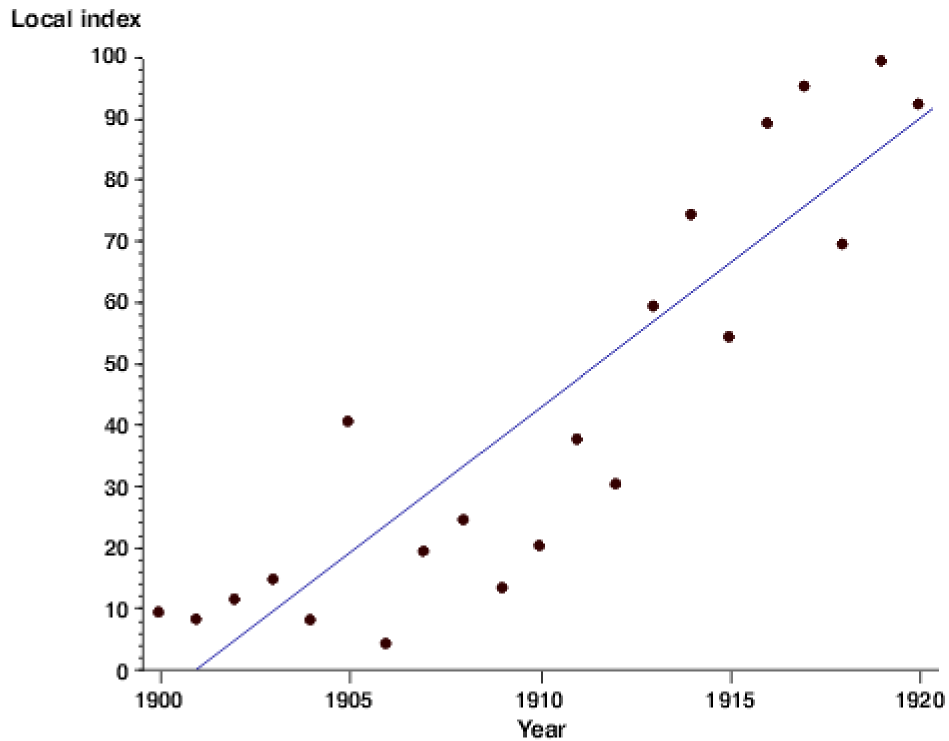
Flexible smoothing with B-splines and Penalties or P-splines

- P-splines = B-splines + Penalization
- Applications : Generalized Linear and non linear Modelling ; Density smoothing
- P-splines have their grounding in Classical regression methods and Generalized linear models
- Regression, Smoothing, Splines?
- B-splines P-splines?

Smoothing, Regression, Splines, B-splines P-splines?

- In statistics, **linear regression** refers to any approach to modeling the relationship between one or more variables denoted *y* and one or more variables denoted *X*, such that the model depends linearly on the unknown parameters to be estimated from the data. Such a model is called a "linear model."

Smoothing, Regression, Splines, B-splines P-splines?



Regression, Smoothing, Splines, B-splines P-splines?

- Linear model

$$y = \alpha + \beta x + \varepsilon$$

- Generalized Linear model where *x′iβ* is the inner product between vectors *xi* and β.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_n x_i^n + e_i$$

Regression, Smoothing, Splines, B-splines P-splines?

- The term *ei* is the residual, . One method of estimation is ordinary least squares. This method obtains parameter estimates that minimize the sum of squared residuals, SSE:

$$e_i = y_i - \hat{y}_i$$

$$SSE = \sum_{i=1}^{N} e_i^2 .$$

Regression, Smoothing, Splines, B-splines P-splines?

