LOCKHEED

ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

JSC-16669
MAY 9 1980

NASA CR-160657

Ref: 643-7947
Contract NAS 9-15800
Job Order 73-302

TECHNICAL MEMORANDUM

A BOUND FOR THE SMOOTHING PARAMETER IN CERTAIN WELL-KNOWN NONPARAMETRIC DENSITY ESTIMATORS

By

George R. Terrell

(NASA-CR-160657) A BOUND FOR THE SMOOTHING PARAMETER IN CERTAIN WELL-KNOWN NONPARAMETRIC DENSITY ESTIMATORS (Lockheed Engineering and Management) 9 p HC A02/MF A01 CSCL 12A G3/65

N81-11747

Unclas 38955

Approved By: T. C. Minter
T. C. Minter, Supervisor
Techniques Development Section

APRIL 1980

LEMSCO-14850

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-16669		

4. Title and Subtitle	5. Report Date
A Bound For The Smoothing Parameter in Certain Well-Known Nonparametric Density Estimators	April 1980
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
George R. Terrell	LEMSCO- 14850
9. Performing Organization Name and Address	10. Work Unit No.
Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058	11. Contract or Grant No. NAS 9-15800
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered Technical Memorandum
National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Two classes of nonparametric density estimators, the histogram and the kernel estimator, both require a choice of smoothing parameter, or "window width." The optimum choice of this parameter is in general very difficult. This note proposes an upper bound to the choices that depends only on the standard deviation of the distribution.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Density estimation, histogram, Parzen kernel, nonparametric methods.	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	11	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75)

NASA -- JSC

CONTENTS

1. INTRODUCTION

The most commonly used methods of clustering and discrimination in picture-element (pixel) data sets from remote sensing problems involve the assumption of normality of subsets of the data. Since the distributions are often very complex, use of this assumption can introduce error. Alternative, so-called nonparametric techniques may be preferable as a first step. This memorandum suggests some simple methods for choosing an initial nonparametric description of such data sets.

2. NONPARAMETRIC DENSITY ESTIMATORS

For the reasons cited in the Introduction, it is often desirable to have an estimate of the unknown density function $f(x)$ of an absolutely continuous random variable. This may be done by ascribing $f(x)$ to a family of functions, such as the normal family, and then estimating those parameters which distinguish the members of the family from each other. If one is unwilling to assert in advance that the underlying distribution for a data set belongs to some parametric family, then one may use nonparametric estimators of density. The most common of these is the histogram; but there are others such as kernel estimators, k-th nearest neighbor estimators, and orthogonal series estimators. We will concentrate on the first two, although similar principles would seem to apply to the others mentioned.

The constant width histogram is obtained by partitioning the domain of the random variable into intervals of a certain width h. Then, if x falls in the interval (x_i, x_i+h),

$$\hat{f}(x) = \frac{1}{nh} \text{ (the number of data points which fall in } (x_i, x_i+h) \text{, out of a sample of } n)$$

Notice that the appropriate "window width" h remains to be determined. If it is too large, features of the distribution may be obscured. If it is too small, peculiarities of the particular sample may dominate the estimate. Using the criterion of minimum Integrated Mean Squared Error, Scott (1979) has determined that the asymptotically optimal value for h is

$$h_n^* = (6/\int f'(x)^2 dx)^{1/3} n^{-1/3}$$

The unknown quantity $\int f'(x)^2 dx$ remains to be estimated from the data. This may be difficult for small n and time consuming for large n. Scott proposes that for a quick choice of h, it be assumed that the distribution is normal, giving

$$h_n^* = 3.49 s n^{-1/3}$$

where s is the sample standard deviation. His empirical studies indicate that $\int f'(x)^2 dx$ is often larger than for a normal distribution, that is, many distributions of interest are less smooth than the normal.

The Parzen (1962) kernel density estimator is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{x_i} k\left(\frac{x-x_i}{h}\right)$$

where k is a standard density function, called the kernel, symmetric with mean zero and variance one. h is once again called the window width, and is a measure of how far away from each data point its influence is felt. Once again, using the Integrated Mean Squared Error criterion, Epanechnikov (1969) established that the optimum choice of h is

$$h_n^* = n^{-1/5}\left(\int f''(x)^2 dx\right)^{-1/5} \left(\int k^2(x)dx\right)^{1/5}$$

He further established the optimum choice for k, and also noted that the choice of kernel makes comparatively little difference to the integrated mean squared error. We cnce again face the problem of estimating a measure of smoothness of the underlying distribution, $\int f''(x)^2 dx$, with only the data to help us. This involves considerable computational and statistical difficulty. We could follow the lead of the previous section and assume that a reasonable value of h would be found by assuming the data is normal. Thus, using the uniform kernel, we get

$$h_n^* = 1.06412sn^{-1/5}$$

This may be a good starting point for the exploratory adjustment of h. The arbitrary nature of the choice is still disquieting.

