ELSEVIER Computational Statistics & Data Analysis 42 (2003) 27–36

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Estimating a density by adapting an initial guess

C.J. Albers*, W. Schaafsma

Department of Mathematics and Computing Science, University of Groningen, P.O. Box 800, 9700 AV Groningen, Netherlands

Received 1 January 2001; received in revised form 1 March 2002

Abstract

De Bruin et al. (Comput. Statist. Data Anal. 30 (1999) 19) provide a unique method to estimate the probability density f from a sample, given an initial guess ψ of f. An advantage of their estimate f_n is that an approximate standard error can be provided. A disadvantage is that f_n is less accurate, on the average, than more usual kernel estimates. The reason is that f_n is not sufficiently smooth. As improvement, a smoothed analogue $f_n^{(m)}$ is considered. The smoothing parameter m (the degree of a polynomial approximation) depends on the supposed quality of the initial guess ψ of f. Under certain conditions, the resulting density estimate $f_n^{(m)}$ has smaller L_1-error, on the average, than kernel estimates with bandwidths based on likelihood cross-validation. The theory requires that the initial guess is made up a priori. In practice, some data peeping may be necessary. The $f_n^{(m)}$ provided look 'surprisingly accurate'. The main advantage of $f_n^{(m)}$ over many other density estimators is its uniqueness (when the procedures developed in this article are followed), another one is that an estimate is provided for the standard error of $f_n^{(m)}$.

Keywords: Density estimation; Nonparametric methods

1. Introduction

A fundamental problem in practical statistics is as follows. *Given* are the outcomes $x_{[1]} < x_{[2]} < \cdots < x_{[n]}$ of an independent random sample $X_1, \ldots, X_n$ from a probability distribution on $\mathbb{R}$ with a density f which is 'smooth' (at least continuous) and strictly positive on a given support (a, b), e.g. $(0, \infty)$, but further unknown. *Required* is an estimate $f_n(x)$ of $f(x)$ in a given point x or, more generally, an estimate f_n of f on (a, b).

* Corresponding author. Tel.: +31-503633962; fax: +31-503633800.
E-mail address: c.j.albers@math.rug.nl (C.J. Albers).

0167-9473/02/$ - see front matter
PII: S0167-9473(02)00124-X

The performance of such density estimate is measured by specifying the 'difference' between the true density and the estimate. Many dissimilarity coefficients can be considered. In this article, we restrict the attention to the total variation distance

$$\|f_n - f\|_1 = \int_a^b |f_n(x) - f(x)|\,\mathrm{d}x$$

because it is invariant under all piecewise continuous (differentiable) bijections and because this dissimilarity coefficient provides the metric which was preferred by a variety of authors (see, e.g., Wand and Devroye (1993), Devroye and Győrfi (1985), and, in particular, De Bruin et al. (1999)).

The empirical distribution and Ferguson's Bayesian modifications of it (see Ferguson (1973)) fail to be continuous estimates of the distribution function F. The total variation distance between these estimates and the true distribution is equal to 2 and consistency is precluded. A plethora of methods exists for specifying estimates of $f(\cdot)$ such that the total variation distance converges to 0 in probability if $n \to \infty$ and f is sufficiently smooth (e.g. Ghosal et al. (1999)). In Ferguson's theory, a Dirichlet prior must be specified by providing a distribution function. The method of De Bruin et al. (1999) is also based on the specification of a distribution function Ψ, now as an initial guess of the true distribution function F. The derivative ψ of Ψ is an initial guess of f. The theory behind this method requires that ψ is chosen a priori. It is assumed that the 'support' $\{x;\ \psi(x) > 0\}$ is specified as the true interval $\{x;\ f(x) > 0\}$ (usually $(0, \infty)$ or $(-\infty, \infty)$). The method provides a special estimate f_n of f.

The underlying rationale is that there are many ways to characterize a probability distribution. One can use the distribution function F, the density function $f = F'$, the quantile function $H = F^{-1}$, the characteristic function, the moment-generating function, etc. Each of these characterizations entails its own estimation approach (Silverman, 1986). The estimates f_n and $f_n^{(m)}$ discussed in this article are based on the characterization by means of H. This involves a peculiar advantage in the sense that the estimates of H derived have positive derivatives.

