# Spline Models

- Introduction to CS and NCS

  Piece-wise polynomials: we divide the range of $x$ into several intervals, and within each interval $f(x)$ is a low-order polynomial, e.g., cubic or quadratic, but the polynomial coefficients change from interval to interval; in addition we require overall $f(x)$ is continuous up to certain derivatives.

- Regression splines

- Smoothing splines

# Cubic Splines

- knots: $a < \xi_1 < \xi_2 < \cdots < \xi_m < b$
- A function $g$ defined on $[a, b]$ is a cubic spline w.r.t knots $\{\xi_i\}_{i=1}^m$ if:

1) $g$ is a cubic polynomial in each of the $m+1$ intervals,

$$g(x) = d_i x^3 + c_i x^2 + b_i x + a_i, \quad x \in [\xi_i, \xi_{i+1}]$$

where $i = 0 : m$, $\xi_0 = a$ and $\xi_{m+1} = b$;

2) $g$ is continuous up to the 2nd derivative: since $g$ is continuous up to the 2nd derivative for any point inside an interval, it suffices to check

$$g^{(0,1,2)}(\xi_i^+) = g^{(0,1,2)}(\xi_i^-), i = 1 : m.$$

- How many free parameters we need to represent $g$? $m + 4$.

  We need 4 parameters $(d_i, c_i, b_i, a_i)$ for each of the $(m+1)$ intervals, but we also have 3 constraints at each of the $m$ knots, so

$$4(m+1) - 3m = m + 4.$$

Suppose the knots $\{\xi_i\}_{i=1}^{m}$ are given.

If $g_1(x)$ and $g_2(x)$ are two cubic splines, so is $a_1 g_1(x) + a_2 g_2(x)$, where $a_1$ and $a_2$ are two constants.

That is, for a set of given knots, the corresponding cubic splines form a linear space (of functions) with dim $(m + 4)$.

- A set of basis functions for cubic splines (wrt knots $\{\xi_i\}_{i=1}^{m}$) is given by

$$h_0(x) = 1;\ h_1(x) = x;$$

$$h_2(x) = x^2;\ h_3(x) = x^3;$$

$$h_{i+3}(x) = (x - \xi_i)_+^3, \quad i = 1, 2, \ldots, m.$$

- That is, any cubic spline $f(x)$ can be uniquely expressed as

$$f(x) = \beta_0 + \sum_{i=1}^{m+3} \beta_j h_j(x).$$

- Of course, there are many other choices of the basis functions. For example, `R` uses the B-splines basis functions.

# Natural Cubic Splines (NCS)

- A cubic spline on $[a, b]$ is a NCS if its second and third derivatives are zero at $a$ and $b$.
- That is, a NCS is linear in the two extreme intervals $[a, \xi_1]$ and $[\xi_m, b]$. Note that the linear function in two extreme intervals are totally determined by their neighboring intervals.
- The degree of freedom of NCS's with $m$ knots is $m$.
- For a curve estimation problem with data $(x_i, y_i)_{i=1}^n$, if we put $n$ knots at the $n$ data points (assumed to be unique), then we obtain a smooth curve (using NCS) passing through all $y$'s.