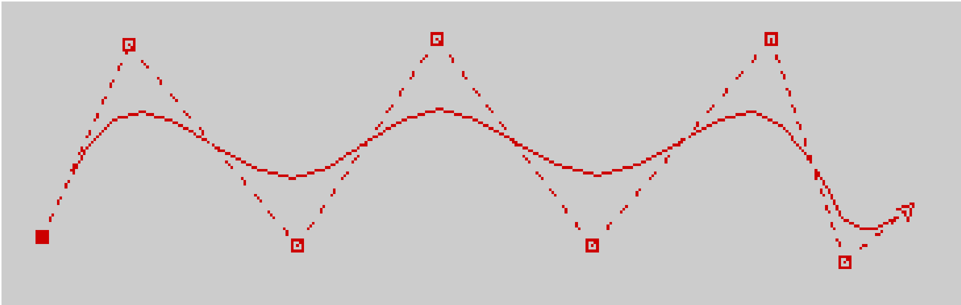
- Smoothing:In statistics and image processing, to **smooth** a data set is to create an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures/rapid phenomena.

Regression, Smoothing, Splines, B-splines P-splines?

- Many different algorithms are used in smoothing. One of the most common algorithms is the "moving average", often used to try to capture important trends in repeated statistical surveys. In image processing and computer vision, smoothing ideas are used in scale-space representations.

Regression, Smoothing, Splines, B-splines P-splines?

- Spline : Originally, a spline tool was a thin flexible strip of wood ,metal or rubber used by draftsman to aid in drawing curved lines.



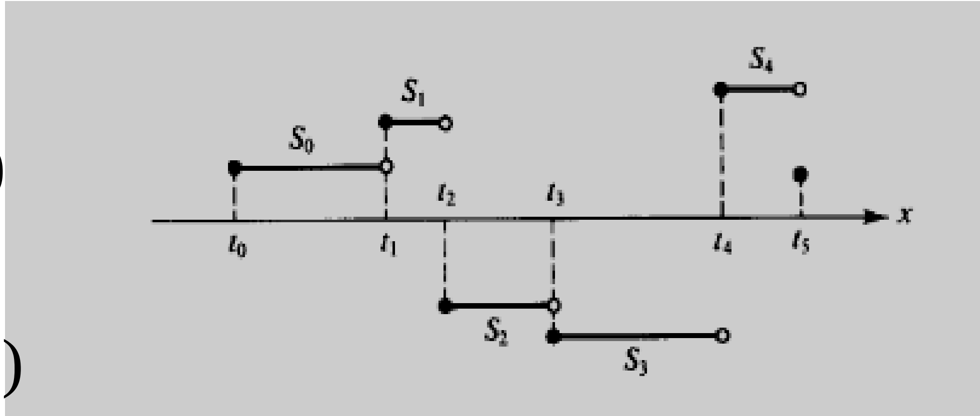
Regression, Smoothing, Splines, B-splines P-splines?

- a **spline** is a special function defined piecewise by polynomials. In interpolating problems, spline interpolation is often preferred to polynomial interpolation because it yields similar results

