Nonparametric Density Estimation: A Piecewise Maximum Likelihood Approach

by

Jianqing Fan
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

Peter J. Bickel
Department of Statistics
University of California
Berkeley, CA 94720

NONPARAMETRIC DENSITY ESTIMATION: A PIECEWISE MAXIMUM LIKELIHOOD APPROACH

Jianqing Fan

Department of Statistics

University of North Carolina

Chapel Hill, N.C. 27514

Peter J. Bickel[1]

Department of Statistics

University of California

Berkeley, CA 94720

Dec. 12, 1990

Abstract

The effort of recovering the features (such as the number of modes and overall shape) of unknown densities leads to nonparametric curve estimation under order restrictions. We introduce a flexible class of nonparametric densities called α-regular shaped densities, and propose the Piecewise Maximum Likelihood Estimate (PMLE) which has an attractive "automatic" bandwidth selection feature. Spline techniques are used to further smooth the PMLE. Simulation studies are presented to illustrate the usefulness of the proposed techniques. A case study of estimating the densities of the size of lipoprotein is also included.

1

1 Introduction

Nonparametric density estimation provides a useful explanatory technique of examining the overall structure of a set of data. Popular approaches include kernel and spline methods. However, the performance of these methods depends strongly on the choice of smoothing parameters. The sensitivity to this choice makes the practical use of the methods time consuming and requires expertise. Data-driven bandwidth selection methods have been studied recently [Rice (1984), Wahba (1985), Härdle, Hall and Marron (1988), and among others]. One tries to minimize the Integrated Square Error (ISE) or the Mean ISE (MISE) or other related objects, and uses one of them as a measure of global effectiveness of a curve estimate. In practical density estimation, shapes such as the number of modes of the curve are very important [Silverman (1986)]. If one is interested in number of modes, then ISE and MISE might not be good criteria. For example, the ISE of two curves can be very small, while the number of modes of the two curves is quite different. Thus, it is clear that a more objective method with variable window sizes determined automatically by the data would greatly enhance the usefulness of practical curve estimation. To achieve this goal, we consider estimating a density under some flexible shape restrictions.

An early paper on shape restricted estimation is Grenander (1956), who estimated a decreasing density by using a maximum likelihood approach. The asymptotic distribution of the MLE was found by Prakasa Rao (1969), and Groeneboom (1985). Recent developments in estimating a monotone density can be found in Birgé (1987a,b). Wegman (1969, 1970a, b) pioneeringly studies the problem of estimating a unimodal density whose mode location is unknown, and finds the asymptotic distribution of his estimator. See also Wegman (1975) for other modifications and development of maximum likelihood methods. Ramsay (1988) gives an illuminating development of the isotonic approach, and its applications including data transformations. Various applications can also be found in Ramsay and Abrahamowicz (1989), Kelly and Rice (1990), Bickel and Fan (1990), among others. Related ideas to our proposal in this paper are independently proposed in Mächler (1989), Mammen (1990). Of

course, there is a long history of isotonic regression and related problems. See Barlow *et al.* (1972) and Robertson *et al.* (1988).

2   Problems and Models

The problem we are concerned with is not only to estimate a density, but also to recover the features (such as the number and locations of modes) of the true density f. In applications, we are interested in modes which are reasonably far apart. Let $x_{j\alpha}^*$ be the $(j\alpha)$-th population quantile. For simplicity of our discussion, we will always assume that $1/\alpha$ is an integer.

Definition. *Call a density α-regular shaped if there is at most one local extremum (minimum or maximum) on $[x_{j\alpha}^*, x_{(j+1)\alpha}^*]$ (for $0 \leq j \leq 1/\alpha - 1$).*

Let f be α-regular shaped. On the interval $[x_{j\alpha}^*, x_{(j+1)\alpha}^*]$ having probability α, the density f has only four possible shapes: increasing, decreasing, U-shaped, and unimodal. From now on, we will use the term "unimodal" to also include monotone (increasing, decreasing) curves, because a monotone curve has maximum over the interval $[x_{j\alpha}^*, x_{(j+1)\alpha}^*]$ at the boundary.

With small α, α-regular shaped densities model almost all interesting densities of practical interest. For example, 10%-regular shaped densities include all increasing, decreasing, unimodal densities, bimodal, trimodal, quadramodal densities whose area between a valley and its nearest peak is at least 10 %, and many *other* bimodal, trimodal, quadramodal densities. Thus, α-regular shaped densities describe many useful density curves.

The objective of this paper is to estimate an α-regular shaped curve. We emphasize that α is not a parameter one can determine from statistical estimation, but it is determined by users from prior knowledge and practical concerns. For a finite set of data at hand, one can not really tell how many modes the data may have: it could come from a density having 5 modes, and could also come from a density having 6 modes or more. In other words, from a statistical viewpoint, one can not give an upper bound on the number of modes of a density, but can give a lower bound on the number of modes (see Donoho (1989)). Thus,

α-regular shaped densities model the unknown density by putting an upper bound $(1/\alpha)$ on the number of modes. This explains why α is not estimable.

In conclusion, the restriction on a density has practical meaning. It includes a variety of interesting curves such monotone, unimodal, and U-shaped curves, as specific examples. It forces the density function to be not too wiggly, and models a "most explainable curve" among the infinite number of curves which are indistinguishable by a finite set of data.

3 Methodologies

Let $X'_1, \cdots, X'_n$ be a random sample from an α-regular shaped density and let $X_1, \cdots, X_n$ be its order statistics. Our approach in estimating f is to first divide the data into $m_\alpha = 1/\alpha$ pieces, and then apply the nonparametric maximum likelihood technique to each corresponding piece to find an estimate of the density for each piece. The Piecewise Maximum Likelihood Estimate (PMLE) puts an estimate on each piece to form an estimate of the whole density curve.

3.1 Piecewise Maximum Likelihood Estimate

Divide the data into m_α pieces: the jth piece contains data $X_{(j-1)n_\alpha+1}, \cdots, X_{jn_\alpha}$, where $n_\alpha = n\alpha$. Note that X_{jn_α} is the $j\alpha$-th sample quantile, which is a good estimate of the population quantile $x^*_{j\alpha}$, an end point of intervals on which the unknown density is either unimodal or U-shaped. Now consider estimators which attempt to maximize the approximate likelihood in the jth piece:

$$\prod_{k=1}^{n_\alpha} f(X_{(j-1)n_\alpha+k}), \tag{3.1}$$

subject to the constraint that $f(\cdot)$ on $[X_{(j-1)n_\alpha+1}, X_{jn_\alpha}]$ is either unimodal or U-shaped with area α.

