相依 t 过程回归模型

王占锋¹, 罗华荣¹, 明瑞星^{2*}, 吴耀华¹

1. 中国科学技术大学统计与金融系, 合肥 230026;

2. 浙江工商大学统计与数学学院, 杭州 310018

E-mail: zfw@ustc.edu.cn, henryluo@mail.ustc.edu.cn, ruixingming@aliyun.com, wuyh@ustc.edu.cn

收稿日期: 2020-08-18; 接受日期: 2021-01-25; 网络出版日期: 2021-09-16; * 通信作者

国家自然科学基金 (批准号: 11971457)、安徽省自然科学基金 (批准号: 1908085MA06) 和浙江省一流学科 A 类 (浙江工商大学统计学) 资助项目

摘要 本文通过引入一个随机效应变量, 使用卷积方法刻画多元响应变量之间的相关性, 进而构造了相依 t 过程 (dependent t -process, DTP). 本文基于 DTP, 提出了相依 t 过程回归 (dependent t -process regression, DTPR) 模型, 并对多元响应变量的函数型数据进行建模. DTPR 保留了 Gauss 过程回归 (Gaussian process regression, GPR) 的可解释性, 而且具有对于离群点的稳健性. 本文研究 DTPR 的稳健性和信息相合性, 并利用数值模拟与实际数据研究验证 DTPR 的可行性.

关键词 相依函数型响应变量 稳健性 信息相合性

MSC (2020) 主题分类 62G35, 62G08

1 引言

近年来, 函数型响应数据的回归方法受到越来越多的关注, 例如, Gauss 过程回归 (Gaussian process regression, GPR) 是其中最为成熟的方法之一 (参见文献 [1–3]). 通常, 在处理函数型响应数据时均假设函数型响应曲线之间是独立的. 然而, 在实际应用中, 函数型数据的观测值之间往往是有相关性的 (如两个相邻城市的气温变化的数据, 或男性与女性随年龄变化的死亡率数据等都是相关的). 这样, 在使用传统的模型对这类数据进行建模并预测时, 会把不同响应变量的观测曲线看作是相互独立的, 这势必会损失信息. 反之, 直观上看, 如果模型考虑了数据之间的相依性, 将会很大程度地提高预测的准确率. 经验表明, 当两个函数型响应变量高度相关且其中一者的数据收集起来有较大困难导致训练数据分布非常稀疏时, 独立的模型预测效果往往会很差. 在这种情形下, 考虑相依性的模型对两者同时进行建模, 借由数据之间的相关信息来进行预测应当是合理且有效的.

基于传统的 Gauss 过程回归, 文献 [4] 提出了相依 Gauss 过程回归 (dependent Gaussian processes regression, DGPR), 通过卷积构造多个有相依性的 Gauss 过程, 进而得到多个函数型响应变量之间的

英文引用格式: Wang Z F, Luo H R, Ming R X, et al. Dependent t -process regression model (in Chinese). Sci Sin Math, 2022, 52: 913–934, doi: 10.1360/SSM-2020-0263

协方差函数. DGPR 可以对多个响应变量的相关性进行建模, 从而使模型有更广泛的适用性. 从卷积过程出发, 文献 [5] 提出了变分诱导函数的概念, 解决了非光滑核函数在卷积过程中带来的困难, 进而解决了真实函数曲线不光滑的问题. 相比于传统单一输出的 GPR 和 DGPR, 文献 [6] 提出了协同多输出 Gauss 过程 (collaborative multi-output Gaussian processes) 模型, 在大数据集上的表现有显著提高. 文献 [7] 将现有的多输出 Gauss 过程回归方法创新性地分为对称型和非对称型, 并对其进行了详细的对比. 然而, Gauss 过程回归对于离群点是不稳健的. 我们用一个示例说明这个问题. 在图 1(a) 中, 曲线 1 (短划线) 和曲线 2 (实线) 为两条有相依性的曲线. 在图 1(b) 中, 空心点和实心点分别代表从曲线 1 和 2 中采样得到的观测点. 可以看出, 数据 2 位于两条竖线之间, 数据 2 没有观测样本点. 对数据 2 的一个样本点的观测值上加了一个扰动, 将该点的响应值减去 1, 使之成为离群值点 (用箭头标出). 在图 1(c) 中, 我们比较了真实曲线 2 (实线)、DGPR 对于数据 2 的拟合曲线 (短划线) 和 GPR 的拟合曲线 (点线). 由于 GPR 在建模时没有使用数据 1, 因此 GPR 从数据缺失的区域 (两条竖线间) 没有获得任何信息, 导致拟合出的曲线离真实曲线 2 有较大偏差. 相比于 GPR, DGPR 利用了数据 1 的信息, 因此其拟合曲线更能贴近真实曲线 2 的趋势. 但是受到离群点的影响, DGPR 的拟合曲线距真实曲线 2 也有大幅度的偏离. 因此, 在面对存在异常点的相依数据时, 寻找一个稳健的方法是很有必要的.

在稳健统计方法的领域, 重尾分布受到了广泛的关注. 文献 [8–10] 分别考虑了在 Gauss 过程回归

图 1 (a) 两条有较强相依性的曲线; (b) 由两条曲线采样得到的数据点; (c) 真实曲线 2、GPR 和 DGPR 对数据 2 的拟合曲线; (d) 真实曲线 2、eTPR 和 DTPR 对数据 2 的拟合曲线

中引入学生 t 分布和 Laplace 分布的噪声来增加模型的稳健性. 文献 [11] 基于两分量噪声模型, 提出了一种稳健的 Gauss 过程方法. 进一步地, 文献 [12–17] 将传统的 Gauss 过程模型推广至其他重尾的过程. 例如, 文献 [18, 19] 证明了 t 型回归估计量的稳健性和相合性. 文献 [12, 13] 将多任务学习模型从 Gauss 过程模型推广至 t 过程模型, 提高了模型在异常点数据上的稳健性. 作为函数曲线的先验, t 过程被证明可以很好地代替 Gauss 过程, 并且保留了 Gauss 过程的很多优良性质^[17]. 文献 [14] 提出了一个扩展 t 过程回归 (extended t -process regression, eTPR) 模型, 它在给出稳健估计的同时, 也继承了 Gauss 过程回归诸多良好的性质, 如可加性和信息相合性等. 文献 [15] 进一步地将 eTPR 推广至误差函数与回归函数独立的情形. 文献 [16] 研究了其他重尾过程, 如受污染的正态 (contaminated-normal) 过程和斜线 (slash) 过程的稳健性. 基于样条方法, 文献 [20] 对损失函数中的正则项进行修改, 提出了一个稳健的函数型回归方法. 借助非凸损失函数, 文献 [21] 给出了函数型线性回归的稳健推断. 文献 [22] 提出了稳健函数型主成分分析和稳健线性回归的两步估计方法, 解决了函数型线性模型中最小二乘法不稳健的缺点. 文献 [23] 提出了一个名为稳健函数型混合模型的 Bayes 方法以提高模型的稳健性. 文献 [24–26] 给出了纵向数据下的稳健推断方法. 然而, 上述这些方法并没有考虑多元函数型响应变量的相依结构. 本文提出相依 t 过程回归 (dependent t -process regression, DTPR) 以提高多元函数型响应变量下的模型稳健性. 图 1(d) 分别给出了 eTPR (点线) 和 DTPR (短划线) 的拟合曲线. 可以看出, eTPR 作为不考虑多个响应变量之间相依性的模型, 不能捕捉到数据 1 的信息, 因此在数据 2 上的拟合表现并不尽如人意. 相比于其他 3 个模型: GPR、DGPR 和 eTPR, 本文提出的 DTPR 的拟合曲线更接近于真实曲线 2 (实线). 这里因为 DTPR 在考虑到两条曲线的相依性的同时, 也具有对异常点的稳健性.