2. The exact definition of f_n

The quantiles $H(p) = F^{-1}(p)$ $(p \in [0,1])$ are estimated by a smoothed version of their sample analogues. The precise definition is as follows. After the replacement of $x_{[1]}, \ldots, x_{[n]}$ by

$$y_{[0]} = 0, \quad y_{[i]} = \Psi(x_{[i]}) \quad (i = 1, \ldots, n), \quad y_{[n+1]} = 1,$$

the quantile function $B = (F \circ \Psi^{-1})^{-1} = \Psi \circ F^{-1}$ of the distribution of $Y_1 = \Psi(X_1)$ on the unit interval is estimated by means of the Bernstein polynomial approximation

$$B_n(p) = \sum_{i=0}^{n+1} y_{[i]} \binom{n+1}{i} p^i (1-p)^{n+1-i}$$

of degree $n+1$ to the empirical quantile function (the inverse of the empirical distribution function). (The notation $B_n(p)$ is preferred over the usual notation $B_{n+1}(p)$ because the underlying sample size is n.) This approximation is very convenient because the corresponding derivative

$$b_n(p) = \sum_{i=0}^{n} (y_{[i+1]} - y_{[i]}) \binom{n}{i} (n+1) p^i (1-p)^{n-i}$$

is strictly positive and continuous. The idea to use polynomial quantile density estimators, such as $B_n(p)$, to estimate $H(p)$ is not new, see, e.g., Muñoz Perez and Fernández Palacín (1987). In Groningen, the study of such estimators started with Dehling et al. (1991), who showed that the Islamic Mean (very similar to $B_n(\frac{1}{2})$) is an interesting alternative to the sample median for estimating the population median.

The estimation of $f(\cdot)$ is performed via back-transformation. The estimate B_n is used to define the estimate $H_n = \Psi^{-1} \circ B_n$. Next $F_n = H_n^{-1} = B_n^{-1} \circ \Psi$ is used to estimate $F = H^{-1}$ and, finally, the estimate $f_n = F_n'$ of f is obtained by numerical differentiation.

If one compares f_n with other nonparametric density estimates such as kernel estimates k_n (see, e.g., Silverman (1986)), then there are some theoretical advantages of f_n but these are outweighed by practical advantages of k_n, even if the initial guess ψ of f is reliable. Some arguments are as follows:

(1) The asymptotic distribution of f_n can be studied via that of b_n (see De Bruin et al. (1999)). This results in the approximation

$$\mathscr{L} n^{1/4}(f_n(x) - f(x)) \approx \mathscr{N}(0, \sigma^2(x)),$$

where

$$\sigma(x) = \frac{f(x)}{\sqrt[4]{4\pi F(x)(1-F(x))}}.$$

The $\approx$-sign can be replaced by a $\rightarrow$-sign when $\Psi = F$. Unfortunately, we were unable to establish this limiting result for $\Psi \neq F$. Since we shall replace f_n by a more complicated estimate $f_n^{(m)}$, details will not be presented here (but in Albers and Schaafsma (2001), and Albers' forthcoming thesis).

(2) The estimate f_n is not sufficiently smooth: the error $f_n' - f'$ in the derivative does not vanish because it is of order $O(n^{1/4})$ (again, see Albers and Schaafsma (2001)).

(3) The rate of convergence of kernel estimates with optimal bandwidths (requiring knowledge of f) is such that $n^{2/5}(k_n(x) - f(x))$ has a limiting distribution, while $k_n' - f'$ vanishes in probability as $n \rightarrow \infty$, at least under regularity conditions (Silverman, 1978).

(4) An extensive numerical comparison of f_n with the kernel estimates k_n with bi-weight as kernel and bandwidths based on likelihood cross-validation provided (for a special density f) that the L_1-error of k_n is significantly smaller than that of f_n, on the average. If the L_1-error of f_n can be decreased by 15% by using a modification of f_n, then k_n would be about equally good as this modification, on the average (De Bruin et al. (1999), Fig. 4).