However, there is no solution to the problem (3.1) because one could always let the density $f(X_{(j-1)n_\alpha+k_0}) = \infty$ at mode position $X_{(j-1)n_\alpha+k_0}$ to maximize the likelihood. Also,

when f is U-shaped, one could let $f(X_{(j-1)n_\alpha+1}) = \infty$ or $f(X_{jn_\alpha}) = \infty$ to maximize the likelihood. Thus, some refinements appear necessary. Let's take $X_{(j-1)n_\alpha+k}$ as a candidate of the location of a peak or valley. Then, consider the following two separate problems:

- *Problem I.* For $1 \leq k \leq n_\alpha$, maximize

$$\prod_{i=1, i \neq k}^{n_\alpha} f(X_{(j-1)n_\alpha+i}), \tag{3.2}$$

subject to the constraint that f is unimodal with mode location $X_{(j-1)n_\alpha+k}$ and the total mass of f on the interval $[X_{(j-1)n_\alpha+1}, X_{jn_\alpha}]$ is α. Denote the solution by $\hat{f}^1_{jk}(x)$, whose value at mode location is undermined. Now, define the log-likelihood of $\hat{f}^1_{jk}(x)$ to be

$$L^1_{jk} = \sum_{i \neq k} \log \left(f^1_{jk}(X_{(j-1)n_\alpha+i}) \right).$$

- *Problem II.* For $2 \leq k \leq n_\alpha - 1$, maximize

$$\prod_{i=2}^{n_\alpha-1} f(X_{(j-1)n_\alpha+i}), \tag{3.3}$$

subject to the constraint that f is U-shaped with valley location $X_{(j-1)n_\alpha+k}$ and the total mass of f on the interval $[X_{(j-1)n_\alpha+1}, X_{jn_\alpha}]$ is α. Denote the solution by $\hat{f}^2_{jk}(x)$, whose values at both end points can not be determined by (3.3). Now, define the log-likelihood of the function by

$$L^2_{jk} = \sum_{i=2}^{n_\alpha-1} \log \left(f^2_{jk}(X_{(j-1)n_\alpha+i}) \right).$$

Let $\hat{f}_j(x)$ be the one of $2(n_\alpha - 1)$ densities

$$\hat{f}^1_{jk}(x) \ (k = 1, \cdots, n_\alpha), \quad \hat{f}^2_{jk}(x) \ (k = 2, \cdots, n_\alpha - 1)$$

having the largest value of the likelihood. Then, it will be shown that the support of $\hat{f}_j(x)$ is $[X_{(j-1)n_\alpha+1}, X_{jn_\alpha}]$. The piecewise maximum likelihood estimate $\hat{f}(x)$ is the estimate that combines $\hat{f}_j$ on each disjoint subinterval $[X_{(j-1)n_\alpha+1}, X_{jn_\alpha}]$:

$$\hat{f}(x) = \sum_{1}^{1/\alpha} \hat{f}_j(x). \tag{3.4}$$

5

Note that the summation in (3.4) just combines $\hat{f}_j(x)$ on each disjoint subinterval $[X_{(j-1)n_\alpha+1},$ $X_{jn_\alpha}]$, as $\hat{f}_j(x)$ has support $[X_{(j-1)n_\alpha+1}, X_{jn_\alpha}]$.

For a more complete discussion on estimating a unimodal density (corresponding to $\alpha = 1$), see a serious of work by Wegman (1969, 1970a,b, 1975). The current method is a continuation and development of his pioneering work.

3.2 Solution to the Optimization Problems

The solutions to the first and the second problems can be found by using isotonic regression techniques [Robertson *et al.* (1988)].

Proposition 1. *The solution to Problem I is given by*

$$
\hat{f}_{jk}^1(x) = \begin{cases} 0, & \text{if } x < y_1 \text{ or } x > y_{n_\alpha} \\ f_i & \text{if } y_i \le x < y_{i+1}, i = 1, \cdots, k-1 \\ f_i & \text{if } y_i < x \le y_{i+1}, i = k, \cdots, n_\alpha - 1 \end{cases} \quad ,
$$

where $y_i = X_{(j-1)n_\alpha+i}(i = 1, \cdots, n_\alpha)$ and f_i is defined by

$$
f_i = \begin{cases} \min_{k \ge t > i} \max_{s \le i} \frac{\alpha(t-s)}{(n_\alpha-1)(y_t-y_s)} & \text{if } i < k \\ \min_{k \le s \le i} \max_{t > i} \frac{\alpha(t-s)}{(n_\alpha-1)(y_t-y_s)} & \text{if } i \ge k \end{cases} \quad . \tag{3.5}
$$

The result of Proposition 1 is contained in Barlow *et al.* (1972), and can be computed by using the "pool-adjacent-violator" algorithm [see Robertson *et al.* (1988), page 8-10].

Proposition 2. *The solution $\hat{f}_{jk}^2$ to Problem II is given by*

$$
\hat{f}_{jk}^2(x) = \begin{cases} 0, & \text{if } x < y_1 \text{ or } x > y_{n_\alpha} \\ f_{i+1} & \text{if } y_i < x \le y_{i+1}, i = 1, \cdots, k-2 \\ f_k & \text{if } y_{k-1} < x < y_{k+1} \\ f_i & \text{if } y_i \le x < y_{i+1}, i = k+1, \cdots, n_\alpha - 1 \end{cases} \quad , \tag{3.6}
$$

6

with

$$f_i = \begin{cases} \min_{s < i} \max_{k \ge t \ge i} \frac{\alpha(t-s)}{(n_\alpha - 2)(y'_t - y'_s)} & \text{for } 1 < i < k \\[2mm] \min_{t > i} \max_{k \le s \le i} \frac{\alpha(t-s)}{(n_\alpha - 2)(y''_t - y''_s)} & \text{for } k < i < n_\alpha \\[2mm] \min \left\{ \min_{s < k} \frac{\alpha(k-s)}{(n_\alpha - 2)(y_{k+1} - y_s)}, \min_{t > k} \frac{\alpha(t-k)}{(n_\alpha - 2)(y_t - y_{k-1})} \right\} & \text{for } i = k \end{cases} \quad (3.7)$$

where $y_i = X_{(j-1)n_\alpha + i}, i = 1, \cdots, n_\alpha,$ *and*

$$y'_i = y_i, \text{ when } i < k, y'_k = y_{k+1}; \quad y''_k = y_{k-1}, y''_i = y_i, \text{ when } i > k.$$

The proof of Proposition 2 is given in the appendix. The estimate $\hat{f}^2_{jk}$ can also be computed by using the "pool-adjacent-violator" algorithm. The C program for computing the PMLE is available from the authors.

With solutions given in Propositions 1 & 2, one could easily find the PMLE (3.4). Even though we have to solve quite a few optimization problems, the computation of PMLE is very fast: it takes a few seconds to compute on a SUN 4.

3.3 Spline Smoothing

The PMLE $\hat{f}$ is a step function (histogram) whose bin width is automatically determined by the data. See Figures 1.1—3.1. The discontinuity of the curve PMLE seems unsatisfactory. Thus, further smoothing of the estimate $\hat{f}$ appears desirable. Let $(x_1, z_1), \cdots, (x_N, z_N)$ denote the midpoints of the histogram estimate $\hat{f}$ (i.e., x_i is the midpoint of the ith histogram bin and z_i is the height). We want to find a smooth density that eventually (in a least squares sense) passes through the midpoints of the histogram.

