



## Using Kernel Density Estimates to Investigate Multimodality

B. W. Silverman

*Journal of the Royal Statistical Society. Series B (Methodological)*, Volume 43, Issue 1 (1981), 97-99.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9246%281981%2943%3A1%3C97%3AUKDETI%3E2.0.CO%3B2-V>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Journal of the Royal Statistical Society. Series B (Methodological)* is published by Royal Statistical Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

---

*Journal of the Royal Statistical Society. Series B (Methodological)*  
©1981 Royal Statistical Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

## Using Kernel Density Estimates to investigate Multimodality

By B. W. SILVERMAN

*University of Bath, U.K.*

[Received August 1980]

### SUMMARY

A technique for using kernel density estimates to investigate the number of modes in a population is described and discussed. The amount of smoothing is chosen automatically in a natural way.

**Keywords:** DENSITY ESTIMATE; MODE; BOOTSTRAP; TOTAL POSITIVITY; CHONDRITES; BUMP HUNTING

### 1. INTRODUCTION

INVESTIGATION of the number of modes or maxima in a density or its derivative has been considered by several authors, for example Cox (1966) and Good and Gaskins (1980). Most methods seem to depend on some arbitrary implicit or explicit choice of the scale of the effects being studied; see the remarks of Silverman (1980). The simple approach based on kernel density estimates described in this note has the virtue of making this choice in an automatic and natural way. The test statistic used is defined in Section 2 below, and in Section 3 a technique for assessing significance is described. Finally in Section 4 an illustrative application is given.

### 2. THE CRITICAL WINDOW WIDTH

A possible test statistic for hypotheses concerning the number of modes in the density can be obtained by constructing kernel density estimates of the data. The kernel density estimate (Rosenblatt, 1956) for window width  $h$  based on univariate observations  $X_1, \dots, X_n$  is defined by

$$\hat{f}(t; h) = n^{-1} h^{-1} \sum_{i=1}^n K\{h^{-1}(t - X_i)\}, \quad (1)$$

where  $K$  is a kernel function, which we shall assume throughout to be the normal density function. Apart from the theoretical advantages of this choice, the use of a normal kernel has strong computational advantages; see Silverman (1981).

The window width  $h$  controls the amount by which the data are smoothed to obtain the kernel estimate. Thus, for example, if the data are strongly bimodal a large value of  $h$  will be needed to obtain a unimodal estimate. Suppose that we wish to test the null hypothesis that the density  $f$  underlying the data has  $k$  modes, against the alternative that  $f$  has more than  $k$  modes; often  $k = 1$ . Define the  $k$ -critical window width  $h_{\text{crit}}$  by

$$h_{\text{crit}} = \inf\{h; \hat{f}(\cdot, h) \text{ has at most } k \text{ modes}\}. \quad (2)$$

Large values of  $h_{\text{crit}}$  will reject the null hypothesis. Silverman (1978) used a critical value of a smoothing parameter in a somewhat different context. The computation of  $h_{\text{crit}}$  in practice is facilitated by the following theorem and corollary.

*Theorem.* Given any fixed  $X_1, \dots, X_n$ , define  $\hat{f}(t; h)$  as in (1) above, using a normal kernel  $K$ . For each integer  $m \geq 0$ , the number of maxima as  $t$  varies in  $\partial^m \hat{f} / \partial t^m$  is a right continuous decreasing function of  $h$ .

The following corollary follows at once.

*Corollary.* Defining  $h_{\text{crit}}$  as in (2) above,  $\hat{f}(\cdot; h)$  has more than  $k$  modes if and only if  $h < h_{\text{crit}}$ .

The corollary shows that  $h_{\text{crit}}$  can be found by a binary search procedure, since for any value of  $h$  we can tell at once whether or not  $h < h_{\text{crit}}$  by counting the number of modes in  $\hat{f}(\cdot; h)$ . The result is also used in Section 3 below.

This section is concluded with the proof of the theorem, which makes use of the theory of total positivity; see, for example, Karlin (1968). Let  $v_{m+1}(h)$  denote the number of sign changes in  $\hat{f}^{(m+1)}(\cdot, h)$ . Since  $(-t)^{m+1} \hat{f}^{(m+1)}(t, h)$  is, for all  $m \geq 0$  and  $h$ , eventually positive as  $t \rightarrow -\infty$  and as  $t \rightarrow \infty$ , it suffices to show that  $v_{m+1}$  is decreasing and right continuous. For  $h_2 > h_1 > 0$ ,  $\hat{f}^{(m+1)}(\cdot, h_2)$  is the convolution of  $\hat{f}^{(m+1)}(\cdot, h_1)$  with a  $N(0, h_2^2 - h_1^2)$  density, and  $\hat{f}^{(m+1)}(\cdot, h_1)$  is continuous and bounded. Thus, by Theorem 2 of Schoenberg (1950),  $v_{m+1}(h_2) \leq v_{m+1}(h_1)$  so that  $v_{m+1}$  is decreasing. Now suppose, for given  $h_0 > 0$ , there exist  $a_1 < b_1 < a_2 < \dots < a_r < b_r$  such that  $\hat{f}^{(m+1)}(a_i, h_0) > 0$  and  $\hat{f}^{(m+1)}(b_i, h_0) < 0$  for all  $i$ . By the continuity of  $\hat{f}^{(m+1)}(t, \cdot)$ , for all sufficiently small  $\varepsilon$  and all  $i$ ,  $\hat{f}^{(m+1)}(a_i, h_0 + \varepsilon) > 0$  and  $\hat{f}^{(m+1)}(b_i, h_0 + \varepsilon) < 0$ . Hence  $\liminf_{h \downarrow h_0} v_{m+1}(h) \geq v_{m+1}(h_0)$ ; the right continuity of  $v_{m+1}$  follows from the fact that  $v_{m+1}$  is known to be decreasing.

