Assessing bandwidth selectors with visual error criteria

J. S. Marron[1]

Department of Statistics, University of North Carolina, Chapel Hill, N.C. 27599-3260, U. S. A.

Summary

"Visual error criteria" have been developed as a means of measuring the distance between curves, in a way that corresponds more closely to "what the eye sees", than the usual norms on function spaces. The main contribution of this paper, is the application of this idea to the comparison of data based bandwidth selection methods. It is seen that which bandwidth selector is "best" is often much different in this sense, but closer to "visual best", than what is obtained from classical methods of measuring error. In addition, new insights into visual error criteria are provided.

Keywords: bandwidth selection, curve estimation, kernel density estimation, visual error measures

1 Introduction

Choice of the smoothing parameter, i.e. bandwidth, is crucial to the effective use of nonparametric curve estimators. The problem is also of current interest because it has close relationships to other important problems, such as model selection. There is a large literature on data based smoothing parameter selection. See Jones, Marron and Sheather (1992) [6] for a recent survey.

The deepest research on data based smoothing parameter selection has been done in the context of kernel density estimation, so that case will be studied here too. However, the main ideas and lessons of this paper clearly apply to other types of curve estimation as well. Kernel density estimation can be viewed as an attempt to use a random sample $X_1, ..., X_n$ from a probability density $f(x)$ to try to recover the curve $f(x)$. Given a kernel function $K(x)$ (taken in all examples here as the Gaussian probability density), and a bandwidth h, the kernel density estimator is:

$$\widehat{f}_h(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i),$$

where $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$. The bandwidth is represented as a separate parameter because "width of the kernel window" is crucial to the performance of the estimator. See Silverman (1986) [12] for illustration of this point, as well as discussion of many other useful aspects of this estimator.

In this paper the bandwidth selection problem is studied with careful attention paid to the issue of how to assess the performance of various bandwidths. There has been considerable past controversy on this issue, mainly centered around whether one should measure error based on the L^1 or on the L^2 norms. However, it is seen in Marron and Tsybakov (1993) [8] that in some situations neither of these is particularly attractive, and indeed "visual choice of bandwidth" (the method used most commonly in actual data analyses) can be quite different than either the L^1 or the L^2 optimal choices. Marron and Tsybakov go on to develop and investigate alternative measures of error which correspond more closely to "what the eye sees". In this paper, their proposed error criterion $VE_2(\widehat{f}_h \to f)$ is studied. Figure 1 shows how $VE_2(\widehat{f}_h \to f)$ assesses bandwidth selectors in a way that is closer to visual perception than L^2.

The error criterion $VE_2(\widehat{f}_h \to f)$ is defined in the next section. Precise formulation of the bandwidth selectors, $LSCV$ (Least Squares Cross Validation), $ROT2$ (Rule Of Thumb 2), $SJPI$ (Sheather Jones Plug In), and $RNSB$ (Root N Smoothed Bootstrap), with suitable references, is given in the appendix. Note

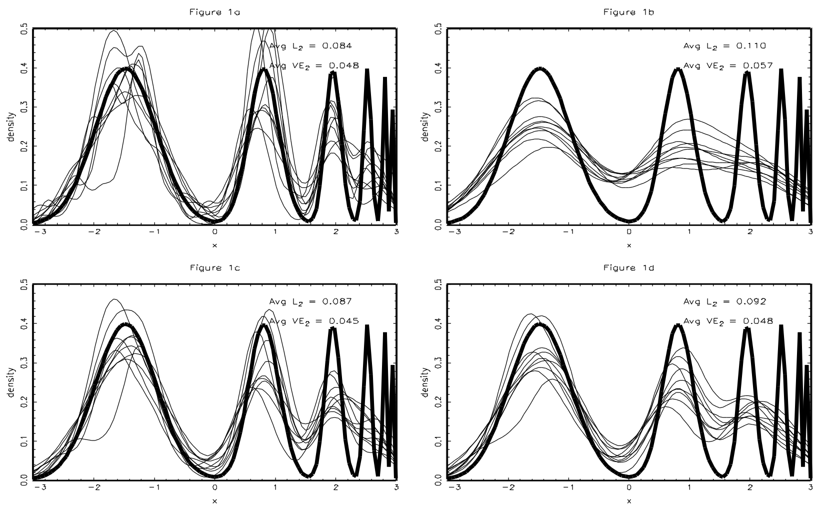

Figure 1: Target density (heavy line) and kernel density estimates (thin lines), from ten simulated data sets (same for each bandwidth selector), each of size $n = 100$, for four major methods of data based bandwidth selection: (a) $LSCV$, (b) $ROT2$, (c) $SJPI$, (d) $RNSB$. Bandwidth selection methods are compared by both L^2 and $VE_2(\widehat{f}_h \rightarrow f)$, averaged over the 10 data sets shown here.

that $LSCV$ is better than $SJPI$ and $RNSB$ in the sense of L^2, but worse with respect to $VE_2(\widehat{f}_h \to f)$. But the visual performances of $SJPI$ and $RNSB$ are clearly superior. $SJPI$ looks slightly better than $RNSB$, which is also reflected by $VE_2(\widehat{f}_h \to f)$. The main point of this paper is to study bandwidth selection methods through criteria which work differently than the usual ones, but are more useful for determining performance in real data situations.