Splines

- Spline of degree zero

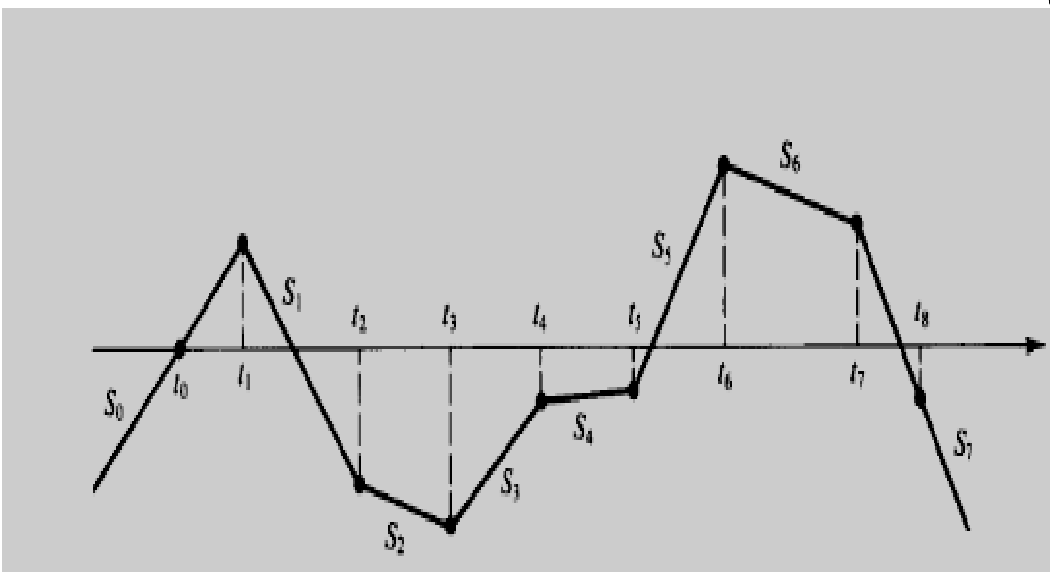
$$S(x)=\begin{cases} S_0 = c_0 & x \in [t_0, t_1) \\ S_1 = c_1 & x \in [t_1, t_2) \\ \vdots & \vdots \\ S_{n-1} = c_{n-1} & x \in [t_{n-1}, t_n) \end{cases}$$



Splines

- A spline of degree 1

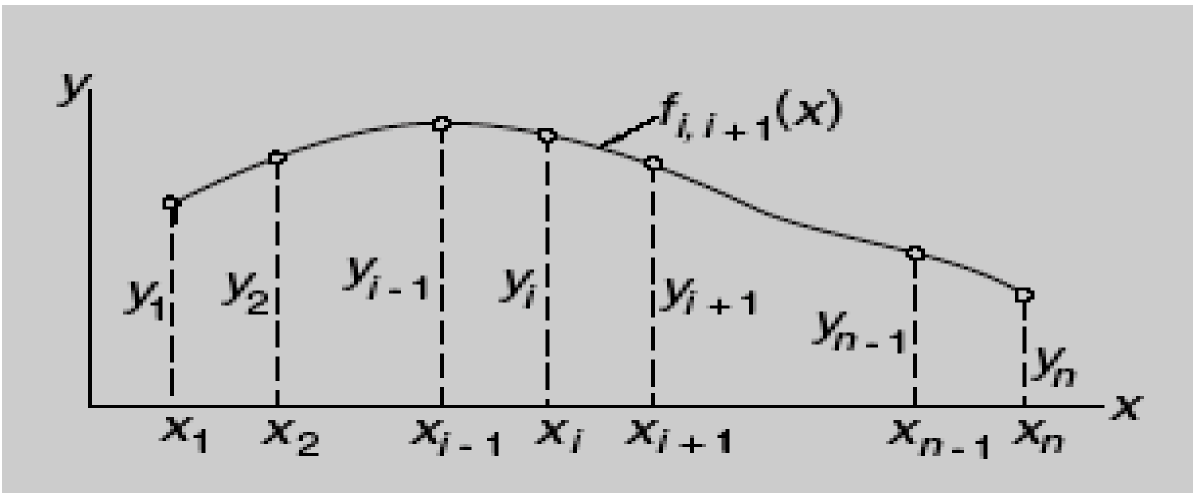
$$S(x) = \begin{cases} S_0 = a_0 x + b_0 & x \in [t_0, t_1) \\ S_1 = a_1 x + b_1 & x \in [t_1, t_2) \\ \vdots & \vdots \\ S_{n-1} = a_{n-1} x + b_{n-1} & x \in [t_{n-1}, t_n) \end{cases}$$



