Smoothing functional data by least squares

12191885 김재겸
12211917 김한별

Functional Data Analysis

4.1~4.4, 4.6

1. Fitting data using a basis system by least squares

2. A performance assessment of least squares smoothing

3. Least squares fits as linear transformations of the data

4. Computing sampling variances and confidence limits

Fitting data using a basis system by least squares

- fit the discrete observations y_j, j = 1, … , n using the model

$$y_j = x(t_j) + \epsilon_j$$

$$x(t) = \sum_k^K c_k \phi_k(t) = \mathbf{c}'\boldsymbol{\phi}.$$

- The vector c of length K , n by K matrix Φ as containing the values $\phi_k(t_j)$

Fitting data using a basis system by least squares

- (1) Ordinary or unweighted least squares fits

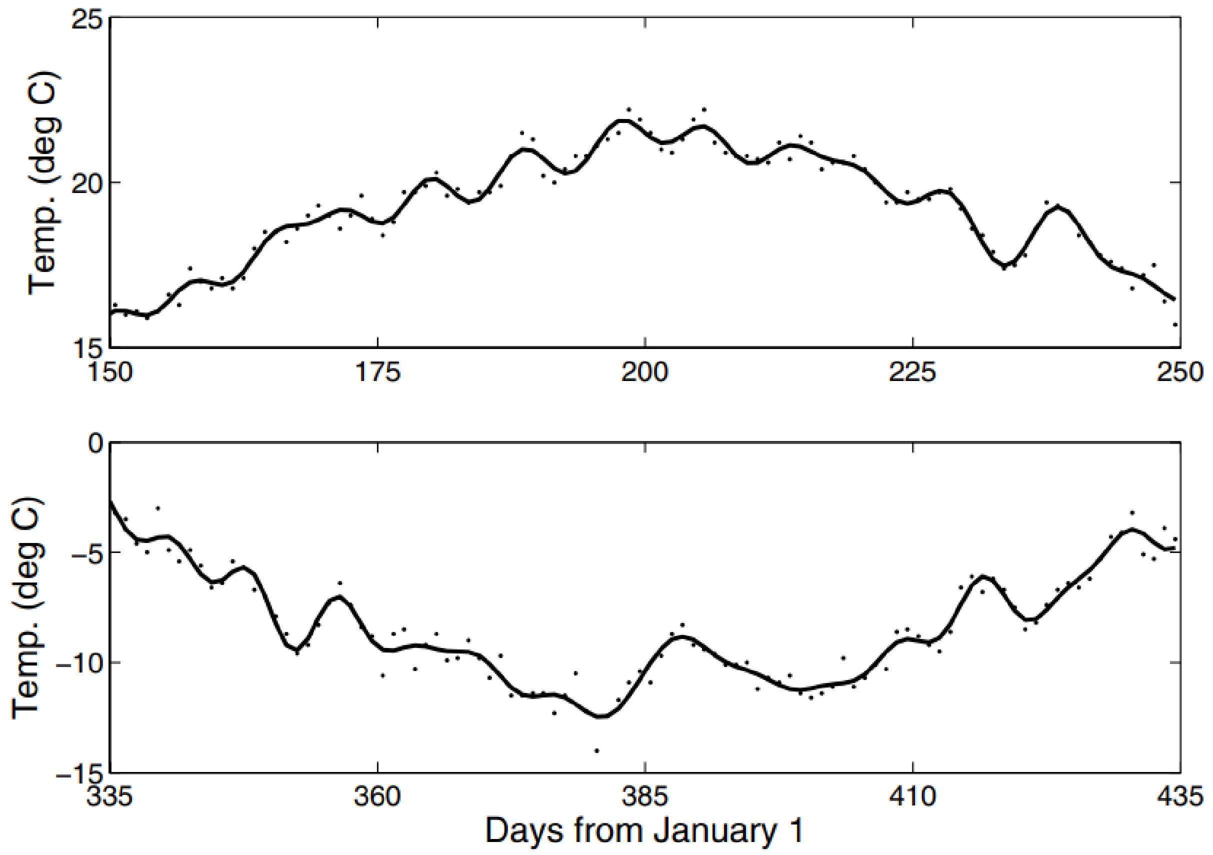
$$\mathtt{SMSSE}(\mathbf{y}|\mathbf{c}) = \sum_{j=1}^{n}[y_j - \sum_{k}^{K} c_k\phi_k(t_j)]^2 = (\mathbf{y} - \mathbf{\Phi c})'(\mathbf{y} - \mathbf{\Phi c}) = \|\mathbf{y} - \mathbf{\Phi c}\|^2$$

