Bootstrap Choice of Bandwidth for Density Estimation

JULIAN J. FARAWAY and MYOUNGSHIC JHUN*

A bootstrap-based choice of bandwidth for kernel density estimation is introduced. The method works by estimating the integrated mean squared error (IMSE) for any given bandwidth and then minimizing over all bandwidths. A straightforward application of the bootstrap method to estimate the IMSE fails because it does not capture the bias component. A smoothed bootstrap method based on an initial density estimate is described that solves this problem. It is possible to construct pointwise and simultaneous confidence intervals for the density. The simulation study compares cross-validation and the bootstrap method over a wide range of densities—a long-tailed, a short-tailed, an asymmetric, and a bimodal, among others. The bootstrap method uniformly outperforms cross-validation. The accuracy of the constructed confidence bands improves as the sample size increases.

KEY WORDS: Cross-validation; Confidence bands; Kernel density estimation; Smoothed bootstrap.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent observations from a density f. Consider kernel density estimates of the form

$$f_n(x; h) = (nh)^{-1} \sum_{i=1}^{n} K((x - X_i)/h), \qquad (1.1)$$

where h is the bandwidth and K is the kernel with $\int_R K(x)\, dx = 1$.

Appropriate selection of h is essential to achieving good estimates. Numerous methods have been proposed to choose h. Silverman (1986) provides a good introduction for readers unfamiliar with these techniques.

A commonly used criterion for judging the efficacy of these techniques is the integrated mean squared error (IMSE) for bandwidth h.

$$\text{IMSE}(h) = \int \text{E}(f_n(x; h) - f(x))^2\, dx. \qquad (1.2)$$

For this criterion, least squares cross-validation is a popular way to select the bandwidth. Large-sample theory for this method was given by Hall (1983), Stone (1984), and Burman (1985). Bowman (1985) conducted a simulation study to compare some of the competing bandwidth selection procedures and concluded that cross-validation was best. Scott and Terrell (1987) discussed some variations on cross-validation.

In this article, we propose a bootstrap-based choice of bandwidth. The idea is to estimate IMSE using the bootstrap and then minimize over h. However, a direct application of the bootstrap fails. The IMSE may be decomposed into bias and variance terms. The regular bootstrap may adequately estimate the variance, but it is not able to estimate the bias. Therefore, we use a different approach. We obtain an initial estimate of the density with the bandwidth chosen by some other procedure and resample from that. Thus we are able to construct an estimate of the IMSE that captures the bias term as well. This smoothed bootstrap procedure tends to improve upon that initial estimate of the density. Pointwise and simultaneous confidence intervals for the density may also be obtained. Another advantage of our method is that it is easily adaptable to criteria other than the IMSE, such as the integrated mean absolute error or even other functionals of the density.

Hall (1990) described a different bootstrap-based bandwidth selection method. During the review of this article, we also discovered a paper by Taylor (1989) that describes a method similar to ours and points out that, with the use of a Gaussian kernel, the bootstrap estimate of the IMSE may be calculated without resampling. The additional computational cost of our method appears to result in better bandwidth selection and confidence intervals, which is a worthwhile advance over mere bandwidth selection.

In Section 2, we describe the implementation of the estimator, and, in Section 3, we describe a simulation study to illustrate the small-sample properties. In Section 4, we discuss the results of the study. Our main conclusion is that our bootstrap method produces superior estimates of the density compared with the cross-validated method and that plausible confidence bands may be constructed.

2. IMPLEMENTATION

The straightforward approach to using the bootstrap method to estimate the IMSE for a given bandwidth would be to resample $X_1^*, \ldots, X_n^*$ from the empirical distribution F_n and then construct bootstrap estimates

$$f_{nj}^*(x; h) = (nh)^{-1} \sum_{i=1}^{n} K((x - X_i^*)/h) \qquad (2.1)$$

for $j = 1, \ldots, B$, where B is the number of bootstrap samples to be taken. Our bootstrapped estimate of the variance of $f_n(x; h)$ would then be

$$B^{-1} \sum_{j=1}^{B} \int (f_{nj}^*(x; h) - \bar{f}_{nj}^*(x; h))^2\, dx,$$

* Julian J. Faraway is Assistant Professor, Department of Statistics, University of Michigan, Ann Arbor, MI 48109. Myoungshic Jhun is Associate Professor, Department of Statistics, Korea University, Seoul, 136-701 Korea. This research was supported by the Korea Science and Engineering Foundation. The authors thank an associate editor and two referees for improving the quality of this article.