There are two popular cubic spline bases

- power bases: $1, x, x^2, x^3, (x - t_j)^3_+$, where t_j is a knot;

- B-spline bases (see de Boor (1978)).

The problem is how to locate knots. There are two popular techniques for locating knots: fixed knots and knot deletion.

fixed knots

Let's take $x_s, x_{2s}, \cdots, x_{ms}$ ($m = [\frac{N-2}{s}]$) as interior knots, where s denotes knot span. Let the corresponding cubic spline bases (e.g. power bases) be $B_1(x), \cdots, B_{m+4}(x)$. We propose to use logsplines [see Kooperberg and Stone (1990)] to smooth the PMLE. Let $\log(f_s(x)) = \sum_1^{m+4} \theta_k B_k(x)$. Find θ_k to minimize

$$\sum_1^N [\log(z_i) - \log(f_s(x_i))]^2 w_i, \tag{3.8}$$

and then renormalize the estimate $\hat{f}_s(x)$ to have area one, where w_i is the area of the histogram estimate on the ith bin.

In principle, the smoothed density need not be α regular. However, since the shape of the unknown density is captured by the PMLE, it is relatively easy to select an s such that the spline smoothed curve has a shape similar to the PMLE. See section 4.3 for more discussion.

An advantage of using logsplines is that the positivity constraints ($f_s(x) \geq 0$) become automatic. Thus, one only needs to use ordinary least squares method instead of using constrained least squares.

knot deletion

Instead of using a fixed number of knots, we can start out with a larger number of knots and then remove those knots that appear to be unessential for the given data.

Let's take $x_3, x_6, \cdots, x_{3m}$ ($m = [\frac{N-1}{3}]$), as initial knots that may be deleted. Let corre-

sponding power bases be

$$\begin{cases} B_j(x) = (x - x_{3j})^3_+, (j = 1, \cdots, m), \\ B_{m+1}(x) = 1, B_{m+2}(x) = x, B_{m+3}(x) = x^2, B_{m+4}(x) = x^3 \end{cases}$$

Let $\log(f_s(x)) = \sum_1^{m+4} \theta_k B_k(x)$. Use the usual least square to find θ_k that minimizes

$$\sum_1^N [\log(z_i) - \sum_1^{m+4} \theta_k B_k(x_i)]^2 w_i, \tag{3.9}$$

where w_j was defined in (3.8).

Denote the least square estimate of (3.9) by $\hat{\theta}_j$ with standard error $SE(\hat{\theta}_j)$. Then, delete the j_0^{th} knot ($1 \le j_0 \le m$) having the smallest absolute t-value: $|\hat{\theta}_j|/SE(\hat{\theta}_j)$, ($1 \le j \le m$).

Repeat the above deleting process (at each step delete one knot). We arrive at a sequence of models indexed by j ($0 \le j \le m$): the jth model contains $m + 4 - j$ free parameters with residuals of sum of square RSS_j. Choose $\hat{j}$ to minimizes the modified Mallow's C_p criterion:

$$C_j = RSS_j + 3(m + 4 - j)\hat{\sigma}^2, \tag{3.10}$$

where $\hat{\sigma}^2$ is the standard deviation at the 0th model (full model).

At the $\hat{j}^{th}$ model, one obtains knots $\hat{x}_1, \cdots, \hat{x}_{\hat{j}}$, with bases $B_j^*(x) = (x - \hat{x}_j)^3_+, j = 1, \cdots, \hat{j}$, and $B_{\hat{j}+1}^*(x) = 1, B_{\hat{j}+2}^*(x) = x, B_{\hat{j}+3}^*(x) = x^2$, and $B_{\hat{j}+4}^*(x) = x^3$, and estimates $\hat{\theta}_j, j = 1, \cdots, \hat{j} + 4$. Now, form the function

$$\hat{f}^*(x) = \exp\left(\sum_1^{\hat{j}+4} \hat{\theta}_j B_j^*(x) \right).$$

Normalize $\hat{f}^*(x)$ to be a density and denote the resulting function by $\hat{f}^{**}(x)$. Then, $\hat{f}^{**}(x)$ is the smoothed version of PMLE by using the knot deletion procedure.

3.4 Comparison with other spline techniques

The usual spline technique involves selecting knots among n possible data points. In our approach, we locates knots at the N possible midpoints of histogram bins. Thus, an advantage of using the PMLE as preliminary estimate is that we reduce the possible number of

knots from n to $N \approx n^{2/3}$, and also get a good preliminary estimate of the density. In our simulations, $n = 200$, and $N \approx 30$. Thus, it is much easier to implement stepwise deletion procedures and to use fixed knots spline techniques.

3.5 Some remarks on the procedures

Remark 1. When $n_\alpha = n\alpha$ is not integer, it is easy to modify our procedure by dividing the data as follows. Let $n_\alpha^* = [n\alpha]$, and $n^* = n - m_\alpha n_\alpha^*$ be the residual, where $m_\alpha = 1/\alpha$. Now, the modified procedure is to let the first n^* pieces contain $(n_\alpha^* + 1)$ data points, and the rest contain n_α^* data. In other words, the jth piece contains data

$$
\begin{cases}
X_{(j-1)(n_\alpha^*+1)+1}, \cdots, X_{j(n_\alpha^*+1)} & (n_\alpha^* + 1 \text{ data points}), \quad \text{when } 1 \le j \le n^* \\
X_{(j-1)n_\alpha^*+n^*+1}, \cdots, X_{jn_\alpha^*+n^*} & (n_\alpha \text{ data points}), \qquad \text{when } n^* < j \le m_\alpha
\end{cases}
$$

Remark 2. A very useful extension of our procedure is to find an α-regular estimate with a certain number, 2 say, of derivatives via the maximum penalized likelihood criterion: find $\hat{f}$ to maximize

$$
\prod_1^n f(X_j) + \beta \int_{-\infty}^{\infty} [f''(x)]^2 dx,
$$

subject to f is α-regular, where β is the penalty. See Good and Gaskins (1971) for further discussion on maximum penalized likelihood estimator. The advantage of this procedure is that the resulting estimate itself is smooth with knots determined by data. In this case, the spline smoothing is not necessary, as the estimate itself is smooth. However, we haven't yet found a good algorithm for the estimate.

Remark 3. These spline smoothing techniques are proposed as supplementary data analysis tools. They can be used to produce a smooth curve estimate. They are optional, if one is satisfied with PMLE estimate.

4 Simulations

4.1 Unimodal Example

A random sample of 200 data is generated from $N(0,1)$. Suppose that we don't know that the data is from a unimodal density, but are interested in recovering 0.25-regular densities. Roughly speaking, we believe the underlying density has no more than 4 modes. Then, we apply the PMLE technique with $\alpha = 1/4$. The resulting estimate is plotted in Figure 1.1. As one can see, there are peaking phenomena at the estimated modes. As we discussed in section 2, for a set of 200 random data, we can not really distinguish whether the data comes from a the density like the one plotted in Figure 1.1 or from a standard normal density, unless we have some other knowledge. If one ignores the thin bins and the valley around zero which are the points where the data are partitioned and are artifacts of PMLE, however, one can see that the estimate is unimodal.