(5) If one accepts the approximation presented under (1) then the following derivation for the expected L_1-error of f_n will be accepted as well (we use the notation

$U \sim \mathcal{N}(0,1)$)

$$\mathbf{E}\int |f_n(x) - f(x)|\,\mathrm{d}x = \int \mathbf{E}\,|f_n(x) - f(x)|\,\mathrm{d}x$$

$$\approx n^{-1/4}\mathbf{E}|U| \int \sigma(x)\,\mathrm{d}x$$

$$= n^{-1/4}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt[4]{4\pi}}\int \frac{\mathrm{d}F(x)}{\sqrt[4]{F(x)(1-F(x))}}$$

$$= n^{-1/4}2(\Gamma(3/4))^2\pi^{-5/4}$$

$$= 0.72n^{-1/4}.$$

Note that this asymptotic behaviour does neither depend on f nor on the initial guess ψ.

3. Replacing f_n by its improvement $f_n^{(m)}$

Under (4), at the end of Section 2, it was suggested that f_n can compete with k_n if its standard deviation can be reduced by something like 15%. How to improve f_n, that is the question. Note the theoretical arguments behind B_n, b_n and, hence, f_n spelled out in De Bruin et al. (1999). A practical argument in favor of these estimates is the strict positivity of b_n mentioned in Section 2. This should certainly not be given up. A natural way of smoothing can be achieved by partitioning the sample into $k=m^{-1}n$ sub-samples of size m each (for the sake of simplicity, assume $n=km$), and taking the arithmetical average of the resulting estimates: for each subsample $y_1^{(h)},\ldots,y_m^{(h)},(h=1,\ldots,k)$ the density estimate $f_{m,h}$ is derived from

$$B_{m,h}(p \mid y_1^{(h)},\ldots,y_m^{(h)}), \quad h=1,\ldots,k$$

in the same way as f_n was derived from B_n. Define $\bar{f}_{n,m}$ as the average of the h estimates $f_{m,h}$. The approximate result

$$\mathscr{L}m^{1/4}(f_{m,h}(x) - f(x)) \approx \mathcal{N}(0,\sigma^2(x))$$

of Section 2 implies that

$$\mathscr{L}m^{1/4}(\bar{f}_{n,m}(x) - f(x)) \approx \mathcal{N}(0,k^{-1}\sigma^2(x)) = \mathcal{N}(0,n^{-1}m\sigma^2(x))$$

or, equivalently, that the standard error of $\bar{f}_{n,m}(x)$ is $m^{-1/4}k^{-1/2}\sigma(x)$, which equals $m^{1/4}n^{-1/2}\sigma(x)$, i.e. $k^{-1/4}$ times that of $f_n(x)$. This suggests that the standard error is reduced by 15% or more if $(m/n)^{1/4} < 0.85$ or, equivalently, if m is less than $< 0.53n$ (see Remark (4) in Section 2).

To remove the permutation dependence (and the assumption $n=km$) a U-statistic symmetrization will be applied. It will have the effect that the degree n of the polynomial expression $b_n(p)$ is lowered to the degree m of the estimate $b_n^{(m)}(p)$ to be derived. The positivity of the quantile density estimate will not be affected.

