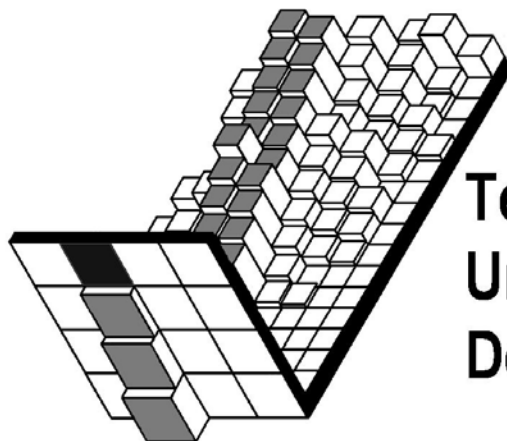


Technical Report

10/2009

Smoothing densities under shape constraints

P. L. Davies, M. Meise



**Technische
Universität
Dortmund**

SFB 475
Komplexitätsreduktion in
multivariaten Datenstrukturen

Sonderforschungsbereich 475
Universität Dortmund
44221 Dortmund
Germany

Smoothing densities under shape constraints

P.L. Davies and M. Meise*
Universität Duisburg-Essen

May 27, 2009

Abstract

In Davies and Kovac (2004) the taut string method was proposed for calculating a density which is consistent with the data and has the minimum number of peaks. The main disadvantage of the taut string density is that it is piecewise constant. In this paper a procedure is presented which gives a smoother density by minimizing the total variation of a derivative of the density subject to the number, positions and heights of the local extreme values obtained from the taut string density.

Key words: Density estimation, total variation
2000 MSC: Primary 62G07

1 Introduction

There are several papers on the shape of densities, see for example Groeneboom (1985), Meyer (2001), Davies and Kovac (2004), Meyer and Woodroffe (2004), Dümbgen and Walther (2008), Dümbgen and Rufibach (2009) and Jongbloed and Van der Meulen (2009), but with the exceptions of Davies and Kovac (2004) and Dümbgen and Walther (2008) they are restricted either to monotone or concave or unimodal densities. Using the results of Dümbgen and Walther (2008) it is possible to derive a lower bound for the number of local extreme values of a density consistent with the data, but the method does not return a candidate density. The taut string procedure of Davies and Kovac (2004) returns a candidate density and in simulations it has proved efficacious for estimating the number, positions and heights of the local extreme values of the generating density (see Davies et al. (2008)). As an example we mention the claw density of Marron and Wand (1992) which is a mixture of six normal distributions and has five peaks. For samples of size $n = 500$ the taut string procedure will correctly identify all the peaks in about 80% of simulated data sets. The left panel of Figure 1 shows one such result. The right panel shows a taut string estimate of the exponential density $f(x) = \exp(-x)$ based on a sample of size $n = 500$. In spite of the longish tale of this density the taut string does not produce superfluous peaks. Figure 1 also shows the main problem with the taut string procedure, namely that the densities it produces are piecewise constant. In this paper a method of smoothing the densities produced by the taut string is described which respects the number, positions and heights of the local extreme values.

Section 2 contains the basic idea of the proposed procedure which is to minimize the total variation of the second derivative subject to linear constraints. This leads to a standard

*This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

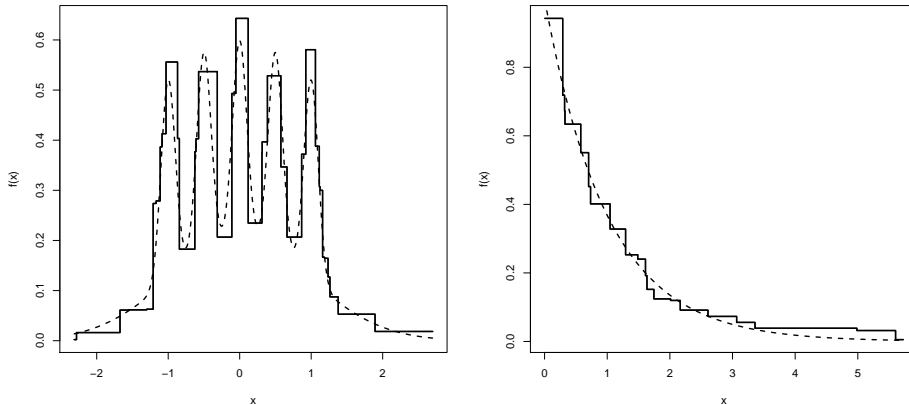


Figure 1: Two examples for the result of the taut string method. Left: Fitted density (solid) and true density (dashed) for a Claw distributed sample. Right: Fitted density (solid) and true density (dashed) for a exponentially distributed sample with parameter $\lambda = 1$. In both cases the sample size is $n = 500$.

linear programming problem. Section 3 gives the results for some simulated and real data sets and Section 4 contains a short discussion of the difficulties of extending the procedure to heavy-tailed distributions and to higher derivatives.