This example is clearly a case where some type of location varying bandwidth is called for. In particular, a rather larger bandwidth is desirable on the left side, while a simultaneously smaller bandwidth will improve performance on the right side. However, such methods are complicated, and are typically used only after a need for them is understood through a preliminary analysis with global bandwidth methods. Hence "best" (albeit out of an unsatisfactory set) choice of global bandwidth is still a vital issue even in situations such as this.

In section 3, previously unknown facts about $VE_2(\widehat{f}_h \to f)$ are described. Different types of comparisons are made between L^2 and $VE_2(\widehat{f}_h \to f)$. One of these is a study of their expected values as functions of n and h. Behavior of their respective minimizers is also studied.

In section 4, it is seen that the measure $VE_2(\widehat{f}_h \to f)$ provides an important new viewpoint for the study of data based bandwidth selection, with some substantially different answers than those obtainable from the L^2 viewpoint. The basic recommendation of Jones, Marron and Sheather (1992) [6], that the $SJPI$ method is very good all around still remains the same. However it is now more clear that $LSCV$ is not appropriate. A surprising feature (at least to the author) is that $ROT2$ is more frequently not so bad as had been previously reported. The $RNSB$ is also now seen to be better than had previously been thought, which is more consistent with its excellent asymptotic properties.

2 Error criteria

The often considered goal of optimizing the L^2 norm is the same as optimizing it's square, the Integrated Squared Error,

$$ISE(h) = \int \left(\widehat{f}_h(x) - f(x)\right)^2 dx.$$

Here and in the following, the integral symbol is understood to mean definite integration over the whole real line. Some researchers prefer to work with the expected ISE, often called the Mean Integrated Squared Error,

$$MISE(h) = E\left[ISE(h)\right].$$

A common notion of "optimal bandwidth" is h_{MISE}, which is a (these need not be unique, as seen in Marron and Wand (1992) [7], but any will suffice in case of ties) minimizer of $MISE(h)$. Minimizers of other criteria will be denoted by similar notation.

As noted above, the visual error criterion $VE_2(\widehat{f}_h \to f)$ was developed in Marron and Tsybakov (1993) [8]. See that paper for motivation and discussion. This is defined as follows. A continuous function $f : [a, b] \to \mathbb{R}$ can be represented by its "graph in $\mathbb{R}^2$",

$$G_f = \{(x, y) : x \in [a, b],\ y = f(x)\} \subset \mathbb{R}^2.$$

A basic building block for constructing "distances" between two graphs is the shortest distance from a point (x, y) to the graph G,

$$d((x, y), G) = \inf_{(x', y') \in G} \left\|(x, y) - (x', y')\right\|_2,$$

where $\|\cdot\|_2$ denotes the usual Euclidean distance. These distances are summarized, in a squared integrated way to yield the *asymmetric* "visual error" criterion

$$VE_2(\widehat{f}_h \to f) = \left[\int_a^b d((x, \widehat{f}_h(x)), G_f)^2 dx\right]^{1/2}.$$

Many variations on this idea are possible, but reasons behind this particular choice are detailed in Marron and Tsybakov (1993) [8].

In that paper, a symmetrized version of $VE_2(\widehat{f}_h \to f)$ also is seen to have useful properties. However, that is not treated here because $VE_2(\widehat{f}_h \to f)$ is seen there to be superior at tracking an experienced data analyst's desire that the estimate should show no more features than can reasonably be obtained from the data.

For computational feasibility, a discretized version of $VE_2(\widehat{f}_h \to f)$, involving both discretized graphs, and also a Riemann sum approximation, over an equally spaced fine grid of 400 points has been used in all examples here. See Marron and Tsybakov (1993) [8] for details.

There has been an interesting controversy over which should be called the "optimal bandwidth in the L^2 sense", between the respective minimizers of ISE and $MISE$. Passions run surprisingly strong on both sides, even to the point of people on *each* side being unwilling to even admit that a controversy exists, feeling that their view is so obviously correct. See Grund, Hall and Marron (1992) [3] for references and discussion. In that paper it is seen that despite important mathematical and conceptual differences, there is often little difference in terms of assessing bandwidth selectors in most practical situations. However $MISE$ does give some what better "resolution" in comparing bandwidth selectors than ISE, so that is used in the rest of this paper. In the empirical work done for this paper, it has been observed that this same lesson holds for $VE_2(\widehat{f}_h \to f)^2$ vs. $EVE_2(\widehat{f}_h \to f)^2$, so the latter will be used for assessment of bandwidth selectors throughout this paper. Explicit details are not presented here, because the main ideas are essentially the same as those of Grund, Hall and Marron (1992) [3], and thus do not seem to be worth the journal space. However detailed summaries are available from the author.

