非参数统计

林红梅
统计与信息学院，乐群楼 205
xuemeijiayi@163.com

November 30, 2020

L5 纵向数据分析

1. 方差函数的估计
2. 纵向数据下非参数回归模型的估计问题

基于残差的估计方法

- Use local linear method and residual-based methods (Fan and Gijbels (1996), Yu and Jones(2004)).
- Specifically, we compute the local linear estimate $f(x)$ based on original observations. Then we compute squared residuals $r_i = (y_i - \hat{f}(x_i))^2$ for $i = 1, \ldots, n$. Then we estimate $\sigma^2(x)$ as $\hat{\alpha}_1$ where

$$(\hat{\alpha}_1, \hat{\alpha}_2) = \arg\min \sum_{i=1}^{n} (r_i - \alpha_1 - \alpha_2(x_i - x))^2 K_h(x_i - x),$$

and the bandwidth h is selected using the cross validation method by Fan and Gijbels (1996).

Suppose that $\{(x_{ij}, y_{ij}), i = 1, \ldots, n, j = 1, \ldots, J_i\}$ is a random sample from the nonparametric regression model

$$y_{ij} = m(x_{ij}) + \varepsilon_{ij}, \tag{2.1}$$

where x_{ij} is a univariate, $m(\cdot)$ is an unknown nonparametric smoothing function and ε_{ij}'s are random errors with $\mathrm{E}(\varepsilon_{ij}) = 0$.

- Here (x_{ij}, y_{ij}) is the jth observation of the ith subject or cluster.
- Combining the idea of the local polynomial modelling (Fan and Gijbels, 1996) with the advantage of the Cholesky decomposition, Yao and Li(2013) proposed a new procedure to estimate the correlation structure and regression function simultaneously for the nonparametric regression model (2.1).

估计过程

For ease of presentation, they start with balanced case with $J_i = J$. Denote $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})^T$ and $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \Sigma$. On the basis of the Cholesky decomposition, there exists a lower triangle matrix Φ with 1's on the main diagonal such that $\mathrm{cov}(\Phi\boldsymbol{\varepsilon}_i) = \Phi\Sigma\Phi^T = \mathbf{D}$, where $\mathbf{D} = \mathrm{diag}(d_1^2, d_2^2, \ldots, d_J^2)$ is a diagonal matrix. Let $\mathbf{e}_i = (e_{i1}, e_{i2}, \ldots, e_{iJ})^T = \boldsymbol{\varepsilon}_i$. Then they have

$$\varepsilon_{ij} = \sum_{k=1}^{j-1} \phi_{j,k}\varepsilon_{ik} + e_{ij}, \quad i = 1, 2, \ldots, n, \quad j = 1, \ldots, J, \qquad (2.2)$$

where $\phi_{j,k}$ is the negative of the (j,k)-element of the Φ and e_{ij}'s are uncorrelated with $\mathrm{var}(e_{ij}) = d_j^2$. By convention, $\sum_{k=1}^{0} \phi_{1,k}\varepsilon_{ik} = 0$ when $j = 1$.

Substituting the equation (2.2) into the model (2.1), they obtain the following partially linear model with uncorrelated error term e_{ij}:

$$y_{ij} = m(x_{ij}) + \sum_{k=1}^{j-1} \phi_{j,k} \varepsilon_{ik} + e_{ij}, \quad i = 1, 2, \ldots, n, \;\; j = 1, \ldots, J. \tag{2.3}$$

In practice, ε_{ik} is not observable. So, they predict them by $\tilde{\varepsilon}_{ik} = y_{ik} - \tilde{m}(x_{ik})$, where $\tilde{m}(x_{ik})$ is a local linear estimate of $m(\cdot)$ based on model (2.1) pretending that the random errors are independent.

Replacing the ε_{ik}'s in (2.3) with $\tilde{\varepsilon}_{ik}$'s, they have

$$y_{ij} \approx m(x_{ij}) + \sum_{k=1}^{j-1} \phi_{j,k} \tilde{\varepsilon}_{ik} + e_{ij}, \quad i = 1, 2, \ldots, n, \;\; j = 1, \ldots, J. \quad (2.4)$$

Because is unknown, for the partial linear model (2.4), they employ the profile least squares techniques(Fan and Li, 2004)to estimate $\phi_{j,k}$ and $m(\cdot)$ in approximation (2.4).

- As demonstrated in Yao and Li (2013), their estimation is more efficient than Lin and Carroll(2000)'s kernel GEE.
- In addition, they prove that the estimator proposed is as asymptotically efficient as if the true covariance matrix were known a priori.
- Compared with the marginal kernel method of Wang(2003), Linton(2003) and so on, their newly proposed procedure may outperform the existing procedures.
- Moreover, the newly proposed procedure can estimate the correlation structure and regression function simultaneously without specifying the correlation structure.
- As they comment, however, when the predictor X is multivariate, their method is less useful due to the so-called curse of dimensionality.

随堂作业

- 首先写出模型的矩阵形式
- 写出具体的估计过程，并给出估计的显示解
- 编写具体的代码


HAPPY END