Bootstrap MSE Estimators to Obtain Bandwidth for Kernel Density Estimation

Jeffrey C. Miecznikowski†
University at Buffalo, Roswell Park Cancer Institute, Buffalo, NY, USA.
Dongliang Wang
University at Buffalo, Roswell Park Cancer Institute, Buffalo, NY, USA.
Alan Hutson
University at Buffalo, Roswell Park Cancer Institute Buffalo, NY, USA.

Summary.
A new method is proposed for estimating mean integrated squared error (MISE) for kernel density estimators (KDEs) . Via the bootstrap we obtain $exact$ estimators for the variance and mean for the KDE and thus we can obtain estimators for MISE. By obtaining estimates for MISE for any given bandwidth, it is possible to obtain an optimal bandwidth that minimizes the estimate of MISE. We provide an overview of other methods to obtain optimal bandwidths and offer a comparison of these methods via a simulation study. The simulation study compares, asymptotic methods, cross-validation and several bootstrap methods over a wide range of densities. In certain situations, our method of estimating an optimal bandwidth yields a smaller MISE than competing methods to compute bandwidths. This procedure is illustrated by an application to two data sets.

Keywords: bandwidth, bootstrap, kernel density, MSE

1. Introduction

Let $X_1, \cdots, X_n$ be independent data points from some distribution with cumulative density function F (and probability density function f) with support over the entire real number line. Let $f(x)$ be the true density (or function) and let $\hat{f}(x)$ be an estimator of $f(x)$. We will define the kernel density estimator (KDE) as:

$$\hat{f}(x;h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right). \tag{1.1}$$

In general, K is the kernel function (e.g. normal density function), n is the number of data points, and h is the bandwidth. A large body of research is devoted to choosing h, essentially the amount of smoothing to apply. Usually, smoothing parameters can be chosen via cross validation or by minimizing a measure of error. A good overview on kernel density estimators is supplied by Silverman (1986); Scott (1992); Mugdadi and Ahmad (2004).

†*Address for correspondence:* Jeffrey C. Miecznikowski, University at Buffalo, Department of Biostatistics, 3435 Main St, Buffalo, NY 14214
E-mail: jcm38@buffalo.edu

To evaluate the kernel density estimators, we first define the error quantities under consideration. Define the integrated squared error (ISE) as:

$$L(f,\hat{f}) = \int (f(x) - \hat{f}(x))^2 dx. \tag{1.2}$$

The risk or mean integrated squared error (MISE) under squared error loss is given by:

$$\text{MISE}(f,\hat{f}) = E\left(L(f,\hat{f})\right), \tag{1.3}$$

where the expectation, E, is taken with respect to the density, f. Note, because we can switch the order of integration, we can re-express the MISE as

$$\text{MISE}(f,\hat{f}) = \int E\left((f(x) - \hat{f}(x))^2\right) dx = \int \text{MSE}(f(x),\hat{f}(x)), \tag{1.4}$$

where the $\text{MSE}(f(x),\hat{f}(x))$ is the mean squared error when using $\hat{f}(x)$ to estimate $f(x)$. Notationally, given squared error loss, the standard decomposition for MSE is given by:

$$\text{MSE}(f(x),\hat{f}(x)) = b^2(x) + v(x), \tag{1.5}$$

where the bias is given by $b(x) = f(x) - E(\hat{f}(x))$ and the variance is given by $v(x) = V(\hat{f}(x))$. Generally, the optimal choice for the bandwidth h is defined as the value of h that minimizes the MISE. The difficulty in practice is that f is usually unknown and so we must estimate the error quantities and subsequently our definition for an "optimal" bandwidth is the value of h that minimizes an estimate of the MISE.

Note, the optimal bandwidth that minimizes MISE, is equivalent to the bandwidth that minimizes the expected value of the quantity:

$$J(f,\hat{f}) = \int \hat{f}^2(x)dx - 2\int \hat{f}(x)f(x)dx. \tag{1.6}$$

Note that $E\left(J(f,\hat{f})\right)$ and MISE differ by the constant term $\int f^2(x)dx$.

Section 2 details currently available estimators for risk and proposes a new method to obtain estimators for MISE. In Section 3 the estimators and the optimal bandwidths from these estimators are compared via simulations. The methods are applied to a real dataset in Section 4. The paper concludes with a discussion and conclusion.

2. Risk Estimators

2.1. Asymptotic Normal

The asymptotic value for h that minimizes the (asymptotic) MISE is given by:

$$h = \left(\frac{\int K^2(x)dx}{n\left(\int K(x)x^2dx\right)^2 \int \left(f''(x)\right)^2 dx}\right)^{1/5}. \tag{2.7}$$

Assuming that f is the normal distribution, then the above form reduces to the estimator h_{an} given by:

$$h_{an} = 1.06\sigma n^{-1/5}, \tag{2.8}$$

where σ is the standard deviation of f; See Scott (1992); Silverman (1986) for a complete review. Hence, using a maximum likelihood estimator (s) to estimate σ, we have a simple data based procedure for choosing the bandwidth. We refer to the bandwidth estimator using s in place of σ in (2.8) as the asymptotic normal method.