We then use the spline smoothing techniques discussed in section 3.3. Figure 1.2 plots the spline smoothed versions with both fixed knots ($s = 5$) and deleted knots. The resulting estimate is clearly a unimodal density.

From this example, we can see that the spline techniques are proposed not only to smooth the discontinuity of the PMLE, but also to repair the artifacts at locations where the data are partitioned.

The following table shows the resulting knot locations produced by our rules of locating knots.

Table 1. Knots resulting from the two selection rules for Figure 1.2—3.2.

	Fixed Knots	Knot deletion
Unimodal	-1.60, -0.59, 0.14, 0.76	None
Bimodal	-3.68, -2.27, -1.08, -0.14, 1.52, 2.70	-0.78, 0.22, 1.64, 2.70
Trimodal	-1.70, -1.27, -0.08, 1.00, 1.91	-2.62, -1.63, -1.39, 0.05, 0.72, 1.51

In the above table, unimodal means the resulting knots for the data set generated from the unimodal density. "None" indicates that all knots are deleted by the knot deletion rule.
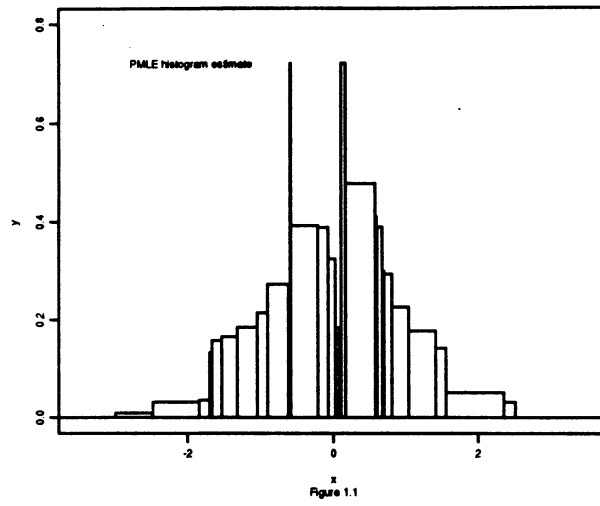
200 data from a unimodal density

PMLE histogram estimate

Figure 1.1

200 data from a unimodal density

true density

fixed knots

- - - deleting knots
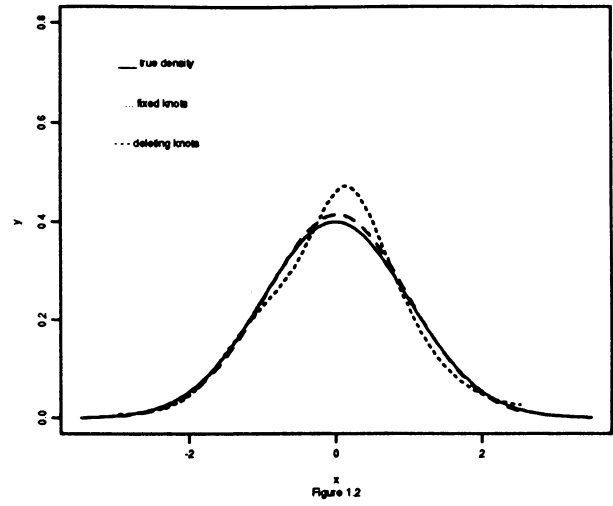
Figure 1.2

200 data from a bimodal density

PMLE histogram estimate
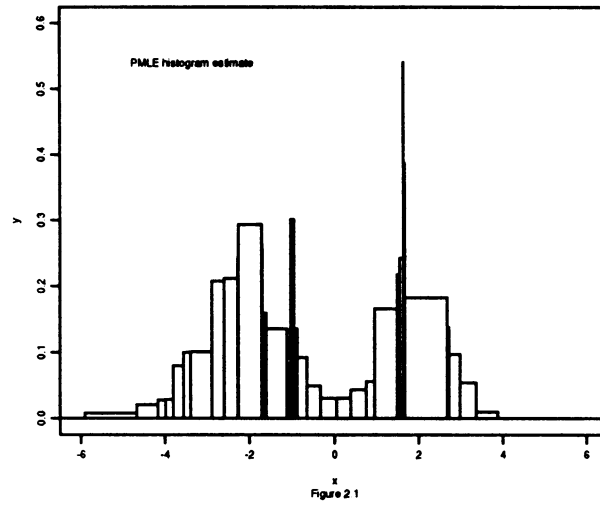
Figure 2.1

200 data from a bimodal density

true density

fixed knots

- - - deleting knots
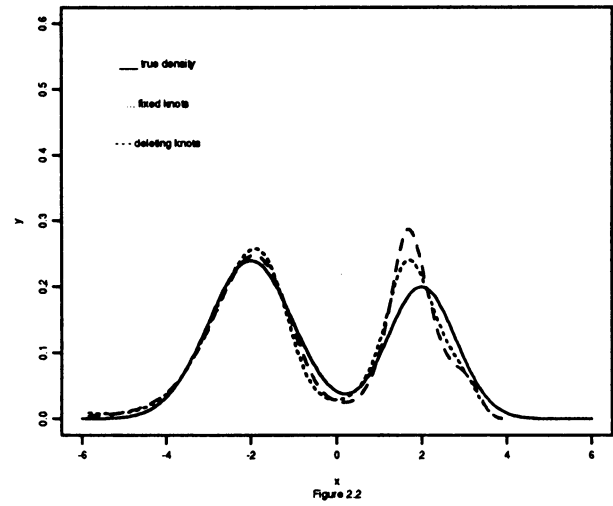
Figure 2.2

200 data from a trimodal density

PMLE histogram estimate
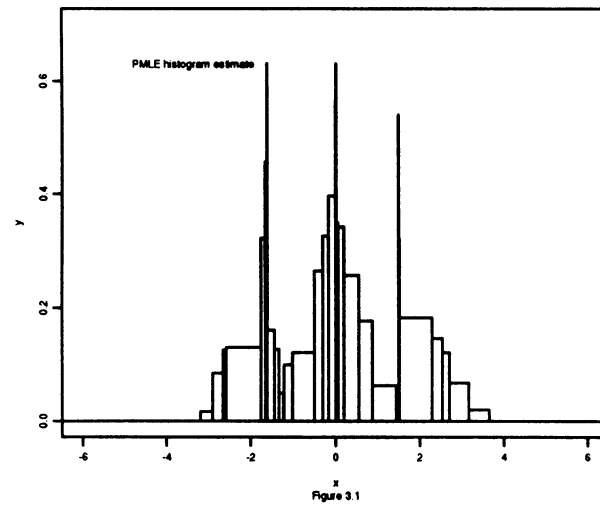
Figure 3.1

200 data from a trimodal density

true density

fixed knots

- - - deleting knots
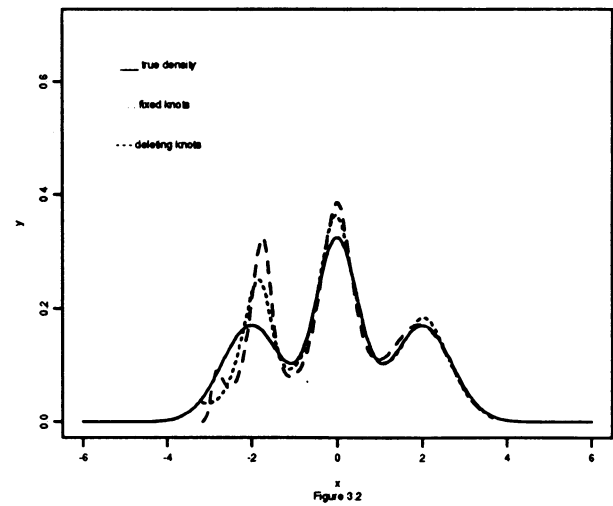
Figure 3.2

Bimodal and Trimodal cases are to be discussed.