Journal of the American Statistical Association
December 1990, Vol. 85, No. 412, Theory and Methods

Table 1. Sample Size = 50

Distribution	Relative efficiency				Sample comparisons		Confidence level (80%)
	Fixed	CV	b1	b2	b1–CV	b2–b1	
Normal	1.09	2.05	1.69	1.50	58–14	33–10	65
Bimodal	1.04	1.39	1.34	1.38	37–46	24–46	59
Contaminated normal	1.06	1.69	1.51	1.47	38–29	24–29	54
Lognormal	1.07	1.45	1.37	1.38	36–47	25–47	29
Cauchy	1.06	1.57	1.41	1.35	39–31	25–25	47
Beta	1.11	1.84	1.50	1.38	62–18	33–19	64

where

$$\bar{f}^*_{nj}(x; h) = B^{-1} \sum_{j=1}^{B} f^*_{nj}(x; h).$$

However, the usual bootstrap estimate of the bias, $f_n(x; h) - \bar{f}^*_{nj}(x; h)$, vanishes. Since the component of bias increases with h and can be substantial, this naive bootstrap method will fail.

This phenomenon was observed in another context in bootstrap selection of bandwidth for quantile estimates under censoring in Padgett and Thombs (1986). Romano (1988) observed the same effect for kernel density estimates of the mode.

We construct an initial estimate of the density $\hat{f}_n(x; h_0)$ and then resample from that. This can be accomplished by adding a random amount $h_0\varepsilon$ to each resampled X^*_j, where ε is distributed with density $K(\cdot)$. So $X^*_j \to X^*_j + h_0\varepsilon$. This is an example of the smoothed bootstrap where smoothing is not only desirable but necessary [see Silverman and Young (1987) for more on this]. We may then construct $f^*_{nj}(x; h)$ as before, estimating the bias by $\hat{f}_n(x; h_0) - \bar{f}^*_{nj}(x; h)$, and estimate the IMSE($h$) as variance + (bias)^2 by

$$\text{BIMSE}(h, h_0) = B^{-1} \sum_{j=1}^{B} \int (f^*_{nj}(x; h) - \hat{f}_n(x; h_0))^2 \, dx. \tag{2.2}$$

We obtain the bootstrap choice of bandwidth $\hat{h}_b$ by minimizing BIMSE(h, h_0) over h.

The L_2 norm has been the most popular criterion for the choice of bandwidth, but the L_1 norm has its advantages [see Devroye and Györfi (1985) for more on the L_1 norm]. One advantage of the bootstrap method is that it can easily be adapted to this criterion by using

$$\text{BIMAE}(h, h_0) = B^{-1} \sum_{j=1}^{B} \int |f^*_{nj}(x; h) - \hat{f}_n(x; h_0)| \, dx. \tag{2.3}$$

We require some way of selecting h_0. The better this selection is, the better our bootstrap method will be. Of course, we hope that the application of the bootstrap will improve upon this initial choice or it will hardly be worth our trouble. The method we have used to make this initial choice is least squares cross-validation, where h_0 is chosen by minimizing

$$\text{CV}(h) = \int \hat{f}_n(x; h)^2 \, dx - 2n^{-1} \sum_{i=1}^{n} f_{n,-i}(X_i \mid h), \tag{2.4}$$

where $f_{n,-i}$ is the density estimate based on all of the data except X_i. It is possible to iterate this method, that is, use the bootstrap choice of bandwidth as the new initial choice and apply the bootstrap method again. This choice appears to work well for the sample sizes we consider here, although for much larger sample sizes theory suggests some upward adjustment of h_0 might be required.

Construction of pointwise or simultaneous confidence bands is a nice by-product of using the bootstrap method. A pointwise confidence interval for $f(x)$ may be constructed from the appropriate quantiles of $f^*_{nj}(x; \hat{h}_b)$. Simultaneous bands involve more computation: For each bootstrap sample compute $b_j = \sup_x |f^*_{nj}(x; \hat{h}_b) - \hat{f}_n(x \mid h_0)|$.