2.2. *Cross Validation*

The application of cross validation methods has been proposed and studied by Rudemo (1982); Bowman (1984); Hall (1983); Stone (1984). The cross validation estimator for $J(f, \hat{f})$ is given by:

$$\hat{J}(f, \hat{f}) = \int \left(\hat{f}(x)\right)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{(-i)}(X_i), \tag{2.9}$$

where $\hat{f}_{(-i)}$ is the estimator obtained after removing the i^{th} observation from the sample. Note $\hat{f}(x)$ may be any estimator of the density, e.g. the kernel density estimator or the empirical density estimator. We refer to the cross-validation bandwidth as the the bandwidth obtained by minimizing (2.9) when $\hat{f}(x)$ is the kernel density estimator defined in (1.1).

The discovery of the bootstrap described in Efron (1979) brought about a new series of methods to obtain estimators for MISE and thus new methods to obtain bandwidths for KDE. An overview of bootstrapping methods are supplied in Efron and Tibshirani (1997) with applications to KDE in Shao and Tu (1995). The following set of estimators for MISE are motivated via bootstrapping methods.

2.3. *Taylor's Method*

Taylor's method Taylor (1989) employs a bootstrap method where the data is resampled from a smoothed distribution $\hat{f}(x)$. Taylor's estimator for MSE is given by:

$$E_*\left(\hat{f}^*(x;h) - \hat{f}(x;h)\right)^2 = E_*\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i^*}{h}\right) - \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right)^2 \tag{2.10}$$

where X_i^* are sampled from the smoothed distribution $\hat{f}(x;h)$ and E_* is the expectation with respect to the distribution of X_i^*.

As expressed in Taylor (1989), integrating the above expression leads to the MISE estimator below:

$$E_*(R^*) = \frac{1}{2n^2h(2\pi)^{1/2}}(\textstyle\sum_{i,j} \exp\left(-\frac{(X_j-X_i)^2}{8h^2}\right) - \frac{4}{3^{1/2}}\sum_{i,j} \exp\left(-\frac{(X_j-X_i)^2}{6h^2}\right) + \tag{2.11}$$

$$2^{1/2}\textstyle\sum_{i,j} \exp\left(-\frac{(X_j-X_i)^2}{4h^2}\right) + n2^{1/2}). \tag{2.12}$$

With certain conditions on $f(x;h)$ and a minor modification of the above expression, $E_*(R^*)$ is a consistent estimator of the MISE.

2.4. Faraway and Jhun's Method

The approach taken in Faraway and Jhun (1990) estimates the MISE by deconstructing the MSE into the bias and variance components. The variance is estimated with a regular bootstrap. The initial estimate of the density (required to estimate the bias) is obtained with a bandwidth chosen by some other procedure (e.g. cross validation) where a resampling procedure is performed on this initial estimate of the density. This smoothed bootstrap procedure tends to improve upon that initial estimate of the density.

In short, construct an initial estimate of the density, $\hat{f}(x;h_0)$, and then resample from that distribution. For example, to each resampled X_j^* add a random amount $h_0\epsilon$ where ϵ is distributed with density $K(\cdot)$. The Faraway and Jhun expression for the variance based on B bootstrapped resamples of the data is:

$$Var(\hat{f}(x;h)) = B^{-1}\sum_{j=1}^{B}\int (f_j^*(x;h) - \bar{f}_j^*(x;h))^2 dx, \tag{2.13}$$

where $\bar{f}_j^*(x;h) = B^{-1}\sum_{j=1}^{B} f_j^*(x;h)$. The bias is given by:

$$\hat{f}(x;h_0) - \bar{f}_j^*(x;h). \tag{2.14}$$

Thus the the Faraway and Jhun estimator for MISE is given by:

$$\widehat{MISE}(h,h_0) = B^{-1}\sum_{j=1}^{B}\int (f_j^*(x;h) - \hat{f}(x;h_0))^2 dx. \tag{2.15}$$

The Faraway and Jhun optimal bandwidth is obtained by minimizing (2.15) over h.

2.5. Bootstrap MSE Method

In this note we proposed a new and novel exact bootstrap estimator for MISE which eliminates the error due to Monte Carlo resampling. In many instances, the Monte Carlo error is non-negligible; e.g. see Ernst and Hutson (2003). For our new method, we recognize that KDEs can be expressed as L-estimators. Then by utilizing the work in Hutson and Ernst (2000), we can obtain exact estimators for the mean and variance of KDEs. Utilizing these estimators for the mean and variance, we can obtain MISE estimates for KDEs and thus obtain "optimal" bandwidths.

An L-estimator is an estimator that is equal to a linear combination of order statistics of the measurements. The median, trimmed mean, and trimean are all examples of L-estimators. Via the bootstrap, exact analytic expressions for the mean and variance for any L-estimator are derived in Hutson and Ernst (2000). The expressions for the bootstrap mean and variance follow from the direct calculation of the bootstrap mean and covariance matrix of the set of order statistics. In order to apply these results, we must show that the kernel density (or function) estimator fits the framework for an L-estimator.