4.2 Bimodal Example

We simulate a random sample of size 200 from $0.6N(-2,1) + 0.4N(2,0.8^2)$ and again implement the PMLE with $\alpha = 1/4$. The resulting PMLE is plotted in Figure 2.1, which looks pretty satisfactory (ignoring artifacts like thin bins.) The smoothed versions of the PMLE are plotted in Figure 2.2. Again, we take $s = 5$ in fixed knots spline smoothing. The resulting knots used in the computation are shown in the third row of Table 1.

4.3 Trimodal Example

In this simulation, a random sample of 200 is taken from a normal mixture:

$$0.3N(-2,0.7^2) + 0.4N(0,0.5^2) + 0.3N(2,0.7^2).$$

The underlying density is trimodal. We again apply the PMLE with $\alpha = 1/4$, and plot the estimate in Figure 3.1. We again use $s = 5$ for the spline smoothed curve with fixed knots, and plot the estimate in Figure 3.2.

In this example, the knots selected by the fixed knot span rule are not quite satisfactory. An artificial mode was produced in the resulting smoothed version. We knew it was artificial because the result looked sufficiently different from the PMLE. With the PMLE, it is relatively easy to locate knots. For example, after examining Figure 3.1, one can choose knots like -2, -1, 0, 1, 2. In Table 1, the fixed knots are determined by using $s = 5$ and then deleting unnecessary knots.

4.4 How Sensitive is the method to the Choice of α?

To examine the effect of α (hence, the model assumptions), we simulate a sample of 800 from the standard normal distribution. We apply the PMLE with $\alpha = \frac{1}{3}, \frac{1}{5}$, and $\frac{1}{7}$. The smaller α is, the higher the number of implicit parameters of the model and the rougher the result. Figure 4.1—Figure 4.3 presents the results.
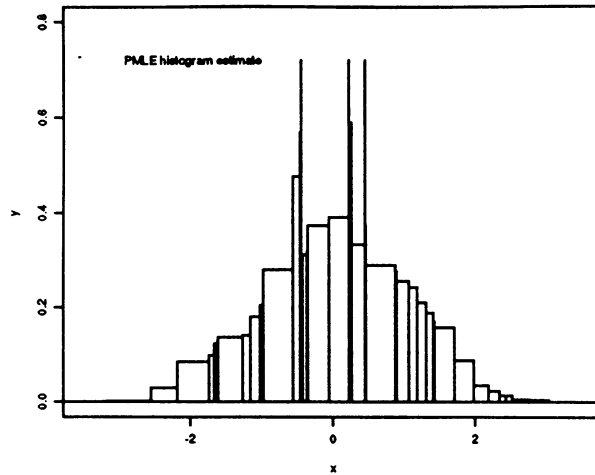
12