Table 2. Sample Size = 50: Bandwidths

Distribution	Fixed Bandwidth	ISE		Cross-validated		Bootstrap one-step		Bootstrap two-step	
		Bandwidth	Standard deviation	Bandwidth	Standard deviation	Bandwidth	Standard deviation	Bandwidth	Standard deviation
Normal	1.11	1.07	.20	1.11	.40	1.13	.31	1.18	.25
Bimodal	.64	.64	.09	.68	.22	.77	.22	.84	.24
Contaminated normal	.79	.76	.14	.82	.33	.91	.31	.98	.29
Lognormal	.49	.46	.12	.51	.22	.59	.22	.66	.22
Cauchy	1.22	1.19	.21	1.27	.48	1.40	.43	1.49	.39
Beta	1.20	1.15	.24	1.10	.39	1.17	.29	1.24	.23

Table 3. Sample Size = 400

Distribution	Relative efficiency				Sample comparisons		Confidence level (80%)
	Fixed	CV	b1	b2	b1–CV	b2–b1	
Normal	1.06	1.51	1.29	1.22	64–16	33–14	80
Bimodal	1.06	1.24	1.13	1.10	59–17	24–12	81
Contaminated normal	1.07	1.49	1.28	1.20	52–16	28–10	72
Lognormal	1.04	1.29	1.17	1.16	52–36	27–44	53
Cauchy	1.05	1.40	1.24	1.19	55–20	28–13	66
Beta	1.10	1.48	1.28	1.20	62–17	34–14	72

Now sort the b_j's and select $b_n(\alpha) \equiv b_{[\alpha B]}$. The $100(1 - \alpha)\%$ confidence band is then $\hat{f}_n(x \mid h_b) \pm b_n(\alpha)$.

3. SIMULATION

For reasons of computational efficiency, we use the Epanechnikov kernel.

$$K(x) = .75(1 - x^2), \quad \text{if } |x| < 1,$$
$$= 0 \quad \text{otherwise.}$$

We consider six test distributions: standard normal; bimodal normal, $\frac{1}{2}N(-1, \frac{1}{4}) + \frac{1}{2}N(1, \frac{1}{4})$; contaminated normal, $\frac{1}{2}N(0, 4) + \frac{1}{2}N(0, \frac{1}{4})$; standard lognormal; Cauchy; and beta(2, 2).

The sample sizes used were 50 and 400. $B = 100$ bootstrap samples were used. This is rather low, but the procedure is somewhat expensive computationally and some economy was required. The numerical integration necessary for the computation of integrated squared errors was done using a grid of 100 points. ISE(h), BIMSE(h), and CV(h) were computed for 20 values of h evenly spaced on an appropriately wide log-scale. A more efficient search for the minimum of these functions is not possible because they are not always perfectly convex. Results were based on 400 replications. Uniform random numbers were generated using a standard linear congruential algorithm and were then transformed to the test distributions using widely available algorithms.

Results of the simulation study are given in Tables 1–4. By "fixed" we mean that fixed choice of bandwidth that minimizes the estimated IMSE. This was empirically chosen. Some previous authors have determined this fixed bandwidth based on asymptotics. Experience has shown that the asymptotics may be misleading as well as unnecessary for our present purposes [see Dodge (1986) for further discussion of this].

By "ISE" we mean that the bandwidth is chosen for any given sample to minimize the integrated squared error, where we presume knowledge of the true underlying density. Hence this estimator is the best one could possibly do with a kernel density estimator given complete knowledge. Thus this estimator is a good benchmark with which to measure the performance of our proposed methods.

We give the ratio of the estimated IMSE of a method to that of IMSE of the ISE choice under the heading "relative efficiency." Estimated error here is around 2%. The bootstrap method based on the cross-validated (CV) initial choice is denoted by "b1," and "b2" is the iterated bootstrap choice.

We also give some statistics to answer the question: "For my sample, will the bootstrap improve my estimate?" The percentage of samples where the method produced a lower ISE than the other is given under "sample comparisons." Note that these percentages do not sum to 100% because we are using a grid of bandwidths and hence the remainder represents that percentage of samples where the two methods produced the same bandwidth choice.