Recall the form of the kernel estimator in (1.1), we can group terms such that:

$$\hat{f}(x;h) = \sum_{i=1}^{n} c_i K\left(\frac{x - X_i}{h}\right) \tag{2.16}$$

where c_i is equal to $1/nh$. Now, define $Y_i = K\left(\frac{x-X_i}{h}\right)$ so that we have,

$$\hat{f}(x;h) = \sum_{i=1}^{n} c_i Y_i \tag{2.17}$$

by rearranging the indices in terms of order statistics, we have:

$$\hat{f}(x;h) = \sum_{j=1}^{n} c_j Y_{j:n} \tag{2.18}$$

where $Y_{1:n} \leq Y_{2:n} \leq ... \leq Y_{n:n}$ are the order statistics corresponding to the sample $Y_i,\ i = 1,...,n$. Hutson and Ernst (2000) specify the form of an L-estimator as:

$$T_n = \sum_{i=1}^{n} c_i X_{i:n}. \tag{2.19}$$

Hence, the general form for our kernel estimator in (2.18) fits the framework for an L-estimator.

With this framework for our kernel estimator, via the methods in Hutson and Ernst (2000) we can now provide exact estimators for the mean and variance of the KDE. Specifically, let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics corresponding to our sample $X_1, \cdots X_n$. Consider an arbitrary random variable $X \sim F$ and define the sample quantile function estimator as

$$\hat{Q}_X(u) = \hat{F}^{-1}(u) = X_{[nu]+1:n} \tag{2.20}$$

where $0 < u < 1$ and $[\cdot]$ denotes the floor function. The goal is to develop expressions for the exact bootstrap mean and variance for $\hat{f}(x;h)$.

From Hutson and Ernst (2000), the $E_{\hat{Q}_X}(Y_{j:n})$ is given by the following lemma.

LEMMA 2.1. *The exact bootstrap estimate of* $\mu_{r:n} = E(Y_{r:n})$, $1 \leq r \leq n$, *is*

$$\hat{\mu}_{r:n} = E_{\hat{Q}_T}(Y_{r:n}) = \sum_{j=1}^{n} w_{j(r)} Y_{j:n}, \tag{2.21}$$

where

$$w_{j(r)} = B_{r,n-r+1}\left(\frac{j}{n}\right) - B_{r,n-r+1}\left(\frac{j-1}{n}\right), \tag{2.22}$$

and

$$B_{a,b}(x) = \int_0^x t^{a-1}(1-t)^{b-1} dt \tag{2.23}$$

is the incomplete beta function.

Proof: The proof is similar to the proof for Theorem 1 in Hutson and Ernst (2000).

THEOREM 2.2. *The exact bootstrap mean of the kernel density estimator* $\hat{f}(x;h)$ *is given* *by*

$$\hat{\mu}_{\hat{f}(x;h)} = E(\hat{f}(x;h)) = \frac{1}{nh}\sum_{j=1}^{n}\hat{\mu}_{j:n} = \frac{1}{nh}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{i(j)} Y_{i:n}, \tag{2.24}$$

where the weights $w_{i(j)}$ *are defined at (2.22).*

Proof: Straightforward application of Lemma 2.1 and (2.18).

Before establishing the variance of $\hat{f}(x;h)$, the following lemmas are required for the variance and covariance of $Y_{i:n}$.

LEMMA 2.3. *The exact bootstrap estimate of* $\sigma^2_{r:n} = Var(Y_{r:n})$ *is given by*

$$\hat{\sigma}^2_{r:n} = Var_{\hat{Q}_X}(Y_{r:n}) = \sum_{j=1}^{n} w_{j(r)}(Y_{j:n} - \hat{\mu}_{r:n})^2, \tag{2.25}$$

where $w_{j(r)}$ *is given in (2.22) and* $\hat{\mu}_{r:n}$ *is given in (2.21) .*

Proof: The proof is similar to that for Theorem 2.1 in Hutson and Ernst (2000).

LEMMA 2.4. *The exact bootstrap estimate of* $\sigma_{rs:n} = Cov(Y_{r:n}, Y_{s:n})$ *for* $r < s$ *is given by*

$$\begin{aligned}\hat{\sigma}_{rs:n} = Cov_{\hat{Q}}(Y_{r:n}, Y_{s:n}) &= \sum_{j=2}^{n}\sum_{i=1}^{j-1} w_{ij(rs)}(Y_{i:n} - \hat{\mu}_{r:n})(Y_{j:n} - \hat{\mu}_{s:n}) \\ &\quad + \sum_{j=1}^{n} v_{j(rs)}(Y_{j:n} - \hat{\mu}_{r:n})(Y_{j:n} - \hat{\mu}_{s:n}), \end{aligned} \tag{2.26}$$

where the weights are given by

$$w_{ij(rs)} = \int_{(j-1)/n}^{j/n}\int_{(i-1)/n}^{i/n} f_{rs}(u_r, u_s)\, du_r\, du_s \tag{2.27}$$

$$v_{j(rs)} = \int_{(j-1)/n}^{j/n}\int_{(j-1)/n}^{u_s} f_{rs}(u_r, u_s)\, du_r\, du_s, \tag{2.28}$$

and

$$f_{rs}(u_r, u_s) = {}_nC_{rs} u_r^{r-1}(u_s - u_r)^{s-r-1}(1 - u_s)^{n-s}, \tag{2.29}$$

is the joint distribution of two uniform order statistics $U_{r:n}$ *and* $U_{s:n}$ *with* ${}_nC_{rs} = n!/[(r-1)!(s-r-1)!(n-s)!]$.