3 New facts on visual error

Part of this research involves studying the error criterion

$$EVE2(h) = EVE_2(\widehat{f}_h \to f)^2$$

in a variety of senses, in examples where some normal mixture densities were used for the underlying f. In particular analogs of all figures have been constructed for all of the 15 target densities of Marron and Wand (1992) [7], but only some representative ones are presented here. The underlying density in Figure 1 is density number 14 from that paper.

Insight into how the *average* curves $MISE(h)$ and $EVE2(h)$ (as functions of h) compare may be obtained from Figure 2.

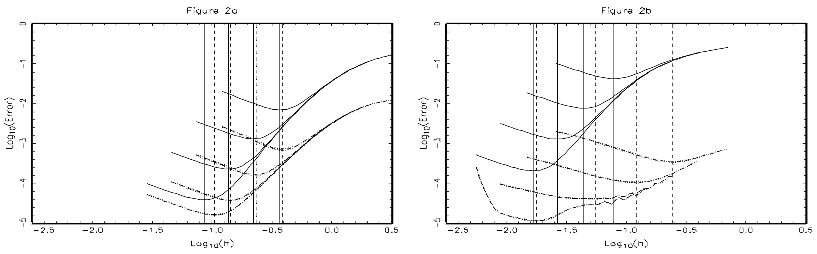

Figure 2: Plots of $MISE(h)$ (solid curves) and $EVE2(h)$ (dashed curves) as functions of the bandwidth h, for $n = 100, 1000, 10,000$ and $100,000$. The underlying densities are: (a) number 2, (b) number 4, from Marron and Wand. For each n, a grid of 49 (logarithmically) equally spaced bandwidths, from $\frac{1}{3}h_{MISE}$ to $9h_{MISE}$ is used. Monte Carlo error in calculation of $EVE_2(\widehat{f}_h \to f)^2$ is seen to be negligible through the inclusion of 95% pointwise confidence bands, as dotted curves. Vertical lines show minimizers of $MISE$ (solid) and $EVE2$ (dotted).

There is no error in the calculation of the $MISE$ curves, since the exact formulas in Marron and Wand (1992) [7] were used. However the $EVE2$ curves were calculated by simulation, in particular by averaging $VE_2(\widehat{f}_h \to f)$ over 500 pseudo data sets (for each n) at each bandwidth. So there is some Monte Carlo error to take into account. This is done in Figure 2, through showing upper and lower 95% pointwise asymptotic confidence bands (constructed using the classical asymptotic normality formula: $\overline{X} \pm 1.96\frac{s}{\sqrt{500}}$). Note that these dotted lines are, at most h locations, nearly indistinguishable from the average curve, so the 500 replications is enough for at least most visual lessons to be correct.

Figure 2a shows this comparison for density number 2. Similar pictures were obtained for density numbers 1, 6, 7, 8, and 9. These are essentially the "easy to estimate densities", where most of the features of the underlying f can be recovered well with only moderate sample sizes. In these situations, there will be little difference between assessing the performance of a data driven bandwidth, say $\widehat{h}$, by $MISE(\widehat{h})$ or by $EVE2(\widehat{h})$. This is in keeping with the findings of Marron and Tsybakov (1993) [8] that in many situations, all error criteria work in essentially the same way. Also, as expected from that paper, $MISE(h)$ is consistently larger than $EVE2(h)$.

Figure 2b shows the results for density number 4, and is fairly representative of the results for the remaining densities. These are mostly characterized as "hard to estimate" densities, which have at least some features which one can not expect to completely recover from a sample of only moderate size. Note that for $n = 100$, 1000 and 10,000, major differences will show up when bandwidth selectors are compared by the 2 methods. However, for $n = 100,000$ the answers will be closer. This is because for such large n, this density no longer has features that are "too hard to obtain", and hence joins the first group, represented in Figure 2a. Note that $EVE2(h)$ tends to be "flatter" in these situations (which means not such large penalties for being further from the minimizer), and also *much* smaller than $MISE(h)$, which also is not surprising in view of the intuition in Marron and Tsybakov (1993) [8].

A disturbing feature of Figure 2b is the "wiggles" on the right hand sides of the $EVE2(h)$ curves for large n. Note that the confidence bands indicate this is systematic, and not due to Monte Carlo variability (as is indicated by the fact that these wiggles "line up" for $n = 10,000$ and $100,000$). This has not been carefully investigated, but seems to be due to the discretization in the numerical calculation of $VE_2(\widehat{f}_h \to f)$. In particular, all of these densities have regions where $|f'(x)|$ is extremely large (the "slopes of the spikes"). In such regions there can be substantial $\mathbb{R}^2$ distance between successive points in the discretized graphs. This appears to make some bandwidths less desirable than others.

