Rank Adapted Kernel Density Estimation

David B. Kim
Department of Mathematical Sciences
U.S. Military Academy
West Point, NY 10996

October, 2001

Abstract

We consider adapting bandwidths of a kernel density estimator according to the ranks of observations. The specifics of bandwidth selection is motivated by a deterministic decomposition of a density into densities of order statistics and their asymptotic behaviors. The resulting estimator has a local bandwidth similar to that of Abramson (1982) and Breiman et al. (1977) with a new feature of rank correction. We investigate its properties and demonstrate that not only it can smooth out the bumps in the tails while maintaining interesting features in data-rich region but also that it can reduce the boundary bias when the support of the target density is compact.

1 Introduction

Kernel density estimation is one of the more intuitive and widely used methods of nonparametric density estimation. Most of the research activities have been focused on the selection of the bandwidth, or smoothing parameter. The simplest implementation of the method uses only one bandwidth over the entire data set. This, however, has many shortcomings that motivated the various strands of researches on adaptive bandwidth selection.

Starting with the works by Breiman et al. (1977) and Abramson (1982), various ways of determining bandwidths depending on the local behavior of the density have been investigated. What one would like to accomplish in adaptive bandwidth selection is to be able to sharply delineate the boundary of a compact support and to be able to smooth out bumps in the tails while depicting interesting features in the density near the mode at the same time. Both Breiman et al. (1977) and Abramson (1982) attempt the latter by adjusting the bandwidth by a factor of a positive power of the reciprocal of the density. Since a larger bandwidth corresponds to a smoother estimate, this approach aims to use larger bandwidths where the density is smaller, and vice versa. Of course, the true underlying density is to be unknown in most cases, so in using this type of approach, one first finds a pilot estimate of the density and use it instead of the true underlying density.

Many researchers have also looked at the former problem using boundary kernels and other approaches (see for example, Müller (1991)). In this report, we look at a new adaptive bandwidth selection scheme where both sharp delineation of the boundary and more smoothing in the region of sparse density can be accomplished at the same time.

2 New Estimator: Motivation and Definition

In this section, we define the new kernel density estimator whose bandwidths are adapted according to the ranks of observations, and we investigate its properties. Let $X_1, \cdots, X_n$ be *iid* observations from an unknown density $f(x)$. Recall that a Kernel density estimator $\hat{f}(x)$ with a bandwidth h identical for all data points can be written as

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

where $K(\cdot)$ is a Kernel function. Kernel functions can be classified by its order k, where the k the order kernel $K(t)$ is such that, for positive integers j,

$$\int t^j K(t)\,dt = 0,\ j < k,\ \text{ and} \int t^k K(t)\,dt \neq 0,$$

where we note that we will assume the integration is over the entire real line $\mathbb{R}$ unless otherwise specified. Higher order kernel functions usually lead to estimators with smaller asymptotic biases, but the kernel functions of order higher than two can no longer be probability density functions (*pdfs*). Having a kernel function that is a *pdf* itself guarantees that the resulting kernel density estimator will again be a *pdf*. For this reason, we shall restrict our attention to the second order kernel which itself is a *pdf*.

The adaptive scheme presented in this paper takes its motivation from the following identity:

$$f(x) = \frac{1}{n} \sum_{j=1}^{n} f_{(j)}(x) \tag{2}$$

where $f_{(j)}(x)$ is the *pdf* of the jth order statistic, $X_{(j)}$. Kim (1999) showed that using the densities of the putative asymptotic densities instead of $f_{(j)}$'s gives a convergent approximation to $f(x)$. The asymptotic normality of the central order statistics, viz,

$$\sqrt{n}\left(X_{(j)} - \xi_{p_j}\right) \overset{\text{a}}{\sim} N\left(0, \frac{p_j(1-p_j)}{f^2(\xi_{p_j})}\right),$$

where $p_j = j/n$ and $\xi_{p_j} = F^{-1}(p_j)$. This can be seen as a consequence of the convergence in *pdfs* (Rao 1973, p. 422), specifically,

$$f_{(j)}(x) \approx \frac{\sqrt{n}}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{n(x-\xi_{p_j})^2}{2\sigma_j^2}\right), \tag{3}$$

where we let $\sigma_j^2 = \frac{p_j(1-p_j)}{f^2(\xi_{p_j})}$.