Proof: This is a re-expression of the results found in Hutson and Ernst (2000).

Thus the following theorem establishes the analytic expression for the variance of $\hat{f}(x;h)$.

Theorem 2.5. *The exact bootstrap estimator for the variance of the kernel density estimator is given by:*

$$\hat{\sigma}^2_{\hat{f}(x)} = \left(\frac{1}{nh}\right)^2 \left(\sum_{j=1}^{n} \hat{\sigma}^2_{j:n} + 2\sum_{j<k}^{n} \hat{\sigma}_{jk:n}\right), \tag{2.30}$$

where $\hat{\sigma}^2_{j:n}$ and $\hat{\sigma}_{jk:n}$ are given in (2.25) and (2.26), respectively.

Proof: Follows as an application of Lemmas 2.3 and 2.4 and (2.18).

Thus we have derived the bootstrap mean and variance for the kernel density estimator. In order to estimate the MSE, we need to estimate the variance term and the bias term. The variance follows directly from Theorem 2.5. From Theorem 2.2, we have an expression for $E(\hat{f}(x))$. Hence, to this point, still must estimate $f(x)$ in order to estimate the bias term.

There are several choices we can use to estimate $f(x)$, and thus estimate the bias. The choices we will consider are the empirical density estimator with a mass of $1/n$ at each data point X_i. The cross validation kernel density estimator $\hat{f}(x; h_{cv})$ where h_{cv} is obtained via cross-validation with the kernel density estimator. We will also consider the KDE using the asymptotic normal bandwidth described in Section 2.1. Using these estimators for $f(x)$, we obtain the following theorem for estimating MSE.

Theorem 2.6. *The bootstrap estimator of the MSE for the kernel density estimator $\hat{f}(x; h)$ is given by*

$$\widehat{MSE}(f(x), \hat{f}(x; h)) = \hat{\sigma}^2_{\hat{f}(x;h)} + \left(\hat{\mu}_{\hat{f}(x;h)} - \hat{f}(x)\right)^2, \tag{2.31}$$

where $\hat{f}(x)$ is either the empirical density estimator, the KDE with bandwidth chosen by the asymptotic normal method (Section 2.1), or the KDE with bandwidth chosen by cross validation (Section 2.2).

Proof. Straightforward application of (1.5), (2.24), (2.30).

Using any of the estimators $\hat{f}(x)$ in Theorem 2.6 will yield an estimator for MISE:

$$\widehat{MISE} = \int_{-\infty}^{\infty} \hat{\sigma}^2_{\hat{f}(x;h)} + \left(\hat{\mu}_{\hat{f}(x;h)} - \hat{f}(x)\right)^2 dx. \tag{2.32}$$

We refer to the estimator BT, as the MISE estimator using the empirical density estimator for $f(x)$, the estimator BT-AN as the MISE estimator using KDE with bandwidth via the asymptotic normal method in (2.32), and BT-CV as the MISE estimator using KDE with bandwidth via cross validation in (2.32).

3. Simulation

For computational simplicity we use the normal density function for our kernel. The objective of the simulation is to compare the different procedures for obtaining the KDE bandwidth under different scenarios. The methods considered were the asymptotic normal method (AN), the cross-validation method (CV), Taylor's method (BT-CT), Faraway and Jhun's bootstrap method (BT-FJ) and our method(s), namely, BT, BT-AN, and BT-CV as defined in Section 2.5. We obtained the minima of the functions using standard numerical integration functions in the R software language (R Development Core Team, 2008). For each sample we calculated the MISE obtained using each method, together with the minimum achievable MISE, corresponding to h_1 where h_1 minimizes the integrated squared error:

$$ISE = \int \left(f(x) - \hat{f}(x)\right)^2 dx. \tag{3.33}$$

The methods were assessed based on the ratios $MISE(h_i)/MISE(h_1)$ and the results are summarized via Tables 1-4. The chosen distributions are the normal, normal bimodal mixture distribution, gamma, and standard lognormal. Each simulation was examined for sample sizes of 25, 50, and 100. Note these simulations contain unimodal, bimodal, symmetric and skew distributions.