Note this effect only occurs on the right side of the curves, where bias is dominant. On the left side, large random variability in the estimates entails that in averaging this effect disappears.

An unpleasant feature of these "wiggles" in $EVE2(h)$ is that many local minima are created. Local minima can also appear in $MISE(h)$, as seen in Marron and Wand (1992) [7], but this happens far less frequently. Even more local minima were encountered in the random curves $VE_2(\widehat{f}_h \to f)$. This is consistent with the fact that ISE tends to have more local minima than $MISE$, as quantified in Hall and Marron (1991) [4].

A completely different way of looking at $VE_2(\widehat{f}_h \to f)$ is to study how its minimizer (where ties are broken by taking the smallest of the set of minimizers), h_{VE2}, compares for given data sets to other notions of "optimal" bandwidth. Here it is compared with h_{ISE}, the minimizer of ISE, and h_{IAE}, the minimizer of the Integrated Absolute Error,

$$IAE = \int \left| \widehat{f}_h(x) - f(x) \right| dx,$$

and also h_{SUP} the minimizer of the supremum norm, $\sup_x \left| \widehat{f}_h(x) - f(x) \right|$. Figure 3 gives an indication of the multivariate distribution of the random vector $\begin{pmatrix} h_{IAE} & h_{ISE} & h_{SUP} & h_{VE2} \end{pmatrix}^T$.

Figure 3a is representative of the "easier to estimate densities", numbers 1, 2, 6, 7, 8, and 9. Note that all notions of "optimal bandwidth" roughly coincide here, with a very strong correlation between them, although less so for h_{SUP}. This is because for these densities all notions of optimal "feel roughly the same aspects of the data and the density" (although this is least true for h_{SUP}).

Figure 3b shows typical behavior for the "harder to estimate densities", numbers 3, 4, 10, 12, 13, 14 and 15. Note that the various optimal bandwidths now behave much differently from each other, both in terms of much smaller correlations, and also different mean behaviors. This is because "each feels different aspects of the variability in the data", in particular h_{SUP} is driven

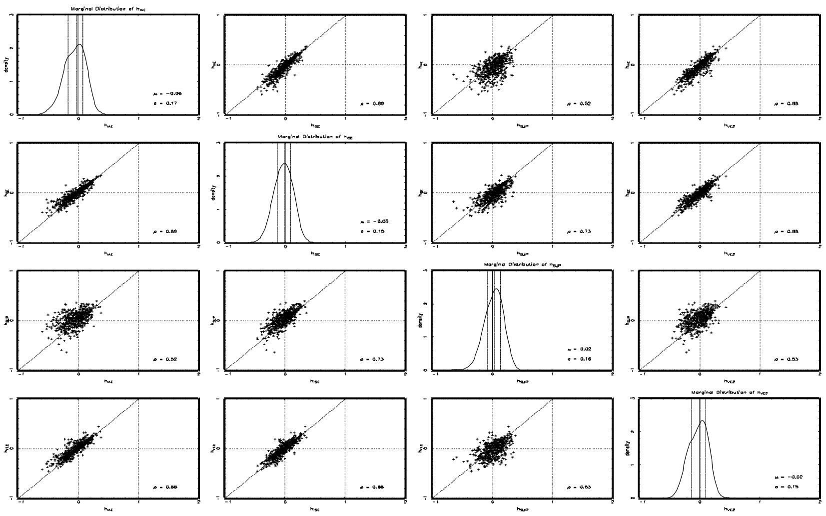

Figure 3: Draftsman's scatterplots of the multivariate distribution of h_{IAE}, h_{ISE}, h_{SUP}, h_{VE2}. From 500 pseudo data sets of size $n = 100$, from density 2. All scatterplot axes are on scale of $\log_3(h) - \log_3(h_{MISE})$. Diagonal plots show kernel density estimates of univariate marginal density for each bandwidth, together with vertical dotted lines showing the quartiles of these distributions.