The second order kernel functions are bona fide densities, and a kernel density estimate using the second order kernels is the arithmetic average of n *pdf*'s centered at n data points:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right), \tag{4}$$

where $K(\cdot)$ is a bona fide density. Silverman (1986) defines an adaptive kernel estimate $\hat{f}_a(x)$ as follows:

$$\hat{f}_a(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h\lambda_i} K\left(\frac{x - X_i}{h\lambda_i}\right), \tag{5}$$

where λ_i is a local bandwidth factor adapted to each observation X_i. Comparing Eq.(2) and Eq.(5) suggests that one way of choosing the local bandwidth is to refer to the standard deviation in Eq.(3). Clearly, we do not know a priori the population quantile, ξ_{p_j}, not to mention the true density f. But $X_{(j)}$ is a consistent point estimator of ξ_{p_j}, so using $X_{(j)}$ in its stead we get the following estimator:

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h\lambda_i} K\left(\frac{x - X_{(i)}}{h\lambda_i}\right), \tag{6}$$

where $\lambda_i = \sqrt{\frac{p_i(1-p_i)}{f^2(X_{(i)})}}$. Even though this estimator still depends on the unknown true density, the dependence is reminiscent of the estimators of Abramson (1982) and Breiman et al. (1977). Still there is one new feature in the local bandwidth not found in the previous estimators: the factor of $\sqrt{p_i(1-p_i)}$, which is presumably responsible for a surprising and unique adaptability which will be shown in the next section. The choice of the global bandwidth h may be made as in using the estimator of Abramson (1982) after choosing a suitable pilot estimator.

3 Examples

In this section, the new rank adapted estimator will be compared to other estimators. Comparison is among rank adapted, Abramson, fixed bandwidth estimators, and the dashed line for the *true density*. The global bandwidth h for each of the estimators was determined using the unbiased cross validation (UCV) criterion. For the rank adapted and Abramson estimators, the pilot estimator using Silverman (1986)'s "rule of thumb" $\hat{h}$. All the computation and the generation of the figures were done in R.

It is well known that the estimators of Abramson (1982) and Breiman et al. (1977) were motivated by the need to smooth out bumps in the tails while depicting interesting features in the density near the mode at the same time, which cannot be achieved with a single uniform bandwidth. The similarity of the new rank adapted estimator and Abramson's estimator leads one to expect the both should behave similarly. In the case of a sample of size 500 from the standard normal distribution, seen in Fig. 1, one indeed observes that.

Fig. 2 shows the comparison of the local bandwidth factors λ_i for Abramson's estimator and the new rank adapted estimator. One observes the attenuation of the growth of the local bandwidth factor in the tails for the new rank adapted estimator–the attenuation is due to the rank correction factor. This is an encouraging sign since the rapid growth of the local bandwidth factor in Abramson's estimator has been shown to be responsible for its rather surprisingly poor performance for a very large sample (Terrell and Scott 1992).

An additional role the rank correction factor plays is demonstrated in the next example where a random sample was drawn from the uniform distribution on (0,1), which has a density with a compact support. Fig. 3 shows that the new rank adapted estimator is more adept at delineating the sharp boundary of the uniform density than other estimators shown (it shows the steepest descent at the boundaries).

In the next example, where the random sample of size 500 is drawn from an exponential distribution (with mean=1), which has a long tail on one end and a sharp boundary on the other end, the dual capabilities of the new rank adapted estimator is demonstrated.

The final two examples show the comparison of the estimators for famous real life data sets (see Silverman (1986) for more details on the data sets). The performance of the new estimator on the suicide data set (Fig. 5), which has similar features to the exponential example seen before, is very encouraging. It demonstrates that the new estimator does an adequate job of sharply delineating the boundary and smoothing in the region of sparse density. Fig. 6 shows the comparison of the estimators for the Yellowstone geyser data set. Again, one observes the competitive performance of the new estimator along with its novel capability where both sharp delineation of the boundary and more smoothing in the region of sparse density can be accomplished at the same time.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates – A square root law. *The Annals of Statistics 10*, 1217–1223.

Breiman, L., W. Meisel, and E. Purcell (1977). Variable kernel esitmates of multivariate densities. *Technometrics 19*, 135–144.

Kim, D. B. (1999). *Quantile Decomposition of a Density*. Ph. D. thesis, University of California, Santa Barbara.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika 78*(3), 521–530.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (Second ed.). John Wiley & Sons, New York.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Terrell, G. R. and D. W. Scott (1992). Variable kernel density estimation. *Ann. Statist. 20*(3), 1236–1265.