Splines

- A cubic spline

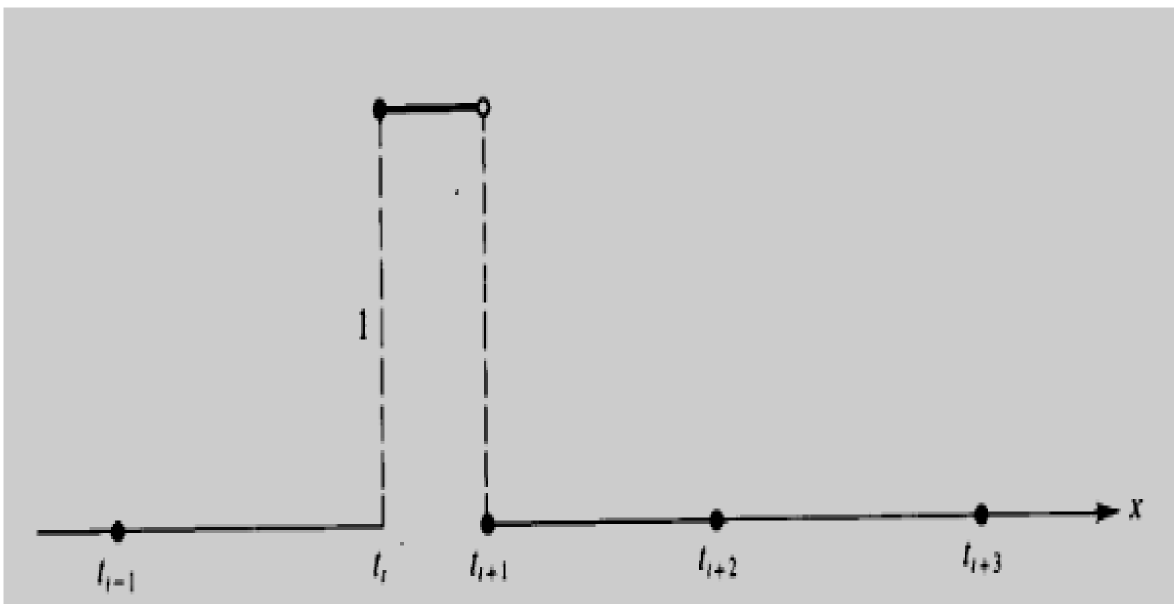
$$S(x) = \begin{cases} x^3 - 1 & x \in [-1, 1/2) \\ 3x^3 - 1 & x \in [1/2, 1) \end{cases}$$



!B-splines!

- B-splines of degree 0

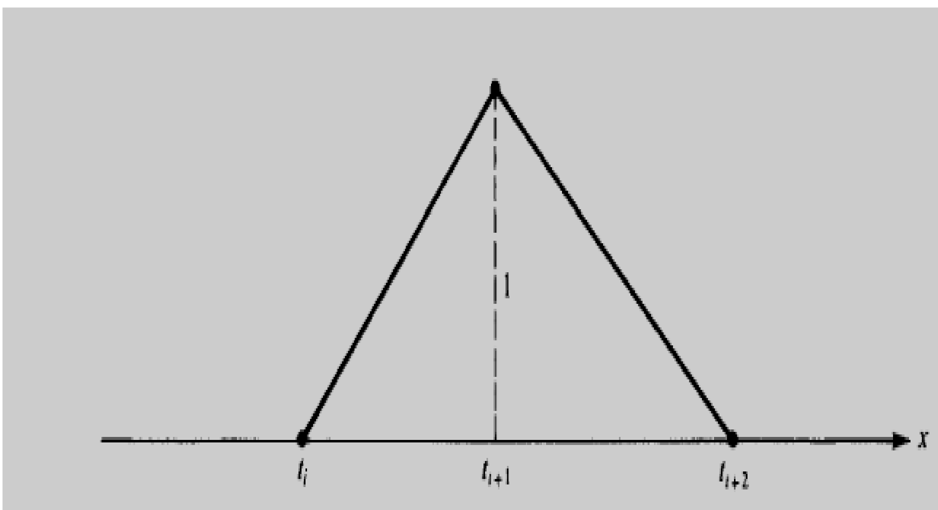
$$B_i^0(x) = \begin{cases} 1 & x \in [t_i, t_{i+1}) \\ 0 & otherwiswe \end{cases}$$