2 Smoothing the density

2.1 Basic idea of the procedure

Given an ordered sample $\mathbf{x}_n = (x_{(1)}, \dots, x_{(n)})^\top$ and a density function f we define

$$\mathbf{f} = (f_1, \dots, f_n)^\top = (f(x_{(1)}), \dots, f(x_{(n)}))^\top$$

where the dependence on the data \mathbf{x}_n has been suppressed. The method is based on the taut string density $f_{n,TS}$ which is calculated as described in Davies and Kovac (2004). The goal is to compute a smooth density f_n which preserves the number, positions and heights of the local extreme values of $f_{n,TS}$. The preservation of these aspects of the taut string density can be expressed by a set of linear inequalities involving \mathbf{f}_n . A smooth density satisfying these constraints can be obtained by solving a minimization problem of the form

$$\min_{f_n} R(f_n) \quad \text{s.t.} \quad \mathbf{A} \mathbf{f}_n \leq \mathbf{b} \quad (1)$$

where $R(f)$ is a measure of roughness of the density f . The matrix \mathbf{A} and the vector \mathbf{b} describing the linear restriction will be specified below.

There are many possible choices for the measure of roughness in (1) some of which lead to non-linear programming problems which are difficult to solve. One example is the Fisher information

$$\int \frac{f^{(1)}(x)^2}{f(x)} dx.$$

If we restrict attention to the total variation of a derivative of f where the partitions are restricted to the data set, then (1) leads to a linear programming problem. For the k th derivative $f^{(k)}$ the total variation is defined as

$$TV(f^{(k)}) = \sum_{i=k+2}^n |f^{(k)}(x_{(i)}) - f^{(k)}(x_{(i-1)})|. \quad (2)$$

As pointed out in Mammen and van de Geer (1997) a function f with minimum $TV(f^{(k)})$ can always be interpolated at the data points by a spline of order k . Thus solving (1) with $R(f) = TV(f^{(k)})$ for $k = 1$ and $k = 2$ results in a piecewise linear and a piecewise quadratic density respectively. Knowledge of the general structure of the solution can be exploited during the computation. Using the vector norm $\|\mathbf{y}\| = \sum_i |y_i|$ the minimization problem can be written as a standard linear programme

$$\min_{\mathbf{f}_n} \|\mathbf{A}^k \mathbf{f}_n\| \quad \text{s.t.} \quad \mathbf{A} \mathbf{f}_n \leq \mathbf{b} \quad (3)$$

where \mathbf{A}^k represents (2).

Smoothing the taut string density $f_{n,TS}$ can cause the integrated distribution function F_n of the smooth density f_n to move away from the empirical distribution function E_n of the data:

$$E_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \leq x\}}. \quad (4)$$

To prevent this we impose the condition

$$d_{ko}(F_n, E_n) \leq \frac{1.36}{\sqrt{n}} \quad (5)$$

where d_{ko} denotes the Kolmogorov metric. The term $1.36/\sqrt{n}$ in (5) is the 0.95-quantile of the Kolmogorov metric. This requirement is usually too weak and below we describe how we iteratively decrease the width of the Kolmogorov tube to obtain a more satisfactory solution. The condition (5) is also expressible as a set of linear inequalities which we incorporate in the matrix \mathbf{A} and the vector \mathbf{b} of (3).

It turns out that it is easier to solve the linear programme in terms of

$$\mathbf{c}^k = (f_1, f_2^{(1)}, \dots, f_k^{(k-1)}, f_{k+1}^{(k)}, \dots, f_n^{(k)})^\top$$

rather than directly in terms of the density itself. This leads to a linear programme of the form

$$\min \|\mathbf{B}^k \mathbf{c}^k\| \quad \text{subject to} \quad \mathbf{E} \mathbf{c}^k \leq \mathbf{b}$$

which is formulated precisely in the next section.

2.2 Computational aspects

After the piecewise constant taut string fit $f_{n,TS}$ has been computed, the number and location of the local extreme values of $f_{n,TS}$ are evaluated. We take the positions of the local extreme points to be the mid-points of the intervals of constancy which correspond to local extreme values (see upper left panel of Figure 2). We denote the locations of the local minima and

maxima by $e_j^- = x_{(s_j)}$, $j = 1, \dots, n_{min}$ and $e_j^+ = x_{(r_j)}$, $j = 1, \dots, n_{max}$ respectively. To guarantee that the solution f_{n,TV_k} of the smoothed problem has the same local extreme values we require

$$f_{n,TV_k}(e_j^-) \leq f_{n,TS}(e_j^-) \quad \text{for } j = 1, \dots, n_{min}, \quad (6)$$

$$f_{n,TV_k}(e_j^+) \geq f_{n,TS}(e_j^+) \quad \text{for } j = 1, \dots, n_{max}. \quad (7)$$

We use here inequality rather than equality signs as this makes the L_1 problem more stable. In practice the overshoot is small due to the regularization. Between the extreme values we require the correct monotonicity which leads to the following inequalities

$$\begin{aligned} f_{n,TV_k}(x_{(i)}) &\leq f_{n,TV_k}(x_{(i+1)}) \quad \text{or} \\ f_{n,TV_k}(x_{(i)}) &\geq f_{n,TV_k}(x_{(i+1)}) \quad \text{for } i = 1, \dots, n \end{aligned} \quad (8)$$

depending on the monotonicity of $f_{n,TS}$ between the local extremes. We note that these are only necessary conditions as they do not guarantee the correct monotonicity of the spline between the data points. We consider this problem below.