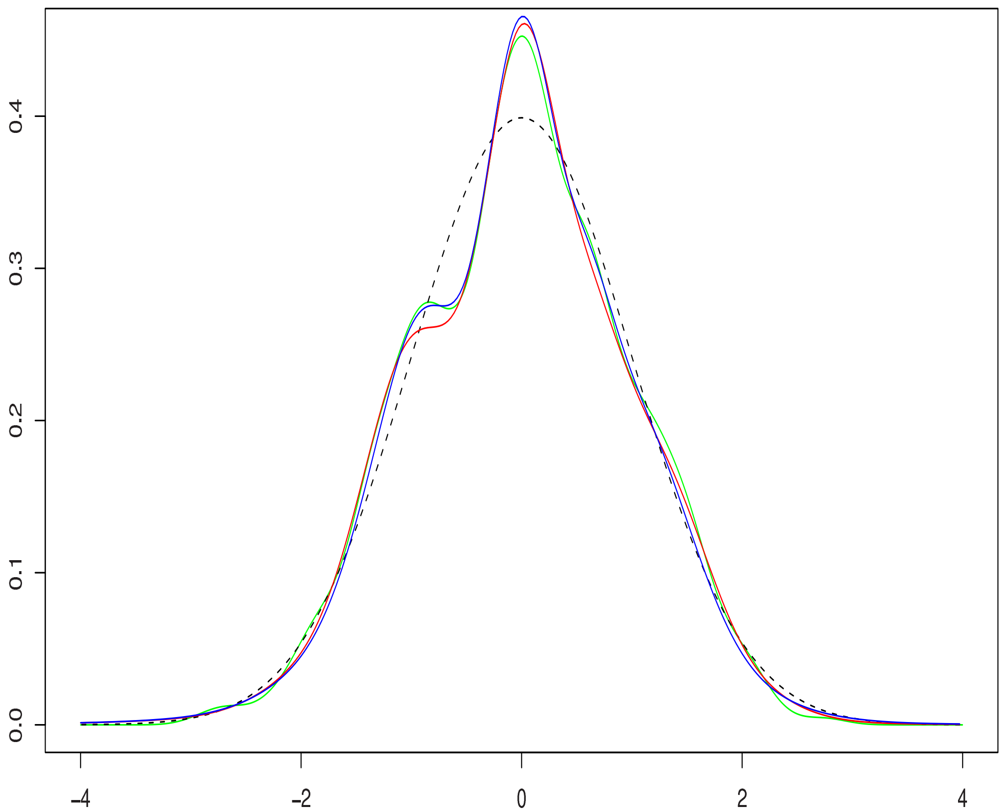

Figure 1: Comparison of the three estimators for a random sample of size 500 from the standard normal distribution.