!B-splines

- A spline of degree 1

$$B_i^0(x) = \begin{cases} \dfrac{x - t_i}{t_{i+1} - t_i} & if \quad x \in [t_i, t_{i+1}) \\ \dfrac{t_{i+2} - x}{t_{i+2} - t_{i+1}} & if \quad x \in [t_{i+1}, t_{i+2}) \\ 0 & otherwise \end{cases}$$



B-splines

FLEXIBLE SMOOTHING

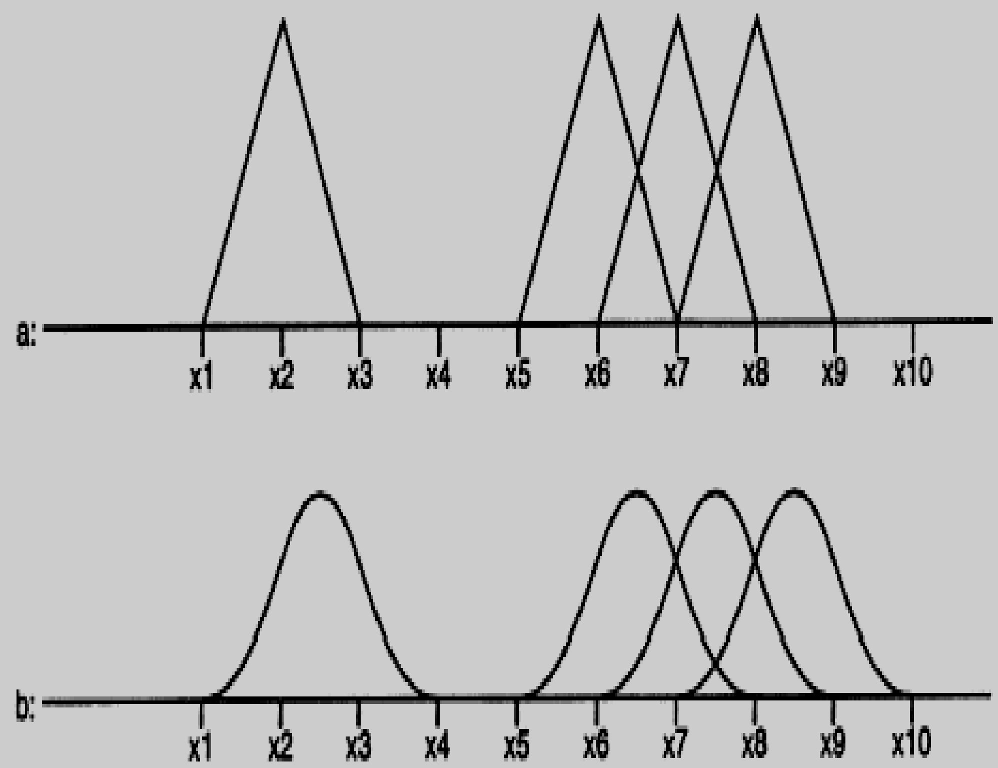


FIG. 1. *Illustrations of one isolated B-spline and several overlapping ones* (a) *degree* 1; (b) *degree* 2.

B-splines

- B-splines are defined recursively

$$B_i^k(x) = \left(\frac{x - t_i}{t_{i+k} - t_i} \right) B_i^{k-1}(x) + \left(\frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} \right) B_{i+1}^{k-1}(x)$$

Properties of P-splines

- No boundary effects
- Are a straightforward extension of (generalized) linear regression models
- Conserve moments like the mean and variances of the data and fit polynomial data exactly
- Computations and cross validation relatively inexpensive

Fitting curves with B-splines

- A fitted curve $\hat{y}\ \ to\left(x_i, y_i\right)$
- is the linear combination $\hat{y} = \sum_{i=1}^{n} \hat{a}_i B_i(x)$