We also computed confidence bands as described in the preceding section. The estimated actual confidence of nominally 80% confidence bands are given. Of course, the discretizations we used previously may cause some additional inaccuracy here. Sample averages and standard deviations for the bandwidths chosen are also given.

4. DISCUSSION

We see that the bootstrap performs almost uniformly better than cross-validation, both in term of relative IMSE and sample comparison. Iterating the method produces

Table 4. Sample Size = 400: Bandwidths

Distribution	Fixed Bandwidth	ISE		Cross-validated		Bootstrap one-step		Bootstrap two-step	
		Bandwidth	Standard deviation	Bandwidth	Standard deviation	Bandwidth	Standard deviation	Bandwidth	Standard deviation
Normal	.70	.71	.13	.69	.18	.69	.11	.71	.08
Bimodal	.41	.40	.06	.40	.09	.41	.05	.42	.03
Contaminated normal	.49	.48	.08	.45	.14	.48	.10	.50	.07
Lognormal	.22	.23	.04	.21	.07	.25	.05	.28	.04
Cauchy	.74	.73	.12	.69	.20	.74	.14	.77	.11
Beta	.70	.71	.14	.67	.20	.70	.13	.73	.10

some further improvement, especially for the larger sample size. Note that the method performs well for the Cauchy density, in contrast to the method of Taylor (1989).

The bootstrap method chooses generally larger bandwidths than cross-validation, but the choice has smaller variation. This smaller variation explains the superior performance. This trend continues when the bootstrap method is iterated.

The actual confidence levels fall somewhat short of the nominal 80%, particularly for the smaller sample size. The intervals, however, are much more accurate for the larger sample size.

A negative correlation was observed between the bandwidth of the ISE choice and the bootstrap choice, larger than that between the ISE choice and the cross-validated choice, as was observed by Scott and Terrell (1987).

We conclude that the bootstrap method is superior to the cross-validated method. Of course, the bootstrap method is computationally more expensive, but this will become less of a disadvantage in time. Confidence bands are an added bonus.

[*Received February 1988. Revised May 1990.*]

REFERENCES

Bowman, A. (1985), "A Comparative Study of Some Kernel Based Nonparametric Density Estimators," *Journal of Statistical Computation and Simulation,* 21, 313–327.

Burman, P. (1985), "A Data Dependent Approach to Density Estimation," *Zeitschrift für Wahrscheinlichkeitstheorie and Verwandte Gebiete,* 69, 609–628.

Devroye, L., and Györfi, L. (1985), *Nonparametric Density Estimation: The L_1 View,* New York: John Wiley.

Dodge, Y. (1986), "Some Difficulties Involving Nonparametric Estimation of a Density Function," *Journal of Official Statistics*, 2, 193–202.

Hall, P. (1983), "Large Sample Optimality of Least Squares Crossvalidation in Density Estimation," *The Annals of Statistics,* 11, 1156–1174.

——— (1990), "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems," *Journal of Multivariate Analysis,* 32, 177–203.

Padgett, W., and Thombs, L. (1986), "Smooth Nonparametric Quantile Estimation Under Censoring: Simulations and Bootstrap Methods," *Communication in Statistics, Part B—Simulation and Computation,* 15, 1003–1025.

Romano, J. (1988), "On Weak Convergence and Optimality of Kernel Density Estimates of the Mode," *The Annals of Statistics,* 16, 629–647.

Scott, D., and Terrell, G. (1987), "Biased and Unbiased Cross-validation in Density Estimation," *Journal of the American Statistical Association,* 82, 1131–1146.

Silverman, B. (1986), "Density Estimation for Statistics and Data Analysis," London: Chapman & Hall.

Silverman, B., and Young, G. (1987), "The Bootstrap: To Smooth or Not To Smooth?," *Biometrika,* 12, 469–479.

Stone, C. (1984), "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," *The Annals of Statistics,* 12, 1285–1297.

Taylor, C. (1989), "Bootstrap Choice of Smoothing Parameter in Kernel Density Estimation," *Biometrika,* 76, 705–712.