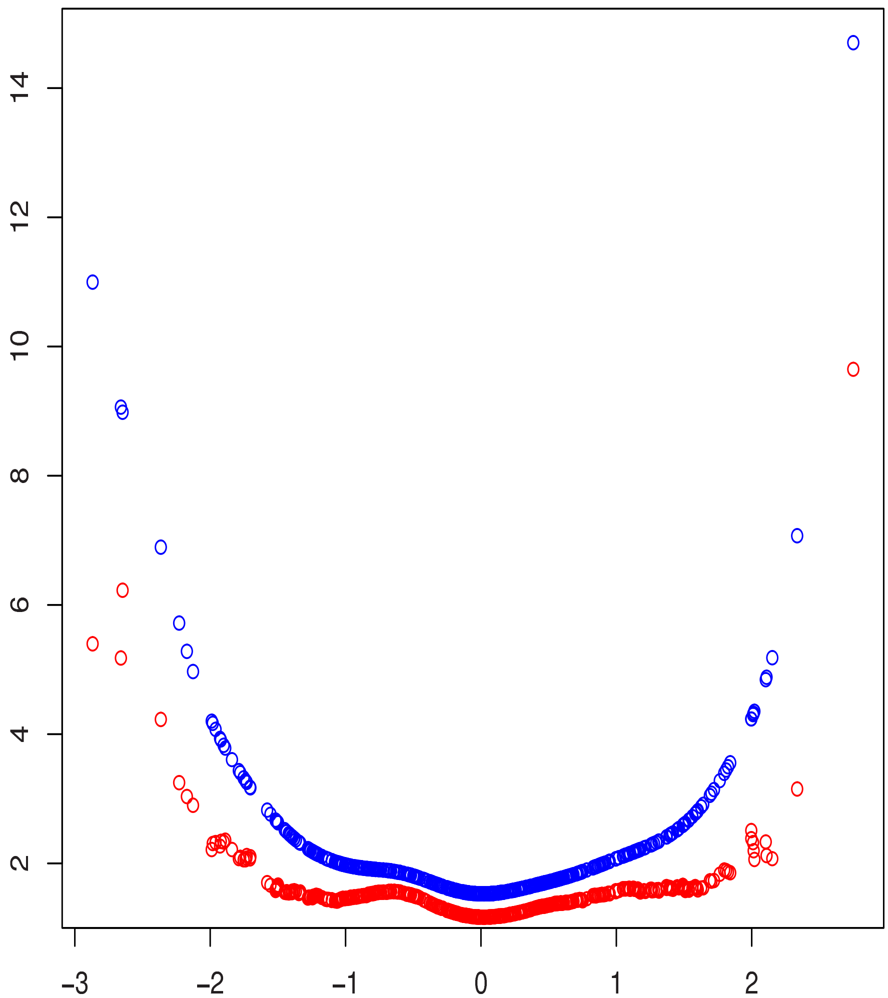

Figure 2: Comparison of the local bandwidth factors λ_i for a random sample of size 500 from the standard normal distribution.

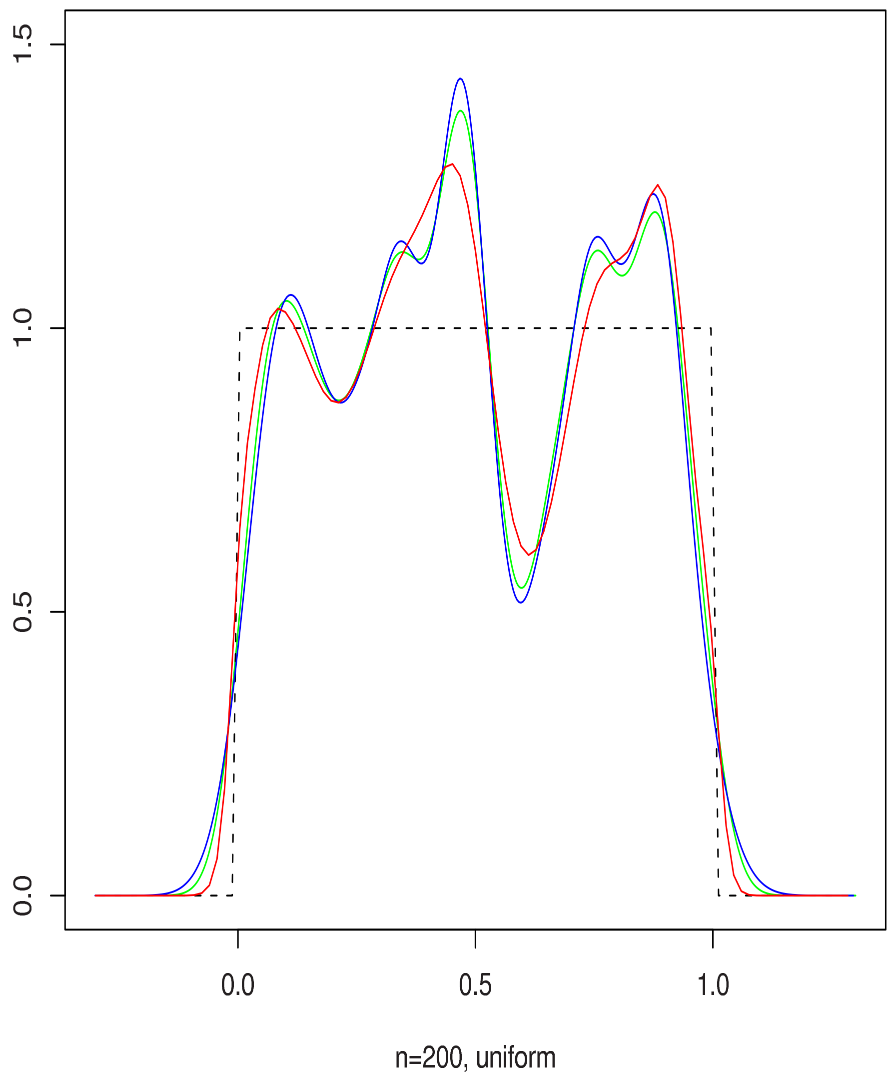

Figure 3: Comparison of the three estimators for a random sample of size 200 from the uniform distribution on (0,1).