4. Real Data Application

We consider the 63 observations of the annual snowfall amounts in Buffalo, New York as observed from 1910/11 to 1972/73 (data in Table 5) ; see, for example, Parzen (1979). The asymptotic rule in (2.8) and the cross-validation method yields the optimal bandwidth of 10.97 and 9.18, respectively. Taylor's exact smoothed bootstrap methods chooses a bandwidth of 16.7. Faraway and Jhun's method estimates the optimal bandwidth at 12 after resampling 1000 times. The main disadvantage of the Monte Carlo methods is that each computation yields different solutions. In this case, we recalculate the optimal bandwidth 100 times for the Faraway and Jhun method where the resulting bandwidths range from 11.74 to 12.30, centered at 12.06 with standard error 0.11. Our BT, BT-AN and BT-CV estimators provide bandwidths of 1.66, 14.6, 13.03, respectively. The optimal bandwidth from our BT estimator is the smallest among all methods mentioned in this note, thus tending to undersmooth the curve. The bandwidth chosen by either BT-AN or BT-CV yields similar density estimates with AN, CV, BT-CT, and BT-FJ (see Figure 1).

In addition, we also examined the dataset given in Table 6 for the determinations of the parallax of the sun (in seconds of a degree) based on the 1761 transit of Venus as described in Stigler (1977). Note that the cross-validation method does not work for this data, tending to select the bandwidth as small as possible. As discussed in Taylor (1989), examples with small sample sizes may occur where there is no global minimum which, according to Taylor, "should make us cautious in our numerical routines." We choose $h = 1e - 04$ artificially. The failure of the CV, BT-FJ, and BT-CV methods results from the numerical problems with the cross-validation method. Nevertheless, the asymptotic rule, Taylor's method and our BT and BT-AN methods select the optimal bandwidth at 0.43, 1.2, 0.1 and 0.55, respectively (see Figure 2).

5. Discussion and Conclusion

As can be seen from Table 1, using our method (BT-AN) to estimate h compares favorably with competing methods in terms of relative MISE. Generally speaking, our BT-AN and BT-CV methods perform favorably over either Taylor's method and Faraway and Jhun's method (Tables 1-4). From studying Tables 1-4, as expected the variance of h decreases as the sample size increases. In general, the BT-AN and BT-CV provide bandwidths with a small variance across all of the simulations. As may be expected (see Taylor (1989)), Standard lognormal densities are the most difficult to estimate optimal bandwidths. Among our methods (BT, BT-AN, and BT-CV), the BT-AN method performs the best relative to MISE, except for the seemingly more difficult lognormal cases, where BT-CV and BT both perform better. The Faraway and Jhun method (BT-FJ) has the largest variance for h in Tables 1-3, while it is fairly competitive for the lognormal densities. When comparing the BT-CV with the CV method and the BT-AN method with the AN method we see that the BT-CV method does better than the CV method except for the lognormal densities. Meanwhile the BT-AN method outperforms the AN method on the normal and lognormal distributions.

Note that using the empirical density function to estimate f in (2.31) will yield a consistent estimator for MISE (with certain conditions on f). In other words, the empirical density function $f(x)$ is a consistent estimator of $f(x)$, and since our mean and variance estimators for the KDE are exact by the work in Hutson and Ernst (2000), we then have that $\widehat{MISE}$ in (2.32) converges in probability to MISE.

For future work it is examining other choices than the normal density for the form of the kernel, e.g. Epanechnikov kernel. In fact, the Epanechnikov kernel is optimal in the sense of yielding the smallest asymptotic mean squared error (Wasserman, 2004). We expect that the results when using the Epanechnikov kernel will not differ greatly than the results presented here using the normal density as the kernel. Also, important to consider is the use of other methods to estimate f in (2.31). We employed the empirical density function and KDE where the bandwidth was chosen either by the asymptotic normal method or the cross validation method. We plan to explore other more sophisticated methods to obtain an estimate for $f(x)$, e.g. an iterative scheme such as described in Faraway and Jhun (1990). The extension of our method to produce confidence intervals for f is also worth exploring. The confidence intervals will require using methods to estimate the variance of our MISE estimates. We speculate that extending the work in Hutson and Ernst (2000) to compute the variance of the estimates for the mean and variance of the L-estimator will allow us to obtain confidence intervals for MISE and f. Of lesser importance, it would also be interesting to explore the application to multivariate densities based on the work in Hutson and Ernst (2000) and to examine the "optimal bandwidths" obtained under different loss functions, e.g. the L_1 norm.

References

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**(2), 353–360.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, **7**(1), 1–26.

Efron, B. and Tibshirani, R. (1997). *An introduction to the bootstrap*. Chapman & Hall.

Ernst, M. D. and Hutson, A. D. (2003). Utilizing a quantile function approach to obtain exact bootstrap solutions. *Statistical Science*, **18**(2), 231–240.

Faraway, J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *Journal of the American Statistical Association*, pages 1119–1122.

Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *The Annals of Statistics*, **11**(4), 1156–1174.

Hutson, A. and Ernst, M. (2000). The exact bootstrap mean and variance of an L-estimator. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pages 89–94.

Mugdadi, A. and Ahmad, I. (2004). A bandwidth selection for kernel density estimation of functions of random variables. *Computational Statistics and Data Analysis*, **47**(1), 49–62.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, pages 105–121.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**(2), 65–78.

Scott, D. (1992). *Multivariate density estimation: theory, practice, and visualization*. Wiley-Interscience.

Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. Springer.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Stigler, S. (1977). Do robust estimators work with real data? *The Annals of Statistics*, **5**(6), 1055–1098.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, **12**(4), 1285–1297.

Taylor, C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, **76**(4), 705.

Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*. Springer Verlag.

Table 1. N(0,1): The methods employed include the asymptotic choice (AN); the cross validation method (CV); our method with empirical PDF (BT); Taylor's smoothed bootstrap method (BT-CT); Faraway and Jhun's bootstrap method (BT-FJ); our method with asymptotic choice for bandwidth choice in KDE (BT-AN) and our method with cross validation bandwidth for KDE (BT-CV). Average over 500 simulations, with standard error in parentheses for a variety of different sample sizes (n). Relative efficiency is the MISE ratio compared with the optimal bandwidth value. The average bandwidth with standard error in parentheses is also shown for each of the different methods.

	Relative efficiency		
	n=25	n=50	n=100
AN	3.85(28.2)	2.14(7.9)	1.54(2.52)
CV	3.73(15.9)	2.55(7.55)	2.86(7.18)
BT	5.98(31.2)	3.52(11.6)	7.32(13.2)
BT-CT	21.9(63)	27.8(63)	1.56(0.972)
BT-FJ	3.02(9.58)	2.28(4.27)	2.45(4.39)
BT-AN	2.66(15.6)	1.78(3.8)	1.48(1)
BT-CV	2.89(9.22)	2.2(3.69)	2.27(3.5)
	Bandwidth		
	n=25	n=50	n=100
AN	0.547(0.0826)	0.481(0.0488)	0.424(0.0296)
CV	0.633(0.198)	0.533(0.136)	0.434(0.13)
BT	0.397(0.118)	0.335(0.0641)	0.173(0.0883)
BT-CT	1.97(0)	1.96(0.0802)	0.577(0.0529)
BT-FJ	0.758(0.223)	0.634(0.149)	0.517(0.144)
BT-AN	0.722(0.122)	0.614(0.0691)	0.532(0.0434)
BT-CV	0.766(0.211)	0.641(0.138)	0.522(0.134)

Table 2. 0.5N(0,1)+0.5N(3,1): The methods employed include the asymptotic choice (AN); the cross validation method (CV); our method with empirical PDF (BT); Taylor's smoothed bootstrap method (BT-CT); Faraway and Jhun's bootstrap method (BT-FJ); our method with asymptotic choice for bandwidth choice in KDE (BT-AN) and our method with cross validation bandwidth for KDE (BT-CV). Average over 500 simulations, with standard error in parentheses for a variety of different sample sizes (n). Relative efficiency is the MISE ratio compared with the optimal bandwidth value. The average bandwidth with standard error in parentheses is also shown for each of the different methods.

	Relative efficiency		
	n=25	n=50	n=100
AN	1.23(0.332)	1.23(0.332)	1.32(0.399)
CV	1.95(1.73)	1.94(1.72)	1.78(2.01)
BT	1.98(1.73)	1.98(1.72)	3.99(4.34)
BT-CT	2.58(1.94)	2.58(1.94)	2.99(2.35)
BT-FJ	1.79(1.11)	1.79(1.11)	1.68(1.31)
BT-AN	1.6(0.876)	1.6(0.876)	1.68(0.759)
BT-CV	1.73(1.04)	1.73(1.03)	1.61(1.11)
	Bandwidth		
	n=25	n=50	n=100
AN	1(0.109)	1(0.11)	0.758(0.041)
CV	0.976(0.417)	0.977(0.417)	0.597(0.219)
BT	0.658(0.283)	0.658(0.283)	0.221(0.107)
BT-CT	2(0)	2(0.025)	1.23(0.119)
BT-FJ	1.22(0.483)	1.22(0.483)	0.742(0.261)
BT-AN	1.34(0.161)	1.34(0.161)	0.906(0.0638)
BT-CV	1.23(0.458)	1.23(0.457)	0.741(0.245)

Table 3. Gamma(2,2): The methods employed include the asymptotic choice (AN); the cross validation method (CV); our method with empirical PDF (BT); Taylor's smoothed bootstrap method (BT-CT); Faraway and Jhun's bootstrap method (BT-FJ); our method with asymptotic choice for bandwidth choice in KDE (BT-AN) and our method with cross validation bandwidth for KDE (BT-CV). Average over 500 simulations, with standard error in parentheses for a variety of different sample sizes (n). Relative efficiency is the MISE ratio compared with the optimal bandwidth value. The average bandwidth with standard error in parentheses is also shown for each of the different methods.

