线性常微分方程参数稳定性的统计推断

周杰*, 吴婷

西安电子科技大学应用数学系, 西安 710071
E-mail: zhoujie@xidian.edu.cn, wuting@mail.xidian.edu.cn

收稿日期: 2010-07-15; 接受日期: 2011-05-09; * 通信作者
国家自然科学基金 (批准号: 60974082, 60874085) 资助项目

摘要 对具有随机误差的观测数据, 讨论了常系数线性常微分方程参数稳定性的统计推断问题. 通过残差项的 Karhunen-Loeve 分解, 给出了变点检验步骤及其在原假设下的极限分布. 在对立假设下定义了变点的估计, 证明了检验以及估计的一致性. 对常系数二阶常微分方程进行了统计模拟, 结果表明原假设下的极限分布是对真实分布非常好的近似; 对立假设下, 即使输入函数的频率存在 0.75% 的变化, 上述检验也能以大概率拒绝原假设. 最后利用上述方法研究了英国中部地区的气温数据, 揭示了数据一些新的特点.

关键词 线性常微分方程 变点 函数型数据 Karhunen-Loeve 分解

MSC (2000) 主题分类 62F99, 62M10

1 引言

微分方程是数学在其他学科应用最为广泛的分支之一, 在很多领域, 微分方程已是标准的分析工具. 在生物种群研究中, 微分方程可以描述种群数量的变化规律; 在金融工程中, 微分方程可以描述金融衍生产品的价格与其他市场变量之间的关系; 在艾滋病治疗的研究中, 微分方程可以描述淋巴细胞与病毒细胞数量之间的动态关系等.

在对给定观测数据建立微分方程模型时, 我们会提出这样一个问题: 是否所有的数据都满足同一个方程? 更简单的提法是在观测数据的范围内, 方程的参数是否是一致的, 有没有发生变化? 譬如在对气候的研究中, 通常利用一个三阶常微分方程来描述年度气温的变化规律, 虽然已有大量文献证实大气温度上升, 但温度方程中的参数变化了吗? 这显然是非常有意义的问题. 此类问题, 若微分方程对于观测数据严格成立, 则在满足参数可识别的条件下, 由数据即可确定出参数的精确值, 从而判断参数是否发生变化. 但和通常的线性回归模型类似, 对于观测数据, 微分方程并不是严格成立的, 其中包含了误差项, 把数据代入方程得到的参数仅仅是真实参数的估计, 这时如何确定参数是否发生了变化?

上述问题可归于统计推断中经典的变点问题. 变点问题涉及到统计学的各个分支, 内容广泛, 这里就不再详述, 可参看 Csörgő 和 Horváth 的论著 [1] 及其所引述的参考文献. 与经典变点问题的不同之处在于, 常微分方程的观测实质上是一条曲线, 即函数型数据. 经典变点问题, 无论是均值变点、方差变点或统计模型中的参数变点, 所用到的样本都是实值随机变量的离散观测值, 在此基础上建立统计推断方法. 经典的变点推断方法不适用于观测值为曲线的微分方程中的变点问题, 对此需要提出新的变点推断理论. 本文针对常系数线性常微分方程参数稳定性问题, 提出一种简单的变点检验和估计方

法. 首先考察残差项的 Karhunen-Loeve 分解, 即所谓函数型主成分分析, 得到特征函数得分估计, 利用得分估计建立累积和统计量, 在累积和统计量的基础上, 建立起变点的检验和估计步骤. 对上述推断方法, 我们得到统计量在原假设下的极限分布, 并在对立假设下, 建立了检验的一致性和变点估计的一致性. 为检验所给推断方法的有效性, 对二阶非齐次常系数常微分方程做了统计模拟, 结果表明对所考察的三种样本容量, 原假设下极限分布对真实分布近似程度非常好; 在对立假设下, 即使输入函数频率变化只有 0.75%, 上述推断方法也能以大概率拒绝原假设. 最后, 我们利用上述的推断方法研究了英国中部地区的 1772 年至 2009 年每日的气温数据, 发现气温除了文献中已经发现的上升现象外, 描述气温的方程也存在着变化, 而这一点是已有文献中还没有发现的.

关于曲线数据稳定性的统计推断问题, 文献中已有的研究包括 Zhou^[2] 对连续时间 Gauss 过程提出了基于极大似然比的稳定性检验, Horváth 等^[3] 对自回归 Hilbert 过程提出了基于累积和的稳定性检验, Berkes 等^[4] 对独立同分布函数型数据的均值函数提出了基于累积和的稳定性检验, Aue 等^[5] 则研究了独立同分布函数型数据变点的估计问题.

本文安排如下: 第 2 节给出预备内容和相关的条件, 第 3 节给出推断的方法和理论结果, 第 4 节是定理的证明, 第 5 节讨论数值分析, 包括对模拟数据的分析和对年度大气温度方程的实例分析, 第 6 节讨论了推断方法可能的改进与推广.

2 预备内容与相关的假设条件

设 $X(t)$ 满足下面的线性常微分方程,

$$D^m X(t) = f(t) + \beta_1 D^{m-1} X(t) + \beta_2 D^{m-2} X(t) + \cdots + \beta_m X(t), \quad t \in [0, 1],$$

其中 D^m 表示 m 阶的微分算子, $f(t)$ 即所谓的输入函数, 本文假定其已知. $[0, 1]$ 仅仅是一个方便的假定, 可以换作 $\mathbb{R}$ 上任何一个紧集. 定义线性微分算子

$$L_m = D^m - \beta_1 D^{m-1} - \cdots - \beta_{m-1} D - \beta_m,$$

则上面的微分方程可以表示为

$$L_m X(t) - f(t) = 0. \quad (2.1)$$

这种表示方法强调 $X(t)$ 经等式左边算子作用之后成为定义于 $[0, 1]$ 的零函数. 在大多数情形下, 对于 $X(t)$ 的独立观测值 $\{X_i(t), 1 \leq i \leq N\}$, 由于各种因素的影响, (2.1) 不是严格地成立, 而且参数 $\{\beta_1, \dots, \beta_m\}$ 也可能在观测过程中发生变化, 所以对于 $X_i(t)$, 我们假定其满足下面等式,

$$L_{mi} X_i(t) - f_i(t) = \epsilon_i(t), \quad (2.2)$$

其中 $\epsilon_i(t)$ 为 $X_i(t)$ 经等式左边算子作用后得到的余项, 微分算子 L_{mi} 定义为

$$L_{mi} = D^m - \beta_{1i} D^{m-1} - \cdots - \beta_{(m-1)i} D - \beta_{mi}.$$

本文的目标包括 (1) 检验参数 β_{ji} ($1 \leq j \leq m$) 在观测范围 $1 \leq i \leq N$ 是否存在变化; (2) 若参数在某点 $1 < n_0 < N$ 存在变化, 给出变化时刻 n_0 的估计. 本文考虑参数仅存在一个变点的情形, 对于多个

结构变化的情形, 即多变点的情形, 可以在单个变点推断基础上得到, 这里就不再考虑. 为检验是否存在变点, 我们给出下面的原假设,

$$H_0: \quad \beta_{j1} = \beta_{j2} = \cdots = \beta_{jN} = \beta_j, \quad 1 \leq j \leq m.$$

而对立假设为 $\exists 1 < n_0 < N$ 满足

$$H_1: \quad \beta_{11} = \cdots = \beta_{1n_0} \neq \beta_{1(n_0+1)} = \cdots = \beta_{1N},$$

$$\vdots$$

$$\text{或 } \beta_{m1} = \cdots = \beta_{mn_0} \neq \beta_{m(n_0+1)} = \cdots = \beta_{mN}.$$

为检验上述的假设, 误差项应满足一定的条件, 这在 (A1) 中给出. 为表述方便, 下文中的定积分和重积分的积分区域都省略, 其中定积分积分区域是 $[0, 1]$, 重积分积分区域为 $[0, 1] \times [0, 1]$.