本文对多元函数型响应数据提出了一个稳健的非参数建模方法. 借助最大后验 (maximum a posteriori, MAP) 概率估计的方法, 我们提出了一个未知参数估计方法. 该方法有如下几个优点: (a) 相依 t 过程 (dependent t -process, DTP) 由独立和相依的两部分组成. 假定每条函数曲线都包含一个刻画其个体特性的 t 过程, 同时多条函数曲线还共用一组有相依性的 t 过程以描述其共同的特性. 这样我们便可以在保留一定独立性的同时也对相依性进行建模. (b) 本文使用基于 DTP 的函数型回归模型对多元函数型响应变量进行拟合, 并给出异常点下回归函数的稳健线性无偏估计. (c) 对 Gauss 核函数引入一个随机效应, 并使用 Gauss 白噪声过程与 Gauss 核进行卷积, 本文提出一个新的构造 DTP 的协方差函数的方法, 其中函数曲线之间的相依程度可以通过对协方差函数中的超参数进行估计, 依数据进行自动调整, 使得模型更具有灵活性, 适用于更多情形下的数据. 相较于 eTPR, 我们无法为 DTPR 指定任意正定的二元函数作为其协方差函数, 因此本文采用卷积的方法构造其协方差函数. 不同于 DGPR 模型, DTPR 模型引入了一个随机效应, 增加了模型的适用性和稳健性. 本文在卷积的过程中添加了一个随机效应, 并以此得到了 DTPR 的协方差函数. DTPR 具有良好的统计性质, 包括稳健性和信息相合性. 相比于 DGPR 和 eTPR, 数值实验的结果显示本文提出的 DTPR 可以给出更精确的预测.

本文余下内容安排如下: 第 2 节介绍两输出的相依 t 过程的构造, 第 3 节引入相依 t 过程回归及其预测步骤, 第 4 节讨论 DTPR 的稳健性和信息相合性, 第 5 节利用数值实验和真实案例说明本文所提方法的表现效果, 第 6 节给出总结和讨论, 附录包含统计性质的证明以及对多个输出的 DTP 情形的推广.

2 两输出情形下的相依 t 过程

考虑联合 t 过程

$$\begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} h_1(\mathbf{x}) + U_1(\mathbf{x}) + V_1(\mathbf{x}) \\ h_2(\mathbf{x}) + U_2(\mathbf{x}) + V_2(\mathbf{x}) \end{pmatrix},$$

其中 h_1 和 h_2 为两个均值函数, $\{U_1, U_2\}$ 和 $\{V_1, V_2\}$ 是两组 t 过程. 假设 U_1 与 U_2 是相关的、 V_1 与 V_2 是不相关的且 $\{U_1, U_2\}$ 和 $\{V_1, V_2\}$ 是不相关的. 因此, f_1 和 f_2 是具有相依性的.

众所周知, 在构造单个 t 过程时, 可以选择任意的二元正定函数作为其协方差函数. 但是对于构造联合 t 过程而言, 这种协方差函数的定义是不可行的^[4]. 因为对于有两个输出的联合 t 过程 $(f_1(\mathbf{x}), f_2(\mathbf{x}))^\top$ 而言, 其协方差矩阵可以分解为三个小矩阵, 分别为 $f_1(\mathbf{x})$ 的自协方差矩阵 (auto-covariance matrix) 和 $f_2(\mathbf{x})$ 的自协方差矩阵以及 $f_1(\mathbf{x})$ 与 $f_2(\mathbf{x})$ 的互协方差阵 (cross-covariance matrix). 对于任意指定的协方差函数, 我们仅能保证每个分块矩阵是正定矩阵, 但是不能确保整个协方差阵是正定的. 借鉴文献 [4, 27] 中的方法, 我们尝试通过卷积方式来定义协方差函数, 具体步骤如下所述.

令 g_1 、 g_2 、 g_3 和 g_4 为定义在输入空间 $\mathcal{X} = \mathbb{R}^p$ 上的 Gauss 核,

$$\begin{aligned} g_1(\mathbf{x}) &= v_1 \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{A}_1 \mathbf{x}\right), & g_2(\mathbf{x}) &= v_2 \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{A}_2 (\mathbf{x} - \boldsymbol{\mu})\right), \\ g_3(\mathbf{x}) &= w_1 \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{B}_1 \mathbf{x}\right), & g_4(\mathbf{x}) &= w_2 \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{B}_2 \mathbf{x}\right), \end{aligned}$$

其中 $v_1, v_2, w_1, w_2 \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$ 且 $\mathbf{A}_i$ 和 $\mathbf{B}_i$ 均为 $p \times p$ 的正定矩阵. 令 r 为一个随机效应, 服从形状参数为 ν 和尺度参数为 ω 的逆伽马分布, 记为 $r \sim \text{IG}(\nu, \omega)$. 我们知道 Gauss 过程可以由一个 Gauss 白噪声过程和一个光滑核函数卷积而得到. 记 W_1 、 W_2 和 W_3 为 3 个独立的 Gauss 白噪声过程. U_i 和 V_i 构造方式如下:

$$\begin{aligned} U_1(\mathbf{x}) &= \sqrt{r}W_1(\mathbf{x}) \star g_1(\mathbf{x}), & U_2(\mathbf{x}) &= \sqrt{r}W_1(\mathbf{x}) \star g_2(\mathbf{x}), \\ V_1(\mathbf{x}) &= \sqrt{r}W_2(\mathbf{x}) \star g_3(\mathbf{x}), & V_2(\mathbf{x}) &= \sqrt{r}W_3(\mathbf{x}) \star g_4(\mathbf{x}), \end{aligned} \tag{2.1}$$

其中 $\mathbf{x} \in \mathcal{X}$, $\star$ 代表卷积运算. 在给定 r 时, 由于共用同一个 Gauss 过程 W_1 , 因此 U_1 和 U_2 是有相依性的, 而 V_1 和 V_2 相互独立. 因此有

$$U_i | r \sim \text{GP}(0, rk_{ii}^U), \quad V_i | r \sim \text{GP}(0, rk_{ii}^V), \quad i = 1, 2,$$

其中 $\text{GP}(h, k)$ 表示一个均值函数为 h 、协方差函数为 k 的 Gauss 过程. 可以看到 rk_{ii}^U 和 rk_{ii}^V 分别是 $U_i | r$ 和 $V_i | r$ 的自协方差函数. 根据文献 [14], 有

$$U_i \sim \text{ETP}(\nu, \omega, 0, k_{ii}^U), \quad V_i \sim \text{ETP}(\nu, \omega, 0, k_{ii}^V), \quad i = 1, 2,$$

其中 $\text{ETP}(\nu, \omega, h, k)$ 表示一个参数为 (ν, ω) 、均值函数为 h 、协方差函数为 k 的扩展 t 过程^[14].

由于上述定义的核函数是平稳的, 为叙述方便, 将协方差函数写作两点之差的形式. 对于固定的向量 $\mathbf{d} \in \mathcal{X}$ 和任意 $\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}$, 若有 $\mathbf{x}_a - \mathbf{x}_b = \mathbf{d}$, 则将关于 $\mathbf{x}_a$ 和 $\mathbf{x}_b$ 的协方差函数记作关于 $\mathbf{d}$ 的函数

$$k_{ii}^U(\mathbf{d}) \doteq k_{ii}^U(\mathbf{x}_a, \mathbf{x}_b), \quad k_{ij}^U(\mathbf{d}) \doteq k_{ij}^U(\mathbf{x}_a, \mathbf{x}_b), \quad k_{ii}^V(\mathbf{d}) \doteq k_{ii}^V(\mathbf{x}_a, \mathbf{x}_b),$$

其中 k_{ij}^U ($i \neq j$) 是 U_i 和 U_j 的互协方差函数. 通过求解 (2.1) 中的卷积, 我们可以得到协方差函数的形式如下:

$$\begin{aligned} k_{ii}^U(\mathbf{d}) &= \frac{\pi^{\frac{p}{2}} v_i^2}{\sqrt{|\mathbf{A}_i|}} \exp\left(-\frac{1}{4} \mathbf{d}^\top \mathbf{A}_i \mathbf{d}\right), \\ k_{12}^U(\mathbf{d}) &= \frac{(2\pi)^{\frac{p}{2}} v_1 v_2}{\sqrt{|\mathbf{A}_1 + \mathbf{A}_2|}} \exp\left(-\frac{1}{2} (\mathbf{d} - \boldsymbol{\mu})^\top \mathbf{G} (\mathbf{d} - \boldsymbol{\mu})\right), \\ k_{21}^U(\mathbf{d}) &= \frac{(2\pi)^{\frac{p}{2}} v_1 v_2}{\sqrt{|\mathbf{A}_1 + \mathbf{A}_2|}} \exp\left(-\frac{1}{2} (\mathbf{d} + \boldsymbol{\mu})^\top \mathbf{G} (\mathbf{d} + \boldsymbol{\mu})\right) = k_{12}^U(-\mathbf{d}), \\ k_{ii}^V(\mathbf{d}) &= \frac{\pi^{\frac{p}{2}} w_i^2}{\sqrt{|\mathbf{B}_i|}} \exp\left(-\frac{1}{4} \mathbf{d}^\top \mathbf{B}_i \mathbf{d}\right), \end{aligned}$$

其中 $\mathbf{G} = \mathbf{A}_1(\mathbf{A}_1 + \mathbf{A}_2)^{-1} \mathbf{A}_2 = \mathbf{A}_2(\mathbf{A}_1 + \mathbf{A}_2)^{-1} \mathbf{A}_1$.