	Relative efficiency		
	n=25	n=50	n=100
AN	1.47(0.532)	1.59(0.58)	1.72(0.649)
CV	1.89(1.56)	1.71(1.08)	1.51(1.03)
BT	2.3(2.09)	2.04(1.34)	2.37(1.53)
BT-CT	2.05(0.92)	2.93(1.35)	2.22(0.943)
BT-FJ	1.74(1.01)	1.6(0.626)	1.48(0.654)
BT-AN	1.68(0.682)	1.93(0.792)	2.01(0.788)
BT-CV	1.7(0.936)	1.55(0.57)	1.44(0.561)
	Bandwidth		
	n=25	n=50	n=100
AN	1.54(0.332)	1.35(0.207)	1.18(0.127)
CV	1.18(0.471)	0.907(0.328)	0.728(0.22)
BT	0.688(0.27)	0.543(0.184)	0.364(0.132)
BT-CT	2.02(0.0841)	1.99(0.0641)	1.37(0.128)
BT-FJ	1.41(0.492)	1.11(0.374)	0.896(0.249)
BT-AN	1.76(0.239)	1.55(0.201)	1.31(0.107)
BT-CV	1.42(0.472)	1.13(0.355)	0.908(0.234)

Table 4. Standard lognormal: The methods employed include the asymptotic choice (AN); the cross validation method (CV); our method with empirical PDF (BT); Taylor's smoothed bootstrap method (BT-CT); Faraway and Jhun's bootstrap method (BT-FJ); our method with asymptotic choice for bandwidth choice in KDE (BT-AN) and our method with cross validation bandwidth for KDE (BT-CV). Average over 500 simulations, with standard error in parentheses for a variety of different sample sizes (n). Relative efficiency is the MISE ratio compared with the optimal bandwidth value. The average bandwidth with standard error in parentheses is also shown for each of the different methods.

	Relative efficiency		
	n=25	n=50	n=100
AN	2.56(1.27)	3.21(1.38)	5.73(2.89)
CV	1.16(0.192)	1.03(0.0658)	1.25(0.341)
BT	1.41(0.443)	1.03(0.0941)	1.21(0.318)
BT-CT	4.51(1.7)	6.21(1.82)	9.41(3.87)
BT-FJ	1.3(0.299)	1.2(0.124)	1.29(0.326)
BT-AN	2.7(1.24)	3.34(1.35)	5.81(2.73)
BT-CV	1.36(0.33)	1.25(0.16)	1.33(0.355)
	Bandwidth		
	n=25	n=50	n=100
AN	1.02(0.475)	0.916(0.35)	0.871(0.299)
CV	0.395(0.119)	0.314(0.0367)	0.187(0.055)
BT	0.498(0.179)	0.31(0.0303)	0.185(0.0493)
BT-CT	2.02(0.116)	2.01(0.0587)	1.47(0.171)
BT-FJ	0.493(0.148)	0.376(0.0512)	0.233(0.063)
BT-AN	1.07(0.396)	0.951(0.32)	0.877(0.229)
BT-CV	0.52(0.145)	0.393(0.053)	0.251(0.0599)

Table 5. The amount of snowfall in Buffalo, New York, for each of 63 winters from 1910/11 to 1972/73. See, for example, Parzen (1979) for more details.

Dataset ($n = 63$)								
126.4	82.4	78.1	51.1	90.9	76.2	104.5	87.4	110.5
25.0	69.3	53.5	39.8	63.6	46.7	72.9	79.6	83.6
80.7	60.3	79.0	74.4	49.6	54.7	71.8	49.1	103.9
51.6	82.4	83.6	77.8	79.3	89.6	85.5	58.0	120.7
110.5	65.4	39.9	40.1	88.7	71.4	83.0	55.9	89.9
84.8	105.2	113.7	124.7	114.5	115.6	102.4	101.4	89.8
71.5	70.9	98.3	55.5	66.1	78.4	120.5	97.0	110.0

Table 6. Short's 1763 determinations of the parallax of the sun (in seconds of a degree) based on the 1761 transit of Venus (Stigler 1977).

Dataset ($n = 18$)					
8.50	8.50	7.33	8.64	9.27	9.06
9.25	9.09	8.50	8.06	8.43	8.44
8.14	7.68	10.34	8.07	8.36	9.71