(A1) 误差项 $\{\epsilon_i(t), 1 \leq i \leq N\}$ 为独立同分布在 $L^2([0, 1])$ 中取值的随机变量, 均值为零, 协方差函数平方可积, 具有有限的四阶矩, 即:

$$E\epsilon_i(t) = 0, \quad E\epsilon_i(s)\epsilon_i(t) = c(s, t), \quad E\left(\int \epsilon_i^2(t)dt\right)^2 < \infty.$$

由 (A1) 和 Mercer 引理 (参见 [6] 中的引理 1.3), 我们有

$$c(s, t) = \sum_{j=1}^{\infty} \lambda_j \psi_j(s) \psi_j(t), \quad s, t \in [0, 1],$$

其中 $\{\psi_j(t)\}_{j=1}^{\infty}$ 为 $L^2([0, 1])$ 中的一组单位正交基, $\{\lambda_j\}_{j=1}^{\infty}$ 为一列单调下降的正数. 再由二阶零均值过程的 Karhunen-Loeve 分解 (参见 [6] 中的定理 1.5),

$$\epsilon_i(t) = \sum_{j=1}^{\infty} \eta_{ij} \psi_j(t), \quad t \in [0, 1],$$

其中对不同的 $1 \leq i \leq N$, $\{\eta_{ij}, j = 1, 2, \dots\}$ 为独立序列, 且满足

$$E\eta_{ij} = 0, \quad E\eta_{ij_1}\eta_{ij_2} = \lambda_{j_1} \delta_{j_1 j_2}, \quad j_1, j_2 \in N,$$

这里若 $j_1 = j_2$, 则 $\delta_{j_1 j_2} = 1$, 否则 $\delta_{j_1 j_2} = 0$. $\{\lambda_j\}_{j=1}^{\infty}$ 和 $\{\psi_j(t)\}_{j=1}^{\infty}$ 即所谓 $\epsilon_i(t)$ 的特征根和特征函数, η_{ij} 为 $\psi_j(t)$ 的得分. 为了得到特征根、特征函数以及得分的估计, 我们需要下面的条件 (A2).

(A2) 对给定的正整数 d , 成立

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1}.$$

为表述方便, 引入下面记号,

$$Y_i(t) = D^m X_i(t) - f(t), \quad \beta = (\beta_1, \beta_2, \dots, \beta_m)^T,$$

$$Z_i(t) = (D^{m-1} X_i(t), D^{m-2} X_i(t), \dots, DX_i(t), X_i(t))^T.$$

由此 $Y_i(t) = Z_i^T(t)\beta + \epsilon_i(t)$. 类似地, 定义

$$Y(t) = (Y_1(t), Y_2(t), \dots, Y_N(t))^T,$$

$$Z(t) = (Z_1(t), Z_2(t), \dots, Z_N(t))^T, \quad \epsilon_t = (\epsilon_1(t), \epsilon_2(t), \dots, \epsilon_N(t))^T,$$

于是

$$Y(t) = Z(t)\beta + \epsilon(t), \quad t \in [0, 1].$$

为了得到原假设下统计量的分布, 除了误差项 $\epsilon(t)$ 之外, 解释变量 $Z(t)$ 也应该满足一定的条件, 这在下面的条件 (A3) 中给出.

(A3) $\frac{1}{N} \sum_{i=1}^N Z_i(s)(Z_i(t))^T$ 在 $L^2([0, 1])$ 中以概率收敛到平方可积矩阵

$$Q(t) \triangleq EZ_1(s)(Z_1(t))^T.$$

这里的平方可积指矩阵中的每一个元素均平方可积, 这是 Merce 引理的条件之一. 为得到未知参数的估计, 定义下面的积分误差平方和,

$$S(\beta) = \int \epsilon^T(t)\epsilon(t)dt = \int (Y(t) - Z(t)\beta)^T(Y(t) - Z(t)\beta)dt.$$

和通常最小二乘法一样, 我们通过最小化 $S(\beta)$ 得到未知参数 β 的估计. 通过简单的运算可得

$$\hat{\beta} = \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t)Y_i(t)dt. \quad (2.3)$$

由此第 i 次观测的残差为

$$\begin{aligned} e_i(t) &= Y_i(t) - Z_i^T(t)\hat{\beta} = Y_i(t) - Z_i^T(t) \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t)Y_i(t)dt \\ &= \epsilon_i(t) - Z_i^T(t) \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t)\epsilon_i(t)dt. \end{aligned} \quad (2.4)$$

定义 $\{\epsilon_i(t), i = 1, 2, \dots, N\}$ 的样本自协方差函数为

$$\hat{c}(s, t) = \frac{1}{N} \sum_{i=1}^N \{\epsilon_i(s) - \bar{e}(s)\} \{\epsilon_i(t) - \bar{e}(t)\},$$

其中 $\bar{e}(t) = \frac{1}{N} \sum_{i=1}^N \epsilon_i(t)$. 对应的样本特征根和样本特征向量 $\{\hat{\lambda}_j, \hat{\psi}_j(t)\}$ 定义为

$$\int \hat{c}(s, t) \hat{\psi}_j(t) dt = \hat{\lambda}_j \hat{\psi}_j(s).$$

定义经验得分为 $\hat{\eta}_{ij} = \int e_i(t) \hat{\psi}_j(t) dt$. 在原假设的条件下, 当样本容量 N 趋于无穷时, 残差 $e_i(t)$ 在 $L^2([0, 1])$ 中会以概率收敛到 $\epsilon_i(t)$, 因此我们利用 $\{\hat{\lambda}_j, \hat{\psi}_j(t), \hat{\eta}_{ij}, i, j = 1, 2, \dots\}$ 估计 $\{\lambda_j, \psi_j(t), \eta_{ij}, i, j = 1, 2, \dots\}$, 这些估计的收敛性在引理 1 中给出.

下面我们考虑在存在变点的情况下, 变点估计及检验的性质. 在对立假设 H_1 下, 设对 $1 \leq i \leq n_0$,

$$L_m^{(1)} X_i(t) - f^{(1)}(t) = \epsilon_i(t),$$

其中

$$L_m^{(1)} = D^m - \beta_1^{(1)} D^{m-1} - \dots - \beta_{m-1}^{(1)} D - \beta_m^{(1)}.$$

对 $n_0 + 1 \leq i \leq N$,

$$L_m^{(2)} X_i(t) - f^{(2)}(t) = \epsilon_i(t),$$

其中

$$L_m^{(2)} = D^m - \beta_1^{(2)} D^{m-1} - \dots - \beta_{m-1}^{(2)} D - \beta_m^{(2)}.$$

未知参数为

$$\beta^{(1)} \triangleq (\beta_1^{(1)}, \dots, \beta_m^{(1)})^T, \quad \beta^{(2)} \triangleq (\beta_1^{(2)}, \dots, \beta_m^{(2)})^T.$$

对 $1 \leq n \leq N$, 我们引入下面的记号

$$\begin{aligned} \Delta &= \beta^{(2)} - \beta^{(1)}, \quad \mu_1(t) = EZ_1(t), \quad \mu_2(t) = EZ_N(t), \\ D_1(s, t) &= E(Z_1(s)(Z_1(t))^T), \quad D_2(s, t) = E(Z_N(s)(Z_N(t))^T), \\ Q_1 &= E \int Z_1(t)(Z_1(t))^T dt, \quad Q_2 = E \int Z_N(t)(Z_N(t))^T dt, \\ U_n^{(1)} &= \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \left(\sum_{i=1}^n \int Z_i(t)(Z_i(t))^T dt \right), \\ U_n^{(2)} &= \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \left(\sum_{i=n+1}^N \int Z_i(t)(Z_i(t))^T dt \right), \\ V_N &= \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t) \epsilon_i(t) dt. \end{aligned}$$

首先我们注意到当 $N \rightarrow \infty$ 时,

$$\begin{aligned} U_{n_0}^{(1)} &\xrightarrow{p} \left(I + \frac{1-\theta}{\theta} Q_1^{-1} Q_2 \right)^{-1} \triangleq U^{(1)}, \\ U_{n_0}^{(2)} &\xrightarrow{p} \left(I + \frac{\theta}{1-\theta} Q_2^{-1} Q_1 \right)^{-1} \triangleq U^{(2)}, \end{aligned}$$

定义

$$\begin{aligned} \tilde{c}(s, t) &= c(s, t) + \theta \Delta^T U^{(2)} \text{Cov}_1 U^{(2)} \Delta + (1-\theta) \Delta^T U^{(1)} \text{Cov}_2 U^{(1)} \Delta \\ &\quad + (1-\theta) \theta \Delta^T (U^{(2)} \mu_1(s) + U^{(1)} \mu_2(s)) (U^{(2)} \mu_1(t) + U^{(1)} \mu_2(t))^T \Delta, \end{aligned}$$

其中 $\text{Cov}_i = D_i(s, t) - \mu_i(s) \mu_i^T(t)$ ($i = 1, 2$) 为 $Z_1(t)$ 和 $Z_N(t)$ 的自协方差函数所组成的矩阵. 由 (A1) 和 (A3), $\tilde{c}(s, t)$ 是对称平方可积的二元函数, 且对 $\forall x(t) \in L^2([0, 1])$ 有

$$\iint \tilde{c}(s, t) x(s) x(t) dt \geq 0,$$

也即 $\tilde{c}(s, t)$ 为协方差函数. 定义 $\tilde{c}(s, t)$ 的特征根和特征函数分别为 $\{\gamma_j, \omega_j(t), j = 1, 2, \dots\}$. 为了得到这些特征量的估计, 我们需要下面的条件.