- the derivative of criterion $\mathtt{SMSSE}(\mathbf{y}|\mathbf{c})$ with repect to c

$$2\mathbf{\Phi\Phi}'\mathbf{c} - 2\mathbf{\Phi}'\mathbf{y} = 0$$

- $\hat{c} = (\Phi'\Phi)^{-1}\Phi' y$
- $\hat{y} = \Phi\hat{c} = \Phi(\Phi'\Phi)^{-1}\Phi' y$
- The residuals about the true curve are independently and identically distributed with mean zero and constant variance σ^2.

Example of simple least squares approximation



Fitting data using a basis system by least squares

- (2) Weighted least squares fits

- To deal with nonstationary and/or autocorrelated errors,

  we may need to bring in a differential weighting of residuals

$$\mathrm{SMSSE}(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})'\mathbf{W}(\mathbf{y} - \mathbf{\Phi c})$$

$\mathbf{W} = \mathbf{\Sigma}_e^{-1}$, we can always set it to $\mathbf{I}$ if the standard model is assumed.

- $\hat{c} = (\Phi' W \Phi)^{-1} \Phi' W y$

- $\hat{y} = \Phi\hat{c} = \Phi(\Phi' W \Phi)^{-1} \Phi' W y$

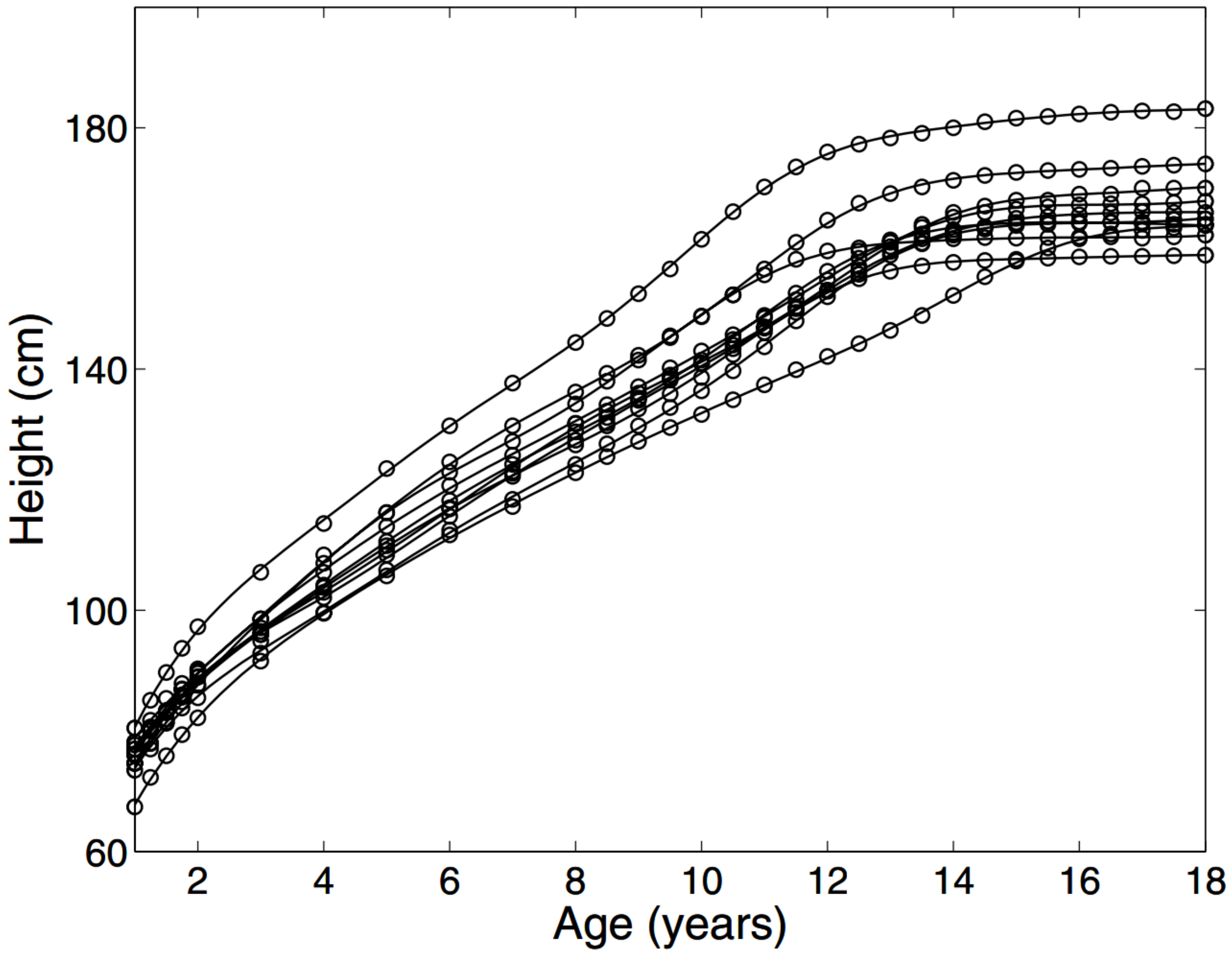


Figure 1.1. The heights of 10 girls measured at 31 ages. The circles indicate the unequally spaced ages of measurement.

A performance assessment of least squares smoothing

- The parametric growth curve proposed by Jolicoeur

$$h(t) = a\frac{\sum_{\ell=1}^{3}[b_\ell(t+e)]^{c_\ell}}{1+\sum_{\ell=1}^{3}[b_\ell(t+e)]^{c_\ell}}$$

eight parameters : a, e and (b, c), l = 1, 2, 3

- average : a = 164.7, e = 1.474, b =(0.3071, 0.1106, 0.0816)’, c = (3.683, 16.665, 1.474)’

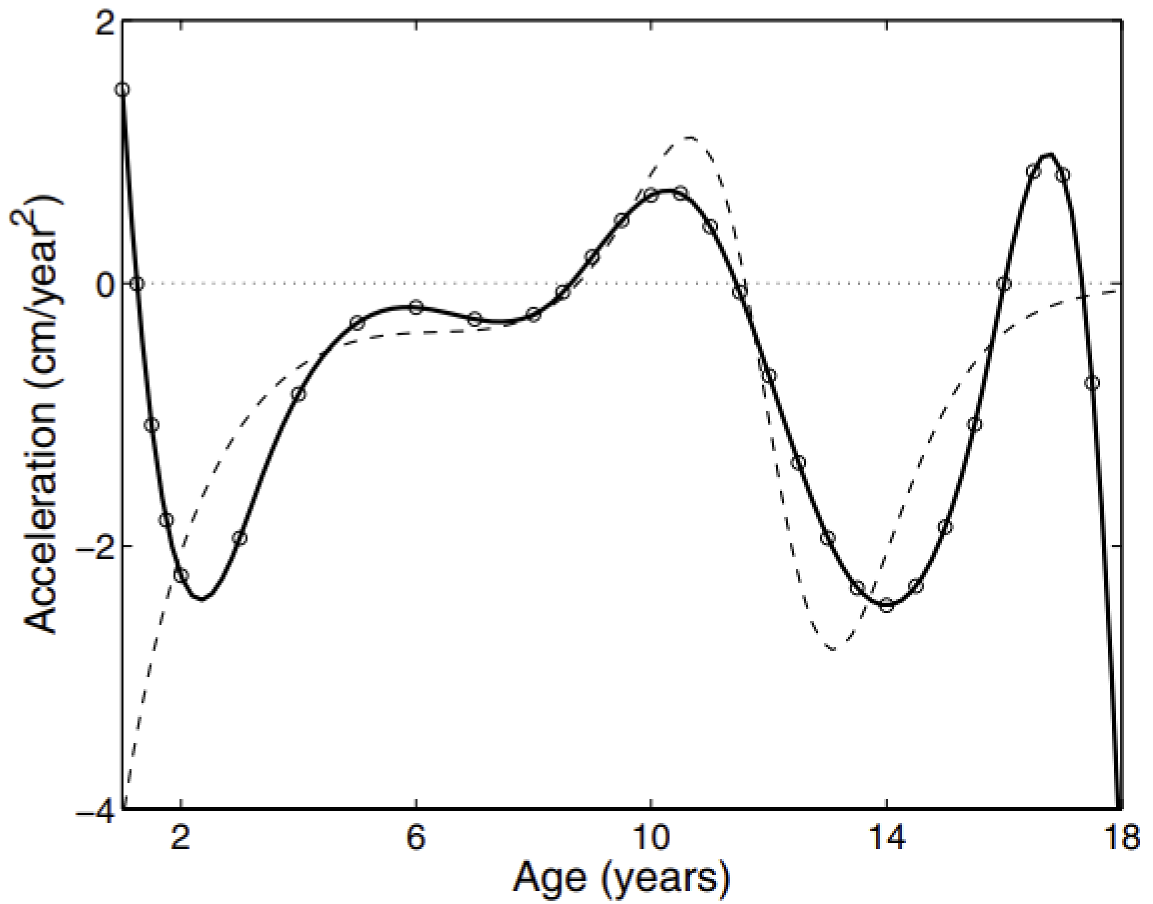


Figure 4.3. The solid curve is the estimated growth acceleration for a single set of simulated data, and the dashed curve is the errorless curve. The circles indicate the ages at which simulated observations were generated.