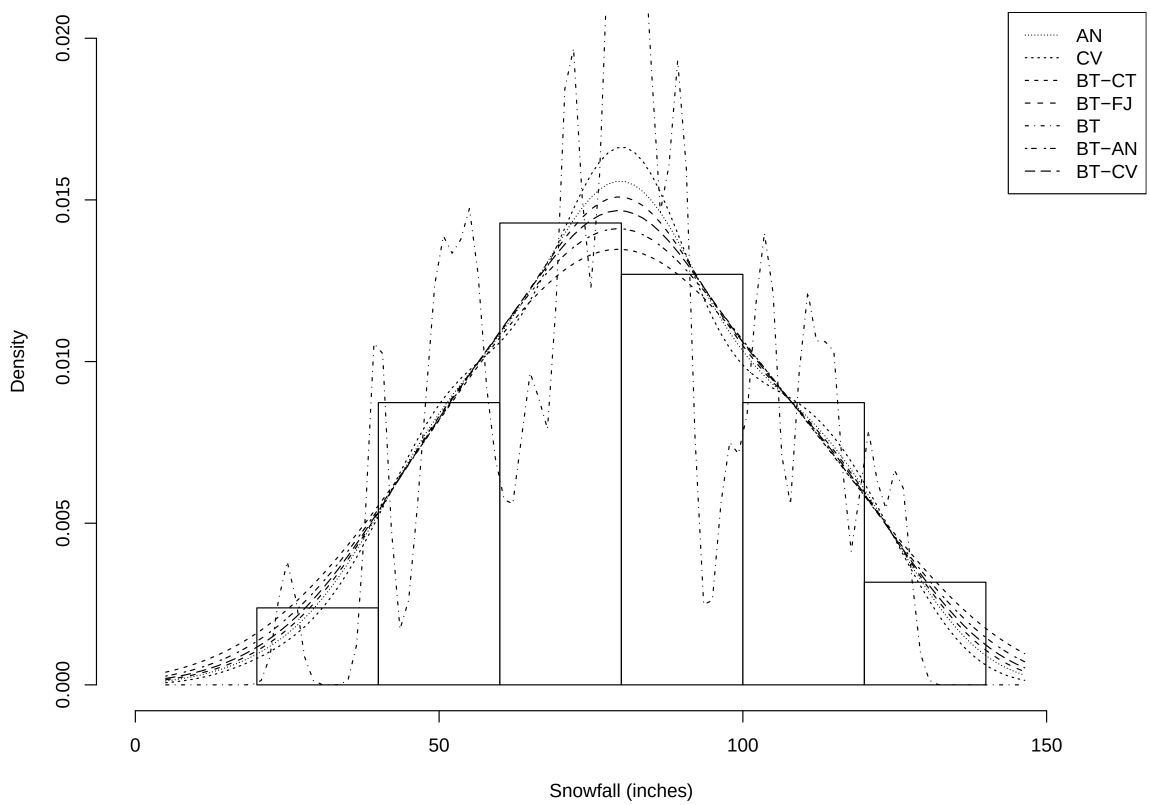

Fig. 1. Kernel density estimation based on Buffalo snowfall data with different methods to obtain the bandwidth.

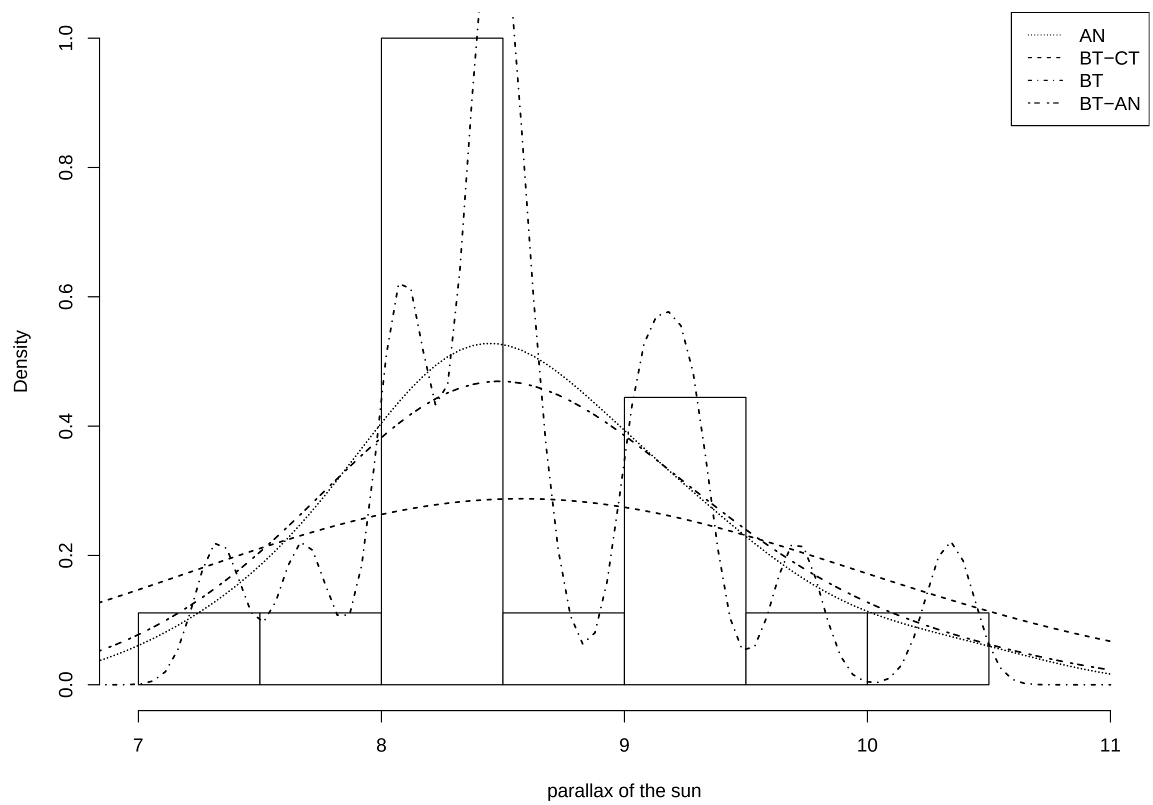

Fig. 2. Kernel density estimation based on Stigler (1977) data with different methods to obtain the bandwidth.