(A4) 对给定的正整数 d , 成立

$$\gamma_1 > \gamma_2 > \dots > \gamma_d > \gamma_{d+1}.$$

与条件 (A3) 对应, 在对立假设下, 解释变量 $Z_i(t)$ 应满足下面的条件.

(A5) 当 $n_0 \rightarrow \infty$ ($N - n_0 \rightarrow \infty$) 时,

$$\frac{1}{n_0} \sum_{i=1}^{n_0} Z_i(s)(Z_i(t))^T \left(\frac{1}{N - n_0} \sum_{i=n_0+1}^N Z_i(s)(Z_i(t))^T \right)$$

在 $L^2([0, 1] \times [0, 1])$ 上以概率收敛至平方可积矩阵 $D_1(s, t)$ ($D_2(s, t)$).

利用上述记号对 $\hat{\beta}$ 化简有

$$\begin{aligned} \hat{\beta} &= \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \left(\sum_{i=1}^N \int Z_i(t)Y_i(t)dt \right) \\ &= \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \\ &\quad \times \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt b^{(1)} + \sum_{i=n_0+1}^N \int Z_i(Z_i)^T dt \Delta + \sum_{i=1}^N \int Z_i(t)\epsilon_i(t)dt \right) \\ &= \beta^{(1)} + \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=n_0+1}^N \int Z_i(t)(Z_i(t))^T dt \Delta \\ &\quad + \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t)\epsilon_i(t)dt \\ &= \beta^{(1)} + U_{n_0}^{(2)} \Delta + V_N. \end{aligned} \quad (2.5)$$

同理有

$$\begin{aligned} \hat{\beta} &= \beta^{(2)} - \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^{n_0} \int Z_i(Z_i)^T dt \Delta \\ &\quad + \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t)\epsilon_i(t)dt \\ &= \beta^{(2)} - U_{n_0}^{(1)} \Delta + V_N. \end{aligned} \quad (2.6)$$

因此当 $1 \leq i \leq n_0$ 时,

$$e_i(t) = Y_i(t) - Z_i^T \hat{\beta} = \epsilon_i(t) - Z_i^T U_{n_0}^{(2)} \Delta - Z_i^T V_N, \quad (2.7)$$

当 $n_0 + 1 \leq i \leq N$ 时,

$$e_i(t) = Y_i(t) - Z_i^T \hat{\beta} = \epsilon_i(t) + Z_i^T U_{n_0}^{(1)} \Delta - Z_i^T V_N, \quad (2.8)$$

(2.5)–(2.8) 被用来证明在对立假设下检验的一致性.

3 推断的步骤与主要结论

对常微分方程的参数进行统计推断的第一步是判断参数是否是可识别的, 即给定一组观测值, 是否存在着唯一的一组参数取值与之对应. 目前可识别性的研究都是针对常参数常微分方程, 包括线性与非线性两种情形, 见 [7–9]. 时变参数常微分方程的可识别性, 目前文献中还没有涉及. 对于本文所

考虑的常参数线性常微分方程, 可以利用高阶导数方法 (HOD, 见 [8]) 或多时间点方法 (MTS, 见 [9]) 等判断模型参数的可识别性.

首先我们注意到无论参数是否存在变化, 都有 $E\epsilon_i(t) = 0$ ($1 \leq i \leq N$). 若原假设成立, 则残差 $e_i(t)$ 以概率收敛至 $\epsilon_i(t)$, 而在对立假设下 $e_i(t)$ 不再收敛到 $\epsilon_i(t)$, 且在变点前后 $e_i(t)$ 的均值也将不会相同, 所以我们考虑利用下面的累积和统计量来对上述的假设进行检验,

$$\begin{aligned} P_n(t) &= \frac{n(N-n)}{N} \left(\frac{1}{n} \sum_{i=1}^n e_i(t) - \frac{1}{N-n} \sum_{i=n+1}^N e_i(t) \right) \\ &= \sum_{i=1}^n e_i(t) - \frac{n}{N} \sum_{i=1}^N e_i(t). \end{aligned}$$

由于 $P_n(t)$ 是定义于 $[0, 1]$ 上的连续函数, 不能直接用作检验统计量, 我们考虑 $P_n(t)$ 在 $\epsilon_i(t)$ 前 d 个特征函数 $\{\psi_j(t), j = 1, 2, \dots, d\}$ 上的投影, 但 $\{\psi_j(t), j = 1, 2, \dots, d\}$ 是无法直接观测到的, 这里我们利用经验特征函数 $\{\hat{\psi}_j(t), j = 1, 2, \dots, d\}$ 代替 $\{\psi_j(t), j = 1, 2, \dots, d\}$, 由此得投影系数为

$$\int \left(\sum_{i=1}^n e_i(t) - \frac{n}{N} \sum_{i=1}^N e_i(t) \right) \hat{\psi}_j(t) dt = \sum_{i=1}^n \hat{\eta}_{ij}(t) - \frac{n}{N} \sum_{i=1}^N \hat{\eta}_{ij}(t).$$

设 $n = [Nx]$, $0 < x < 1$, 定义下面的统计量

$$T_n(x) = \frac{1}{N} \sum_{j=1}^d \hat{\lambda}_j^{-1} \left(\sum_{i=1}^n \hat{\eta}_{ij} - x \sum_{i=1}^N \hat{\eta}_{ij} \right)^2. \quad (3.1)$$

设 $B_1(\cdot), \dots, B_d(\cdot)$ 为独立的 Brown 桥过程, 下面的定理给出了 $T_n(x)$ 的分布, 由此可以得到检验假设的统计量.

定理 1 设原假设及条件 (A1)–(A3) 成立, 则在 Skorohod 空间 $D[0, 1]$ 中成立

$$T_n(x) \xrightarrow{d} \sum_{j=1}^d B_j^2(x), \quad 0 \leq x \leq 1, \quad (3.2)$$

这里 $\xrightarrow{d}$ 表示 $D[0, 1]$ 中的弱收敛.

由定理 1, 对任何的连续泛函 $M : D[0, 1] \rightarrow \mathbb{R}$, 我们有 $M(T_n) \xrightarrow{d} M\{\sum_{j=1}^d B_j^2(\cdot)\}$. 对变点问题, 文献中最常用的泛函包括最大值泛函^[10], 积分泛函即 Cramer-von Mises 泛函^[4]. 这里我们考察 Cramer-von Mises 泛函, 故有

$$\int T_n(x) dx \xrightarrow{d} \int \sum_{j=1}^d B_j^2(x) dx,$$

等价地,

$$S_{N,d} \triangleq \frac{1}{N^2} \sum_{j=1}^d \hat{\lambda}_j^{-1} \sum_{n=1}^N \left(\sum_{i=1}^n \hat{\eta}_{ij} - \frac{k}{N} \sum_{i=1}^N \hat{\eta}_{ij} \right)^2 \xrightarrow{d} \int \sum_{j=1}^d B_j^2(x) dx. \quad (3.3)$$

当检验统计量 $S_{N,d}$ 超过所给定的临界值时, 就拒绝原假设 H_0 , 认为参数存在变化, 这里的临界值由 (3.3) 右边随机变量 $\int \sum_{j=1}^d B_j^2(x) dx$ 所确定, 其中 Kiefer^[11] 和 Horváth 等^[12] 讨论了临界值的计算方法并给出了一些常见的临界值.