Apart from the above inequalities we also have $2n$ inequalities which derive from the condition (5). As the final solution has to be a density this gives rise to a further n non-negativity constraints. Finally we may also impose boundary conditions on the values of the density and its derivatives. In all we have of the order of $4n$ linear restrictions. This makes it impossible to solve the L_1 problem by standard methods except for relatively small samples sizes. Our method is to take a sub-sample of the original sample which includes all the points of e_j^+ and e_j^- of (6) and (7). Since it is known that f_{n,TV_k} is a spline of order k a fit calculated only on a sub-sample can be extended to the entire data set without difficulty. It can then be checked if this solves the original problem. It turns out that a reasonable number of observations which can be treated relatively fast is about 200 but even sub-samples of size 500 can also be accommodated at the cost of an increase in computer time. Hence for small data sets all observations $x_{(1)}, \dots, x_{(n)}$ can be used but for larger sample sizes, especially for $n \geq 500$, a subset $\tilde{x}_{(1)}, \dots, \tilde{x}_{(\tilde{m})}$ of size m is chosen which is given by

$$\tilde{x}_{(j)} = x_{(1 + \lfloor \frac{(j-1)(n-1)}{\tilde{m}-1} \rfloor)} \quad j = 1, \dots, \tilde{m}.$$

After the inclusion of the e_j^+ and e_j^- the reduced sample is of size at most $m = \tilde{m} + n_{min} + n_{max}$.

The optimization problem $\min TV(\mathbf{f}^{(k)})$ can be written in the standard form of a linear programme: $\min \|\mathbf{A}^k \mathbf{f}\|$ which is replaced by the equivalent problem $\min \|\mathbf{B}^k \mathbf{c}^k\|$. Using now the reduced sample we define the $(m - k - 1) \times m$ matrix \mathbf{B}^k and the vector \mathbf{c}^k as follows:

$$\mathbf{B}^k = \left(\begin{array}{c|ccc} & -1 & 1 & \mathbf{0} \\ \mathbf{0}_{(m-k-1) \times k} & & \ddots & \ddots \\ & \mathbf{0} & & -1 & 1 \end{array} \right), \quad \mathbf{c}^k = \begin{pmatrix} f_1 \\ f_2^{(1)} \\ \vdots \\ f_k^{(k-1)} \\ f_{k+1}^{(k)} \\ \vdots \\ f_m^{(k)} \end{pmatrix}.$$

The definition of \mathbf{c}^k and hence the entire setup of the optimization problem uses that $f^{(k)}$ of the solution is piecewise constant. Based on this definition we define the $m \times m$ matrices \mathbf{E}^i , $i = 1, 2, 3$ to state all necessary linear side conditions:

$$\mathbf{E}^1 \mathbf{c}^k = \begin{pmatrix} f_1 \\ f_2^{(1)} \\ \vdots \\ f_m^{(1)} \end{pmatrix}, \quad \mathbf{E}^2 \mathbf{c}^k = \mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} \quad \text{and} \quad \mathbf{E}^3 \mathbf{c}^k = \mathbf{F} = \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}.$$

To guarantee that the resulting approximation \mathbf{f}_{n,TV_k} is a density vector with distribution \mathbf{F}_{n,TV_k} and $|F_n(\tilde{x}_i) - E_n(\tilde{x}_i)| \leq \frac{1.36}{\sqrt{n}}$ we need the side conditions

$$\begin{aligned} (\mathbf{E}^2 \mathbf{c}^k)_i &\geq 0 & i = 1, \dots, m \\ (\mathbf{E}^3 \mathbf{c}^k)_i &\leq u_i & i = 1, \dots, m-1 \\ (\mathbf{E}^3 \mathbf{c}^k)_i &\geq l_i & i = 1, \dots, m-1 \\ (\mathbf{E}^3 \mathbf{c}^k)_m &= 1 \end{aligned}$$

where $\mathbf{u} = (u_1, \dots, u_m)^\top$ and $\mathbf{l} = (l_1, \dots, l_m)^\top$ are the upper and the lower bound respectively of the 0.95%-Kolmogorov tube and hence

$$u_i = E_n(x_i) + \frac{1.36}{\sqrt{n}} \quad \text{and} \quad l_i = E_n(x_i) - \frac{1.36}{\sqrt{n}}, \quad i = 1, \dots, m.$$

To keep notation simple we set $u_m = l_m = 1$ and then have $\mathbf{E}^3 \mathbf{c}^k \leq \mathbf{u}$ and $-\mathbf{E}^3 \mathbf{c}^k \leq -\mathbf{l}$.

Following (6) and (7) we require to guarantee the monotonicity

$$\begin{aligned} (\mathbf{E}^2 \mathbf{c}^k)_{\tilde{r}_j} &\geq f_{n,TS}(e_j^+) & j = 1, \dots, n_{max} \\ (\mathbf{E}^2 \mathbf{c}^k)_{\tilde{s}_j} &\leq f_{n,TS}(e_j^-) & j = 1, \dots, n_{min} \end{aligned}$$

where \tilde{r}_j and \tilde{s}_j are the corresponding indices of r_j and s_j in the reduced sample. The values $f_{n,TS}(e_j^+)$, $j = 1, \dots, n_{max}$ and $f_{n,TS}(e_j^-)$, $j = 1, \dots, n_{min}$ are stored in the vectors $\mathbf{f}_{n,TS}^{e^+}$ and $\mathbf{f}_{n,TS}^{e^-}$ respectively. Additionally we have (cf. (8)) conditions of the type