注意到回归函数的形式为 $f_i = h_i + U_i + V_i$. 由扩展 t 过程的可加性, 有 $f_i \sim \text{ETP}(\nu, \omega, h_i, k_{ii})$, 其中 $k_{ii} = k_{ii}^U + k_{ii}^V$ 是 f_i 的自协方差函数. 此外, 由于 V_1 和 V_2 的互协方差为 0, 而 U_1 和 U_2 的协方差函数为 k_{12}^U , 因此 f_1 和 f_2 的协方差函数恰为 k_{12}^U , 简记作 k_{12} . 因而可得 f_1 和 f_2 服从如下的联合 t 过程, 或称为相依 t 过程:

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \sim \text{DTP}\left(\nu, \omega, \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}\right),$$

其中

$$\begin{aligned} k_{11}(\mathbf{d}) &= k_{11}^U(\mathbf{d}) + k_{11}^V(\mathbf{d}), & k_{12}(\mathbf{d}) &= k_{12}^U(\mathbf{d}), \\ k_{22}(\mathbf{d}) &= k_{22}^U(\mathbf{d}) + k_{22}^V(\mathbf{d}), & k_{21}(\mathbf{d}) &= k_{21}^U(\mathbf{d}). \end{aligned}$$

当 $\nu \rightarrow \infty$ 且 $\omega/\nu = \lambda$ (λ 为常数) 时, 我们可以证明 $\lim_{\nu \rightarrow \infty} \text{DTP}(\nu, \omega, h, k) = \text{DGP}(h, \lambda k)$ (参见文献 [14]), 其中 $\text{DGP}(h, k)$ 为均值函数为 h 、协方差函数为 k 的相依 Gauss 过程 (dependent Gaussian process, DGP) [4].

记第 i 条曲线上的样本点为 $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^\top$, 其中 $\mathbf{x}_{il} \in \mathcal{X}$, $i = 1, 2, l = 1, \dots, n_i$. 对于所有的样本点集合 $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$, 令 $\mathbf{f} \doteq f(\mathbf{X}) = (f_1(\mathbf{X}_1)^\top, f_2(\mathbf{X}_2)^\top)^\top$ 为 $f = (f_1, f_2)^\top$ 在 $\mathbf{X}$ 上的一个实现, 其中 $f_i(\mathbf{X}_i) = (f_i(\mathbf{x}_{i1}), \dots, f_i(\mathbf{x}_{in_i}))^\top$ 为单个过程 f_i 的样本实现. 由 t 过程的定义, $\mathbf{f}$ 服从一个相依多元 t 分布 (dependent multivariate t -distribution, DMTD): $\mathbf{f} \sim \text{DMTD}(\nu, \omega, \mathbf{h}, \mathbf{K})$, 其概率密度函数为

$$p(\mathbf{X}) = |2\pi\omega\mathbf{K}|^{-1/2} \frac{\Gamma(N/2 + \nu)}{\Gamma(\nu)} \left(1 + \frac{(\mathbf{X} - \mathbf{h})^\top \mathbf{K}^{-1} (\mathbf{X} - \mathbf{h})}{2\omega}\right)^{-(N/2 + \nu)},$$

其中 $N = n_1 + n_2$, $\Gamma(\cdot)$ 为伽马函数, $\mathbf{h} = (\mathbf{h}_1^\top, \mathbf{h}_2^\top)^\top$, $\mathbf{h}_i = (h_i(\mathbf{x}_{i1}), \dots, h_i(\mathbf{x}_{in_i}))^\top$ 为 $\mathbf{X}_i$ 的均值向量,

$$\begin{aligned} \mathbf{K} &= \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix}, \\ \mathbf{K}_{ij} &= \begin{pmatrix} k_{ij}(\mathbf{x}_{i1} - \mathbf{x}_{j1}) & \dots & k_{ij}(\mathbf{x}_{i1} - \mathbf{x}_{jn_j}) \\ \vdots & \ddots & \vdots \\ k_{ij}(\mathbf{x}_{in_i} - \mathbf{x}_{j1}) & \dots & k_{ij}(\mathbf{x}_{in_i} - \mathbf{x}_{jn_j}) \end{pmatrix}, \end{aligned}$$

为 $\mathbf{X}_i$ 和 $\mathbf{X}_j$ 的协方差阵.

令 $\mathbf{u} \in \mathcal{X}$ 为第 i 条曲线输入空间上的一个新观测点, 记在 $\mathbf{u}$ 与所有已知数据点 $\mathbf{X}$ 之间的协方差阵 $\zeta_{\mathbf{u}}$ 为

$$\zeta_{\mathbf{u}} = (k_{i1}(\mathbf{u}, \mathbf{x}_{11}), \dots, k_{i1}(\mathbf{u}, \mathbf{x}_{1n_1}), k_{i2}(\mathbf{u}, \mathbf{x}_{21}), \dots, k_{i2}(\mathbf{u}, \mathbf{x}_{2n_2}))^\top,$$

那么在给定样本实现 $\mathbf{f}$ 时, 过程 $(f_1, f_2)^\top$ 有如下的 DTP:

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \bigg| \mathbf{f} \sim \text{DTP} \left(\nu^*, \omega^*, \begin{pmatrix} h_1^* \\ h_2^* \end{pmatrix}, \begin{pmatrix} k_{11}^* & k_{12}^* \\ k_{21}^* & k_{22}^* \end{pmatrix} \right),$$

其中

$$\begin{aligned} \nu^* &= \nu + \frac{N}{2}, \quad \omega^* = \omega + \frac{N}{2}, \\ h_i^*(\mathbf{u}) &= \zeta_{\mathbf{u}}^\top \mathbf{K}^{-1}(\mathbf{f} - \mathbf{h}) + h_i(\mathbf{u}), \\ k_{ij}^*(\mathbf{u}, \mathbf{v}) &= \frac{2\omega + (\mathbf{f} - \mathbf{h})^\top \mathbf{K}^{-1}(\mathbf{f} - \mathbf{h})}{2\omega + N} (k_{ij}(\mathbf{u}, \mathbf{v}) - \zeta_{\mathbf{u}}^\top \mathbf{K}^{-1} \zeta_{\mathbf{v}}), \end{aligned}$$

这里 $\mathbf{v} \in \mathcal{X}$ 为第 j 条曲线输入空间中的一个新样本点.

值得注意的是, 当随机效应 r 为常数 1 时, DTP 会退化为 DGP. 因此也可以认为 DGP 是 DTP 的一种特殊情形.

3 相依 t 过程回归模型

记第 i 条曲线的样本点 $\mathbf{X}_i$ 所对应的响应变量为 $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$, 其中 $y_{il} \in \mathbb{R}$. 不失一般性, 令每条曲线上的样本点数目相等, 即 $n_1 = n_2 = n$, 且有 $N = n_1 + n_2 = 2n$. 将观测数据集合记为 $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$, 其中 $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$, $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top)^\top$.

考虑如下的函数型回归模型:

$$y_{il} = f_{0i}(\mathbf{x}_{il}) + \epsilon_{il}, \quad i = 1, 2, \quad l = 1, \dots, n, \quad (3.1)$$

其中 f_{0i} 为第 i 条未知的曲线, ϵ_{ij} 为误差项. 为了估计未知的函数 f_{0i} , 考虑如下的过程回归模型:

$$y_i(\mathbf{x}) = f_i(\mathbf{x}) + \epsilon_i(\mathbf{x}), \quad i = 1, 2, \quad (3.2)$$

其中 f_i 为第 i 个输出的随机函数, $\epsilon_i(\mathbf{x})$ 为误差过程.