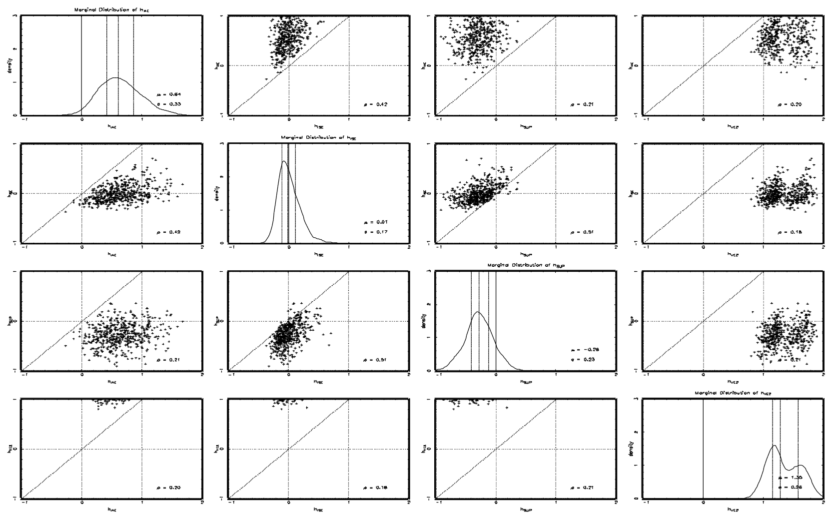

Figure 4: Same as Figure 3, but from density 3.

by behavior at the peak, while h_{VE2} is more strongly influenced by the flat parts of density, while h_{ISE} and h_{IAE} feel different things in between these extremes.

One interesting exception to the above pattern was density number 5. This is not shown explicitly, to save space, but the selectors h_{IAE}, h_{ISE}, and h_{SUP} behave much the same as in Figure 3a, with very strong correlation and similar means. However the distribution of h_{VE2} is substantially larger than the others, and also has essentially no correlation, because this method feels behavior in the tails of distribution.

Another interesting exception is density number 11, in the case $n = 100$. Here h_{IAE}, h_{ISE}, and h_{VE2} work much as in Figure 3a, since with respect to these criteria, this density is essentially the same as number 6. however the sup norm feels behavior of the estimate at the thin tall spikes almost exclusively. Hence it is not surprising that h_{SUP} is uncorrelated with the others, and also spread over a very wide range.

4 Comparison of Bandwidth Selectors

A criterion which has been used for the simulation comparison of data driven bandwidth selectors, see for example Jones, Marron and Sheather (1992) [6], is

$$MISE - CRIT = \underset{\widehat{h}}{avg} \frac{MISE(\widehat{h}) - MISE(h_{MISE})}{MISE(h_{MISE})},$$

where the average is taken over simulated realizations of $\widehat{h}$. A direct analog of this, based on $EVE2(h)$ is

$$EVE2 - CRIT = \underset{\widehat{h}}{avg} \frac{EVE2(\widehat{h}) - EVE2(h_{EVE2})}{EVE2(h_{EVE2})},$$

where h_{EVE2} is the h to minimize $EVE2(h)$. These two ways of comparing data driven bandwidths are compared in Figures 5, 6 and 7. These are visual summaries, intended to provide more ready comparisons of several types than the usual set of tables (whose presentation would be quite voluminous and tedious to interpret in this particular context).

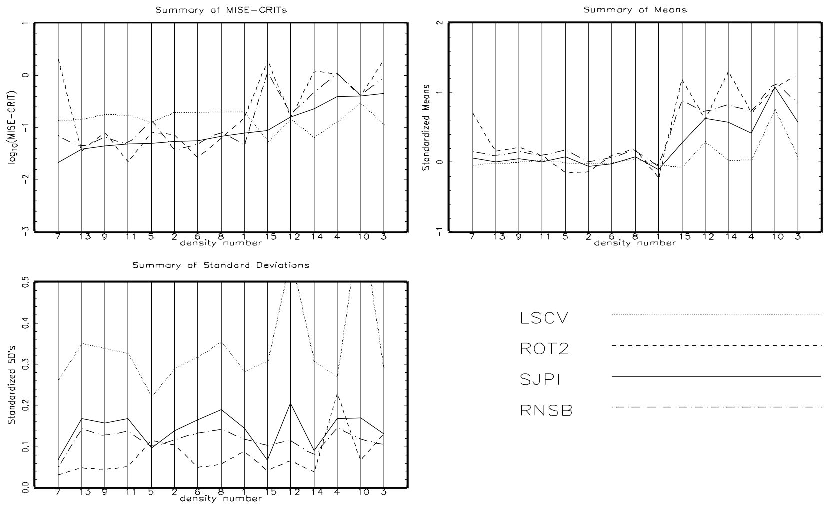

Figure 5: Visual displays of summary statistics for comparison of bandwidth selectors, over all 15 target densities in Marron and Wand, for $n = 100$. In each case the densities are ordered by $MISE - CRIT$ for $SJPI$. Means and standard deviations are on the scale of $\log_3(\widehat{h}) - \log_3(h_{MISE})$.