为了实施上述检验, 必须确定其中所涉及的特征函数的个数 d . 对此问题, 文献中已给出了几种选择方法, 其中应用最广泛的方法基于累积方差比 (CPV, 参见 [4]). 对前 d 个函数型主成分, 累积方差比定义为

$$\text{CPV}(d) = \frac{\sum_{k=1}^d \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}.$$

对于事前给定的值, 譬如 85%, 选择使得 $\text{CPV}(d)$ 大于 85% 的最小的 d . Cattell^[13] 提出了一种基于图形的选择方法 (scree test). 首先画出特征值序列图, 在序列图下降平缓处的最小的 d 作为特征函数个数的最终选择. Biau 等^[14] 在研究函数型数据 K 近邻分类问题中, 提出了一种基于数据集分割的选择方法. 将数据分为训练集和验证集, 选择使得错误分类概率达到最小的 d 作为特征函数个数的选择. Yao 等^[15] 采用了拟 AIC 和 one-curve-leave-out 交叉验证的方法选择 d . 在本文的数据分析部分, 我们采用了累积方差比的方法选择特征函数的个数.

下面考察在对立假设下统计量 $S_{N,d}$ 的大样本性质. 我们将证明若参数存在变点, 则 $S_{N,d}$ 将以概率收敛至无穷大. 设真实的变点为 $n_0 = [N\theta]$, 其中 $0 < \theta < 1$. 在引理 2 中, 我们证明了, 除一个符号函数外, 经验特征根和特征函数 $\{\hat{\lambda}_j, \hat{\psi}_j(t), j = 1, 2, \dots\}$ 以概率收敛至 $\tilde{c}(s, t)$ 的特征根和特征函数 $\{\gamma_j, \omega_j(t), j = 1, 2, \dots\}$. 对 $1 \leq j \leq d$, 定义

$$g_j(x) = \begin{cases} x(\theta - 1)\Delta^T \left(U^{(2)} \int \mu_1(t)\omega_j(t)dt + U^{(1)} \int \mu_2(t)\omega_j(t)dt \right), & 0 \leq x \leq \theta, \\ \theta(1 - x)\Delta^T \left(U^{(2)} \int \mu_1(t)\omega_j(t)dt + U^{(1)} \int \mu_2(t)\omega_j(t)dt \right), & \theta < x \leq 1. \end{cases}$$

于是我们有下面的定理 2 成立.

定理 2 给定条件 (A1), (A4) 和 (A5), 在对立假设 H_1 下我们有

$$\sup_{0 < x < 1} \left| \frac{1}{N} T_N(x) - \sum_{j=1}^d \gamma_j^{-1} g_j^2(x) \right| = o_p(1).$$

由定理 2 和 (3.3) 式,

$$\frac{1}{N} S_{N,d} \xrightarrow{p} \sum_{j=1}^d \gamma_j^{-1} \int_0^1 g_j^2(x) dx.$$

故在对立假设 H_1 下, 若 $\exists 1 \leq j \leq d$ 满足 $g_j(x) \neq 0$, 则 $S_{N,d} \xrightarrow{p} \infty$, 即检验是一致的. 反之, 若对所有 $1 \leq j \leq d$, 有 $g_j(x) = 0$, 则检验就不是一致的. 而 $g_j(x) = 0$ 等价于

$$U^{(2)} \int \mu_1(t)\omega_j(t)dt + U^{(1)} \int \mu_2(t)\omega_j(t)dt = 0.$$

由 (2.7) 和 (2.8), 上式等价于

$$\lim_{N \rightarrow \infty} \int (Ee_1 - Ee_N)\omega_j = 0.$$

由此可知, 检验的一致性依赖于变点前后残差的渐近均值之差: 若其正交于前 d 个特征函数 $\{\omega_j(t), j = 1, \dots, d\}$ 所生成的空间, 则检验就不是一致的; 反之检验就是一致的.

下面考察变点 $n_0 = [N\theta]$ 的估计. 定义

$$D_N(x) = \frac{1}{N} T_N(x), \quad D(x) = \sum_{j=1}^d \gamma_j^{-1} g_j^2(x),$$

则 θ 的估计定义为

$$\hat{\theta} = \inf_{0 < x < 1} \left\{ x, D_N(x) = \max_{1 < [Nx] < N} D_N(x) \right\}.$$

对上述定义的 $\hat{\theta}$, 有下面的结论成立.

定理 3 给定条件 (A1), (A4) 和 (A5), 在对立假设 H_1 下有 $\hat{\theta} \xrightarrow{P} \theta$.

证明 定义

$$A_j = \Delta^T \left(U^{(2)} \int \mu_1(t) w_j(t) dt + U^{(1)} \int \mu_2(t) w_j(t) dt \right),$$

于是

$$g_j(x) = \begin{cases} x(\theta - 1)A_j, & 0 \leq x \leq \theta, \\ \theta(1 - x)A_j, & \theta < x \leq 1. \end{cases}$$

由此

$$D(x) = \begin{cases} (\theta - 1)^2 x^2 \sum_{j=1}^d \gamma_j^{-1} A_j^2, & 0 \leq x \leq \theta, \\ \theta^2 (1 - x)^2 \sum_{j=1}^d \gamma_j^{-1} A_j^2, & \theta < x \leq 1. \end{cases}$$

易知 $x = \theta$ 为 $D(x)$ 的唯一的最大值点. 由定理 2,

$$\sup_{0 < x < 1} |D_N(x) - D(x)| \xrightarrow{P} 0,$$

故 $D_N(x)$ 的最大点以概率收敛到 $D(x)$ 的最大点, 命题得证.

4 主要定理的证明

引理 1 设条件 (A1)–(A3) 成立, 定义 $\hat{c} = \text{sgn}\{\int \hat{\psi}(t)\psi(t)dt\}$, 则在原假设 H_0 下对 $j \geq 1$, 我们有

$$\hat{\lambda}_j \xrightarrow{P} \lambda_j, \quad \int (\hat{\psi}(t) - \hat{c}\psi(t))^2 dt \xrightarrow{P} 0. \quad (4.1)$$

证明 定理的证明建立在 [6] 中的引理 4.2 和 4.3 之上. 利用引理 4.2 和 4.3 可以把特征根和特征函数收敛性的考察转化为对协方差算子收敛性的考察. 由此仅需要证明下面的结论成立即可,

$$\iint (\hat{c}(s, t) - c(s, t))^2 ds dt \xrightarrow{P} 0,$$

其中

$$\hat{c}(s, t) = \frac{1}{N} \sum_{i=1}^N (e_i(s) - \bar{e}(s))(e_i(t) - \bar{e}(t)), \quad \bar{e}(t) = \frac{1}{N} \sum_{i=1}^N e_i(t).$$

对向量 V_N , 由 $L^2([0, 1])$ 中独立同分布随机变量的极限定理 (参见 [6, 定理 2.4 和 2.7])

$$V_N = \left(\sum_{i=1}^N \int Z_i(t)(Z_i(t))^T dt \right)^{-1} \sum_{i=1}^N \int Z_i(t)\epsilon_i(t) dt = O\left(\frac{1}{\sqrt{N}}\right),$$

再由 (2.4),

$$e_i(t) = \epsilon_i(t) - Z_i^T(t)V_N, \quad \bar{e}(t) = \bar{\epsilon}(t) - \frac{1}{N} \sum_{i=1}^N Z_i^T(t)V_N, \quad (4.2)$$

所以

$$\begin{aligned} \hat{c}(s, t) &= \frac{1}{N} \sum_{i=1}^N (\epsilon_i(s) - \bar{\epsilon}(s))(\epsilon_i(t) - \bar{\epsilon}(t)) - \frac{1}{N} \sum_{i=1}^N (\epsilon_i(s) - \bar{\epsilon}(s))(Z_i(t) - \bar{Z}(t))^T V_N \\ &\quad - \frac{1}{N} \sum_{i=1}^N (\epsilon_i(t) - \bar{\epsilon}(t))(Z_i(s) - \bar{Z}(s))^T V_N + \frac{1}{N} \sum_{i=1}^N V_N^T (Z_i(s) - \bar{Z}(s))(Z_i(t) - \bar{Z}(t))^T V_N. \end{aligned}$$

上式右边第一项在 $L^2([0, 1] \times [0, 1])$ 中以概率收敛至 $c(s, t)$, 后面三项在 $L^2([0, 1] \times [0, 1])$ 中以概率收敛至零, 结论得证.

