Calculation of Cubic Smoothing Splines for Equally Spaced Data

David Culpin
CSIRO, Division of Mathematics and Statistics, P.O. Box 218, Lindfield, N.S.W., Australia

Summary. A method is described for fitting cubic smoothing splines to samples of equally spaced data. The method is based on the canonical decomposition of the linear transformation from the data to the fitted values. Techniques for estimating the required amount of smoothing, including generalized cross validation, may easily be integrated into the calculations. For large samples the method is fast and does not require prohibitively large data storage.

Subject Classifications: AMS(MOS): 65D07, 65D10; CR: G1.2.

1. Cubic Smoothing Splines for Arbitrarily Spaced Data

Let data points be \((x_i, y_i)\) for \(i = 1, 2, \ldots, n, \ n \geq 3\). A cubic smoothing spline may be written

\[
f(x) = a_t + b_t(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3
\]

for \(x_i \leq x \leq x_{i+1}, \ i = 1, \ldots, n-1\). It is a continuous function with continuous first and second derivatives. These continuity conditions, together with the constraints \(f''(x_1) = f''(x_n) = 0\), make the coefficients \(b_i, c_i, d_i\) all linear functions of the \(a_i\)'s (see Sect. 3 for the formulae). Thus the determination of the smoothing spline rests on the determination of the \(a_i\)'s. Note that \(a_i = f(x_i)\). Thus if the spline is required to pass through the data points, then \(a_i = y_i\) and the spline is determined.

A smoothing spline need not pass through the data points, but is determined by requiring a certain balance between its smoothness and its closeness to the data. Let

\[
C = \frac{A^3}{6} \int_{x_1}^{x_n} f''(x)^2 \, dx \quad \text{and} \quad D = \sum_{i=1}^{n} (y_i - a_i)^2,
\]

(1.2)
where $\Delta = \max\{\delta_1, \ldots, \delta_{n-1}\}$, $\delta_i = x_{i+1} - x_i$. The quantities $C$ and $D$ measure, inversely, the smoothness of the spline and its closeness to the data, respectively. (The factor $\Delta^3/6$ in $C$ is introduced merely for convenience of exposition.) The spline may be determined by choosing the $a_i$'s so as to minimize

$$sC + D,$$

where $s$ is some fixed number whose size governs the smoothness of the spline as against its closeness to the data. $s$ may be termed the "smoothness parameter"; the larger it is, the smoother the spline. Note that $C$ is a quadratic form in $a_1, \ldots, a_n$, where $a_n$ is defined to equal $f(x_n)$. Thus we may write

$$C = a^TQa,$$

(1.3)

where $a = (a_1, \ldots, a_n)'$ and $Q$ is a positive definite symmetric matrix depending on $x_1, \ldots, x_n$. Minimizing $sC + D$ with respect to $a$ yields

$$a = A(s)y,$$

(1.4)

where $y = (y_1, \ldots, y_n)'$ and

$$A(s) = (I + sQ)^{-1}.$$

(1.5)

To effect the calculation of $a$, two problems have to be met; the choice of $s$ and the calculation of $A(s)$. The next section deals briefly with the first of these. The remainder of the paper is concerned with the second problem.

2. Choice of Smoothness Parameter

A criterion for judging methods of choosing the smoothness parameter $s$ is that the coefficients $a_i$ that it yields should be close to the true values of the function which the $y_i$'s represent. Thus, if

$$y_i = \mu_i + e_i, \quad i = 1, \ldots, n,$$

where $\mu_i$ is the true value and $e_i$ is an error, then the $a_i$'s must be close to the $\mu_i$'s. When the $e_i$'s are independent random errors with zero expectations (means) and equal variances $\sigma^2$, define the function $R(s)$, the mean squared error, by

$$R(s) = E\left[\sum_{i=1}^{n} (a_i - \mu_i)^2\right]$$

(here $E$ denotes expectation); then an optimum value of $s$ is that for which $R(s)$ is least, and methods for choosing $s$ may be judged by how well they estimate that optimum. Three methods for choosing $s$ are the following.

Method 1. Reinsch's recommendation when $\sigma^2$ is known (Reinsch [3]). Choose $s$ so that

$$\|y - a\|^2 = n\sigma^2.$$