$$f_i \leq f_{i+1} \quad \text{or} \quad f_i \geq f_{i+1}$$

for all $i = 1, \dots, m-1$ depending on the monotonicity of $\mathbf{f}_{n,TS}$ of the according interval. All such inequalities are stored in the matrix \mathbf{M} and result in the following additional conditions

$$(\mathbf{M} \mathbf{c}^k)_i \leq 0 \quad i = 1, \dots, m-1.$$

Finally the constrained optimization problem can be stated as

$$\min \|\mathbf{B}^k \mathbf{c}^k\| \quad \text{s.t.} \quad \mathbf{E} \mathbf{c}^k \leq \mathbf{b} \tag{9}$$

with

$$\mathbf{E} = \begin{pmatrix} -\mathbf{E}^2 \\ \mathbf{E}^3 \\ -\mathbf{E}^3 \\ -(\mathbf{E}^2_{\tilde{r}_1, \dots, \tilde{r}_{n_{max}}}) \\ \mathbf{E}^2_{\tilde{s}_1, \dots, \tilde{s}_{n_{min}}} \\ \mathbf{M} \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} \mathbf{0}_m \\ \mathbf{u} \\ \mathbf{l} \\ -\mathbf{f}_{n,TS}^{e^+} \\ \mathbf{f}_{n,TS}^{e^-} \\ \mathbf{0}_{m-1} \end{pmatrix}.$$

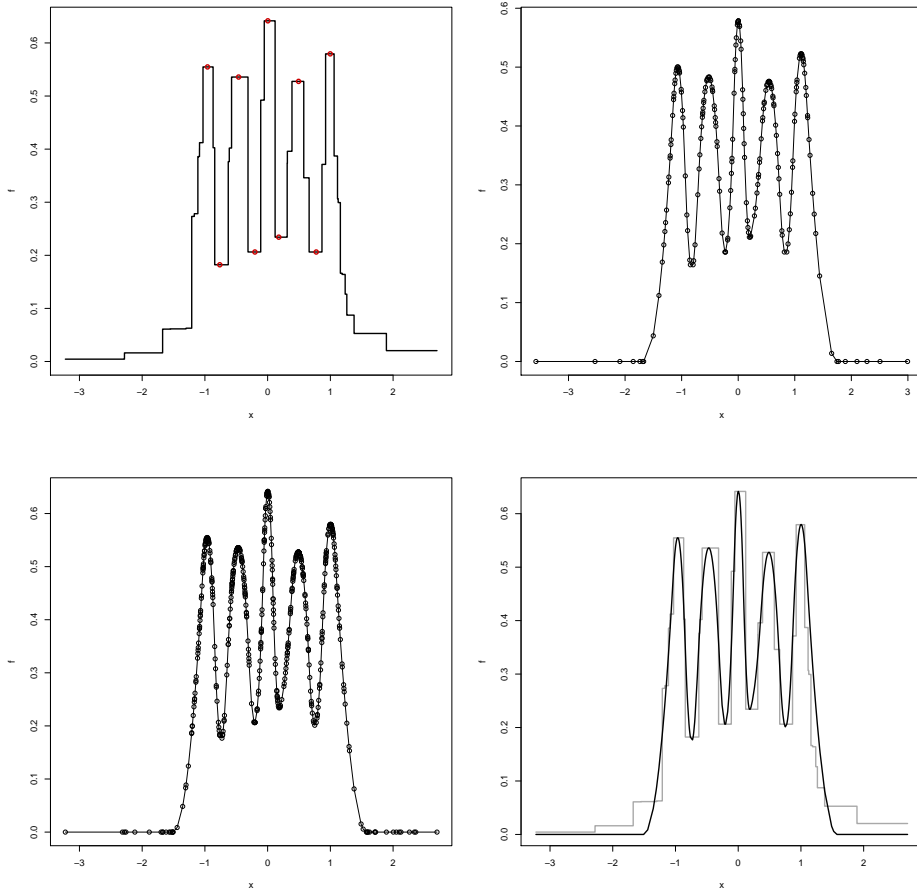


Figure 2: Top left: $\mathbf{f}_{n,TS}$ with marked extreme values, top right: \mathbf{f}_{n,TV_2} on the reduced sample, bottom left: \mathbf{f}_{n,TV_2} extended to the entire sample and bottom right: $\mathbf{f}_{n,TS}$ (gray) and \mathbf{f}_{n,TV_2} (black).

The optimization problem (9) is solved using the simplex algorithm of Barrodale and Roberts (1978, 1980). In our limited experience interior point methods have problems with the large number of monotonicity conditions and are less stable than simplex algorithms.

The solution of (9) gives the density estimate \mathbf{f}_{n,TV_k} on the design points $\tilde{x}_1, \dots, \tilde{x}_m$. This is then extrapolated to the entire sample x_1, \dots, x_n . As mentioned before, we use $k = 2$ and hence \mathbf{f}_{n,TV_2} consists of quadratic pieces. It may happen that the extended density does not satisfy the required inequalities in some intervals between two data points. In particular it may not have the correct monotonicity behaviour. In such cases the mid-point of the interval is included in the reduced sample and the entire procedure is then repeated until \mathbf{f}_{n,TV_2} satisfies all inequalities.