为了刻画 y_1 和 y_2 的相依性, 令 f_1 和 f_2 为第 2 节所定义的相依 t 过程. 假定 f_1 、 f_2 、 ϵ_1 和 ϵ_2 服从如下的过程:

$$\begin{pmatrix} f_1 \\ f_2 \\ \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim \text{DTP} \left(\nu, \omega, \begin{pmatrix} h_1 \\ h_2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} & 0 & 0 \\ k_{21} & k_{22} & 0 & 0 \\ 0 & 0 & k_{\epsilon_1} & 0 \\ 0 & 0 & 0 & k_{\epsilon_2} \end{pmatrix} \right), \quad (3.3)$$

其中 $k_{\epsilon_i}(\mathbf{x}_a, \mathbf{x}_b) = \sigma^2 I(\mathbf{x}_a = \mathbf{x}_b)$ 为误差过程的协方差函数, 这里 $I(\cdot)$ 为示性函数, $\sigma^2 > 0$.

为叙述方便, 记

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}, \quad \mathbf{k}_\epsilon = \begin{pmatrix} k_{\epsilon_1} & 0 \\ 0 & k_{\epsilon_2} \end{pmatrix}, \quad (3.4)$$

则有 $\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon} \sim \text{DTP}(\nu, \omega, \mathbf{h}, \mathbf{k} + \mathbf{k}_\epsilon)$, 即为所谓的相依 t 过程回归模型.

3.1 预测分布

依照文献 [14], 不妨令均值函数 $\mathbf{h} = \mathbf{0}$. 记 $\boldsymbol{\Sigma}_{ii} = \mathbf{K}_{ii} + \sigma^2 \mathbf{I}$ 和 $\boldsymbol{\Sigma} = \mathbf{K} + \sigma^2 \mathbf{I}$. 由 (3.3) 可知, 在给定 $\mathbf{X}$ 时, $f_1(\mathbf{X}_1)$ 、 $f_2(\mathbf{X}_2)$ 、 $\mathbf{y}_1$ 和 $\mathbf{y}_2$ 服从如下的相依多元 t 分布:

$$\begin{pmatrix} f_1(\mathbf{X}_1) \\ f_2(\mathbf{X}_2) \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \middle| \mathbf{X} \sim \text{DMTD} \left(\nu, \omega, \mathbf{0}, \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{21} & \mathbf{K}_{22} \\ \mathbf{K}_{11} & \mathbf{K}_{12} & \boldsymbol{\Sigma}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right).$$

进一步地, 有如下的分层模型:

$$\begin{aligned} f(\mathbf{X}) \mid \mathbf{X} &\sim \text{DMTD}(\nu, \omega, \mathbf{h}, \mathbf{K}), \\ \mathbf{y} \mid f, \mathbf{X} &\sim \text{DMTD}(\nu, \omega, f(\mathbf{X}), \sigma^2 \mathbf{I}), \\ \mathbf{y} \mid \mathbf{X} &\sim \text{DMTD}(\nu, \omega, \mathbf{h}, \boldsymbol{\Sigma}). \end{aligned} \quad (3.5)$$

在给定观测数据 $\mathcal{D}$ 时, $f(\mathbf{X})$ 服从条件 DMTD:

$$f(\mathbf{X}) \mid \mathcal{D} \sim \text{DMTD} \left(\frac{N}{2} + \nu, \frac{N}{2} + \omega, \boldsymbol{\mu}, \mathbf{C} \right),$$

其中

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{E}(f(\mathbf{X}) \mid \mathcal{D}) = \mathbf{K} \boldsymbol{\Sigma}^{-1} \mathbf{y}, \\ \mathbf{C} &= \text{Cov}(f(\mathbf{X}) \mid \mathcal{D}) = s_0 \sigma^2 \mathbf{K} \boldsymbol{\Sigma}^{-1}, \\ s_0 &= \mathbf{E}(r \mid \mathcal{D}) = \frac{\mathbf{y}^\top \boldsymbol{\Sigma} \mathbf{y} + 2\omega}{N + 2(\nu - 1)}. \end{aligned}$$

可以发现 DTPR 的预测值 $\hat{\mathbf{f}} = \boldsymbol{\mu} = \mathbf{E}(f(\mathbf{X}) \mid \mathcal{D})$ 是给定数据下 $f(\mathbf{X})$ 的后验均值, 而且是 $\mathbf{y}$ 的线性函数. 作为线性混合模型中最佳线性无偏预测 (best linear unbiased prediction, BLUP) [28] 的推广, 在给定参数下, $\hat{\mathbf{f}}$ 为 DTPR 模型中 $f(\mathbf{X})$ 的 BLUP. 当 $\nu \rightarrow \infty$ 且 ω/ν 趋于某个常数 λ 时, $s_0 = (\mathbf{y}^\top \boldsymbol{\Sigma} \mathbf{y} + 2\omega)/[N + 2(\nu - 1)]$ 随着 $\nu \rightarrow \infty$ 而趋向于常数, 即影响因子 s_0 退化为常数 λ . 类似前文所说的 DTP 退化为 DGP, 此时 DTPR 模型就退化成 DGPR 模型.

对于第 i 条曲线上的一个新的数据点 $\mathbf{u} \in \mathcal{X}$, 给定 $\mathbf{X}$, 我们有

$$\begin{pmatrix} f_i(\mathbf{u}) \\ \mathbf{y} \end{pmatrix} \middle| \mathbf{X} \sim \text{DMTD} \left(\nu, \omega, \mathbf{0}, \begin{pmatrix} k_{ii}(\mathbf{u}, \mathbf{u}) & \boldsymbol{\zeta}_\mathbf{u}^\top \\ \boldsymbol{\zeta}_\mathbf{u} & \boldsymbol{\Sigma} \end{pmatrix} \right).$$

进一步地,

$$f_i(\mathbf{u}) \mid \mathcal{D} \sim \text{DMTD}\left(\frac{N}{2} + \nu, \frac{N}{2} + \omega, \mu_i^*, C_i^*\right),$$

其中均值和方差分别为

$$\begin{aligned} \mathbb{E}(f(\mathbf{u}) \mid \mathcal{D}) &= \mu_i^* = \boldsymbol{\zeta}_u^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}, \\ \text{Cov}(f(\mathbf{u}) \mid \mathcal{D}) &= C_i^* = s_0(k_{ii}(\mathbf{u}, \mathbf{u}) - \boldsymbol{\zeta}_u^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta}_u). \end{aligned}$$

类似地, 给定样本 $\mathcal{D}$ 下, 过程函数 f_i 为

$$f_i \mid \mathcal{D} \sim \text{DTP}\left(\frac{N}{2} + \nu, \frac{N}{2} + \omega, h_i^*, k_{ii}^*\right),$$

其中均值函数和协方差函数分别为 $h_i^*(\mathbf{u}) = \mu_i^*$ 和 $k_{ii}^*(\mathbf{u}, \mathbf{v}) = s_0(k_{ii}(\mathbf{u}, \mathbf{v}) - \boldsymbol{\zeta}_u^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta}_v)$, 这里 $\mathbf{v}$ 为曲线 i 的输入空间上的另一数据点.

进一步地有

$$y_i(\mathbf{u}) \mid \mathcal{D} \sim \text{DMTD}\left(\frac{N}{2} + \nu, \frac{N}{2} + \omega, \mu_i^*, C_i^* + s_0\sigma^2\right),$$

其中 $\mathbb{E}(y_i(\mathbf{u}) \mid \mathcal{D}) = \mu_i^*$, $\text{Var}(y_i(\mathbf{u}) \mid \mathcal{D}) = C_i^* + s_0\sigma^2$. 因此对于第 i 条曲线上的新样本点 $\mathbf{u}$, 我们有预测值和预测的方差分别为 $\hat{y}_i(\mathbf{u}) = \mathbb{E}(y_i(\mathbf{u}) \mid \mathcal{D}) = \mathbb{E}(f_i(\mathbf{u}) \mid \mathcal{D})$ 和 $\text{Var}(\hat{y}_i(\mathbf{u}) \mid \mathcal{D}) = C_i^* + s_0\sigma^2$.