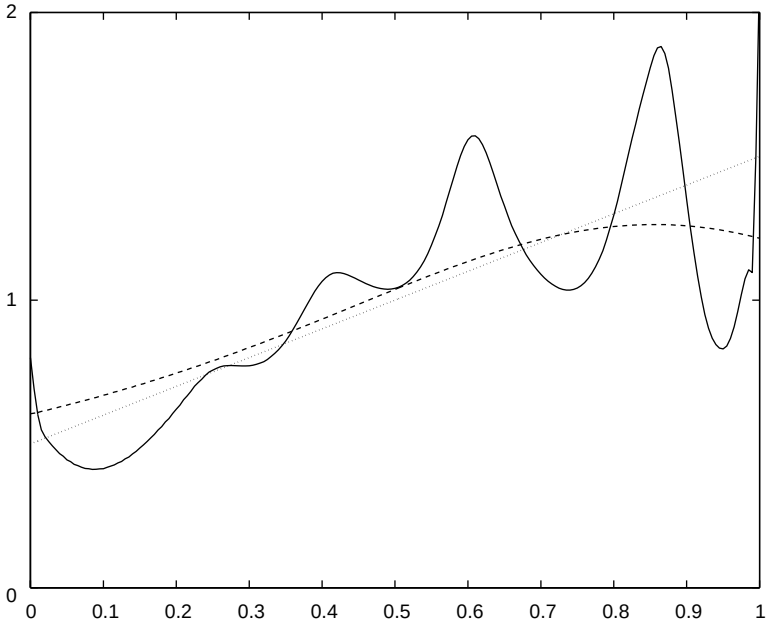

Fig. 1. Visualization of the improvement discussed in Section 3. A sample of size $n = 100$ is taken from the distribution on $(0,1)$ with density $f(x) = \frac{1}{2} + x$ (the dotted line). Using the uniform density as initial guess, the densities $f_n^{(m)}$ $(m = 0,\ldots,100)$ were computed and those corresponding to $m = 6$ (dashed line) and $m = 100$ (solid line) are displayed.

Let $B_m(p \mid y_1,\ldots,y_m)$ denote the approximation introduced in the previous section, applied to the m observations $y_1,\ldots,y_m$ (with $y_i = \Psi(x_i)$). Note that the degree of the polynomial is thus lowered from $n+1$ to $m+1$. The smoothed analogue is the U-statistic

$$B_n^{(m)}(p) = \binom{n}{m}^{-1} \sum_{1 \leqslant \alpha_1 < \cdots < \alpha_m \leqslant n} B_m(p \mid y_{\alpha_1},\ldots,y_{\alpha_m}),$$

which can be rewritten as the L-statistic

$$p^{m+1} + \sum_{j=1}^{m} \binom{m+1}{j} p^j (1-p)^{m+1-j} \sum_{i=j}^{n-m+j} \frac{\binom{i-1}{j-1}\binom{n-i}{m-j}}{\binom{n}{m}} y_{[i]}.$$

So, $B_n^{(m)}$ is the average of the quantile function estimates based on all size-m subsets of the sample. From this, we get $F_n^{(m)} = B_n^{(m)^{-1}} \circ \Psi$, $f_n^{(m)} = F_n^{(m)'}$, etc. For $m \to \infty$, the $B_m(p \mid \ldots)$ are consistent estimators of $(\Psi^{-1} \circ H)(p)$. For m fixed, $B_n^{(m)}(p)$ is an unbiased estimate of $\mathbf{E}(B_m(p \mid X_1,\ldots,X_m))$ which, of course, depends on the underlying distribution function of X_1. In principle, the (exact) theory of Hoeffding (1948) about

U-statistics is applicable but the complexity of the formulas and the dependence on f, F, H, etc. precludes application in practice. However, the rate of convergence is better than $n^{-3/8}$, which arises for $\bar{f}_{n,m}$ when $k = n^{1/2}$ is taken. The next section will contain a numerical elaboration.

The example in Fig. 1 gives an illustration of errors involved in using f_n and $f_n^{(m)}$ and suggests that improvement indeed is gained by lowering the degree of the estimating polynomial.

4. Choosing *m*

In the current context of density estimation, it is obvious that m should depend on the sample size and on the reliability of the initial guess. To investigate the structure of this dependence, extensive Monte Carlo simulations have been performed. Results will be presented and discussed in this section.

Given some density f, consider (estimates of) the expected value of the L_1-distance $\|f_n^{(m)} - f\|_1$ as a function of m. The optimal m, i.e. the value for which this expected L_1-distance is minimum, depends on the sample size and reliability of ψ. This optimal m is studied using the special densities

$$f_a(x) = (1 - a) + 2ax, \quad |a| \leqslant 1, \ \ 0 \leqslant x \leqslant 1,$$

and the standard-uniform density $\psi(x)=1$ as initial guess. Fig. 1 provides a visualization for $a = \frac{1}{2}$.

Computations were performed for a variety of values of n, as well of a. Variations in a lead to variations in the reliability of the initial guess through the relation $\|f_a - \psi\|_1 = \frac{1}{2}|a|$. For each pair (n,a) in Table 1, a random sample (of size n) is drawn from f_a. For all possible values of m ($m=0,1,\ldots,n$), $\|f_a - f_n^{(m)}\|_1$ is calculated. This process is replicated 249 times and Table 1 displays those m for which the average total variation distance between true and estimated density was smallest, for the $R=250$ simulations. These simulations suggest that the optimal m is approximately proportional to $n^{1/2}$ and $\|f_a - \psi\|_1$.