Figure 2 shows the different steps of the procedure for a simulated data set \mathbf{X}_n^F with F being the Claw distribution and $n = 500$. The upper left panel shows $\mathbf{f}_{n,TS}$ with marked extreme values, the upper right panel shows \mathbf{f}_{n,TV_2} on the reduced sample. The lower left panel the version extended to the whole data set, the lower right panel shows $\mathbf{f}_{n,TS}$ and \mathbf{f}_{n,TV_2} together.

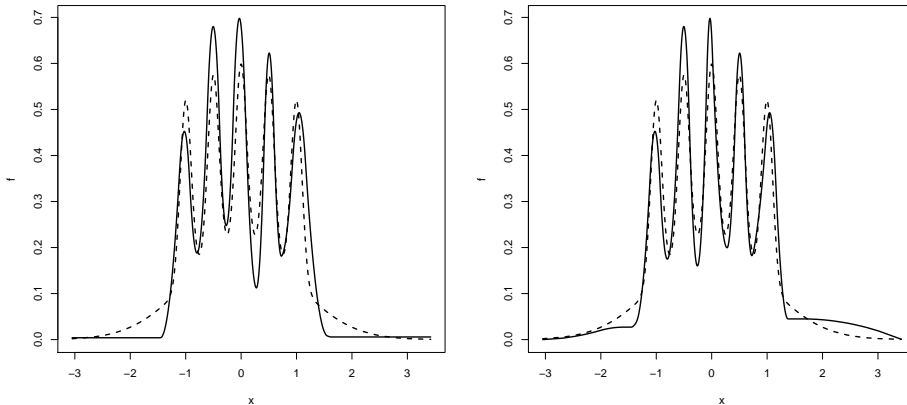


Figure 3: Final result using the Kuiper criterion for the Claw-distributed sample \mathbf{X}_n^F of length $n = 2000$ of Figure 3. Solid: fitted density function \mathbf{f}_{n,TV_2} , dashed: true density function f .

Thus we finally have a smooth density fit \mathbf{f}_{n,TV_2} with the same monotonicity properties as $\mathbf{f}_{n,TS}$ such that the fitted distribution function \mathbf{F}_{n,TV_2} is close to the empirical distribution.

2.3 Decreasing the tube width

The method described so far has problems adapting to the tails of the claw density, at least for the sample sizes used here (cf. left panel of Figure 3). This is due to the relatively wide tube width of the Kolmogorov ball which is used to force the integrated density to be close to the empirical distribution function of the data (5). Better results can be obtained by applying the Kuiper criterion as used in Davies and Kovac (2004): Squeezing the Kolmogorov tube gives a sequence of densities whose modality is nondecreasing. If at some stage the taut string distribution function is $F_{n,TS}$ the transformed data $F_{n,TS}(x_i), i = 1, \dots, n$ is tested for a uniform distribution. If the test is passed, the squeezing stops and is otherwise continued. The test for uniformity is based on the increments of Kuiper distances d_{ku}^κ up to order $\kappa = 19$ (default value) as defined by

$$d_{ku}^\kappa(F, G) = \max \left\{ \sum_{i=1}^{\kappa} |F(b_i) - F(a_i) - (G(b_i) - G(a_i))| : a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_\kappa < b_\kappa \right\}. \quad (10)$$

The present version of the taut string method also allows for testing for a uniform distribution using the multiscale inference of Dümbgen and Walther (2008). This allows small concentrated peaks of lower power to be detected whereas the test based on the Kuiper metrics can only detect peaks containing $O(\sqrt{n})$ observations. For smoothing $f_{n,TS}$ the order κ of the Kuiper criterion is chosen depending on the number of peaks detected by the taut string procedure: if the taut string density $f_{n,TS}$ has k peaks then we use the increments of the Kuiper distances up to order $\kappa = 2k - 1$, see Davies and Kovac (2004). For large data sets this causes the Kolmogorov tube to become very small which may cause numerical problems when solving the linear programme. In such a case we terminate the iterations at the smallest tube width

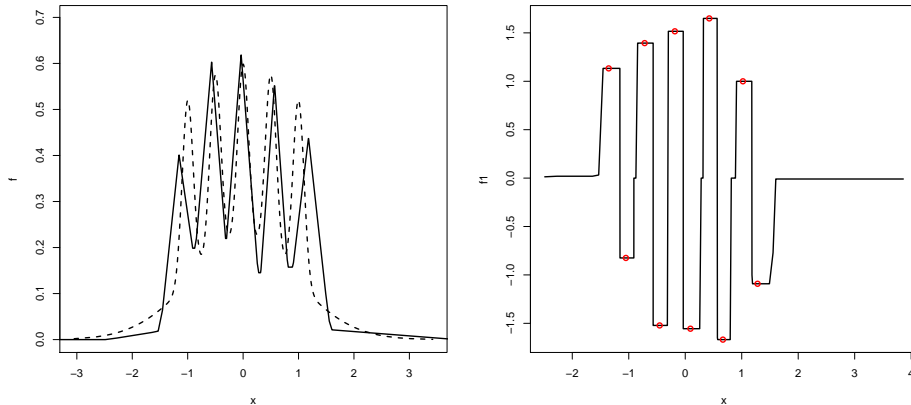


Figure 4: Left: f_{n,TV_1} (solid) together with the true density function f (dashed), right: $f_{n,TV_1}^{(1)}$ for the sample of Fig. 3

which does not lead to numerical problems. The result of doing this is shown in the right panel of Figure 3. The fit is clearly better but is still not satisfactory because the resulting density is now concave in the tails rather than convex. We consider this problem in the following section.