引理 2 给定条件 (A1), (A4) 和 (A5), 在对立假设 H_1 下我们有

$$\iint (\hat{c}(s, t) - \tilde{c}(s, t))^2 ds dt \xrightarrow{P} 0.$$

证明 由 (2.7) 和 (2.8),

$$\begin{aligned} \bar{e}(t) &= \frac{1}{N} \left\{ \sum_{i=1}^{n_0} e_i(t) + \sum_{i=n_0+1}^N e_i(t) \right\} \\ &= \frac{1}{N} \left\{ \sum_{i=1}^N \epsilon_i(t) - \left(\sum_{i=1}^{n_0} Z_i(t) \right)^T U_{n_0}^{(2)} \Delta + \left(\sum_{i=n_0+1}^N Z_i(t) \right)^T U_{n_0}^{(1)} \Delta - \left(\sum_{i=1}^N Z_i(t) \right)^T V_N \right\}, \end{aligned}$$

因此

$$\begin{aligned} \hat{c}(s, t) &= \frac{1}{N} \sum_{i=1}^{n_0} (e_i(t) - \bar{e}(t))(e_i(s) - \bar{e}(s)) + \frac{1}{N} \sum_{i=n_0+1}^N (e_i(t) - \bar{e}(t))(e_i(s) - \bar{e}(s)) \\ &= \frac{1}{N} \sum_{i=1}^{n_0} \left\{ \epsilon_i(t) - \bar{\epsilon}(t) - \left(Z_i(t) - \frac{1}{N} \sum_{i=1}^{n_0} Z_i(t) \right)^T U_{n_0}^{(2)} \Delta \right. \\ &\quad \left. - \frac{1}{N} \sum_{i=n_0+1}^N Z_i^T(t) U_{n_0}^{(1)} \Delta - (Z_i(t) - \bar{Z}(t))^T V_N \right\} \\ &\quad \times \left\{ \epsilon_i(s) - \bar{\epsilon}(s) - \left(Z_i(s) - \frac{1}{N} \sum_{i=1}^{n_0} Z_i(s) \right)^T U_{n_0}^{(2)} \Delta \right. \\ &\quad \left. - \frac{1}{N} \sum_{i=n_0+1}^N Z_i^T(s) U_{n_0}^{(1)} \Delta - (Z_i(s) - \bar{Z}(s))^T V_N \right\} \\ &\quad + \frac{1}{N} \sum_{i=n_0+1}^N \left\{ \epsilon_i(t) - \bar{\epsilon}(t) + \left(Z_i^*(t) - \frac{1}{N} \sum_{i=n_0+1}^N Z_i(t) \right)^T U_{n_0}^{(1)} \Delta \right. \\ &\quad \left. - \frac{1}{N} \sum_{i=1}^{n_0} Z_i^T(t) U_{n_0}^{(2)} \Delta - (Z_i(t) - \bar{Z}(t))^T V_N \right\} \\ &\quad \times \left\{ \epsilon_i(s) - \bar{\epsilon}(s) + \left(Z_i(s) - \frac{1}{N} \sum_{i=n_0+1}^N Z_i(s) \right)^T U_{n_0}^{(1)} \Delta \right. \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{N} \sum_{i=1}^{n_0} Z_i^T(s) U_{n_0}^{(2)} \Delta - (Z_i(s) - \bar{Z}(s))^T V_N \Big\} \\
& \triangleq A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9 + o_p(1),
\end{aligned}$$

其中

$$\begin{aligned}
A_1 &= \frac{1}{N} \sum_{i=1}^N (\epsilon_i(t) - \bar{\epsilon}_i(t)) (\epsilon_i(s) - \bar{\epsilon}_i(s)), \\
A_2 &= \frac{1}{N} \Delta^T (U_{n_0}^{(2)})^T \sum_{i=1}^{n_0} \left(Z_i(t) - \frac{1}{N} \sum_{i=1}^{n_0} Z_i(t) \right) \left(Z_i(s) - \frac{1}{N} \sum_{i=1}^{n_0} Z_i(s) \right)^T U_{n_0}^{(2)} \Delta, \\
A_3 &= \frac{n_0}{N} \frac{1}{N^2} \Delta^T (U_{n_0}^{(1)})^T \sum_{i=n_0+1}^N Z_i(t) \sum_{i=n_0+1}^N Z_i^T(s) U_{n_0}^{(1)} \Delta, \\
A_4 &= \frac{1}{N^2} \Delta^T (U_{n_0}^{(1)})^T \sum_{i=n_0+1}^N Z_i(t) \sum_{i=1}^{n_0} \left(Z_i(s) - \frac{1}{N} \sum_{i=1}^{n_0} Z_i(s) \right)^T U_{n_0}^{(2)} \Delta, \\
A_5 &= \frac{1}{N^2} \Delta^T (U_{n_0}^{(1)})^T \sum_{i=n_0+1}^N Z_i(s) \sum_{i=1}^{n_0} \left(Z_i(t) - \frac{1}{N} \sum_{i=1}^{n_0} Z_i(t) \right)^T U_{n_0}^{(2)} \Delta, \\
A_6 &= \frac{1}{N} \Delta^T (U_{n_0}^{(1)})^T \sum_{i=n_0+1}^N \left(Z_i(t) - \frac{1}{N} \sum_{i=n_0+1}^N Z_i^*(t) \right) \left(Z_i(s) - \frac{1}{N} \sum_{i=n_0+1}^N Z_i(s) \right)^T U_{n_0}^{(1)} \Delta, \\
A_7 &= \frac{N-n_0}{N} \frac{1}{N^2} \Delta^T (U_{n_0}^{(2)})^T \sum_{i=1}^{n_0} Z_i(t) \sum_{i=1}^{n_0} Z_i^T(s) U_{n_0}^{(2)} \Delta, \\
A_8 &= \frac{1}{N^2} \Delta^T (U_{n_0}^{(2)})^T \sum_{i=1}^{n_0} Z_i(t) \sum_{i=n_0+1}^N \left(Z_i(s) - \frac{1}{N} \sum_{i=n_0+1}^N Z_i(s) \right)^T U_{n_0}^{(1)} \Delta, \\
A_9 &= \frac{1}{N^2} \Delta^T (U_{n_0}^{(2)})^T \sum_{i=1}^{n_0} Z_i(s) \sum_{i=n_0+1}^N \left(Z_i(t) - \frac{1}{N} \sum_{i=n_0+1}^N Z_i(t) \right)^T U_{n_0}^{(1)} \Delta + o_p(1).
\end{aligned}$$

由大数定理,

$$\begin{aligned}
A_1 &\xrightarrow{p} c(s, t), \\
A_2 &\xrightarrow{p} \theta \Delta^T (U^{(2)})^T (D_1 - \theta(2 - \theta) \mu_1^T(t) \mu_1(s)) U^{(2)} \Delta, \\
A_3 &\xrightarrow{p} \theta(1 - \theta)^2 \Delta^T (U^{(1)})^T \mu_2^T(t) \mu_2(s) U^{(1)} \Delta, \\
A_4 &\xrightarrow{p} \theta(1 - \theta)^2 \Delta^T (U^{(2)})^T \mu_2^T(t) \mu_1(s) U^{(1)} \Delta, \\
A_5 &\xrightarrow{p} \theta(1 - \theta)^2 \Delta^T (U^{(2)})^T \mu_1^T(t) \mu_2(s) U^{(1)} \Delta, \\
A_6 &\xrightarrow{p} (1 - \theta) \Delta^T (U^{(1)})^T (D_2 - (1 - \theta^2) \mu_2^T(t) \mu_2(s)) U^{(1)} \Delta, \\
A_7 &\xrightarrow{p} \theta^2(1 - \theta) \Delta^T (U^{(2)})^T \mu_1^T(t) \mu_1(s) U^{(2)} \Delta, \\
A_8 &\xrightarrow{p} \theta^2(1 - \theta) \Delta^T (U^{(1)})^T \mu_1^T(t) \mu_2(s) U^{(2)} \Delta, \\
A_9 &\xrightarrow{p} \theta^2(1 - \theta) \Delta^T (U^{(1)})^T \mu_2^T(t) \mu_1(s) U^{(2)} \Delta.
\end{aligned}$$

把这些表达式代入 $\hat{c}(t, s)$ 有

$$\hat{c}(s, t) \xrightarrow{p} c(s, t) + \theta \Delta^T U^{(2)} \text{Cov}_1 U^{(2)} \Delta + (1 - \theta) \Delta^T U^{(1)} \text{Cov}_2 U^{(1)} \Delta$$

$$+(1-\theta)\theta\Delta^T(U^{(2)}\mu_1(s)+U^{(1)}\mu_2(s))(U^{(2)}\mu_1(t)+U^{(1)}\mu_2(t))^T\Delta \\ \triangleq \tilde{c}(s,t).$$

注 由引理 2, $\hat{c}(s,t)$ 在 $L^2([0,1] \times [0,1])$ 中以概率收敛至 $\tilde{c}(s,t)$, 所以和引理 1 同样道理, 除以一个符号函数外, $\hat{c}(s,t)$ 的特征根和特征函数 $\{\hat{\lambda}_j, \hat{\psi}_j(t), j=1,2,\dots\}$ 以概率收敛至 $\tilde{c}(s,t)$ 对应的特征根和特征函数 $\{\gamma_j, \omega_j(t), j=1,2,\dots\}$.