Fitting curves with splines

- The corresponding SSE
- (quadratic error) is

$$S(x) = \sum_{i=1}^{m} \{y_i - \hat{y}\}^2 = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} \hat{a}_j B_j(x) \right\}^2$$

O'Sullivan penalty

- Between 1986 and 1988 O'Sullivan introduced a penalty on the second derivative of the fitted curve;

$$S(x) = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} \hat{a}_j B_j(x) \right\}^2 + \lambda \int_{x\min}^{x\max} \left\{ \sum_{j=1}^{n} \hat{a}_j B_j''(x) \right\}$$

Eilers and MarxPenalities

- Eilers /Marx penalty proposal based on finite differences

$$S(x)=\sum_{i=1}^{m}\left\{y_i-\sum_{j=1}^{n}\hat{a}_j B_j(x)\right\}^2+\lambda\sum_{j=k+1}^{n}\left(\Delta^k\hat{a}_j\right)^2$$

Applications

- Generalized linear modeling
- Density Smoothing
- Example 1 : Motorcycle crash helmet impact simulation data (Härdle 1994) head acceleration in g units at different times after impact. Smoothed with B-splines of degree 3 and a second order penalty

Example 1 Graph

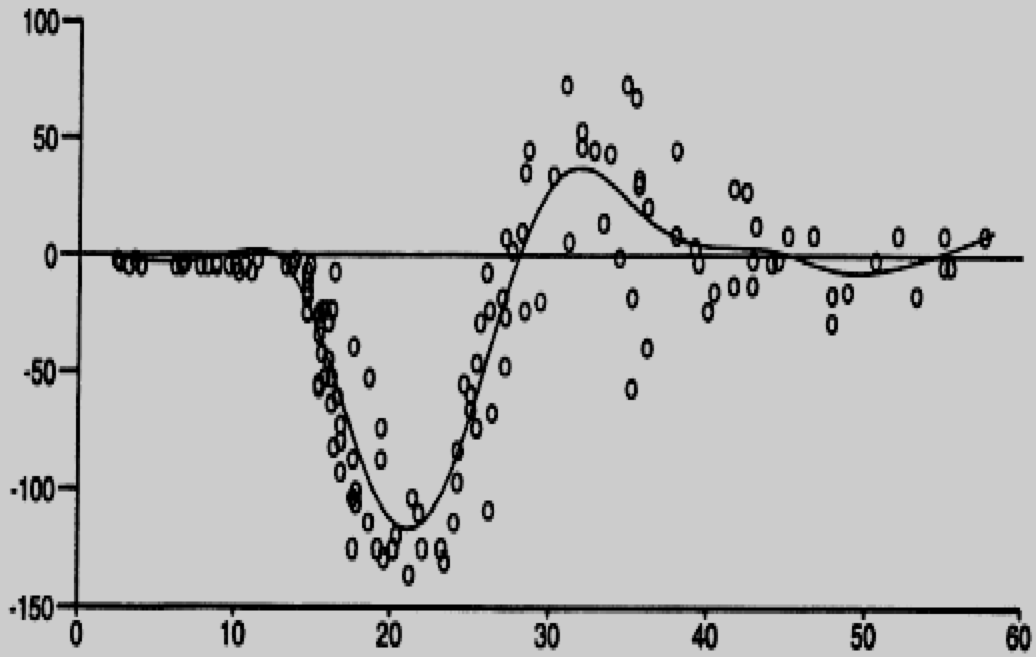


FIG. 2. *Motorcycle crash helmet impact data: optimal fit with B-splines of third degree, a second-order penalty and* $\lambda = 0.5$.