2.4 Convexity conditions

The concavity/convexity behaviour of the solution can be regulated as follows. The monotonicity behaviour of the smooth density was derived from that of the taut string density. A similar reference density is required for the concavity/convexity behaviour and there are good reasons for using the density f_{n,TV_1} which minimizes the total variation of the first derivative $k = 1$ subject to the monotonicity constraints. This density is piecewise linear and is shown in the left panel of Figure 4 for the data of Figure 3: the first derivative is shown in the right panel. At each change-over of the density from convex to concave and from concave to convex there is a common straight-line section which is a local extreme value of the first derivative. The mid-point of this interval is used to delineate the intervals of convexity and concavity as shown in the right panel of Figure 4.

The concavity/convexity conditions lead to further linear inequalities of the form

$$f_i^{(1)} \leq f_{i+1}^{(1)} \quad \text{or} \quad f_i^{(1)} \geq f_{i+1}^{(1)}, \quad i = 2, \dots, m-1.$$

Using a matrix \mathbf{C} these can be expressed as

$$(\mathbf{C}\mathbf{c}^k)_i \leq 0 \quad i = 1, \dots, m-2.$$

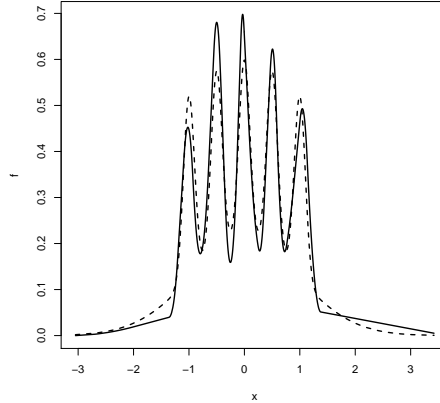


Figure 5: $f_{n,TS}$ (gray) and f_{n,TV_2} (black) using additional convexity conditions from f_{n,TV_1} for the sample of Fig. 3.

and then included in the matrix \mathbf{E} of (9) by putting

$$\mathbf{E} = \begin{pmatrix} -\mathbf{E}^2 \\ \mathbf{E}^3 \\ -\mathbf{E}^3 \\ -(\mathbf{E}^2)_{\tilde{r}_1, \dots, \tilde{r}_{n_{max}}}, \\ (\mathbf{E}^2)_{\tilde{s}_1, \dots, \tilde{s}_{n_{min}}}, \\ \mathbf{M} \\ \mathbf{C} \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} \mathbf{0}_m \\ \mathbf{u} \\ \mathbf{l} \\ -\mathbf{f}_{TS}^+ \\ -\mathbf{f}_{TS}^- \\ \mathbf{0}_{m-1} \\ \mathbf{0}_{m-2} \end{pmatrix}.$$

Figure 5 shows the density estimate for the sample of Figure 3 which is obtained by solving the optimization problem with additional convexity conditions.

3 Results

3.1 Simulated data

Figure 6 shows the results of applying the procedure to a $N(0,1)$ density (upper left), an exponential density (upper right), a mixture of Γ -densities (lower left) and the uniform density on $[0, 1]$ (lower right). In all cases the sample size was $n = 2000$. In all cases the resulting densities have the correct monotonicity and concavity/convexity behaviour and are smooth, at least to the eye at the level of the densities themselves. The first derivatives of the densities are continuous and piecewise linear. The second derivatives are piecewise constant with, of necessity, jumps.

3.2 Real data

The results for four real data sets are shown in Figure 7. The first three data sets are data of Richardson and Green (1997). The top left panel shows the enzyme data with sample size $n = 245$. The top right panel shows the acidity data with sample size $n = 155$. The centre