定理 1 的证明 在定理 1 和 2 的证明中, 因为所涉及的取整运算不影响结论, 所以为表述方便, 不妨假定 $[nx] = nx$. 设 $\eta_i = (\eta_{i1}, \eta_{i2}, \dots, \eta_{id})^T$. 由于 $\{\epsilon_i(t), i=1,2,\dots\}$ 是 $L^2([0,1])$ 中独立同分布的随机变量且 $E\epsilon_i(t) = 0$, 故 $\{\eta_i, i=1,2,\dots\}$ 为独立同分布实值 d 维随机变量且满足

$$E\eta_i = 0, \quad \text{Cov}\eta_i \triangleq \sum_d = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d),$$

由此及泛函中心极限定理, 在 Skorohod 空间 $D^d[0,1]$ 中我们有

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{Nx} \eta_i \xrightarrow{d} W_d(x), \quad 0 \leq x \leq 1,$$

其中 W_d 为 d 维标准 Brown 运动, 由此收敛性我们可得

$$\tilde{D}_N(x) \triangleq \frac{1}{N} \left(\sum_{i=1}^{Nx} \eta_i - x \sum_{i=1}^N \eta_i \right)^T \sum_d^{-1} \left(\sum_{i=1}^{Nx} \eta_i - x \sum_{i=1}^N \eta_i \right) \xrightarrow{d} \sum_{i=1}^d B_i^2(x),$$

其中 $\{B_i(x), i=1,\dots,d\}$ 为 $[0,1]$ 上独立的 Brown 桥. 为证明定理结论, 我们只需证明 $D_N(x)$ 和 $\tilde{D}_N(x)$ 具有相同的弱极限即可, 为此只需证明下面两个等式成立,

$$\sum_d \xrightarrow{p} \sum_d, \quad (4.3)$$

$$\frac{1}{\sqrt{N}} \left\| \hat{c}_j \left(\sum_{i=1}^{Nx} \eta_i - x \sum_{i=1}^N \eta_i \right) - \left(\sum_{i=1}^{Nx} \hat{\eta}_i - x \sum_{i=1}^N \hat{\eta}_i \right) \right\|_d = o(1), \quad (4.4)$$

这里 $\|\cdot\|_d$ 表示 d 维 Euclid 空间的范数. 由 (4.1) 显然可得 (4.3). 对 (4.4),

$$\begin{aligned} & \frac{1}{\sqrt{N}} \left| \hat{c}_j \left(\sum_{i=1}^{Nx} \eta_{ij} - x \sum_{i=1}^N \eta_{ij} \right) - \left(\sum_{i=1}^{Nx} \hat{\eta}_{ij} - x \sum_{i=1}^N \hat{\eta}_{ij} \right) \right| \\ &= \frac{1}{\sqrt{N}} \left| \sum_{i=1}^{Nx} (\hat{c}_j \eta_{ij} - \hat{\eta}_{ij}) - x \sum_{i=1}^N (\hat{c}_j \eta_{ij} - \hat{\eta}_{ij}) \right| \\ &= \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{Nx} (\langle \epsilon_i, \hat{c}_j \psi_j \rangle - \langle \epsilon_i, \hat{\psi}_j \rangle + \langle \epsilon_i, \hat{\psi}_j \rangle - \langle e_{ij}, \hat{\psi}_j \rangle) \right. \\ & \quad \left. - x \frac{1}{\sqrt{N}} \sum_{i=1}^N (\langle \epsilon_i, \hat{c}_j \psi_j \rangle - \langle \epsilon_i, \hat{\psi}_j \rangle + \langle \epsilon_i, \hat{\psi}_j \rangle - \langle e_{ij}, \hat{\psi}_j \rangle) \right| \\ &\leq \frac{1}{\sqrt{N}} \left| \left\langle \sum_{i=1}^{Nx} \epsilon_i, \hat{c}_j \psi_j - \hat{\psi}_j \right\rangle - x \left\langle \sum_{i=1}^N \epsilon_i, \hat{c}_j \psi_j - \hat{\psi}_j \right\rangle \right| \\ & \quad + \frac{1}{\sqrt{N}} \left| \left\langle \sum_{i=1}^{Nx} (\epsilon_i - e_i), \hat{\psi}_j \right\rangle - x \left\langle \sum_{i=1}^N (\epsilon_i - e_i), \hat{\psi}_j \right\rangle \right| \end{aligned}$$

$$= I + II,$$

其中

$$\begin{aligned} I &\leq 2 \sup_{0 < x < 1} \left| \int \left(\hat{c}_j \psi_j - \hat{\psi}_j \right) \left(N^{-1/2} \sum_{i=1}^{Nx} \epsilon_i \right) dt \right| \\ &\leq 2 \sup_{0 < x < 1} \left(\int \left(\hat{c}_j \psi_j - \hat{\psi}_j \right)^2 dt \right)^{1/2} \left(\int \left(N^{-1/2} \sum_{i=1}^{Nx} \epsilon_i \right)^2 dt \right)^{1/2}. \end{aligned}$$

由引理 1, 上式中第一项以概率收敛至零, 而第二项中的被积函数 $N^{-1/2} \sum_{i=1}^{Nx} \epsilon_i(t)$ 在

$$D\{[0, 1], L^2([0, 1])\}$$

弱收敛, 故以概率有界, 即有

$$\sup_{0 < x < 1} \left[\int \left\{ N^{-1/2} \sum_{i=1}^{Nx} \epsilon_i(t) dt \right\}^2 \right] = O_p(1), \quad (4.5)$$

因此 I 以概率收敛至零. 对于 II , 由 (4.2),

$$\begin{aligned} II &= \frac{1}{\sqrt{N}} \left| \int \sum_{i=1}^{Nx} V_N^T Z_i^*(t) \hat{\psi}_j(t) dt - x \int \sum_{i=1}^N V_N^T Z_i^*(t) \hat{\psi}_j(t) dt \right| \\ &= \frac{1}{\sqrt{N}} \left| V_N^T \left(\int \sum_{i=1}^{Nx} (Z_i^*(t) - \mu(t)) \hat{\psi}_j(t) dt - x \int \sum_{i=1}^N (Z_i^*(t) - \mu(t)) \hat{\psi}_j(t) dt \right) \right| \\ &\leq 2 \sup_{0 < x < 1} \left| V_N^T \int N^{-1/2} \sum_{i=1}^{Nx} (Z_i^*(t) - \mu(t)) \hat{\psi}_j(t) dt \right| \\ &\leq 2 |V_N^T| \sup_{0 < x < 1} \left[\int \left(N^{-1/2} \sum_{i=1}^{Nx} (Z_i^*(t) - \mu(t)) \right)^2 dt \right]^{1/2}, \end{aligned}$$

其中上式第一项 $V_N^T = O(\frac{1}{\sqrt{N}})$, 和 (4.5) 相同的理由知第二项以概率有界, 故 II 以概率收敛至零, 由此定理成立.