fit 800 normal data with alpha = 1/3

PMLE histogram estimate

Figure 4.1a

fit 800 normal data with alpha = 1/3

_____ true density

... fixed knots

· · · deleting knots
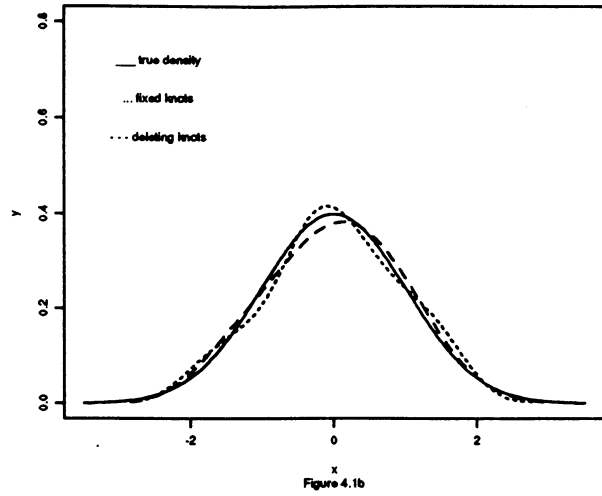
Figure 4.1b

fit 800 normal data with alpha = 1/5

PMLE histogram estimate
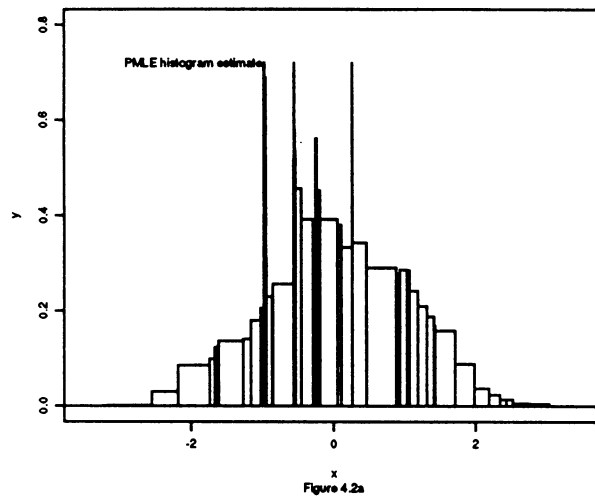
Figure 4.2a

fit 800 normal data with alpha = 1/5

_____ true density

... fixed knots

· · · deleting knots
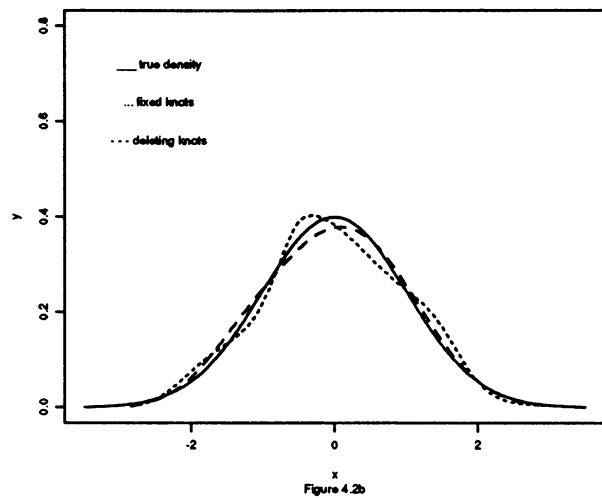
Figure 4.2b

fit 800 normal data with alpha = 1/7

PMLE histogram estimate
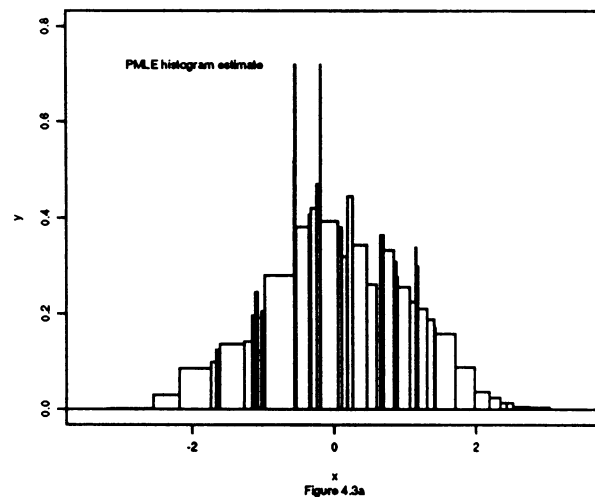
Figure 4.3a

fit 800 normal data with alpha = 1/7

_____ true density

... fixed knots
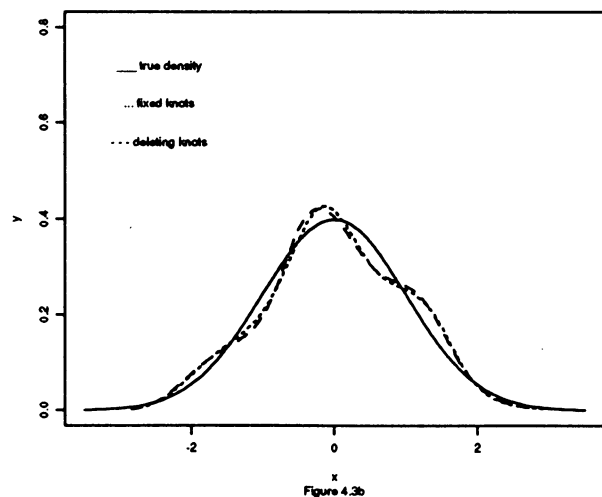
· · · deleting knots

Figure 4.3b

fit 800 normal data with alpha = 1

PMLE histogram estimate

Figure 4.4a

fit 800 normal data with alpha = 1

true density
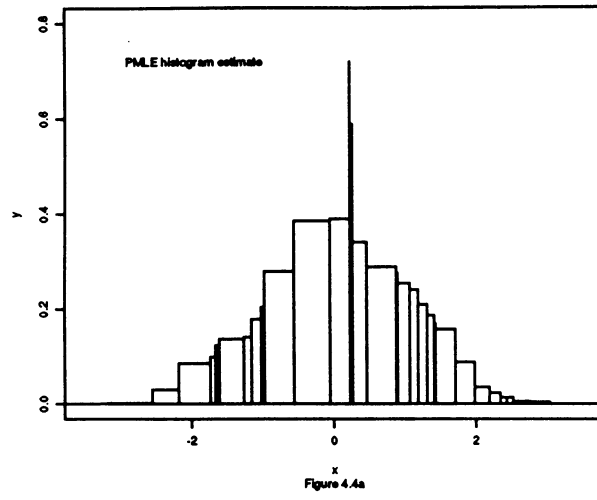
... fixed knots

· · · deleting knots
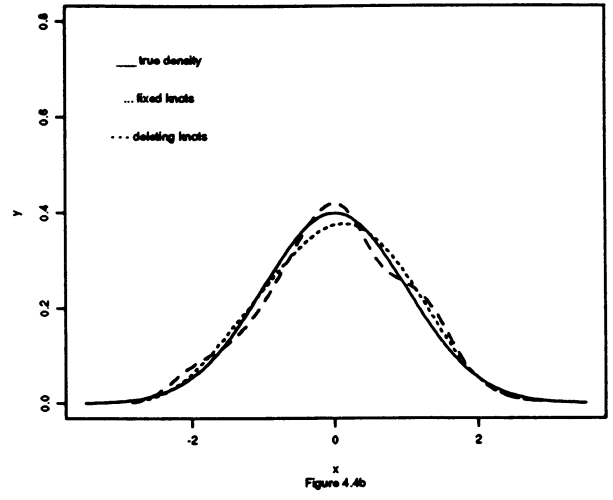
Figure 4.4b

In spline smoothing with fixed knot spans, we use $s = 9, 9, 11$ for the cases of $\alpha = 1/3, 1/5, 1/7$. The following table shows the resulting knot locations.

Table 2. Knots resulting from the two selection rules for Figure 4.1b—4.4b.

	Fixed Knots	Knots deletion
$\alpha = 1/3$	-1.09, -0.36, 0.89, 2.26	-1.09
$\alpha = 1/5$	-1.09, -0.56, -0.24, 0.90, 1.85	-0.85
$\alpha = 1/7$	-1.05, -0.21, 0.67, 1.57	-1.00, -0.57, 0.67, 1.37
$\alpha = 1$	-1.09, 0.67, 2.09	-1.62, 0.23, 0.67, 1.42

It seems that there are too many knots remained in the knots deletions for the case $\alpha = 1/7$. With a PMLE estimate, we could easily delete some unnecessary knots by hand or add more weight to penalize the number of variables (e.g. change 3 to 4 in (3.10)).

Figure 4.1-4.3 suggest that the underlying density is unimodal. Thus, we can estimate the density by using a more informative model—unimodal ($\alpha = 1$). The results are shown in Figure 4.4a–4.4b with $s = 9$ for the fixed knots rule.

In summary, with a PMLE as preliminary estimate, one can easily determine where to locate important knots. If knots determined by either a fixed knots rule or a knot deletion rule turns out unsatisfactory so that the resulting spline smoothing curve is qualitatively differs too much from the PMLE, suitable modifications are required and these can easily be done by hand with the help of PMLE. It should also stress that PMLE itself provides informative knowledge about the structure of data sets.

5 Applications to Lipoprotein Data

In this section, we apply the PMLE techniques and spline smoothing techniques to the two sets of lipoprotein data.