3.2 参数估计

本小节考虑更一般的成组数据. 将模型 (3.2) 拓展至组数为 M 的函数型回归模型

$$y_{mil} = f_{0mi}(\mathbf{x}_{mil}) + \epsilon_{mil}, \quad m = 1, \dots, M, \quad i = 1, 2, \quad l = 1, \dots, n_{mi}, \quad (3.6)$$

其中 f_{0mi} 为第 m 组中第 i 条曲线的未知真实函数, y_{mil} 为第 m 组中第 i 条曲线的第 l 个样本点, ϵ_{mil} 代表数据中的噪声. 令 $\mathbf{X}_{mi} = (\mathbf{x}_{mi1}, \dots, \mathbf{x}_{min})^\top$, $\mathbf{y}_{mi} = (y_{mi1}, \dots, y_{min})^\top$. 为叙述方便且不失一般性, 不妨令每条曲线的协变量均相同, 即有 $\mathbf{x}_{m1l} = \mathbf{x}_{m2l} \doteq \mathbf{x}_l$, $n_{m1} = n_{m2} = n$. 记成组的观测数据为 $\mathcal{D} = \{\mathbf{X}_{mi}, \mathbf{y}_{mi}, m = 1, \dots, M, i = 1, 2\}$.

为拟合 (3.6), 考虑如下的过程回归模型:

$$y_{mi}(\mathbf{x}) = f_{mi}(\mathbf{x}) + \epsilon_{mi}(\mathbf{x}), \quad m = 1, \dots, M, \quad i = 1, 2,$$

其中 f_{mi} 为随机函数, ϵ_{mi} 为误差项.

对于某个固定的 m , 假定 f_{m1} 、 f_{m2} 、 ϵ_{m1} 和 ϵ_{m2} 为如下的 DTP:

$$\begin{pmatrix} f_{m1} \\ f_{m2} \\ \epsilon_{m1} \\ \epsilon_{m2} \end{pmatrix} \sim \text{DTP} \left(\nu, \omega, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} k_{m11} & k_{m12} & 0 & 0 \\ k_{m21} & k_{m22} & 0 & 0 \\ 0 & 0 & k_{\epsilon_{m1}} & 0 \\ 0 & 0 & 0 & k_{\epsilon_{m2}} \end{pmatrix} \right), \quad (3.7)$$

其中 k_{mij} 为过程 f_{mi} 和 f_{mj} 的协方差函数, $k_{\epsilon_{mi}}$ 为误差过程 ϵ_{mi} 的协方差函数. 对于该成组函数型模型, 所有第 3.1 小节中提出的性质仍旧成立.

为方便起见, 将第 m 组中两条曲线的观测数据合并为一个向量, 记作 $\mathbf{X}_m = (\mathbf{X}_{m1}^\top, \mathbf{X}_{m2}^\top)^\top$, $\mathbf{y}_m = (\mathbf{y}_{m1}^\top, \mathbf{y}_{m2}^\top)^\top$. 同时记第 m 组所有协变量上的一个样本实现为

$$f_m(\mathbf{X}_m) = (f_{m1}(\mathbf{X}_{m1})^\top, f_{m2}(\mathbf{X}_{m2})^\top)^\top,$$

则可以将之前给出的分层模型 (3.5) 拓展至成组数据的情形

$$\begin{aligned} f_m(\mathbf{X}_m) &\sim \text{DMTD}(\nu, \omega, 0, \mathbf{K}_m), \\ \mathbf{y}_m | f_m, \mathbf{X}_m &\sim \text{DMTD}(\nu, \omega, f_m(\mathbf{X}_m), \sigma^2 \mathbf{I}), \\ \mathbf{y}_m | \mathbf{X}_m &\sim \text{DMTD}(\nu, \omega, 0, \Sigma_m), \end{aligned} \quad (3.8)$$

其中

$$\begin{aligned} \mathbf{K}_m &= \begin{pmatrix} \mathbf{K}_{m11} & \mathbf{K}_{m12} \\ \mathbf{K}_{m21} & \mathbf{K}_{m22} \end{pmatrix}, \\ \mathbf{K}_{mij} &= \begin{pmatrix} k_{mij}(\mathbf{x}_{mi1} - \mathbf{x}_{mj1}) & \dots & k_{mij}(\mathbf{x}_{mi1} - \mathbf{x}_{mjn}) \\ \vdots & \ddots & \vdots \\ k_{mij}(\mathbf{x}_{min} - \mathbf{x}_{mj1}) & \dots & k_{mij}(\mathbf{x}_{min} - \mathbf{x}_{mjn}) \end{pmatrix}, \\ \Sigma_m &= \mathbf{K}_m + \sigma^2 \mathbf{I}. \end{aligned}$$

记 $\boldsymbol{\theta}_m = (v_{m1}, v_{m2}, w_{m1}, w_{m2}, \mathbf{A}_{m1}, \mathbf{A}_{m2}, \mathbf{B}_{m1}, \mathbf{B}_{m2}, \mu_m)^\top$ 为包含 f_m 中所有超参数的列向量, $\boldsymbol{\beta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_M^\top, \sigma)^\top$ 为模型中所有超参数的向量, 其中 σ 为噪声项的参数. 将向量 $\boldsymbol{\beta}$ 的第 k 个分量记作 β_k . 在给定第 m 组数据的协变量条件下, 第 m 组观测变量服从 DMTD: $\mathbf{y}_m | \mathbf{X}_m \sim \text{DMTD}(\nu, \omega, \mathbf{0}, \Sigma_m)$. 记 $N = n_{m1} + n_{m2}$. 我们可以写出其概率密度函数

$$p_{\boldsymbol{\beta}}(\mathbf{y}_m | \mathbf{X}_m) = |2\pi\omega\Sigma_m|^{-1/2} \frac{\Gamma(N/2 + \nu)}{\Gamma(\nu)} \left(1 + \frac{\mathbf{y}_m^\top \Sigma_m^{-1} \mathbf{y}_m}{2\omega}\right)^{-(N/2 + \nu)}.$$

考虑如下的得分估计方程:

$$\sum_{m=1}^M \frac{\partial \log p_{\boldsymbol{\beta}}(\mathbf{y}_m | \mathbf{X}_m)}{\partial \beta_k} = \frac{1}{2} \sum_{m=1}^M \text{Tr} \left((s_{1m} \Sigma_m^{-1} \mathbf{y}_m (\Sigma_m^{-1} \mathbf{y}_m)^\top - \Sigma_m^{-1}) \frac{\partial \Sigma_m}{\partial \beta_k} \right) = 0, \quad (3.9)$$

其中

$$s_{1m} = \frac{N + 2\nu}{2\omega + \mathbf{y}_m^\top \Sigma_m^{-1} \mathbf{y}_m}.$$

类似相依 Gauss 过程回归参数估计的过程 [4], 我们可以通过求解 (3.9) 来得到参数 $\boldsymbol{\beta}$ 的极大似然估计 $\hat{\boldsymbol{\beta}}$.

需要注意当 $M = 1$ 时, 随机效应 r 仅有一个样本 r_1 , 会与协方差矩阵中的参数混淆在一起, 从而导致无法估计 ν 和 ω [14]. 借鉴文献 [14, 29] 中的方法, 令 $\omega = \nu - 1$, 则有

$$\text{Var}(f_i) = \frac{\omega k_{ii}}{\nu - 1} = k_{ii}.$$

可以看到此时随机函数 f_i 的方差仅与其协方差函数 k_{ii} 有关. 另一方面, 如前文所述,

$$\lim_{\nu \rightarrow \infty} \text{DTP}(\nu, \omega, h, k) = \text{DGP}(h, \lambda k),$$

这意味着当 ν 趋于无穷时, DTP 的表现与 DGP 无异, 因此 DTP 也不存在稳健性的优势. 所以我们倾向于选择一个较小的 ν 值. 根据文献 [14] 可知, $\nu = 1.05$ 是一个兼顾了稳健性与 $\text{Var}(f_i)$ 的定义的选择. 故本文沿用该设定, 当数据组数为 1 时, 令 $\nu = 1.05$, $\omega = \nu - 1 = 0.05$.

当 $M > 1$ 时, ν 可以估计. 参数 ν 的得分估计函数为

$$\begin{aligned} \sum_{m=1}^M \frac{\partial \log p_{\beta}(\mathbf{y}_m | \mathbf{X}_m)}{\partial \nu} &= -\frac{1}{2} \sum_{m=1}^M \left(\frac{N}{\nu-1} + 2 \log \left(1 + \frac{\mathbf{y}_m^{\top} \Sigma_m^{-1} \mathbf{y}_m}{2(\nu-1)} \right) \right. \\ &\quad \left. - \frac{(N+2\nu) \mathbf{y}_m^{\top} \Sigma_m^{-1} \mathbf{y}_m}{2(\nu-1)^2 + (\nu-1) \mathbf{y}_m^{\top} \Sigma_m^{-1} \mathbf{y}_m} - 2\psi \left(\frac{N}{2} + \nu \right) + 2\psi(\nu) \right) \\ &= 0, \end{aligned}$$

其中 $\psi(\cdot)$ 满足 $\psi(t+1) = \psi(t) + 1/t$.