Since the density numbering in Marron and Wand (1992) [7] was not intended to be relevant for bandwidth selection, various reorderings have been experimented with. Most informative seems to be to order in terms of $MISE - CRIT$ for a bandwidth selector of interest. Note that in the $MISE$ sense, the methods are not comparable, with no method being dominant in all situations. However, Jones, Marron and Sheather (1992) [6] suggest $SJPI$ as a reasonable "all around favorite", because it is never too far from the best (while each of the others is far from best for at least some densities). Reasons for the various performances are seen clearly by studying the means and variances of the given methods. For example, when $LSCV$ gives poor performance in

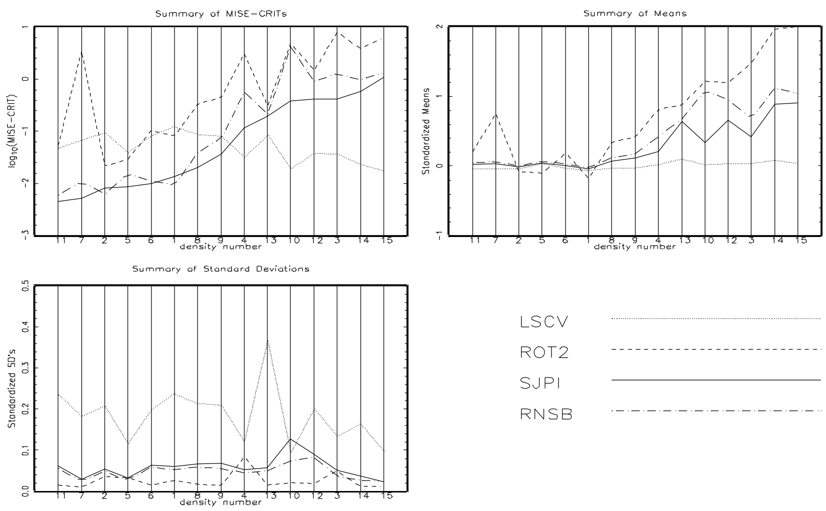

Figure 6: Same as Figure 5, but for $n = 1000$.

the $MISE$ sense, it is because it has too much variability (which is already well known). When $LSCV$ works well, it is because it has substantially less bias than the others. The bandwidth selector $ROT2$ gives generally poor performance because, while it has smaller than usual variance (because the only estimation is of "scale" versus the more challenging estimation that is at least implicitly attempted by the other methods). Further insight into this "variance - bias" effect comes from Figure 7.

Note that some of the $MISE - CRIT$ summaries in Figures 5 and 6 are not particularly useful, because the $MISE$ sense of a good bandwidth is different from visual impression. This is roughly the same phenomenon observed in Figure 1. To see how the lessons of Figures 5 and 6 change, when $MISE - CRIT$ is replaced by $EVE2 - CRIT$, see Figures 8 and 9.

The means and standard deviations are again displayed, because the ordering of the densities is now different from that of

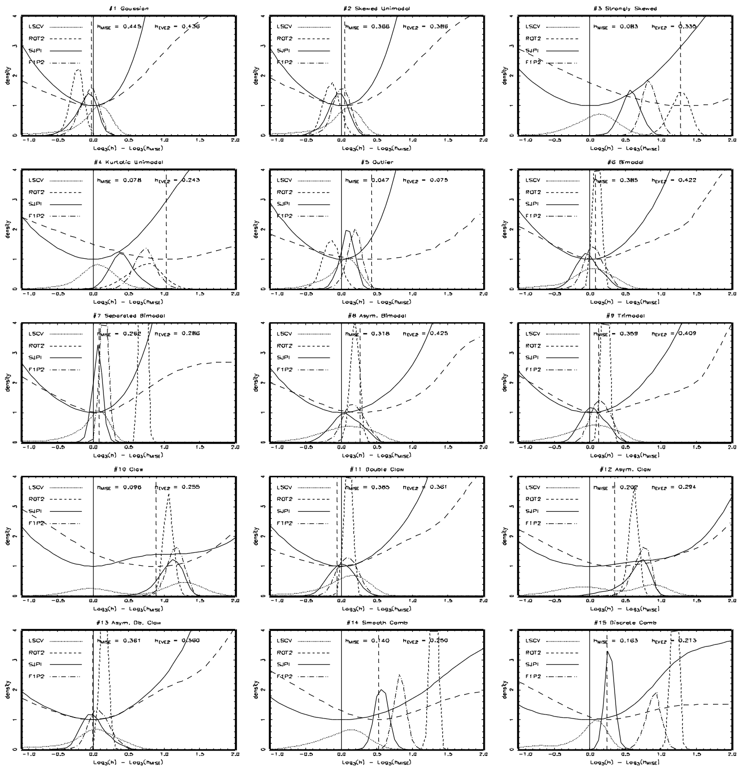

Figure 7: Kernel density estimates showing the distributions of data driven bandwidths (with same line types as in Figures 5 and 6), for $n = 100$, and each of the target densities, again standardized to the scale $\log_3(\widehat{h}) - \log_3(h_{MISE})$. The overlayed solid curve is $MISE(h)$ and the dashed curve is $EVE2(h)$. $MISE - CRIT$ and $EVE2 - CRIT$ behavior is understood by "plugging the densities into the corresponding curve".