The measurements are as follows. Each individual gave a blood sample and a suitably treated fraction of the plasma was subjected to gradient gel electrophoresis and stained
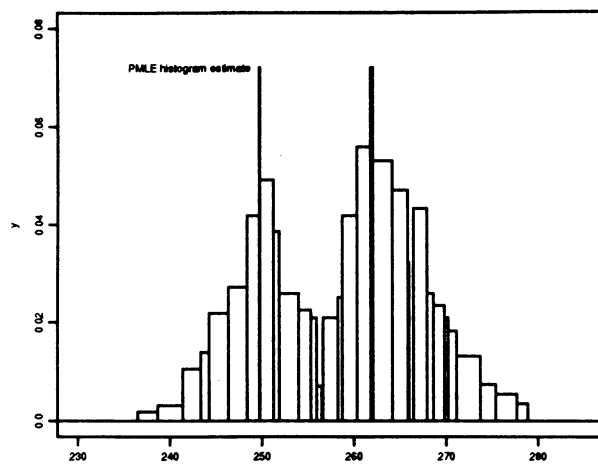
Boston Lipoprotein Data, n=244

PMLE histogram estimate

Figure 5.1

Boston Lipoprotein Data, n=244

fixed knots

···· deleting knots
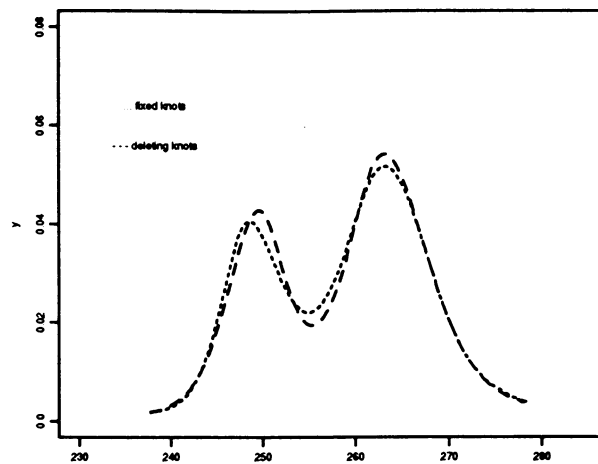
Figure 5.2

Montreal Lipoprotein Data, n = 684

PMLE histogram estimate
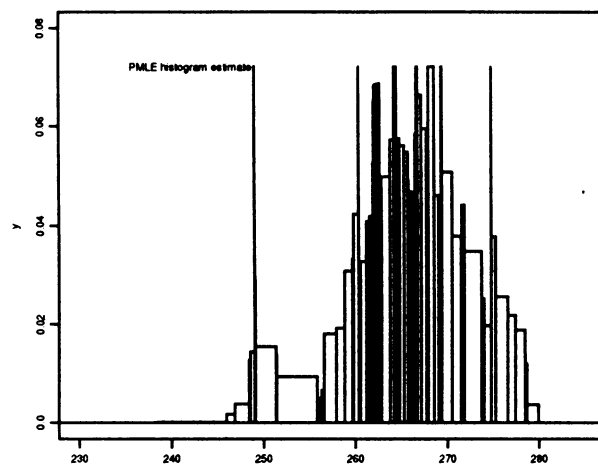
Figure 6.1

Montreal Lipoprotein Data, n = 684

fixed knots

···· deleting knots
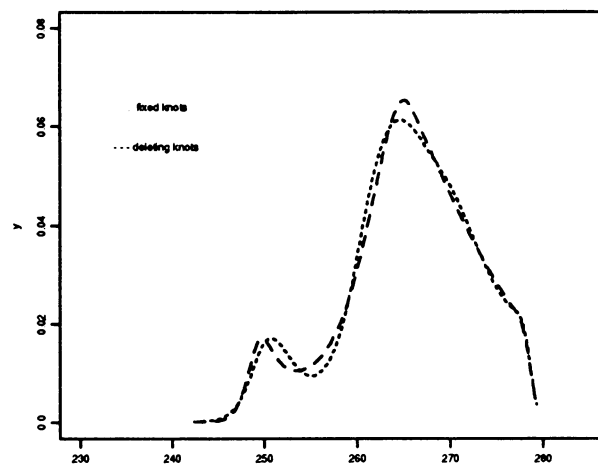
Figure 6.2

Montreal Lipoprotein Data, n = 684

PMLE histogram estimate
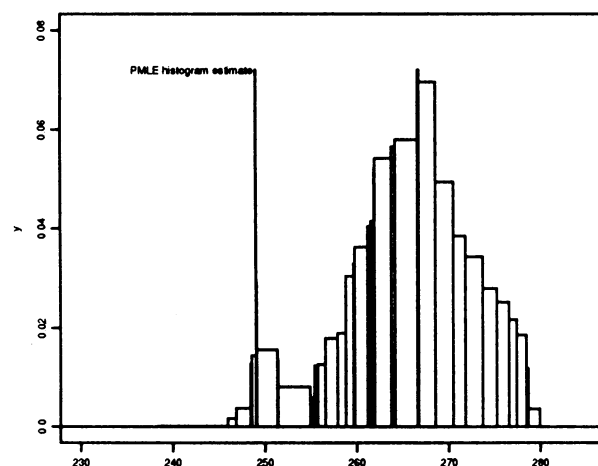
Figure 6.1a

Montreal Lipoprotein Data, n = 684

fixed knots

···· deleting knots
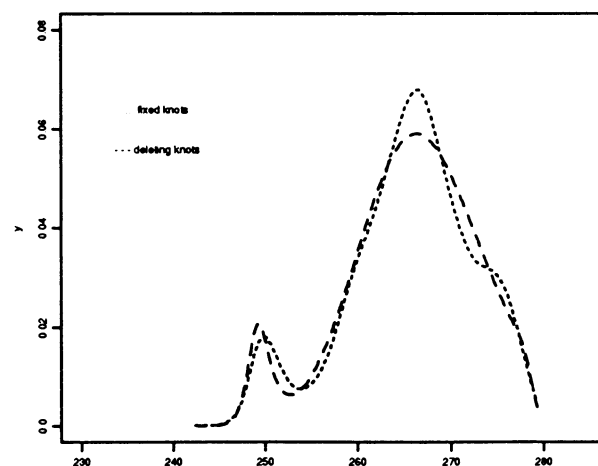
Figure 6.2a

with a protein stain. The measurements are diameters (in Angstroms) of the location of the major peak in the low-density lipoprotein region, the values being obtained by calibrating a densitometric scan.

Figure 5.1 shows the PMLE of the major sizes of lipoprotein for 244 people in Boston area. The estimate is computed with $\alpha = 1/5$: we allow an unknown population density having up to 5 modes. The result shows that the underlying density is bimodal with modal locations around 250 and 265. The estimate is smoothed by using the spline techniques described in section 3.3. In the fixed knots case, we use knot span $s = 6$. The following table shows the knots used in the computation.

Table 3. Knots resulting from the two selection rules for Figure 5.2-6.2.

	Fixed knots	Knot deletion
Boston	247.40, 254.65, 261.05, 267.23, 274.55	250.53, 254.65, 261.05
Montreal	249.50, 255.10, 260.33, 267.74, 271.24, 277.01	249.04, 251.35, 265.48, 266.29, 277.02

Similar measurements were used in the Montreal area, where the blood plasma of 684 individuals were measured. The PMLE with $\alpha = 1/10$ for the Montreal data set is plotted in Figure 6.1. In this case, we specified a very flexible model which allows the unknown density to have up to 10 modes. Ignoring the thin bins which are usual the artificial mode produced by the PMLE, the resulting estimate still looks bimodal. The spline smoothed versions are plotted in Figure 6.2. In the spline technique with fixed knot span, we initially take $s = 9$, and then delete some unnecessary knots. The resulting knots are shown in Table 3.