Density Smoothing

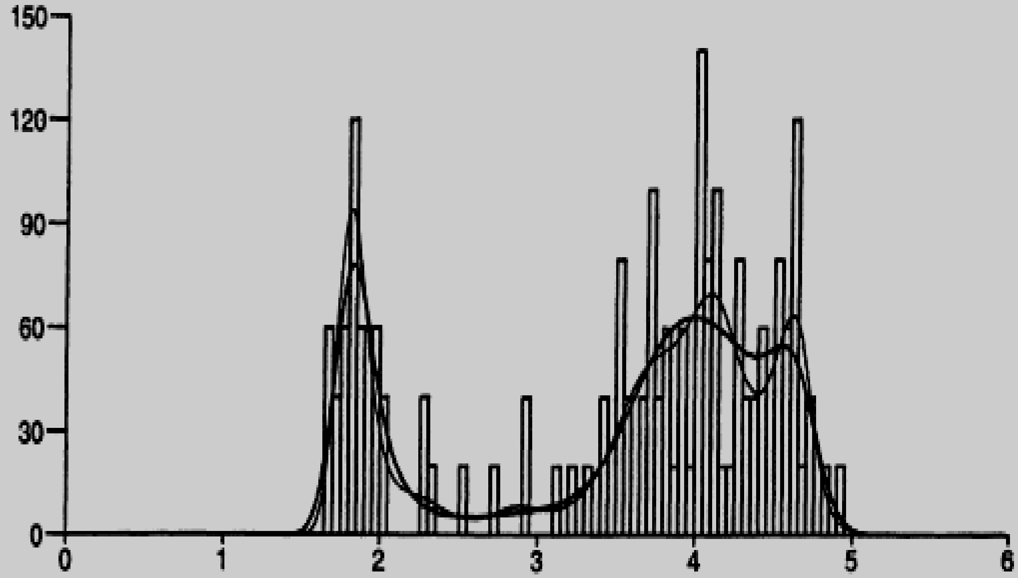


FIG. 6. *Density smoothing of durations of Old Faithful geyser eruptions: density histogram and fitted densities; thin line, third-order penalty with* $\lambda = 0.001$(AIC = 84.05); *thick line, optimal* $\lambda = 0.05$, *with* AIC = 80.17; *B-splines of degree* 3 *with* 20 *intervals on the domain from* 1 *to* 6.