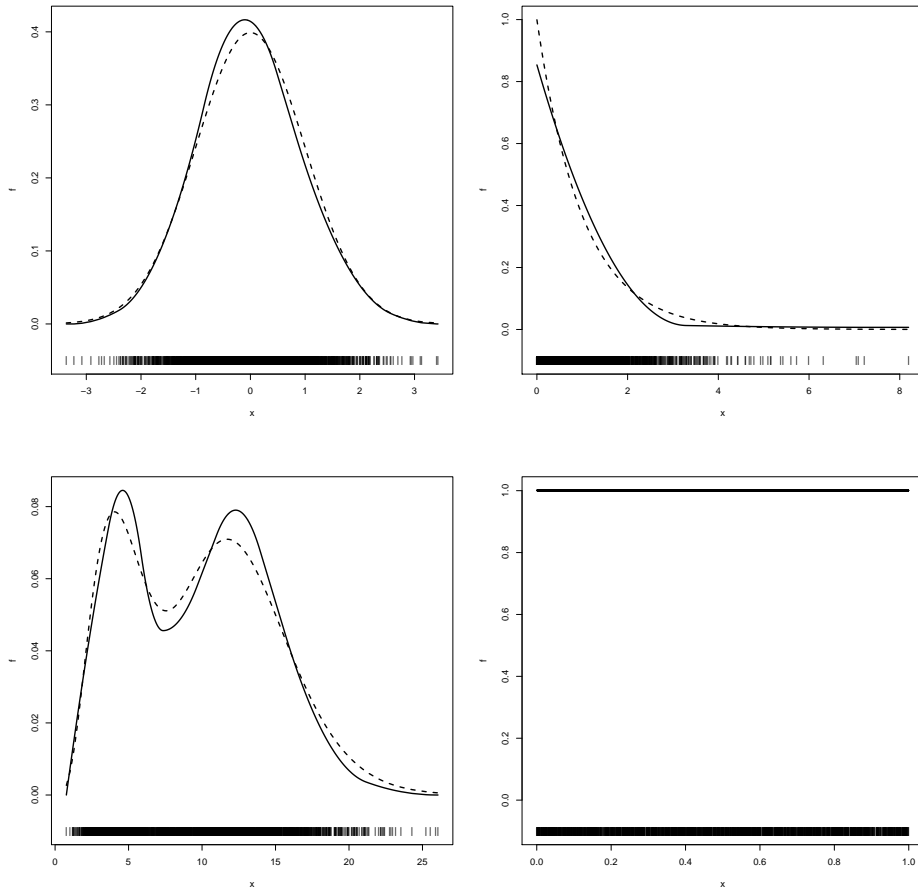


Figure 6: The true density function f (dashed) and \hat{f}_{n,TV_2} (solid) together with the data points for a normal (upper left), an exponential (upper right), a mixed gamma (lower left) and a claw distributed (lower right) sample. The size is $n = 2000$ and the results are computed using monotonicity and convexity conditions.

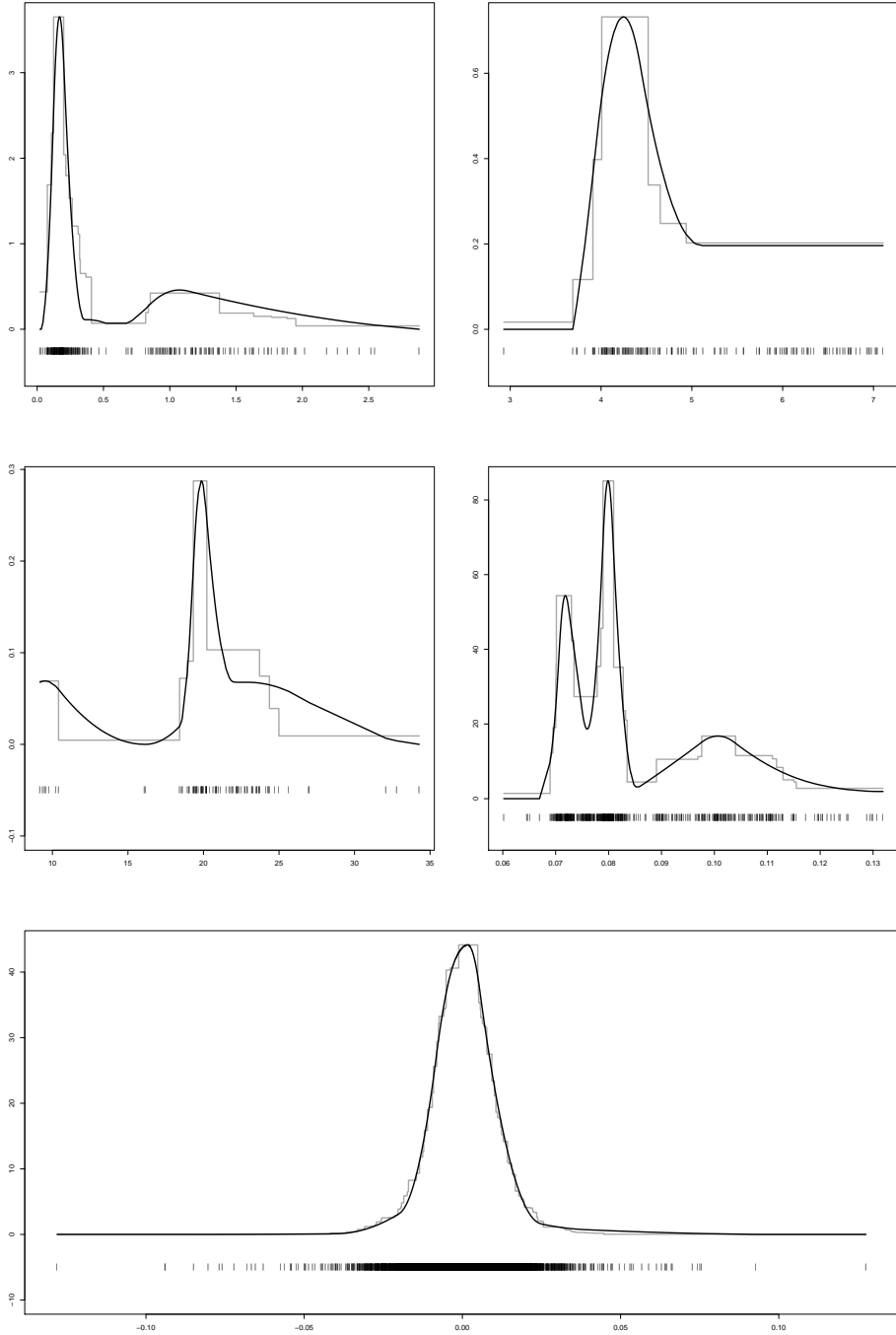


Figure 7: Top left: Enzyme data, top right: acidity data. Centre left: galaxy data, centre right: stamps data. Bottom: DAX data (relative increments, 4th of Jan. 1960 – 28th of Dec. 2001). The original taut string is shown in grey, the smoothed version in black.