From the analysis above, we can fit a more informative model: the population density is unimodal in the intervals (230,255) and (255,280), as suggested by Figure 6.1-6.2. Thus, we separate the data into two subsets: data no larger than 255 (59 cases) and data larger than 255. We apply the PMLE with $\alpha = 1$ to each data set. The resulting combined density estimates are shown in Figure 6.1a-6.2a. Note that this more informative model makes Figure 6.2a looks nicer than Figure 6.2.

It is of interest to note that both populations have the same intrinsic features: bimodal

14

with mode locations around 250 and 265 and a valley at 255. However, there are a larger proportion of people having lower sizes (≤ 255) of lipoprotein in Boston area (about 37.5%) than in Montreal area (about 8.6%), while the dispersion around their modes are almost the same. To explain these results, one might need to study the diets and ethnic groups in both areas, which might be of interest in the Public Health.

The data sets were made available by Dr. Ronald M. Krauss of Lawrence Berkeley Laboratory via Professor Terry Speed.

6 Concluding Remarks

The major conceptual innovation of this paper is the introduction of the notion of regularly shaped densities. The PMLE is then introduced to take the shape restrictions into account. It is an intuitively appealing method, having the attractive feature of determining the smoothing parameters automatically. The PMLE itself is very informative. With the PMLE as a preliminary estimate, it is easy to used the fixed-knot spline technique to produce a smoothed version. We expect that the method will become a useful exploratory data analysis tool after some further modification and study.

7 Appendix—Proof of Proposition 2

First, let's show why the solution should have the form (3.6). For any U-shaped curve $g(x)$ with total area α on the interval $[y_1, y_{n_\alpha}]$, one can always define another U-shaped curve $g^*(x)$ by (3.6) with $f_i = g(y_i)$. Then, the likelihood (3.3) remains the same for both g and g^*, but the total area under the curve g^* on the interval $[y_1, y_{n_\alpha}]$ is no larger than that under g: (because $g^*(x) \leq g(x)$ due to the U-shaped constraints)

$$\int_{y_1}^{y_{n_\alpha}} g^*(x)dx \leq \int_{y_1}^{y_{n_\alpha}} g(x)dx.$$

Now, define a new function

$$g^{**}(x) = \frac{\int_{y_1}^{y_{n_\alpha}} g(x)dx}{\int_{y_1}^{y_{n_\alpha}} g^*(x)dx} \times g^*(x) \geq g^*(x).$$

Then, g and g^{**} have the same area in the interval $[y_1, y_{n_\alpha}]$, but

$$\prod_{i=2}^{n_\alpha-1} g(y_i) = \prod_{i=2}^{n_\alpha-1} g^*(y_i) \leq \prod_{i=1}^{n_\alpha-1} g^{**}(y_i).$$

This completes the first part of the proof.

Now, let's show why f_i is given by (3.7). Since the solution must have form (3.6), Problem II is equivalent to finding the solution to the problem

$$\max \sum_{i=2}^{n_\alpha-1} \log f_i, \tag{7.1}$$

$$\text{subject to}: \quad (U\text{-}shape) \quad f_2 \geq \cdots \geq f_{k-1} \geq f_k \leq f_{k+1} \cdots \leq f_{n_\alpha-1}, \tag{7.2}$$

$$(area\ \alpha) \quad \sum_{i=2}^{n_\alpha-1} w_i f_i = \alpha. \tag{7.3}$$

where

$$w_i = y_i - y_{i-1}, i = 2, \cdots, k-1; w_k = y_{k+1} - y_{k-1}; w_i = y_{i+1} - y_i, i = k+1, \cdots, n_\alpha - 1.$$

Denote $g_i = \frac{\alpha}{(n_\alpha-2)w_i}$. Then, the equality constraint (7.3) can be written as

$$\sum_{2}^{n_\alpha-1} (f_i - g_i)w_i = 0. \tag{7.4}$$

Consider the isotonic regression problem

$$\min_f \sum_{1}^{n} (f_j - g_j)^2 w_j \tag{7.5}$$

with a partial order $2 \succeq \cdots \succeq k-1 \succeq k \preceq k+1 \preceq \cdots \preceq n_\alpha - 1$. Then, the solution to the problem (7.5) is given by (3.7) (see page 23 of Robertson *et al.* (1988)) and satisfies also (7.4) [Theorem 1.3.6 of Robertson *et al.* (1988)]. Finally, by applying Theorem 1.5.1 of Robertson *et al.* (1988), the solution to the problem (7.5) is also the solution to the problem (7.1). The conclusion follows.

16

References

[1] Barlow, R. E., Barthlomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. John Wiley and Sons, London.

[2] Bickel, P. J. and Fan, J. (1990). Some problems on the estimation of densities under shape restrictions. *Technical Report 258*, Dept. Statist., Univ. of California, Berkeley.

[3] Birgé, L. (1987a). Estimating a density under the order restrictions: Non-asymptotic minimax risk. *Ann. Statist.*, **15**, 995-1012.

[4] Birgé, L. (1987b). On the risk of histograms for estimating decreasing densities. *Ann. Statist.* **15**, 1013-1022.

[5] De Boor, C (1978). *A Practical Guide to Splines*. Springer, New York.

[6] Donoho, D. L. (1989). One sided inference about functionals of a density. *Ann. Statist.*, **16**, 1390–1420.

[7] Good, I.J. and Gaskins, R.A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, **58**, 255-277.

[8] Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skand. Akt.*, **39**, 125–153.

[9] Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, (L. M. Le Cam and R. A. Olshen, eds.), 539–555.

[10] Härdle, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen smoothing parameter selectors from their optimum? *J. Amer. Statist. Assoc.*, **83**, 86–101.

[11] Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose response curves and the assessment of synergism, *Manuscript.*

[12] Kooperberg, C. and Stone, C. J. (1990). A study of logspline density estimation. *manuscript.*

[13] Mächler, M.B. (1989). 'Parametric' smoothing quality in nonparametric regression: shape control by penalizing inflection points. *Dissertation,* Swiss Federal Institute of Technology, Zurich.

[14] Mammen, E. (1990). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.,* to appear.

[15] Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A,* **31**, 23-26.

[16] Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.,* **3**, 425–441.

[17] Ramsay, J. O. and Abrahamowicz, M. (1989). Binomial regression with monotone splines: a psychometric applications. *J. Amer. Statist. Assoc.* **84**, 906-915.

[18] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.,* **12**, 1215-1230.

[19] Robertson, T., Wright, F.T., Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* John Wiley, New York.

[20] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[21] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.* **13**, 1378-1402.

[22] Wegman, E. J. (1969). A note on estimating a unimodal density. *Ann. Math. Statist.,* **40**, 1661-1667.

[23] Wegman, E. J. (1970a). Maximum likelihood estimation of a unimodal function. *Ann. Math. Statist.,* **40**, 457–471.

[24] Wegman, E. J. (1970b). Maximum likelihood estimation of a unimodal function, II. *Ann. Math. Statist.*, **40** 2169–2174.

[25] Wegman, E. J. (1975). Maximum likelihood estimation of a probability density function. *Sankhyā* Ser. A, **37**, 211-224.