4 稳健性与信息相合性

4.1 稳健性

文献 [14] 证明了 eTPR 模型相比于 GPR 模型对于输出空间上的异常点更加稳健. 直观上看, 在相依过程回归的框架下, 本文提出的 DTPR 模型也应具有稳健性.

为叙述方便, 令组数 $M = 1$, 从而省略下标中表示组数的 m . 对于第 i 条曲线中的协变量 $\mathbf{u}$, 令 $\hat{f}_{iT}(\mathbf{u}) = \hat{\mu}_i^* = \mu_i^*|_{\beta=\hat{\beta}}$ 和 $V_{iT} = \hat{C}_i^* = C_i^*|_{\beta=\hat{\beta}}$ 分别为 DTPR 模型下 $f_i(\mathbf{u})$ 的均值和方差, 其中 $\hat{\beta}$ 为超参数 β 的估计. 我们仅需令 (3.9) 中的 s_1 为常数 1, 便可得到 DGPR 模型下 $f_i(\mathbf{u})$ 的均值 $\hat{f}_{iG}(\mathbf{u})$ 和方差 V_{iG} . 在原假设 $f_i(\mathbf{u}) = f_{0i}(\mathbf{u})$ 下, 定义两个 t 统计量 $M_{iT} = (\hat{f}_{iT}(\mathbf{u}) - f_{0i}(\mathbf{u}))/\sqrt{V_{iT}}$ 和 $M_{iG} = (\hat{f}_{iG}(\mathbf{u}) - f_{0i}(\mathbf{u}))/\sqrt{V_{iG}}$. 对于有界的协方差函数, 当某个观测值 $y_{il} \rightarrow \infty$ 时, 可得 M_{iG} 趋于无穷, 而 M_{iT} 仍是有界的.

对于 DTPR 的得分函数而言 (见附录 A), 当 $y_{il} \rightarrow \infty$ 时, 其影响因子 s_1 限制了其得分并使之保持有界. 然而在 DGPR 中, 影响因子 s_1 为常数 1, 因此当 $y_{il} \rightarrow \infty$ 时其得分趋于 ∞ . 因此在参数估计中以及对新数据点进行预测时, DTPR 会表现得更加稳健.

命题 4.1 若协方差函数 k_{ij} ($i = 1, 2, j = 1, 2$) 有界且关于其参数连续可微, 则给定 ν 时, DTPR 参数的极大似然估计 $\hat{\beta}$ 有有界的影响函数, 而 DGPR 不具有该性质.

4.2 信息相合性

在本小节中, 为方便叙述, 将两条曲线的数据合并记为一个向量, 如 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{n_1+n_2})^{\top}$ 和 $\mathbf{y} = (y_1, \dots, y_{n_1+n_2})^{\top}$. 除非特殊声明, 我们将忽略下标 i , 而只保留表示样本点次序的下标 l .

给定 $\mathbf{X}$ 时, 观测数据 $\mathbf{y}$ 服从联合概率密度函数 $p_{\sigma_0}(\mathbf{y} | f_0, \mathbf{X})$, 其中 σ_0 为参数 σ 的真实值, 而 f_0 为模型 (3.1) 中未知的真实函数. 如 (3.4), 将随机函数向量 $(f_1, f_2)^{\top}$ 的协方差函数简记为 k , 而记 $\boldsymbol{\theta}$ 为 k 中的超参数向量. 对于一组固定的参数 $\boldsymbol{\theta}$, 密度函数 $p_{\boldsymbol{\theta}}(f)$ 可以被看作是随机过程 $f \in \mathcal{F}$ 的测

度, 其中 $\mathcal{F} = \{f(\cdot) : \mathcal{X} \rightarrow \mathbb{R}\}$ 为函数空间. 令

$$p_{\sigma, \theta}(\mathbf{y} | \mathbf{X}) = \int_{\mathcal{F}} p_{\sigma}(\mathbf{y} | f, \mathbf{X}) dp_{\theta}(f)$$

为在 DTPR 模型 (3.3) 下 $\mathbf{y}$ 的条件概率密度函数. 相应地, 在 DTPR 模型下, $p_{\sigma, \hat{\theta}}(\mathbf{y} | \mathbf{X})$ 为给定参数估计量 $\hat{\theta}$ 时的条件概率密度. 我们使用 Kullback-Leibler (KL) 散度来刻画两个概率分布之间的相似度, 其中概率密度 p_1 与 p_2 之间的 KL 散度定义为

$$D[p_1, p_2] = \int (\log p_1 - \log p_2) dp_1.$$

我们给出如下的结论.

命题 4.2 在给定协方差函数 k 下, 记 $\|f_0\|_k$ 为真实函数 f_0 在函数空间中的范数. 假设范数 $\|f_0\|_k$ 有界且关于 $\mathbf{X}$ 分布的期望 $E_{\mathbf{X}}(\log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}|) = o(n_1 + n_2)$, 则在引理 B.1 (附录 B) 中的条件下, 当 $n_1 \rightarrow \infty$ 和 $n_2 \rightarrow \infty$ 时, 有

$$\frac{1}{n_1 + n_2} E_{\mathbf{X}}(D[p_{\sigma_0}(\mathbf{y} | f_0, \mathbf{X}), p_{\sigma_0, \hat{\theta}}(\mathbf{y} | \mathbf{X})]) \rightarrow 0.$$

文献 [30, 31] 分别证明了 Gauss 过程回归模型和广义 Gauss 过程回归模型的信息相合性, 文献 [14] 证明了 eTPR 也满足信息相合性. 我们基于 DTPR 模型, 得到了函数估计的信息相合性.

5 数值实验

5.1 模拟实验

本小节研究相依 Gauss 过程回归、扩展 t 过程回归和本文提出的相依 t 过程回归在模拟数据上的表现. 我们采用如下的模型来生成相依数据:

$$y_i(\mathbf{x}) = h(\mathbf{x}) + f_i(\mathbf{x}) + \epsilon_i(\mathbf{x}), \quad i = 1, 2,$$

其中两组数据共用同一个均值函数 $h \sim \text{GP}(0, k_0)$, 模型中 (f_1, f_2) 为相依 Gauss 过程

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \sim \text{DGP} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \right),$$

并且误差项服从独立的正态分布 $\epsilon_i \sim N(0, \sigma^2)$. 取协方差函数 k_0 为

$$k_0(\mathbf{d}) = \frac{\pi^{\frac{1}{2}} v_0^2}{\sqrt{\mathbf{A}_0}} \exp \left(-\frac{1}{4} \mathbf{d}^\top \mathbf{A}_0 \mathbf{d} \right),$$

而 (f_1, f_2) 的核函数 k_{ij} 取第 2 节引入的 Gauss 核函数. 记模型中包含的参数为 $\boldsymbol{\beta} = (v_0, \mathbf{A}_0, v_1, v_2, w_1, w_2, \mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2, \mu, \sigma)^\top$. 我们在生成的两条曲线上各取 100 个样本点, 即 $n \doteq n_1 = n_2 = 100$, 其中协变量 $\mathbf{x}$ 在实数轴 $[-1.0, 1.0]$ 上等间隔分布. 在每组实验中, 我们从每条曲线的 n 个样本点中等间隔地再取 n_{train} 个点作为训练数据, 并将其余的样本点作为测试数据. 这里分别考虑训练集大小 n_{train} 为 20、30 和 40 的情形. 我们分别给出如下三组不同的参数设定:

- (1) $\boldsymbol{\beta} = (1.5, \exp(3.5), 0.3, 0.3, 0.1, 0.1, \exp(3), \exp(3), \exp(1), \exp(1), 0, 0.2);$

(2) $\beta = (2, \exp(2), 0.2, 0.2, 0.1, 0.1, \exp(2), \exp(2), \exp(1), \exp(1), 0, 0.2);$

(3) $\beta = (3, \exp(4), 0.2, 0.2, 0.3, 0.3, \exp(2), \exp(2), \exp(3), \exp(3), 0, 0.2).$

本小节的模拟实验均重复 500 次.

