2.34\sigma n^{-1/5}$

density. Section 5 suggests that this does not cause serious problems for data with normal tails. For very heavy-tailed data, this will no longer be the case; but since the difficulty will be obvious from inspection, Lemma 2 suggests that a trimmed range may be used to choose the bin width. So long as the trimming proportion goes to zero as the sample size goes to infinity, our results still apply.

8. RELATIONSHIP TO BIN WIDTHS

The theorems may be restated in terms of upper bounds on the bin widths h rather than lower bounds on the number of bins (see Sec. 9). The variance of the oversmoothed density f_k in (4.1) is $\sigma_k^2 = [4(2k + 3)]^{-1}$. If f_k is rescaled to have variance σ^2, the new interval of support is $[-(2k + 3)^{1/2}\sigma, (2k + 3)^{1/2}\sigma]$ with length $D_k = 2(2k + 3)^{1/2}\sigma$. Substituting D_1 into (4.3) and D_2 into (4.5) gives the inequalities of the theorems in terms of the bin widths of histograms and frequency polygons, respectively. In Table 3, these bounds may be compared to the optimal bin widths (2.1) and (2.2) for a Gaussian density, which apparently is very smooth.

We may extend this argument to the nonnegative kernel estimator given by

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - x_i}{h}\right),$$

where K is a zero-mean probability density function with variance σ_K^2. Under conditions given in Parzen (1962), the optimal choice of the parameter h is

$$h_2 = \left[\int K(x)^2\,dx \Big/ n\sigma_K^4 \int \{f''(x)\}^2\,dx\right]^{1/5}.$$

Using the inequality (4.4) and D_2 as defined, we obtain the bound given in Table 3 for the particular choice $K(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}(x)$, the Epanechnikov (1969) kernel.

9. CONCLUSIONS

The general approach exemplified by this article is to choose a measure of scale for the random variable under study and find the smoothest distribution with that measure to place a bound on the smoothing parameter. The standard deviation was used as the measure of scale in Terrell (1980) with results nearly identical to the simple heuristic of Section 8. In this article, we have focused on the range as a scale parameter because it is so easily estimated from the data and readily translated into the dimensionless number of bins and because Theorem 1 takes such an elegant form. If following Freedman and Diaconis (1981), we use the interquartile range as a measure of scale, we conclude that $h_0 \leqq 2.60 \times$ [interquartile range] $\times\, n^{-1/3}$, by using $e_1 = e_2 = \frac{1}{4}$ in Lemma 2. They suggest a constant of 2 rather than 2.60, which is consistent with this result.

Wahba (1975) also considered bounds on smoothing parameters in terms of prior knowledge of an upper limit on the norm of the density in the appropriate Sobolev space. These bounds are equivalent to upper bounds on the number of bins as opposed to the lower bounds of our theorems.

APPENDIX A

Conditions A (Freedman and Diaconis 1981). Assume that $f'(x)$ is absolutely continuous, $\int_{-\infty}^{\infty} f'(x)^2\,dx > 0$, and $\int_{-\infty}^{\infty} f''(x)^2\,dx < \infty$.

Conditions B (Scott 1983). Assume that $f''(x)$ is absolutely continuous, $\int_{-\infty}^{\infty} f''(x)^2\,dx > 0$, and $\int_{-\infty}^{\infty} f'''(x)^2\,dx < \infty$.

APPENDIX B: PROOF OF LEMMA 2

We may restrict ourselves to functions that are absolutely continuous, since otherwise the quantity to be minimized is infinite. Therefore we consider variations $\eta(x)$, which, because of the constraints on f, have the properties $\int_{-\infty}^{a} \eta(x) = \int_{a}^{b} \eta(x) = \int_{b}^{\infty} \eta(x) = 0$ and $\eta(x) \geqq 0$ wherever $f(x) = 0$. Let $f(x)$ be our proposed solution. Then the Gateaux variation of $\int f'(x)^2$ is

$$2\int \eta' f' = 2\int_a^\beta \eta' f' = -2\int_a^\beta \eta f''$$

$$= -2k_1\int_a^a \eta - 2k_2\int_a^b \eta - 2k_3\int_b^\beta \eta.$$

The middle term is zero, so

$$= -2k_1\int_{-\infty}^{a} \eta + 2k_1\int_{-\infty}^{a} \eta$$

$$- 2k_3\int_b^\infty \eta + 2k_3\int_\beta^\infty \eta.$$

The first and third terms are zero; further, since f is continuously differentiable at β, then k_1 and k_3 are greater than zero. Since $f(x) = 0$ for $x \geqq \beta$ and $x \leqq a$, then $\eta(x) \geqq 0$ on those intervals. Therefore $2\int \eta' f' \geqq 0$ for all feasible variations η. Our constraint set is convex, and the operator $\int f'(x)^2$ is strictly convex on the set of densities $\int f(x) = 1$; therefore our solution is a true minimum.

The proof of part (b) is similar.

[*Received September 1982. Revised April 1984.*]

REFERENCES

Cencov, N. N. (1962), "Evaluation of an Unknown Distribution Density From Observations," *Soviet Mathematics*, 3, 1559–1562.

Duin, R. P. W. (1976), "On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions," *IEEE Transactions on Computers*, C-25, 1175–1179.

Epanechnikov, V. A. (1969), "Nonparametric Estimators of a Multivariate Probability Density," *Theory of Probability and Its Applications*, 14, 153–158.

Freedman, D., and Diaconis, P. (1981), "On the Histogram as a Density

Estimator: L_2 Theory," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 453–476.

Habbema, J. D. F., Hermans, J., and van den Broek, K. (1974), "A Stepwise Discriminant Analysis Program, Using Density Estimation," in *COMPSTAT 1974, Proceedings in Computational Statistics*, ed. G. Bruckmann, Wien: Physica Verlag, pp. 101–110.

Kendall, M. G., and Stuart, A. (1969), *The Advanced Theory of Statistics* (Vol. 1, 3rd ed.), London: Hafner Press.

Parzen, E. (1962), "On the Estimation of Probability Density Function and Mode," *Annals of Mathematical Statistics*, 33, 1065–1076.

Scott, D. W. (1979), "On Optimal and Data-Based Histograms," *Biometrika*, 66, 605–610.

——— (in press), "Frequency Polygons: Theory and Application," *Journal of the American Statistical Association*, 80.

Scott, D. W., and Factor, L. E. (1981), "Monte Carlo Study of Three Data-Based Nonparametric Density Estimators," *Journal of the American Statistical Association*, 76, 9–15.

Scott, D. W., Tapia, R. A., and Thompson, J. R. (1977), "Kernel Density Estimation Revisited," *Journal of Nonlinear Analysis*, 1, 339–372.

Silverman, B. W. (1978), "Choosing a Window Width When Estimating a Density," *Biometrika*, 65, 1–11.

Sturges, H. A. (1926), "The Choice of a Class Interval," *Journal of the American Statistical Association*, 21, 65–66.

Tarter, M. E., and Kronmal, R. A. (1976), "An Introduction to the Implementation and Theory of Nonparametric Density Estimation," *The American Statistician*, 30, 105–112.

Terrell, G. R. (1980), "A Bound for the Smoothing Parameter in Certain Well-Known Nonparametric Density Estimators," Technical Memorandum No. 14850, Houston: Lockheed/emsco.

Wahba, G. (1975), "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation," *Annals of Statistics*, 3, 15–29.

——— (1981), "Data-Based Optimal Smoothing of Orthogonal Series Density Estimates," *Annals of Statistics*, 9, 146–156.