The densities f_a studied were very regular with only one sign change of $f_a - \psi$. That is why a second simulation study was carried out involving $f_{\frac{1}{2}} = f_1$ and four

Table 1
Optimal m for the cases f_a with various a (horizontal axis) and n (vertical)

	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
25	0	0	1	2	2	3	4	5	5	6	7
50	0	1	2	3	4	5	6	7	8	9	10
75	0	1	2	3	4	6	7	9	10	11	11
100	0	1	2	4	5	7	8	10	11	13	14
150	0	2	3	5	7	8	10	12	14	16	17
200	0	2	4	6	8	10	12	14	16	18	20
250	0	2	4	6	9	11	13	15	17	19	21

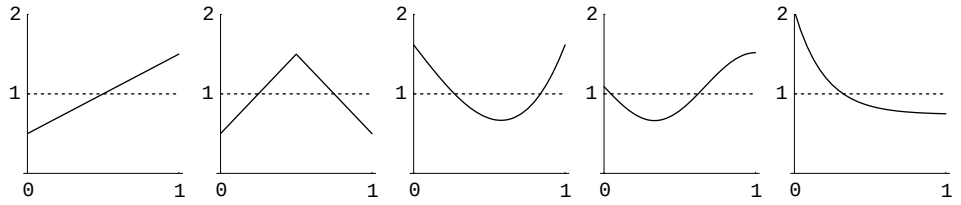

Fig. 2. From left to right: densities f_{I} up to f_{V}, discussed in Section 4.

alternative densities, all with the unit interval as support, $\psi(x)=1$, and $\|f-\psi\|_1=0.25$. The examined densities are

$$f_{\mathrm{I}}(x) = 0.5 + x,$$

$$f_{\mathrm{II}}(x) = 1.5 - |2x - 1|,$$

$$f_{\mathrm{III}}(x) = 1.616 - 2.464x(1 - x^2),$$

$$f_{\mathrm{IV}}(x) = 1.091 - 0.427\sin(4.712x),$$

$$f_{\mathrm{V}}(x) = 0.741 + 1.301\mathrm{e}^{-5x}.$$

See Fig. 2. We believe that these densities are 'representative' for distributions occurring in practice. Analyzing these five densities will give insight in the dependence of the optimal m on the shape or regularity of the density.

For cases I and V, deviations $\psi(x) - f(x)$ are more regular (involving only one crossing of 0) than for the cases II to IV where two crossings occur. It is reasonable to expect that the optimal m is smallest in cases I and V.

To quantify the optimal m in these 5 cases, a sample of size $n = 100$ is drawn from each density, and the corresponding $f_{100}^{(m)}$ and $\|f_{100}^{(m)} - f\|_1$ are computed for all possible values of m ($m = 0, 1, \ldots, 100$).

This process is performed $R = 500$ times. The average L_1-errors for a deliberate choice of values of m is reported in Table 2. The last row of this table contains the values $0.72m^{1/4}n^{-1/2}$, derived under (5) in Section 1. These values provide a first, conservative, approximation to the expected L_1-errors, especially those for the optimal m (the boldfaced ones), since they correspond to the average L_1-error of the less accurate estimates $\bar{f}_{n,m}$ studied at the beginning of Section 3. The approximation in case $m = 0$ is deleted because it involves a degeneracy (if $m = 0$ is actually used, then the value 0.250 appears as the constant true value).

