Linear Smoothing: Conclusion and Extensions V

Example: Height of Girls

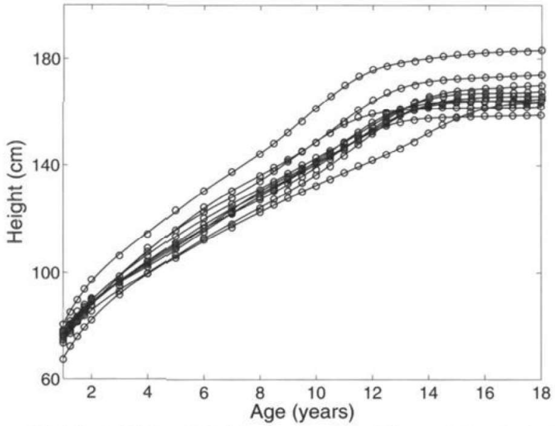

Figure 1.1. The heights of 10 girls measured at 31 ages. The circles indicate the unequally spaced ages of measurement.

Linear Smoothing: Conclusion and Extensions VI

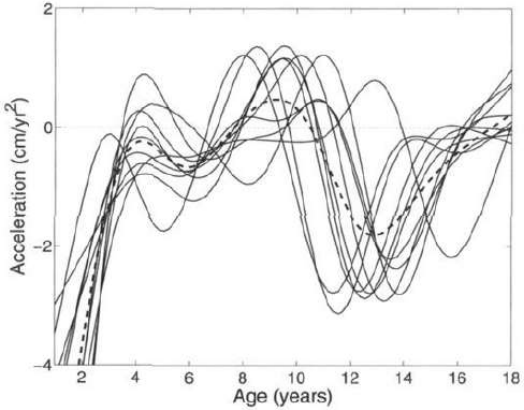

Figure 1.2. The estimated accelerations of height for 10 girls, measured in centimeters per year. The heavy dashed line is the cross-sectional mean, and is a rather poor summary of the curves.

Functional PCA I

Definition (Mean and variance functions)

Consider the functions $\{\mu_j(t)\}_{j=1,\cdots,J}$ that are observed instances of the random function $s(t)$, the mean function is the average of the functions:

$$\mathbb{E}[s(t)] = \overline{\mu}(t) = \frac{1}{J}\sum_{j=1}^{J}\mu_j(t)$$

and the variance function is:

$$\mathbb{V}\text{ar}[s(t)] = \frac{1}{J-1}\sum_{j=1}^{J}(\mu_j(t) - \overline{\mu}(t))^2$$

Functional PCA II

Example (Height of Girls)

$$\overline{\text{Height}}(t) = \frac{1}{10}\sum_{i=1}^{10}\text{Height}_i(t)$$

$$\text{Var}_{\text{Height}}(t) = \frac{1}{10-1}\sum_{i=1}^{10}(\text{Height}_i(t) - \overline{\text{Height}}(t))^2$$

Functional PCA III

FPCA, Principal Component Analysis for functions:

- We compute the mean $\overline{\mu}(t)$ and substract it to each curve:

$$\tilde{\mu}_j(t) = \mu_j(t) - \overline{\mu}(t)$$

- Let $v(t_1, t_2) = \mathbb{E}[\tilde{\mu}(t_1)\ \tilde{\mu}(t_2)]$ be the sample covariance function, it can be estimated by:

$$\hat{v}(t_1, t_2) = \frac{1}{J} \sum_{j=1}^{J} \tilde{\mu}_j(t_1)\ \tilde{\mu}_j(t_2)$$

- An eigencurve $u(t)$ is then computed such that:

$$\forall t_1,\ \int \hat{v}(t_1, t)\ u(t)\ dt = \lambda\ u(t_1) \quad \text{subject to} \quad \int u(t)^2 dt = 1$$

Functional PCA IV

- A simple approach is to discretise the functions $\tilde{\mu}_j(t)$ to a fine grid of N sites equally spaced on the time line. This allows to treat these functions as finite dimensional vectors and standard PCA can be applied to estimate the eigenvectors.
- The continuous principal component $u(t)$ is recovered by interpolating the discrete eigenvector.
- The choice of interpolation method does not matter when recovering the eigencurve from the eigenvector when the space between the N sampling sites is small.
- This approach is the earliest approach to functional principal component analysis.