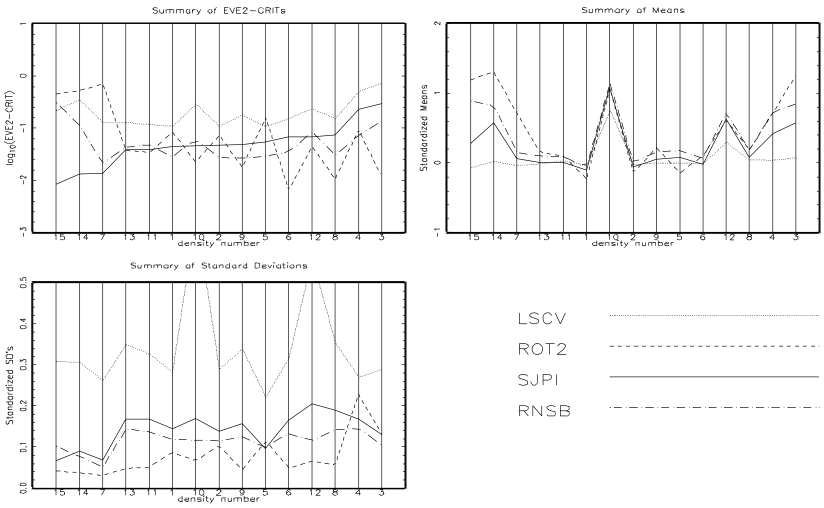

Figure 8: Visual displays of summary statistics for comparison of bandwidth selectors, over all 15 target densities in Marron and Wand, for $n = 100$. In each case the densities are ordered by $EVE2-CRIT$ for $SJPI$. Means and standard deviations are on the scale of $\log_3(\widehat{h}) - \log_3(h_{MISE})$.

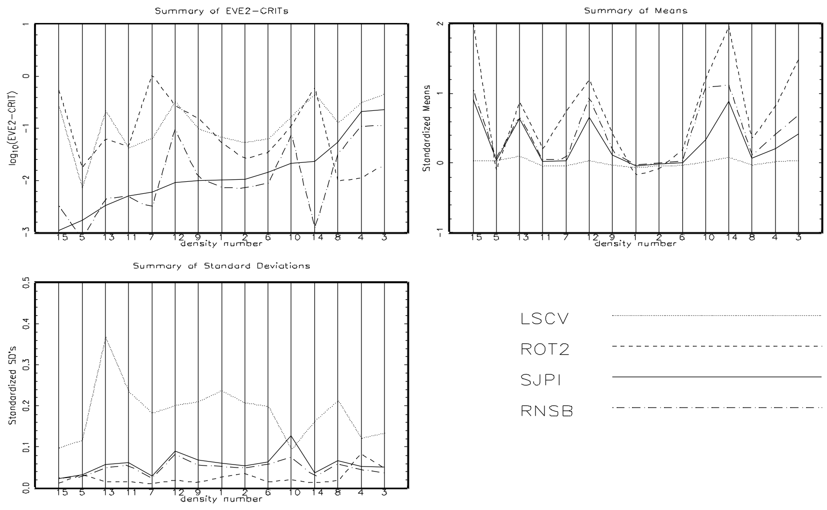

Figure 9: Same as Figure 8, but for $n = 1000$.

Figures 5 and 6. Note that $SJPI$ is still never too far from best, so the previous recommendation of this as a general purpose default seems to remain acceptable. However, the relative standing of $LSCV$ has changed a lot. It goes from being best a fair number of times in the $MISE$ sense, to being almost uniformly worst in the $EVE2$ sense, and often by quite a lot. Figure 7 shows that it's difficulties are a combination of "unbiased with respect to h_{MISE}" actually being a drawback in this sense, and it's very large variance. This interpretation of the performance of $LSCV$ is much more in keeping with intuitive and practical opinion of its behavior, see for example Sheather (1992) [10]. Another noticeable difference between Figures 5, 6, 8 and 9, is that $ROT2$ looks much better now, at least for $n = 100$. This is also in keeping with anecdotal evidence several researchers have expressed in personal conversation. An inspection of Figure 7 shows that the main reason for this good performance is the low variance of $ROT2$ combined with its often being fortunately close to h_{EVE2}. This closeness disappears for $n = 1000$, where this good performance falls off significantly. Finally note that $RNSB$ now looks better with respect to $SJPI$ than it did in the $MISE$ sense. This is interesting because $RNSB$ has superior asymptotic performance, and it has been previously thought that these asymptotics required unreasonably large sample sizes before the superior performance of $RNSB$ appeared.

A fascinating area for future research is to design bandwidth selection methods which specifically target the criterion $EVE2$.