为了体现模型的稳健性, 我们在每条曲线的训练集上各添加了一个离群点. 在每一次重复实验中, 首先从均匀分布 $U(0.2, 0.8)$ 中抽取两个样本 α_1 和 α_2 . 这里 α_i ($i = 1, 2$) 表示异常点在数据中的位置, 我们将 α_i 乘以 n_{train} 再取整, 便得到第 i 条曲线的一个训练数据的下标. 得到下标后, 我们在该训练数据的观测值上添加 1.5 的扰动 ε , 使该点成为数据中的离群点. 值得注意的是, DGPR 和 DTPR 两个相依模型是基于两条曲线的数据共同训练的, 而独立的 eTPR 模型只能分别对于两条曲线的训练集进行训练. 我们分别用均方误差 (mean squared error, MSE) $\text{MSE} = \sum_{i=1}^2 \sum_{l=1}^{n_{\text{test}}} (\hat{f}_i(\mathbf{x}_{il}^*) - f_{0i}(\mathbf{x}_{il}^*))^2 / (2n_{\text{test}})$ 和预测误差 (prediction error, PE) $\text{PE} = \sum_{i=1}^2 \sum_{l=1}^{n_{\text{test}}} (\hat{f}_i(\mathbf{x}_{il}^*) - y_{il}^*)^2 / (2n_{\text{test}})$ 两个指标对三个模型进行比较, 其中 $\{\mathbf{x}_{il}^*, y_{il}^*\}$ ($i = 1, 2, l = 1, \dots, n_{\text{test}}$) 是测试数据. 具体的实验结果见表 1. 可以发现 DTPR 在所有实验中均有最小的均方误差和预测误差, 而 eTPR 的均方误差和预测误差大体上优于 DGPR 模型. 由于 DGPR 模型对于异常点不具有稳健性, 因此, 在这组添加离群点的实验中, DGPR 的结果表现最不好. 而 DTPR 在考虑到两组数据的相依性的同时, 也表现出对离群点的稳健性, 故在多组实验中均有较小的误差. 我们也在表 1 记录了在不同模拟数据下三个模型计算的平均耗时. 可以看出, DTPR 耗时最长, DGPR 耗时最短.

我们也研究了缺失数据下三个模型的表现. 在本实验中, 取噪声项的参数 $\sigma = 0.1$ 且不对数据人为地添加离群点. 其余的模型及参数设定均与上个实验相同, 故不再进行赘述. 在确定训练数据和测试数据后, 我们删除了第二条曲线的部分训练数据. 将下标位于 $0.3 \times n_{\text{train}}$ 和 $0.7 \times n_{\text{train}}$ 的数据点从训练数据中删去. 具体的示例可以参照图 1(b), 不过与图 1 不同的是, 本实验中并没有添加离群点. 本组实验的预测误差、均方误差和模型计算的平均时间见表 2. 可以发现 DTPR 的表现略好于 DGPR, 而 eTPR 在对第二条曲线的测试数据进行预测时无法参考第一条曲线的信息, 因此在缺失数据部分有较大的偏差. 计算平均耗时与之前的模型设置下的结果类似, DTPR 耗时最长, DGPR 耗时最短.

考虑到真实的数据往往并不是稠密排列的, 我们研究了在稀疏不规则的采样数据点下三个模型的表现. 这里以参数设定 (1) 的情形为例说明此问题. 令在离群点上扰动的值 ε 为 1.5 不变, 分别令 n_{train}

表 1 在存在异常点的相依数据上, DGPR、eTPR 和 DTPR 的预测误差、均方误差、平均训练耗时及其标准差 (括号内)

n_{train}	模型	预测误差			均方误差			平均耗时 (秒)		
		DGPR	eTPR	DTPR	DGPR	eTPR	DTPR	DGPR	eTPR	DTPR
20	(1)	13.3 (0.027)	10.7 (0.032)	8.3 (0.017)	9.3 (0.026)	6.6 (0.030)	4.3 (0.015)	0.461	0.580	1.924
	(2)	9.4 (0.021)	8.5 (0.025)	7.1 (0.014)	5.4 (0.020)	4.5 (0.023)	3.1 (0.012)	0.363	0.609	2.149
	(3)	19.5 (0.092)	14.5 (0.097)	10.0 (0.019)	15.5 (0.092)	10.5 (0.095)	6.0 (0.017)	0.309	0.558	1.305
30	(1)	8.9 (0.017)	7.4 (0.015)	6.2 (0.010)	4.9 (0.015)	3.4 (0.012)	2.2 (0.007)	0.791	0.996	3.359
	(2)	6.7 (0.014)	6.3 (0.013)	5.6 (0.009)	2.7 (0.011)	2.3 (0.012)	1.6 (0.007)	0.545	0.899	2.939
	(3)	10.9 (0.020)	8.8 (0.022)	7.2 (0.012)	6.9 (0.018)	4.8 (0.020)	3.2 (0.009)	0.490	1.244	1.777
40	(1)	7.0 (0.011)	6.2 (0.012)	5.5 (0.009)	3.0 (0.009)	2.2 (0.008)	1.5 (0.005)	1.157	1.585	4.909
	(2)	5.8 (0.010)	5.8 (0.051)	5.1 (0.008)	1.7 (0.006)	1.8 (0.049)	1.0 (0.004)	0.794	1.340	4.706
	(3)	8.4 (0.015)	7.1 (0.016)	6.2 (0.010)	4.3 (0.012)	3.1 (0.014)	2.1 (0.006)	0.757	1.998	2.869

注: 表中的预测误差和均方误差均为乘以 100 后的数值.

表 2 在有缺失的相依数据上, DGPR、eTPR 和 DTPR 的预测误差、均方误差、平均训练耗时及其标准差 (括号内)

n_{train}	模型	预测误差			均方误差			平均耗时 (秒)		
		DGPR	eTPR	DTPR	DGPR	eTPR	DTPR	DGPR	eTPR	DTPR
20	(1)	2.3 (0.011)	12.3 (0.132)	1.8 (0.007)	1.3 (0.010)	11.4 (0.132)	0.8 (0.007)	0.295	0.424	1.299
	(2)	1.9 (0.008)	6.6 (0.089)	1.5 (0.005)	0.9 (0.007)	5.6 (0.090)	0.5 (0.004)	0.512	0.659	2.117
	(3)	4.3 (0.023)	41.2 (0.564)	2.8 (0.014)	3.3 (0.023)	40.2 (0.565)	1.8 (0.014)	0.749	0.960	2.911
30	(1)	1.9 (0.008)	10.0 (0.101)	1.5 (0.004)	0.9 (0.007)	9.0 (0.102)	0.5 (0.004)	0.239	0.457	1.105
	(2)	1.7 (0.012)	3.8 (0.038)	1.4 (0.005)	0.7 (0.013)	2.8 (0.039)	0.4 (0.004)	0.361	0.572	1.666
	(3)	3.7 (0.024)	32.7 (0.352)	2.4 (0.011)	2.8 (0.023)	31.7 (0.351)	1.4 (0.011)	0.565	0.799	2.424
40	(1)	1.7 (0.014)	8.8 (0.083)	1.4 (0.003)	0.7 (0.014)	7.8 (0.082)	0.4 (0.003)	0.209	0.316	0.717
	(2)	1.6 (0.009)	3.2 (0.036)	1.3 (0.003)	0.6 (0.009)	2.2 (0.036)	0.3 (0.002)	0.369	0.723	1.148
	(3)	3.4 (0.021)	29.7 (0.308)	2.1 (0.010)	2.4 (0.021)	28.7 (0.307)	1.1 (0.009)	0.568	1.100	1.635

注: 表中的预测误差和均方误差均为乘以 100 后的数值.

为 11 和 6 以获得更为稀疏的训练样本. 在 $n_{\text{train}} = 11$ 时, 我们分别取下标 1, 10, 20, $\dots$, 90, 100 作为训练数据下标初始集合. 为了使得样本点的采样不规则, 我们对除 1 和 100 以外的其他 9 个下标加上一个随机数作为下标的扰动, 该随机数独立同分布采样于离散均匀分布 $\text{DiscreteU}(-3, 3)$. 在每次重复实验中, 对每一条曲线都这样随机生成一组下标集合, 抽取这些下标所对应的样本点作为训练数据, 其余的数据作为测试数据. 而在 n_{train} 为 6 时, 我们取 1、20、40、60、80 和 100 作为训练集下标的初始集合. 类似地, 我们在 20、40、60 和 80 上分别加一个同样的随机数, 得到稀疏不规则的下标. 重复 500 次实验后得到的预测误差和均方误差结果见表 3 中的模型设定 M_1 . DTPR 的预测误差和均方误差优于 eTPR, eTPR 的结果优于 DGPR.