The first column of Table 2 provides the values $\|f - \psi\|_1 = 0.25$ because if $m = 0$ then the estimates correspond to the initial guess ψ. It follows from the first row that,in case $f = f_{\mathrm{I}}$, the expected L_1 error is minimum if m is 6 or 7 (the fourth row of Table 1 provides the value 7 if $a = \frac{1}{2}$). The second, third and fourth row provide that m should be about twice as large if the behaviour of f is less regular. Row five shows, as expected, that m must be chosen substantially smaller if these f's are replaced by the more regular one f_{V}. Both simulation studies taken together indicate that m should

Table 2
Performance of f_{I}, f_{II} and f_{III} and a theoretical approximation. Displayed values are $1/R\Sigma\|f_{\cdot} - f_{\cdot,n}^{(m)}\|_1$, the bold-faced values are the minima for each $f_{\cdot}$.

m	0	4	5	6	7	8	10	15	20	30
f_{I}	0.250	0.096	0.091	**0.089**	**0.089**	0.090	0.093	0.097	0.106	0.122
f_{II}	0.250	0.182	0.167	0.155	0.145	0.138	0.129	**0.122**	0.124	0.134
f_{III}	0.250	0.184	0.168	0.156	0.146	0.138	0.128	**0.123**	0.125	0.134
f_{IV}	0.250	0.147	0.138	0.132	0.127	0.124	0.119	**0.116**	0.119	0.130
f_{V}	0.250	0.104	0.096	0.092	0.090	**0.089**	0.090	0.098	0.107	0.124
Theor.	0.250	0.102	0.108	0.113	0.117	0.121	0.128	0.141	0.152	0.169

depend on the sample size n, the value v of $\|f-\psi\|_1$ where optimality is required, and the (expected) regularity of f. The latter will be quantified by the symbol w, where w is equal to 1 if the number of sign changes of $\psi - f$ is equal to 1, like in the cases of f_a, f_{I} and f_{V}, and it is equal to 2 if two or more sign changes are occurring (cases f_{II} to f_{IV}).

5. Conclusion

The simulation studies performed can be summarized by stating that the optimal m is approximately equal to

$$m^* = 2.6n^{1/2}\, vw.$$

Note that $m^* = 6.5$ in the case $n = 100$, $v = 0.25$, $w = 1$ studied in the fourth row of Table 1 and the first row of Table 2. The choice $m^* = 13$ is indicated for f_{II}, f_{III} and f_{IV}. In practice, f is unknown and the research worker will have to choose the value v of $\|f-\psi\|_1$ and the value $w \in \{0,1\}$ where 'optimality is required'.

6. Discussion

The nonparametric density estimates f_n and $f_n^{(m)}$ require that an initial guess ψ of the true density f is made. In principle, the initial guess should be made *a priori* because the theory is based on this assumption. We believe that in practice some data peeping may be allowed. The subjectivity involved in the specification of the initial guess remains visible if m is small. If one chooses $m=n$ such that $f_n^{(n)} = f_n$ corresponds to the estimator studied in De Bruin et al. (1999), then one lets the data speak almost exclusively and the information provided by ψ is almost completely ignored. That is why f_n cannot compete with estimates that take more information into account. The situation becomes more interesting, but also more complicated, if one accepts the idea that ψ is 'reliable' in the sense that the difference $\|\psi - f\|_1$ between ψ and the true but unknown f may be expected to be less than some constant c which is substantially smaller than the upper bound 2. If, e.g., one is willing to believe that $\|\psi - f\|_1$ is < 0.50 then one will try to choose m such that m is 'optimal' if $\|\psi - f\|_1$ is 0.25. Constructing m in such a way, enables one to incorporate the supposed reliability,

Table 3
A random sample of size $n = 100$ is drawn from $f(x) = \frac{1}{2} + x$ on $(0,1)$, with $\psi(x) = 1$. On basis of this sample, estimates $f_{100}^{(m)}$, with $m = 4, 6, 7, 8, 12, 20$ are constructed. This process is repeated 249 times, the averages of the $R = 250$ L_1-distances $\|f_{100}^{(m_1)} - f_{100}^{(m_2)}\|_1$ are reported