Least squares fits as linear transformations of the data - (1) How linear smoothers work

- A linear smoother estimates the function value y_j by a linear combination of the discrete observations

$$\hat{x}(t_j) = \sum_{\ell=1}^{n} S_j(t_\ell) y_\ell \quad , \qquad \hat{x}(\mathbf{t}) = \mathbf{S}\mathbf{y} = \Phi\hat{c} = \Phi(\Phi'\Phi)^{-1}\Phi' y$$

- In the uweighted least squares case : $\mathbf{S} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'.$

- Weighted least squares : $\mathbf{S} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{W}$

- property of the smoothing matrix
  - Projection matrix : it creates an image of data vetor y on the space spanned by the columns of matrix Φ

    => residual vector e is orthogonal to the fit vector $\hat{y}$: $(\mathbf{y} - \hat{\mathbf{y}})'\hat{\mathbf{y}} = 0 \quad , \qquad (\mathbf{y} - \hat{\mathbf{y}})'\mathbf{W}\hat{\mathbf{y}} = 0$
  - idempotency : SS = S

Least squares fits as linear transformations of the data - (1) How linear smoothers work

$$\mathbf{S}(a\mathbf{y} + b\mathbf{z}) = a\mathbf{S}\mathbf{y} + b\mathbf{S}\mathbf{z}$$

- The linearity property is important for working out various properties of the smooth representation, and the simplicity of the smoother implies relatively fast computation.

- But, some nonlinear smoothers may be more adaptive to different behavior in different parts of the range of observation
  - ex) wavelet transform

Least squares fits as linear transformations of the data - (2) The degrees of freedom of a linear smooth

- the number of parameters is the length K of the coefficient vector c
- degrees of freedom for error : n – K
- degrees of freedom of the smooth fit
  - $df = \mathrm{trace}\,\mathbf{S}$
  - $df = \mathrm{trace}\,(\mathbf{S}\mathbf{S}')$

Computing sampling variances and confidence limits - (1) Sampling variance estimates

- the model for the data vector y, in this case $x(t)$, is regarded as a fixed effect having zero variance
  - $y = x(t) + \varepsilon$, $\therefore \Sigma_e$만 고려
- $\mathtt{Var}[\mathbf{c}] = (\mathbf{\Phi}'\mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{W}\mathbf{\Sigma}_e\mathbf{W}\mathbf{\Phi}(\mathbf{\Phi}'\mathbf{W}\mathbf{\Phi})^{-1}$ $\hat{c} = (\Phi' W \Phi)^{-1} \Phi' W y$
- When the standard model is assumed, $\Sigma_e = \sigma^2 \mathrm{I}$, and if unweighted least squares is used,

$$\mathtt{Var}[\mathbf{c}] = \sigma^2(\mathbf{\Phi}'\mathbf{\Phi})^{-1}$$

Computing sampling variances and confidence limits - (1) Sampling variance estimates

- the sampling variance of the fit to the data defined by $x(t) = \phi(t)'\mathbf{c}$

$$\texttt{Var}[\hat{x}(t)] = \phi(t)'\texttt{Var}[\mathbf{c}]\phi(t)$$

$$\texttt{Var}[\hat{\mathbf{y}}] = \mathbf{\Phi}\texttt{Var}[\mathbf{c}]\mathbf{\Phi}'$$

- In the standard model / unweighted least squares case

$$\texttt{Var}[\hat{\mathbf{y}}] = \sigma^2\mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}' = \sigma^2\mathbf{S}$$

Computing sampling variances and confidence limits - (2) Estimating Σ_e

- $$s^2 = \frac{1}{n-K}\sum_j^n (y_j - \hat{y}_j)^2$$

- strategy of choosing K
  - Add basis function until s^2 fails to decrease substantially.