Note that Schoenberg's theorem does not apply for general kernels. Indeed, the convolution of unimodal densities need not be unimodal; see Feller (1966, p. 164).

### 3. ASSESSING SIGNIFICANCE

For any particular  $k$ -modal simple null hypothesis, it is easy to assess, by simulation, the significance of the value of the critical window width obtained from the data. Suppose the null hypothesis is that the true density is  $g$  and that the value of  $h_{\text{crit}}$  obtained from the data is  $h_0$ . Then the theory of Section 2 implies that

$$\text{pr}_g(h_{\text{crit}} > h_0) = \text{pr} \{ \hat{f}(\cdot; h_0) \text{ has more than } k \text{ modes} \mid \{X_1, \dots, X_n\} \text{ is drawn from } g \}.$$

Thus, in order to assess the significance of  $h_0$  for sample size  $n$ , it is only necessary to calculate the single density estimate  $\hat{f}(\cdot; h_0)$  for each sample of size  $n$  generated from  $g$ ; there is no need to find  $h_{\text{crit}}$  for each replication.

The hypothesis that the true density is at most  $k$ -modal is of course a compound hypothesis. To provide a conservative assessment of the significance of  $h_0$ , an appealing choice of the representative  $g_0$  from which to simulate is obtained by rescaling  $\hat{f}(\cdot, h_0)$ , as constructed from the data, to have variance equal to the sample variance. The theory of Section 2 shows that  $g_0$  is indeed at most  $k$ -modal; it is, in a sense, the most extreme  $k$ -modal density consistent with the data. It is extremely easy to simulate from  $g_0$ ; Efron (1979) pointed out that independent observations  $y_i$  from  $g_0$  are given by

$$y_i = (1 + h_0^2/\sigma^2)^{-\frac{1}{2}}(X_{I(i)} + h_0 \varepsilon_i),$$

where  $X_{I(i)}$  are sampled uniformly, with replacement, from the data  $X_1, \dots, X_n$ ,  $\sigma^2$  is the sample variance of the data, and  $\varepsilon_i$  is an independent sequence of standard normal random variables.

Simulating from  $g_0$  to assess significance is an example of a smoothed bootstrap procedure as defined by Efron (1979). However, Efron's procedure contains an implicit arbitrary choice of smoothing parameter, since his  $\sigma_\varepsilon^2$  is essentially arbitrary. In our case, the amount of smoothing is automatically determined in a natural way.

Finally, it should be pointed out that the theory and procedure of finding a critical window width and simulating from a rescaled density estimate constructed using this window width carries over immediately, *mutatis mutandis*, to the investigation of maxima in the first or higher derivative of the data. Both Cox (1966) and Good and Gaskins (1980) show a preference for seeking maxima in the density derivative.

### 4. AN APPLICATION

We illustrate the method by analysing a small data set of observations on chondrite meteors. These data consist of 22 observations which are given in Table 2 of Good and Gaskins (1980).

TABLE 1  
*Chondrite data: critical window widths and their  
estimated significance levels*

<i>Number of modes</i>	<i>Critical window width</i>	<i>P</i>
1	2.39	0.92
2	1.83	0.95
3	0.68	0.21
4	0.47	0.07

The data have been considered by several authors; see Good and Gaskins (1980) for details. In this analysis the raw values of the observations were used. Table 1 gives critical window widths and significance levels for tests of the null hypothesis that the underlying density has at most  $k$  modes against the alternative that it has more than  $k$  modes. The  $p$ -values are computed by simulating from a critical density as described above; 100 replications of 22 observations were used in each case.

These results must of course be interpreted as a hierarchical set of significance tests. All other things being equal, considerations of parsimony perhaps suggest that we should test successively for an increasing number of modes until we find a number that is accepted. Particularly bearing in mind the small sample size, the results clearly indicate the trimodal nature of the population; Good and Gaskins (1980) also arrived at this conclusion.

5. ACKNOWLEDGEMENTS

I am most grateful to Professor J. B. Copas for a discussion which initiated this work, and to the Sonderforschungsbereich 123, University of Heidelberg, for their support.

REFERENCES

COX, D. R. (1966). Notes on the analysis of mixed frequency distributions. *Brit. J. Math. Statist. Psychol.*, **19**, 39–47.  
 EFRON, B. (1979). Bootstrap methods—another look at the jack-knife. *Ann. Statist.*, **7**, 1–26.  
 FELLER, W. (1966). *An Introduction to Probability Theory and its Applications, Volume II*. New York: Wiley.  
 GOOD, I. J. and GASKINS, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *J. Amer. Statist. Ass.*, **75**, 42–56.  
 KARLIN, S. (1968). *Total Positivity*. Stanford: Stanford University Press.  
 ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832–837.  
 SCHOENBERG, I. J. (1950). On Pólya frequency functions. II: Variation diminishing integral operators of the convolution type. *Acta Scientiarum Mathematicarum Szeged*, **12B**, 97–106.  
 SILVERMAN, B. W. (1978). Density ratios, empirical likelihood and cot death. *Appl. Statist.*, **27**, 26–33.  
 — (1980). Comment on Good and Gaskins (1980). *J. Amer. Statist. Ass.*, **75**, 67–68.  
 — (1981). Density estimation for univariate and bivariate data. In *Interpreting Multivariate Data* (V. Barnett, ed.) Chichester: Wiley.