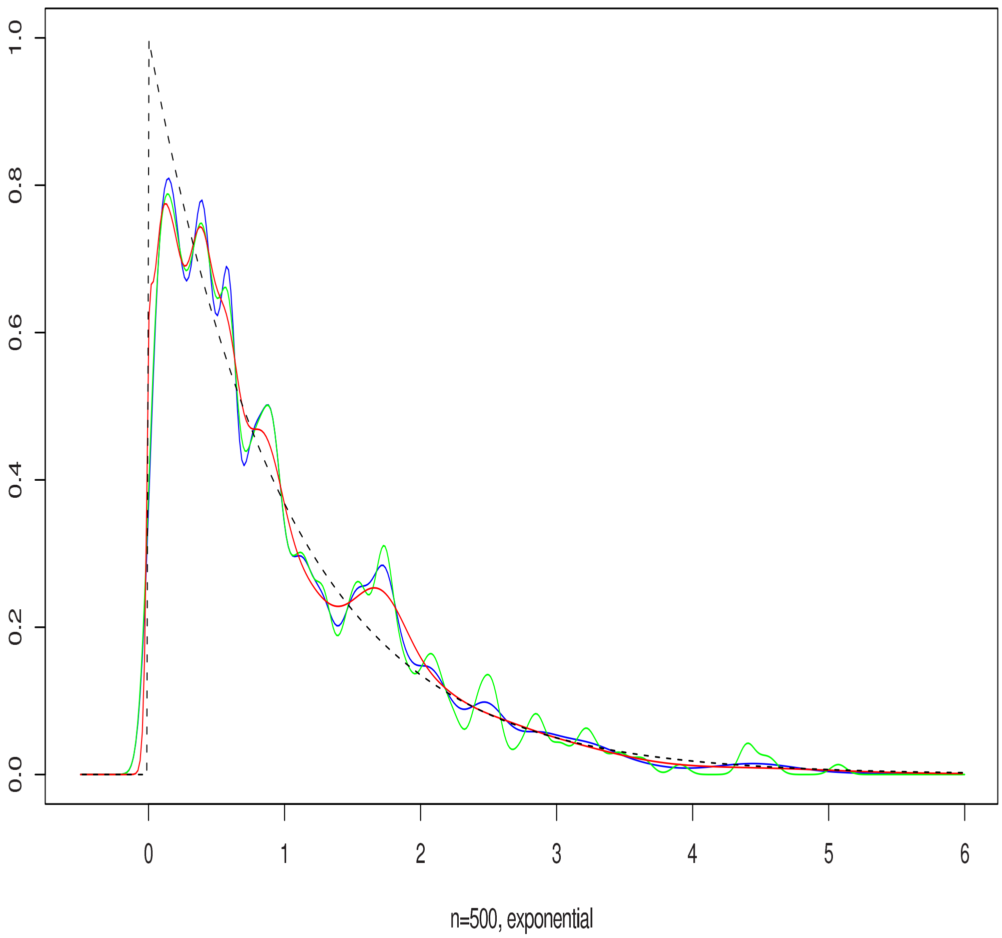

Figure 4: Comparison of the three estimators for a random sample of size 500 from the exponential distribution with mean=1.