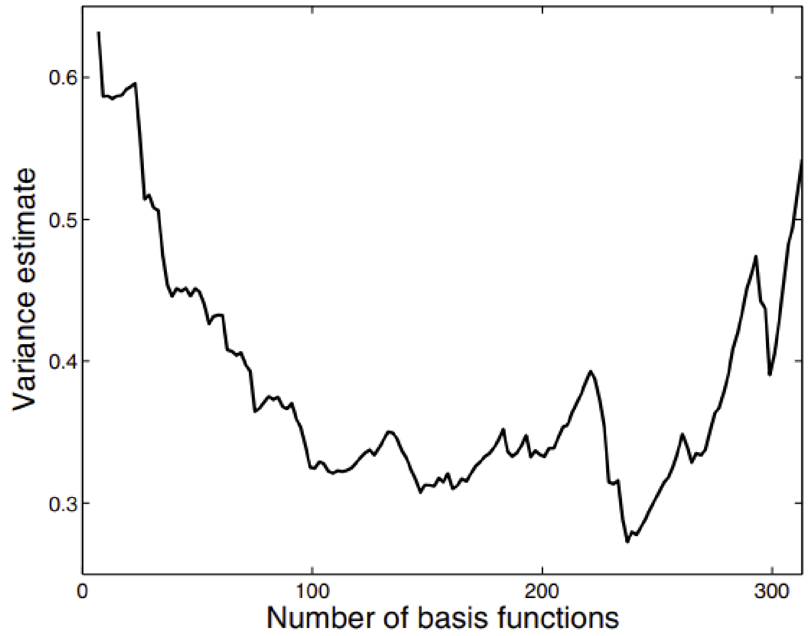


Figure 4.6. The relation between the number of Fourier basis functions and the unbiased estimate of the residual variance (4.22) in fitting the Montreal temperature data.

Computing sampling variances and confidence limits - (3) Confidence limits

actual fit $\pm$ s.e. $* \boxed{}$

s.e. = $\sqrt{Var(\hat{x}(t))} = \sqrt{\phi(t)'Var(c)\phi(t)}$

Confidence limits on fit computed in this way are called “point-wise”.

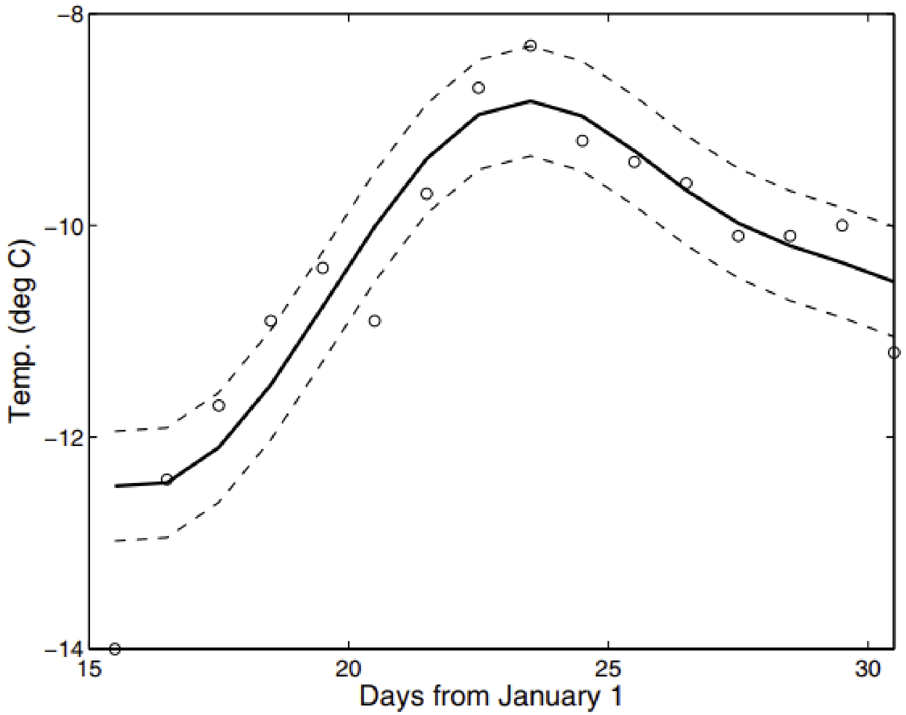


Figure 4.7. The temperatures over the mid-winter thaw for the Montreal temperature data. The solid line is the smooth curve estimated in Figure 4.1 and the lower and upper dashed lines are estimated 95% point-wise confidence limits for this fit.

Q&A