	4	6	7	8	12	20
4	0.000	0.025	0.033	0.041	0.063	0.090
6		0.000	0.009	0.017	0.039	0.068
7			0.000	0.008	0.031	0.060
8				0.000	0.023	0.054
12					0.000	0.031

of the initial guess in the construction of the density estimate. The rule of thumb $m^* = 2.6n^{1/2}\,vw$ mentioned at the end of Section 4 will then provide $m^* = 6.5w$ if $n = 100$ and, hence, $m^* = 6.5$ if exactly one sign change of $\psi - f$ is expected. This implies that the 'uniqueness' of f_n is preserved, at least to some extent. The 'unique' nonparametric density estimate $f_n^{(m)}$ thus defined has an expected L_1 error which is considerably smaller than that of f_n if the true density f is not too much different from the initial guess ψ. It is natural that, under these conditions, $f_n^{(m)}$ will be superior to conventional kernel estimates with bandwidths based on likelihood cross-validation. This is supported by the simulations. However, the advantage of $f_n^{(m)}$ to general kernel estimates, is partly because of the additional knowledge incorporated in $f_n^{(m)}$. The situation may change if the kernel and bandwidth are also chosen on the basis of the initial guess ψ. In practice, we recommend the choice $m = 2.6n^{1/2}vw$ and it will depend on the true density f whether $f_n^{(m)}$ has the smallest expected L_1 error or some kernel estimate with a bandwidth determined such that it is optimal for $f = \psi$.

Table 3 indicates that $\|f_n^{(m_1)} - f_n^{(m_2)}\|_1$ is satisfactorily small if m_1 and m_2 are 'not much different'. This is, of course, of considerable importance, because it would have been unfortunate if minor variations in the smoothing parameter m would induce major variations in the resulting density estimate.

We claim that $f_n^{(m)}$ with $m = 0.65n^{1/2}w$ (the case $v = 0.25$), possibly with $w = 1$, is quite reasonable in practice, as is indicated by simulations, and that the formula

$$m^{1/4}n^{-1/2}\hat{\sigma}(x) = \frac{(0.65w)^{1/4}}{n^{3/8}} \frac{f_n^{(m)}(x)}{\sqrt[4]{4\pi F_n^{(m)}(x)(1 - F_n^{(m)}(x))}}$$

for the corresponding standard error is not unreasonable, though perhaps a bit too conservative (see the beginning of Section 3). The first part of this claim is supported by extensive applications to problems from practice (e.g. the determination of the distribution of concentrations of chemical substances in surface water), providing estimates $f_n^{(m)}$ that are very appealing. These applications required some data peeping for constructing reliable initial guesses. It is obvious that situations exist where data peeping is not necessary because, e.g., experiences of a previous investigation providing ψ have to be adapted to the current state of nature. The new estimator can be used in statistical inference, e.g. in the testing of the goodness of fit hypothesis $H_0 : f = \psi$. In that case, the choice of m has to be reconsidered: it should be much smaller (work in progress).

Acknowledgements

The authors thank the referees and associate editor for their helpful comments.

References

Albers, C.J., Schaafsma, W., 2001. Details on the standard error of a special density estimate. Technical Report IWI-2001-5-04, University of Groningen. *Note*: part of Albers' forthcoming. Ph.D. Thesis. Available from http://www.math.rug.nl/∼casper.

De Bruin, R., Salomé, D., Schaafsma, W., 1999. A semi-Bayesian method for nonparametric density estimation. Comput. Statist. Data Anal. 30, 19–30.

Dehling, H.G., Kalma, J.N., Moes, C., Schaafsma, W., 1991. The Islamic Mean: a peculiar *L*-statistic Studia Sci. Math. Hungar. 26, 297–308.

Devroye, L., Győrfi, L., 1985. Nonparametric Density Estimation: the L_1 View. Wiley, New York.

Ferguson, T., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist. 1 (2), 209–230.

Ghosal, S., Ghosh, J.K., Ramamoorthi, R.V., 1999. Posterior consistency of Dirichlet mixtures in density estimation. Ann. Statist. 27 (1), 143–158.

Hoeffding, W., 1948. A Class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293–325.

Muñoz Perez, J., Fernández Palacín, A., 1987. Estimating the quantile function by Bernstein polynomials. Comput. Statist. Data Anal. 5, 391–397.

Silverman, B.W., 1978. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6 (1), 177–184.

Silverman, B.W., 1986. Denstiy Estimation for Statistics and Data Analysis. Chapman & Hall, London.

Wand, M.P., Devroye, L., 1993. How easy is a given density to estimate? Comput. Statist. Data Anal. 16 (3), 311–323.