left panel shows the galaxy data with sample size $n = 82$. The centre right panel shows the stamp data of Izenman and Sommer (1988) with sample size $n = 485$. Finally the bottom panel shows the density derived from the daily returns of the German DAX-index from the 4th of Jan. 1960 – 28th of Dec. 2001. The sample size is $n = 10515$. The results are in all cases satisfactory but some comments on the stamp data are in order.

The stamp data are given in integer multiples of 0.001 and are characterized by a high degree of discreteness. For example there are 42 observations with the value 0.079 and 37 observations with the value 0.080. This degree of discreteness is incompatible with a density and any self-respecting density estimator should refuse to provide an estimator, as does the taut string. One possibility is to explicitly allow for rounding to the nearest multiple of 0.001. This would mean that the 42 observations with the value 0.079 can be taken as uniformly distributed over the interval $[0.0785, 0.0795]$. One possibility is to do this at random by adding $0.001(U_i - 0.5)$ to the observation x_i with the U_i uniformly distributed over $(0, 1)$ If we now apply the taut string the results are not stable. Without the multiscale inference of Dümbgen and Walther (2008) the taut string returns one or two peaks and occasionally three peaks. If the multiscale inference of Dümbgen and Walther (2008) is applied then taut string returns mostly the three peaks shown in Figure 7 but it can also return 6, 7 and 8 peaks. If the discreteness is broken by placing the points equidistantly on the interval $(x_i - 0.0005, x_i + 0.0005)$ rather than randomly then the result is again the three peaks shown in Figure 7. We refer to Izenman and Sommer (1988) for an interesting discussion of this data set.

4 Discussion

As with all procedures the one presented here has limits. It will not perform well for data with a large number of outlying observations such as are generated by the Cauchy distribution. Smoothing the very long tails whilst respecting the monotonicity constraints deriving from the taut string causes problems. The same applies to densities with very high and narrow peaks. Here the values of the second derivative may differ by orders of magnitude again making the problem numerically highly unstable.

In theory it is possible to include higher derivatives in the definition of smoothness. Density estimation itself is an ill-posed problem and the higher the number of derivatives, the more ill-posed becomes the problem. For relative benign data sets such as those coming from a normal distribution then third and even fourth derivatives are possible. In general however the linear programme becomes numerically instable. At the moment we use the simplex method of Barrodale and Roberts (1978, 1980) which is a general method. It may be possible to improve numerical stability by developing a tailor-made for the sort of minimization problem we consider.

Although we have not done so, it is possible to include boundary conditions such as $f_1 = f_n = 0$ and $f_1^{(1)} = f_n^{(1)} = 0$. These are to some extent subjective unless there are well-grounded reasons for imposing them. In general they cause no problem as long as they are consistent with the data. If not, they will introduce numerical instabilities.

5 Acknowledgment

The authors gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475 of the Fakultät Statistik of the Technische Universität Dortmund.

References

- 1 Barrodale, I. and Roberts, F. D. K. (1978). An efficient algorithm for discrete l_1 linear approximation with linear constraints. *SIAM J. Numer. Anal.*, 15(3):603–611.
- 2 Barrodale, I. and Roberts, F. D. K. (1980). Algorithm 552: Solution of the constrained l_1 linear approximation problem. *ACM Transactions on Mathematical Software (TOMS)*, 6(2):231 – 235.
- 3 Davies, P., Gather, U., Nordman, D., and Weinert, H. (2008). A comparison of automatic histogram constructions. *ESAIM: Probability and Statistics*, to appear.
- 4 Davies, P. L. and Kovac, A. (2004). Densities, spectral densities and modality. *Annals of Statistics*, 32(3):1093–1136.
- 5 Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, 15:40–68.
- 6 Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Annals of Statistics*, 36:1758–1785.
- 7 Groeneboom, P. (1985). Estimating a monotone density. In LeCam, L. and Olshen, R., editors, *Proceedings, Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume II, Monterey, CA. Wadsworth.
- 8 Izenman, A. J. and Sommer, C. J. (1988). Philatelic mixtures and multimodal densities. *Journal of the American Statistical Association*, 83:941–953.
- 9 Jongbloed, G. and Van der Meulen, F. H. (2009). Estimating a concave distribution function from data corrupted with additive noise. *The Annals of Statistics*, 37:782–815.
- 10 Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Annals of Statistics*, 25:387–413.
- 11 Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *Annals of Statistics*, 20:712–736.
- 12 Meyer, M. C. (2001). An alternative unimodal density estimator with a consistent estimate of the mode. *Statistica Sinica*, 11(4):1159–1174.
- 13 Meyer, M. C. and Woodroffe, M. (2004). Estimation of a unimodal density using shape restrictions. *Canadian Journal of Statistics*, 32(1):85–100.
- 14 Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, 59:731–792.