3. SMOOTHEST DISTRIBUTIONS

We will proceed to show that, under certain mild restrictions, there exists a smoothest distribution from the point of view of the histogram and a smoothest distribution from the point of view of the kernel estimators. These will immediately provide us with an upper bound to the choice of smoothing parameter h.

Consider distributions f with two continuous derivations on the whole real line. We will construct an f with fixed variance that has minimum $\int f'(x)^2 dx$. We will use a variational argument very similar to that used by Epanechnikov (1969) in constructing the optimal kernel. Assuming f has variance one, we will consider a slightly varied distribution $f+\delta$ such that $\int\delta = 0$ $\int x\delta = 0$ $\int x^2\delta = 0$. Thus $\int(f'+\delta')^2 - \int f'^2$ will be negligible compared to δ, so $\int f'\delta' = 0$. Integrating by parts, this becomes $\int f''\delta = 0$. Since δ is arbitrary up to the given restrictions, f'' must be a quadratic polynomial, and so f is a quartic polynomial. By an argument very close to Epanechnikov's, the distribution with minimal $\int f'(x)^2 dx$ is

$$f(x) = \begin{cases} \frac{15}{16}(1-x^2)^2 & \text{on } [-1,1] \\ 0 & \text{elsewhere} \end{cases}$$

up to an affine change of variables. It may be noted that $\int f'(x)^2 dx$ for the unit normal distribution is .14105 whereas for the smoothest distribution with variance one it is .1157.

With the criterion $\int f''(x)^2 dx$ for smoothness in the expression for the optimal kernel width we can do a similar computation. Among all distributions with four continuous derivatives, there is one for which $\int f''(x)^2 dx$ is minimal given a fixed variance. We can derive it just as above, to get

$$f(x) = \begin{cases} \frac{35}{32}(1-x^2)^3 & \text{on } [-1,1] \\ 0 & \text{elsewhere} \end{cases}$$

up to an affine change of variables. $\int f''(x)^2 dx$ for the unit normal is .21157 and for the minimum valued function with variance one it is .14403.

4. APPLICATIONS

We can now construct an upper bound to the bin width h for a histogram in terms of the sample standard deviation s.

$$h_{max} \cong 3.71 s n^{-1/3}$$

Any histogram with bin width much wider than this is presumably oversmoothed. Notice that this is only seven percent wider than the normal optimum proposed by Scott (1979). According to his results on sensitivity to optimum width, the maximum width histogram would only be one half of one percent larger than the optimum in integrated mean squared error. Thus, the maximum width histogram is a good starting point for graphical and iterative methods of seeking optimal representation. It carries with it the additional reassurance that adjustments in bin width need proceed in only one direction, toward smaller values.

Similarly, the upper bound for the optimal choice of smoothing parameter h for a kernel density estimate is

$$h_{max} \cong 1.15 s n^{-1/5}$$

using the sample standard deviation, the uniform kernel, and $\int f''(x)^2 dx$ for our smoothest density. Again, it is only slightly larger than the optimal value for a normal distribution, and so is a plausible initial value in the downward search for the optimum of an unknown density.

Clearly this variational approach may be generalized to other density estimation procedures in which the optimum value of the smoothing parameter depends on some related measure of density smoothness. For example, Terrell and Scott (1980) propose an estimator with a higher rate of convergence than a Parzen estimator whose optimal smoothing parameter depends upon the second and fourth derivative of the density. It should be noted that our density-free choices for smoothing parameter may be less useful in the case of higher-order methods than the kernel method because of the increasing sensitivity of such methods to deviation from the optimal window width.

5. REFERENCES

1. Epanechnikov, V. A.: Nonparametric Estimates of a Multivariate Probability Density. Theory of Probability and Its Applications, 14, pp. 153-158, 1969.

2. Parzen, E.: On The Estimation of Probability Density and Mode. Annals of Mathematical Statistics, 33, pp. 1065-1070, 1962.

3. Scott, D. W.: On Optimal and Data Based Histograms. Biometrika, 66, pp. 605-610, 1979.

4. Terrell, G. R.; and Scott, D. W.: On Improving Convergence Rates For Nonnegative Kernel Density Estimators. Annals of Statistics (to appear), 1980.