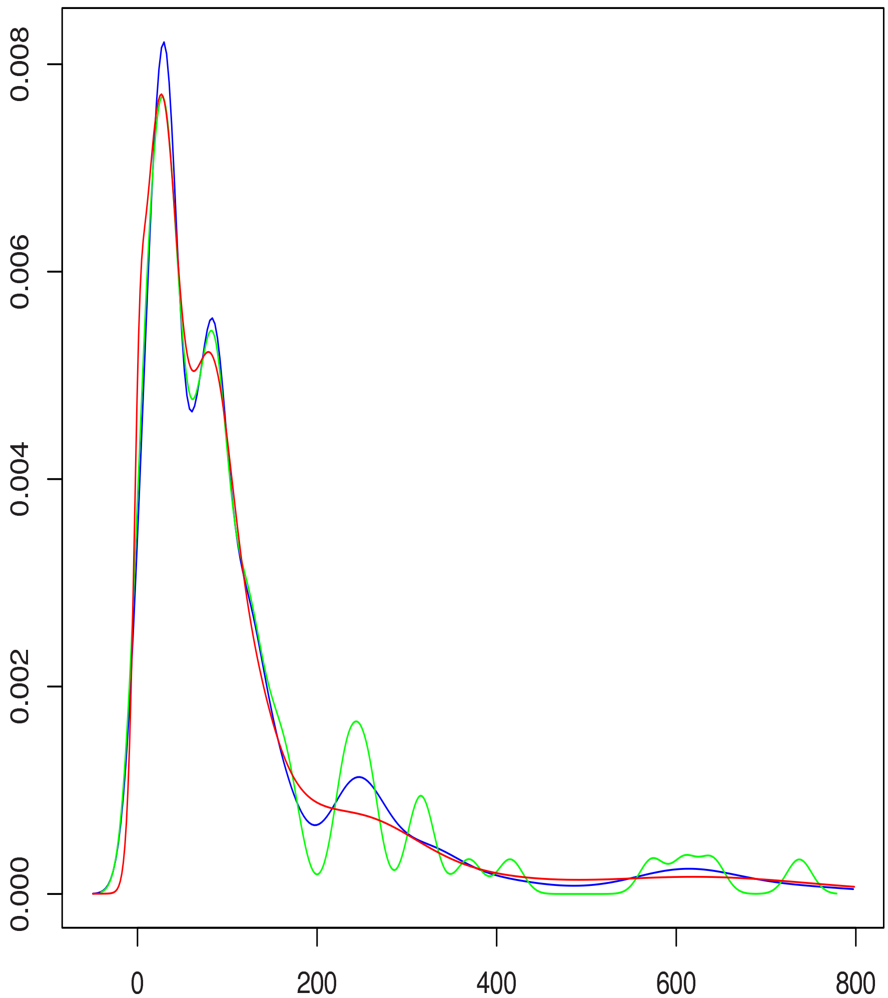

Figure 5: Comparison of the three estimators for the suicide data set.

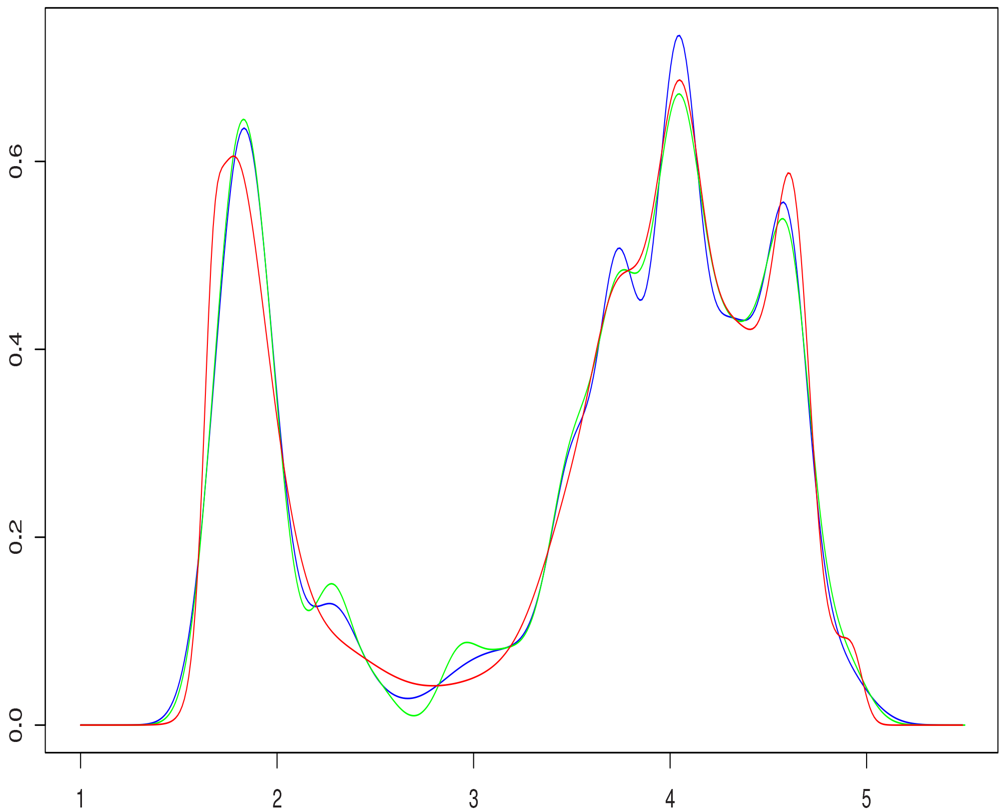

Figure 6: Comparison of the three estimators for the geyser data set.