Density Smoothing

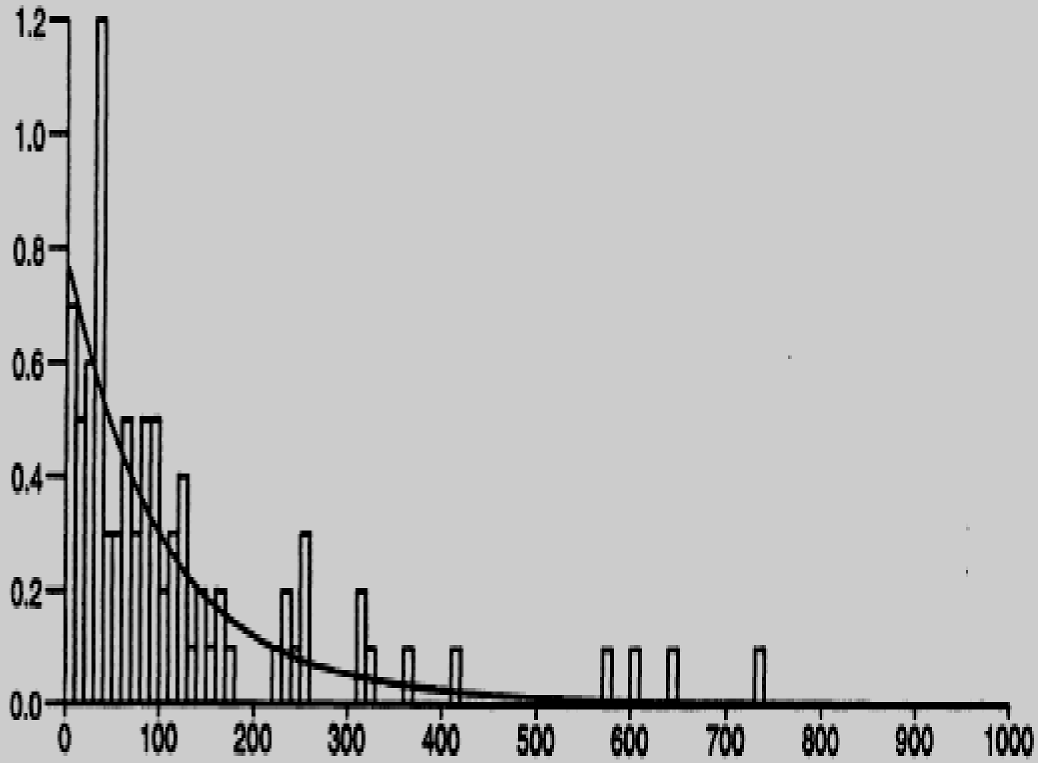


FIG. 7. *Density smoothing of suicide data: positive domain (0–1,000); B-splines of degree 3, penalty of order 2, 20 intervals,* λ = 100, AIC = 69.9.

Density Smoothing

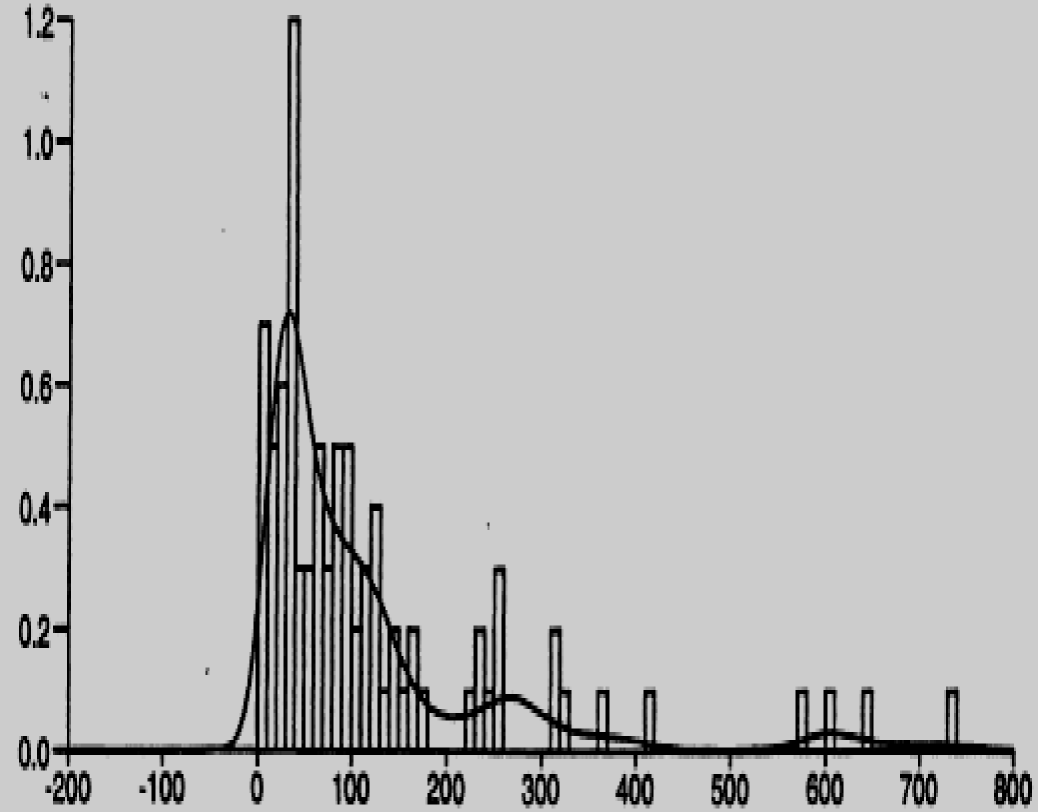


FIG. 8. *Density smoothing of suicide data: the domain includes negative values* (−200–800); *B-splines of degree* 3, *penalty of order* 2, 20 *intervals*, $\lambda = 0.01$, AIC = 83.6.