References

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

[2] Fan, J. and Marron, J. S. (1993) Fast implementations of nonparametric curve estimators, unpublished manuscript.

[3] Grund, B., Hall, P. and Marron, J. S. (1992) Loss and risk in smoothing parameter selection, unpublished manuscript.

[4] P. Hall and Marron, J. S. (1991) Local minima in cross-validation functions, *Journal of the Royal Statistical Society, Series B*, 53, 245-252.

[5] Jones, M. C., Marron, J. S. and Park, B.-U. (1991) A simple root n bandwidth selector, *Annals of Statistics*, 19, 1919-1932.

[6] Jones, M. C., Marron, J. S. and Sheather, S. J. (1992) Progress in data-based bandwidth selection for density estimation, unpublished manuscript.

[7] Marron, J. S. and Wand, M. P. (1992) Exact mean integrated squared error, *Annals of Statistics*, 20, 712-736.

[8] Marron, J. S. and Tsybakov, A. B. (1993) Visual error criteria for qualitative smoothing, unpublished manuscript.

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

[10] Sheather, S. J. (1992) The performance of six popular bandwidth selection methods on some real data sets, *Computational Statistics*, 7, 225-250.

[11] Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society, Series B*, 53, 683-690.

[12] Silverman, B. W. (1986) *Density Estimation for statistics and data analysis*, Chapman and Hall, New York.

5 Appendix: Bandwidth selectors

In all cases, calculations were done using the ideas behind the fast "binned" implementations described in Fan and Marron (1993) [2]. This involves evaluation at an equally spaced grid of points, which was taken to be 400 points on [-3,3].

For each of these bandwidth selectors, only a quick definition, and a few important facts are mentioned. See Jones, Marron and Sheather (1992) [6] for further discussion.

5.1 Least Squares Cross Validation

This method was proposed by Rudemo (1982) [9] and Bowman (1984) [1]. This bandwidth is taken to be the minimizer (in the case of multiple local minimizers, the largest h value is used here, as that seems to give the best performance) of the function

$$LSCV(h) = \int \widehat{f}_h(x)^2 \, dx - 2n^{-1} \sum_{i=1}^{n} \widehat{f}_{h,i}(x)$$

where $\widehat{f}_{h,i}(x)$ is the kernel density estimate based on the sample with X_i deleted. This minimizer has an intuitively attractive connection to h_{ISE}. However, it is known to have rather poor practical and asymptotic properties.

5.2 Sheather Jones Plug In

This method was proposed in Sheather and Jones (1991) [11]. This bandwidth selector is the solution (taken to be the largest h value when there are multiple solutions) of the fixed point equation (in the variable h)

$$h = \left[\frac{\int K^2}{\left(\int x^2 K\right)^2 \int \left(\widehat{f}''_{g_2(h)}\right)^2 n} \right]^{1/5}$$

where

$$g_2(h) = C(K) D(\widehat{f}_{g_1}) h^{5/7}$$

where $C(K)$ and $D(f)$ are appropriate constants, and $\widehat{f}_{g_1}$ is a pilot estimator, whose bandwidth is chosen by the normal reference distribution method. This bandwidth selector is known to have much better asymptotic and also applied performance than h_{LSCV}.

5.3 Rule Of Thumb 2

This method is proposed in Silverman (1986) [12].This is defined as

$$h_{ROT2} = 0.9 \left[\frac{\int K^2}{\left(\int x^2 K\right)^2 \int \left(\varphi''_{\widehat{\sigma}}\right)^2 n} \right]^{1/5}$$

where $\varphi_{\widehat{\sigma}}$ denotes the normal probability density with variance $\widehat{\sigma}^2 = \min\left(s^2, \frac{\widehat{IQR}}{IQR_\varphi}\right)$, where s^2 is the usual sample variance, $\widehat{IQR}$ is the sample Inter Quartile Range, and IQR_φ is the Inter Quartile Range of the standard normal distribution. Asymptotically, this method is far worse than any of the others here, because (for f different from the Gaussian in shape) it does not converge to any notion of "optimal bandwidth". However, it has the advantage that it has less variability, because only the scale is estimated.

5.4 Root N Smoothed Bootstrap

The version of this used here is that of Jones, Marron and Park (1991) [5]. This is the minimizer (taken as the global minimizer in the case of local minima) of the function

$$RNSB(h) = \frac{\int K^2}{nh} + \int \left[K_h * f_{g_2(h)} - f\right]^2$$

where

$$g_2(h) = C(K) D(\widehat{f}_{g_1}) n^{-23/45} h^{-2}$$

where $C(K)$ and $D(f)$ are different appropriate constants, and $\widehat{f}_{g_1}$ is a pilot estimator, whose bandwidth is chosen by a normal reference method. This bandwidth selector has excellent asymptotic properties, even a fast "root n" rate of convergence usually encountered only in parametric contexts.