定理 2 的证明 记

$$\hat{g}_j(x) = \frac{1}{N} \left(\sum_{i=1}^{Nx} \hat{\eta}_{ij} - x \sum_{i=1}^N \hat{\eta}_{ij} \right),$$

由 (2.7) 和 (2.8), 当 $1 < i \leq n_0$ 时,

$$\begin{aligned} \hat{\eta}_{ij} &= \int e_i(t) \hat{\psi}_j(t) dt \\ &= \int \epsilon_i(t) \hat{\psi}_j(t) dt - \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(2)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N. \end{aligned}$$

当 $n_0 + 1 \leq i < N$ 时,

$$\begin{aligned} \hat{\eta}_{ij} &= \int e_i(t) \hat{\psi}_j(t) dt \\ &= \int \epsilon_i(t) \hat{\psi}_j(t) dt + \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(1)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N. \end{aligned}$$

因此对于 $0 < x < \theta$,

$$\begin{aligned}\hat{g}_j(x) = & \frac{1}{N} \left\{ \sum_{i=1}^{Nx} \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt - \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(2)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \right. \\ & - x \sum_{i=1}^{N\theta} \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt - \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(2)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \\ & \left. - x \sum_{i=n\theta+1}^N \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt + \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(1)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \right\},\end{aligned}$$

而在 $L^2([0, 1])$ 中,

$$\begin{aligned}\frac{1}{N} \sum_{i=1}^{Nx} \int \epsilon_i(t) \hat{\psi}_j(t) dt & \xrightarrow{p} 0, \quad \frac{1}{N} \sum_{i=1}^{N\theta} \int \epsilon_i(t) \hat{\psi}_j(t) dt \xrightarrow{p} 0, \\ \frac{1}{N} \sum_{i=n\theta+1}^N \int \epsilon_i(t) \hat{\psi}_j(t) dt & \xrightarrow{p} 0, \quad \frac{1}{N} \sum_{i=1}^{Nx} \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \xrightarrow{p} 0, \\ \frac{1}{N} \sum_{i=1}^{N\theta} \int Z_i^T(t) \hat{\psi}_j(t) dt V_N & \xrightarrow{p} 0, \quad \frac{1}{N} \sum_{i=N\theta+1}^N \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \xrightarrow{p} 0.\end{aligned}$$

因此由引理 2,

$$\hat{g}_j(x) \xrightarrow{p} -x(1-\theta)\Delta^T \left(U^{(2)} \int \mu_1(t) \omega_j(t) dt + U^{(1)} \int \mu_2(t) \omega_j(t) dt \right).$$

同理对于 $\theta < x < 1$,

$$\begin{aligned}\hat{g}_j(x) = & \frac{1}{N} \left\{ \sum_{i=1}^{N\theta} \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt - \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(2)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \right. \\ & + \sum_{i=N\theta+1}^{Nx} \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt + \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(1)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \\ & - x \sum_{i=1}^{N\theta} \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt - \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(2)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \\ & \left. - x \sum_{i=n\theta+1}^N \left(\int \epsilon_i(t) \hat{\psi}_j(t) dt + \int Z_i^T(t) \hat{\psi}_j(t) dt U_{n_0}^{(1)} \Delta - \int Z_i^T(t) \hat{\psi}_j(t) dt V_N \right) \right\} \\ & \xrightarrow{p} \theta(x-1)\Delta^T \left(U^{(2)} \int \mu_1(t) \omega_j(t) dt + U^{(1)} \int \mu_2(t) \omega_j(t) dt \right),\end{aligned}$$

由此及 $T_N(x)$ 的定义即得定理结论.

5 数值分析

5.1 模拟数据分析

考虑用以描述简谐振动的常系数非齐次二阶常微分方程

$$L_2 X(t) - 17 \sin(\phi t) = \epsilon(t),$$

其中

$$L_2 = D^2 X(t) - \beta_1 D - \beta_2,$$

表 1 经验 I 型错误概率

N	名义检验水平 α								
	0.1			0.05			0.01		
	200	500	1000	200	500	1000	200	500	1000
$d = 1$	0.088	0.102	0.098	0.054	0.057	0.048	0.007	0.011	0.012
$d = 2$	0.089	0.091	0.096	0.057	0.041	0.047	0.008	0.009	0.011
$d = 3$	0.098	0.099	0.107	0.053	0.047	0.046	0.012	0.009	0.012
$d = 4$	0.091	0.091	0.093	0.052	0.053	0.052	0.014	0.015	0.013

表 2 存在变点时经验拒绝概率

δ		名义检验水平 α								
		0.1			0.05			0.01		
		0.005	0.01	0.015	0.005	0.01	0.015	0.005	0.01	0.015
$N=200$	$d = 1$	0.269	0.648	0.886	0.156	0.504	0.824	0.050	0.258	0.622
	$d = 2$	0.245	0.627	0.900	0.164	0.473	0.843	0.052	0.238	0.639
	$d = 3$	0.250	0.650	0.938	0.148	0.483	0.879	0.052	0.259	0.731
	$d = 4$	0.241	0.664	0.962	0.153	0.514	0.930	0.045	0.299	0.801
$N=500$	$d = 1$	0.493	0.946	1	0.368	0.902	0.990	0.173	0.744	0.977
	$d = 2$	0.418	0.941	1	0.304	0.892	1	0.134	0.717	0.985
	$d = 3$	0.424	0.932	1	0.303	0.884	1	0.124	0.73	0.990
	$d = 4$	0.404	0.933	1	0.298	0.881	1	0.117	0.742	0.992
$N=1000$	$d = 1$	0.766	0.996	1	0.660	0.993	1	0.409	0.982	1
	$d = 2$	0.697	0.998	1	0.578	0.996	1	0.325	0.970	1
	$d = 3$	0.661	0.998	1	0.546	0.994	1	0.303	0.974	1
	$d = 4$	0.641	0.997	1	0.513	0.996	1	0.278	0.974	1

β_1 反映了空气阻力系数, β_2 反映了弹簧的弹性系数. 而 $17 \sin(\phi t)$ 为输入函数, 反映了施加在物体上的外力的变化. $\epsilon(t)$ 为 Gauss 白噪声过程, 表示其他不可观测的影响物体运动的外力.

首先考察在原假设下统计量 $S_{N,d}$ 极限分布对真实分布的近似. 取 $\beta_1 = -2$, $\beta_2 = -5$, $\phi = 2$, 初始条件为 $X(0) = 0$, $DX(0) = 0$. 由 Xia 等^[8] 提出的高阶导数方法容易验证参数 (β_1, β_2, ϕ) 在 $(-2, -4, 2)$ 的邻域内是局部可识别的. 为了得到上述方程的解, 我们把区间 $[1, 20]$ 等分为 500 个小区间, 在第 j 个分点 t_j 根据 Runge-Kutta 方法得到 $X_i(t_j)$, 其中 $\epsilon_{ij} \sim N(0, 1)$. 对这些离散的观测值, 因为是二阶微分方程的解, 故利用带粗糙惩罚的 6 阶 B 样条进行光滑, 粗糙度利用 4 阶导数平方的积分来度量, 光滑参数利用 GCV (广义交叉验证) 准则选取, 由此得到第 i 次模拟观测 $X_i(t)$. 把上述步骤重复 N 次, 就得到了 N 个模拟样本曲线, 即 $1 \leq i \leq N$.

选取 11 个周期为 π 的 Fourier 基函数将输入函数展开; 关于特征函数个数的选取, 利用累积方差比法, 取 $d = 4$, 它们解释了数据中 92% 的方差. 考察了三种常见的样本容量 $N = 200, 500, 1000$ 下检

验犯 I 型错误的概率, 每种样本容量下检验重复的次数都为 1000, 所得结果列于表 1.

由表 1, 对所考察的三种样本容量, 理论检验水平与经验检验水平都相当接近, 其次随着样本容量的增大, 二者之间的差别也越来越小, 所以对上述的三种样本容量, 极限分布可以作为真实分布合理的近似.

我们也考察了当对立假设成立时检验的功效. 设参数在 $n_0 = [0.3N]$ 处发生了变化, 其中当 $1 \leq i \leq n_0$ 时, 参数为

$$\beta_1 = -2, \quad \beta_2 = -5, \quad f(t) = 17 \sin(2t),$$

当 $n_0 + 1 \leq i \leq N$ 时, 考虑三种外力频率变化但振幅不变的情形:

$$\text{I: } \beta_1 = -2, \beta_2 = -5, f(t) = 17 \sin(2 + 0.005)t,$$

$$\text{II: } \beta_1 = -2, \beta_2 = -5, f(t) = 17 \sin(2 + 0.01)t,$$

$$\text{III: } \beta_1 = -2, \beta_2 = -5, f(t) = 17 \sin(2 + 0.015)t.$$

在变点前后初始条件都为 $X(0) = 0, DX(0) = 0$. 对所有的三种情形, 变点前后方程的解差别都极小, 在加了噪声之后很难通过观测发现方程的参数是否以及何时发生了变化. 表 2 给出了检验的经验势函数取值, 也即拒绝 H_0 的概率. 由表 2, 考虑 $\alpha = 0.05$ 的情形, 当周期改变量为 0.25% (对应 $\delta = 0.005$), 则检验拒绝 H_0 的概率有 50% 左右, 当周期改变量增至 0.75% 时 (对应 $\delta = 0.015$), 检验对 $N = 200, 500, 1000$ 就以大概率拒绝 H_0 , 而这样小的参数改变量即使在一些经典的变点问题中也是很难检验出来的, 譬如重尾随机变量均值变点的检验 [16].

5.2 实例分析

下面利用上述推断方法考察英国中部地区 1772 年至 2009 年气温的变化规律, 共 238 年, 每一年有 365 个观测数据, 每一个数据代表某一天的日平均气温, 已有多文献对此数据或其子集做了研究, 可参见 [2, 4, 12, 17]. 已有文献对此数据的研究集中于揭示在过去的两百年里年度平均气温是否发生了变化, 而对温度变化的机制, 即其满足的微分方程是否变化, 目前还是研究领域内的空白.

众所周知, 年度气温可以利用下述线性常微分方程描述 [18]:

$$D^3 \text{Temp}(t) - \beta_1 D^2 \text{Temp}(t) - \beta_2 D \text{Temp}(t) - \beta_3 \text{Temp}(t) - f(t) = \epsilon(t),$$

其中 $\text{Temp}(t)$ 表示 t 时刻的气温, $f(t)$ 反映了外在因素对于气温的影响, $\epsilon(t)$ 是误差项. 这个方程还可以化简, 但对本文没有影响, 这里就用这种一般形式. 对每一年的 365 个数据, 利用带粗糙惩罚的 7 阶 B 样条做光滑, 其中惩罚项采用 5 阶导数平方的积分来表示, 光滑参数利用 GCV 选取. $f(t)$ 利用 12 个周期为 365 的 Fourier 基函数展开. 在统计量 $S_{N,d}$ 的计算中, 选取 $d = 8$ 个特征函数, 解释了数据 84% 的方差, 由此得到 $S_{N,d} = 3.652$, 而与检验水平 $\alpha = 0.01$ 所对应的临界值为 $S_0 = 2.5892$, 所以方程的参数变化是非常显著的. 利用在第 3 节所给出的变点的估计方法我们得到方程参数的变化时刻为 $\hat{n} = [N\theta] = 154$, 也即大约在 1926 年附近参数的变化最明显. 再考察 1772 年至 1926 年和 1927 年至 2009 年两个子区间, 所得到的统计量都不再显著, 也即没有证据表明在这两个区间内方程参数发生了改变.

把此结论与 [12] 以及 [4] 中的结果相比较发现非常有意思的现象. 这两篇文献研究了平均气温是否存在变化, 其中前者发现了两个变点 1815 和 1926, 后者发现了 4 个变点, 1808, 1850, 1926 和 1992.

上述结果说明 1926 年左右气温的变化不仅仅是平均气温的变化, 也导致了温度方程中参数的改变, 而其他时间点处平均气温的变化并没有导致方程中参数的改变.

6 相关讨论

本文对常系数线性常微分方程中的参数稳定性问题做了初步的探讨, 给出了参数稳定性统计推断方法及相关的理论结果. 论文中一个隐含的条件是可以直接得到观测曲线 $X_i(t)$ 及其导数, 实际中这些曲线是通过核光滑或样条光滑的方法得到的, 不可避免地会带有误差, 尤其在观测值稀疏的时候误差会更大. 由曲线光滑所产生的误差是否会对本文提出的统计推断方法产生影响需要进一步的研究. 已有研究中, Liang 和 Wu^[19] 对常参数非线性常微分方程研究了核光滑产生的误差对参数估计的影响, Chen 和 Wu^[20, 21] 研究了核光滑所产生的误差对时变参数估计的影响. 其次, 实际问题中非线性常微分方程应用更为广泛, 对非线性模型也存在着参数稳定性的统计推断问题 (譬如根据测量数据判断艾滋病病人体内病毒是否发生了变异), 本文所提出的方法不能直接应用于此类问题, 对此需要提出新的统计推断方法.

参考文献

- 1 Csörgő M, Horváth L. Limit Theorems in Change-point Analysis. Wiley Series in Probability & Statistics. New York: Wiley, 1998
- 2 Zhou J. Maximum likelihood ratio test for the stability of sequence of Gaussian random processes. Comput Statist Data Anal, 2011, 55: 2114–2127
- 3 Horváth L, Hušková M, Kokoszka P. Testing the stability of the functional autoregressive process. J Multivariate Anal, 2010, 101: 352–367
- 4 Berkes I, Gabrys R, Horváth L, et al. Detecting changes in the mean of functional observations. J R Stat Soc Ser B Stat Methodol, 2009, 71: 1–20
- 5 Aue A, Gabrys R, Horváth L, et al. Estimation of a change-point in the mean function of functional data. J Multivariate Anal, 2009, 100: 2254–2269
- 6 Bosq D. Linear Processes in Function Spaces. New York: Springer, 2000
- 7 Audoly S, D'Angiò L, Saccomani M P, et al. Global identifiability of linear compartmental models—a computer algebra algorithm. IEEE Trans Bio-Med Eng, 1998, 45: 36–47
- 8 Xia X, Moog C H. Identifiability of nonlinear systems with applications to HIV/AIDS models. IEEE Trans Control, 2003, 48: 330–336
- 9 Wu H L, Zhu H H, Miao H Y, et al. Parameter identifiability and estimation of HIV/AIDS dynamic models. Bull Math Biol, 2008, 70: 785–799
- 10 Bai J S. Least square estimation of a shift in linear process. J Time Ser Anal, 1994, 15: 453–472
- 11 Kiefer J. K-sample analogus of the Kolmogorov-Smirnov and Cramer-v. Mises tests. Ann Statist, 1959, 30: 420–447
- 12 Horváth L, Kokoszka P S, Steinebach J. Testing for changes in multivariate dependent observations with applications to temperature changes. J Multivariate Anal, 1999, 68: 96–119
- 13 Cattell R B. The scree test for the number of factor. J Multivariate Behav Res, 1966, 1: 245–276
- 14 Biau G, Bunea F, Wegkamp M H. Functional classification in Hilbert spaces. IEEE Trans Inform Theory, 2005, 51: 2163–2172
- 15 Yao F, Müller H G, Wang J L. Functional data analysis for sparse longitudinal data. J Amer Statist Assoc, 2005, 100: 577–590
- 16 Zhou J, Liu S Y. Inference for mean change-point in infinite variance AR process. Statist Probab Lett, 2009, 79: 6–15
- 17 Hosking J R. Modeling persistence in hydrological time series using fractional differencing. Water Resour Res, 1984, 20: 1898–1908
- 18 Ramsay J O, Silverman B W. Functional Data Analysis, 2nd ed. New York: Springer Verlag, 2005

- 19 Liang H, Wu H L. Parameter estimation for differential equation models using a framework of measurement error in regression models. *J Amer Statist Assoc*, 2008, 103: 1570–1583
- 20 Chen J W, Wu H L. Estimation of time varying parameters in deterministic dynamic models. *Statist Sin*, 2008, 18: 987–1006
- 21 Chen J W, Wu H L. Efficient local estimation for time varying coefficients in deterministic dynamic models with applications to HIV-I dynamics. *J Amer Statist Assoc*, 2008, 103: 369–384

Statistical inference for the stability of the parameters in linear ordinary differential equation

ZHOU Jie & WU Ting

Abstract With the noisy observations, the statistical inferences are investigated for the stability of the parameters in the linear ordinary differential equation. Based on the Karhunen-Loeve expansion of random process, the test and estimating procedures for the change point are proposed. Under the null hypothesis, the limit distribution for the test statistic is established, and the procedures are shown to be consistent under the alternative hypothesis. For the linear second-order ordinary differential equation, simulation studies show that the limit null distribution is a nice approximation to the empirical null distribution, while under the alternative hypothesis, the procedures can detect the 0.75 percent changes in the frequency of the input function with high probability. The inference procedures are also applied to the temperature data of central England and some new characteristics in the data are discovered.

Keywords: linear ODE, change point, functional data, Karhunen-Loeve expansion

MSC(2000): 62F99, 62M10

doi: 10.1360/012010-537