对于不同离群点的数目以及不同离群点上的扰动值 ε 大小对模型稳健性的影响, 我们仍取 $n_{\text{train}} = 20$, 数据仍由参数设定 (1) 生成, 并沿用生成离群点的方法, 在每条曲线上生成 n_{outliers} 个离群点. 对于离群点, 它们的响应值加以扰动 ε . 在表 3 中, 模型 M_2 展示了 $\varepsilon = 1.5$ 以及离群点数目 n_{outliers} 分别取 1、2 和 3 时三个模型的预测误差和均方误差的结果. 另外, 保持 n_{outliers} 为 1, 仅改变扰动值的大

表 3 在不同设定下的相依数据上, DGPR、eTPR 和 DTPR 的预测误差和均方误差及其标准差 (括号内)

模型设定		预测误差			均方误差		
		DGPR	eTPR	DTPR	DGPR	eTPR	DTPR
M_1	$n_{\text{train}} = 6$	76.0 (0.611)	48.7 (0.177)	39.7 (0.320)	72.1 (0.614)	44.7 (0.175)	35.7 (0.324)
	$n_{\text{train}} = 11$	68.4 (0.452)	28.2 (0.104)	17.1 (0.058)	64.4 (0.454)	24.3 (0.103)	13.1 (0.057)
M_2	$n_{\text{outliers}} = 1$	13.1 (0.025)	10.6 (0.029)	8.3 (0.017)	9.1 (0.025)	6.6 (0.027)	4.3 (0.015)
	$n_{\text{outliers}} = 2$	37.5 (0.131)	27.2 (0.135)	18.5 (0.058)	33.5 (0.128)	23.2 (0.134)	14.5 (0.057)
	$n_{\text{outliers}} = 3$	56.2 (0.166)	40.2 (0.178)	28.1 (0.079)	52.2 (0.166)	36.3 (0.178)	24.1 (0.077)
M_3	$\varepsilon = 1.5$	13.1 (0.025)	10.6 (0.029)	8.3 (0.017)	9.1 (0.025)	6.6 (0.027)	4.3 (0.015)
	$\varepsilon = 3$	34.0 (0.084)	21.8 (0.096)	16.0 (0.042)	30.1 (0.082)	17.8 (0.095)	12.0 (0.040)
	$\varepsilon = 10$	316.1 (0.983)	126.7 (1.338)	60.5 (0.232)	312.2 (0.982)	122.8 (1.340)	56.4 (0.231)

注: M_1 为稀疏不规则训练数据的情形, M_2 为考虑不同数目的离群点 n_{outliers} 的情形, M_3 为考虑离群点上不同扰动值 ε 的情形. 表中的预测误差和均方误差均为乘以 100 后的数值.

小, 分别令 ε 等于 1.5、3 和 10 以研究不同扰动程度的离群点对于模型稳健性的影响. 3 个模型在不同离群点扰动下的表现见表 3 中的模型设定 M_3 . 可以看出, 无论在不同数目的离群点下, 还是在不同扰动程度的离群点下, DTPR 均有着最小的预测误差和均方误差, 表现出更强的稳健性.

现在研究噪声项服从不同分布时 3 个模型的表现情形. 仍旧保持参数设定 (1) (除了 σ), 等间隔地选取 $n_{\text{train}} = 20$ 个训练样本, 并将其余 80 个样本作为测试数据. 从每条曲线的训练样本中随机选取一个样本点作为离群点, 添加 1.5 的扰动值. 我们分别假定噪声项服从正态分布、 t 分布和卡方分布. 3 个模型的预测误差和均方误差见表 4. 可以看出, DTPR 仍然在多种噪声分布设定下具有最小的预测误差和均方误差.

5.2 真实案例

本小节将 3 个模型应用于澳大利亚的死亡率数据. 澳大利亚的死亡率数据由澳大利亚统计局 (Australian Bureau of Statistics) 整理, 并且可以直接从 R 包 “fda” 中获取. 这份数据集记录了自 1901 年至 2003 年间澳大利亚及部分州的死亡率, 并分别记录了每年由新生儿至 100 岁不同年龄段的对数形式的死亡率. 在本文中, 这份数据集的男性和女性的死亡率数据被看作是相依的函数型响应变量. 我们研究了不同模型在以下地区的数据集上的表现: 澳大利亚 (AUS)、新南威尔士州 (NSW)、昆士兰州 (QLD)、维多利亚州 (VIC)、西澳大利亚 (WA)、南澳大利亚 (SA)、北领地 (NT)、塔斯马尼亚岛 (TAS) 和澳大利亚首都领地 (ACT). 我们在男性和女性数据中各随机选取了 20% 的观测数据作为训练集, 并将其余的数据作为测试集. 我们选取了数据集中 2001 年至 2003 年的数据作为组数为 3 的成组数据. 使用 DGPR、DTPR 和 eTPR 对训练数据进行拟合, 并计算模型在测试数据上的预测误差. 此过程重复 100 次.

图 2 展示了澳大利亚死亡率数据的其中一次实验的结果. 图 2(a) 展现了澳大利亚 2003 年的男性和女性随年龄变化的对数死亡率数据. 可以观察到对数死亡率的总体趋势是随着年龄增长而上升的, 但是新生儿有着较高的对数死亡率. 图 2(b) 为随机采样得到的训练数据, 其中女性对数死亡率的训练数据 (实心点) 并没有采集到新生儿的数据, 而男性对数死亡率的训练数据 (空心点) 则采集到了年龄为 0 的样本点. 图 2(c) 分别为三个模型对女性对数死亡率数据的拟合情形, 实心点为女性对数死亡率测试数据. 可以看出, eTPR 由于仅考虑了女性的数据, 因此并没有从训练集中获得关于新生儿对数死

表 4 在噪声项服从不同的分布时, DGPR、eTPR 和 DTPR 的预测误差和均方误差及其标准差 (括号内)

噪声项服从分布	预测误差			均方误差		
	DGPR	eTPR	DTPR	DGPR	eTPR	DTPR
$N(0, 0.1)$	8.6 (0.021)	6.1 (0.017)	4.6 (0.011)	7.6 (0.020)	5.1 (0.017)	3.6 (0.010)
$N(0, 0.2)$	13.1 (0.025)	10.6 (0.029)	8.3 (0.017)	9.1 (0.025)	6.6 (0.027)	4.3 (0.015)
$N(0, 0.3)$	21.2 (0.038)	18.3 (0.052)	14.5 (0.026)	12.3 (0.035)	9.3 (0.048)	5.5 (0.022)
$t(5)$	274.2 (0.652)	231.3 (0.674)	199.5 (0.436)	106.4 (0.522)	63.3 (0.539)	30.6 (0.185)
$t(10)$	210.4 (0.392)	175.1 (0.387)	152.1 (0.236)	82.8 (0.297)	47.9 (0.303)	25.5 (0.126)
$t(30)$	175.4 (0.271)	150.8 (0.322)	129.5 (0.182)	69.5 (0.224)	43.5 (0.277)	22.8 (0.115)
$\chi^2(1)$	328.5 (1.016)	267.8 (0.837)	240.7 (0.678)	241.4 (1.189)	159.8 (0.833)	138.7 (0.616)
$\chi^2(3)$	972.6 (2.222)	761.6 (2.114)	708.8 (1.418)	1317.8 (3.599)	1033.5 (3.051)	989.9 (2.537)
$\chi^2(5)$	1605.9 (3.330)	1234.1 (2.738)	1173.6 (2.025)	3197.1 (7.153)	2704.8 (5.885)	2648.6 (5.559)

注: 表中的预测误差和均方误差均为乘以 100 后的数值.

亡率的信息, 其拟合曲线 (点线) 在最左端不能够贴近实际的趋势. 此外, 女性对数死亡率数据在横坐标为 20 的附近非常稀疏, 而 DGPR 和 DTPR 的拟合曲线 (分别为短划线和实线) 在横坐标为 20 处均有波动, DGPR 甚至出现了明显的上升趋势. 这与图 1 所展示的示例十分相似, 在对女性数据建模时, DGPR 和 DTPR 会参考男性数据的趋势, 而男性对数死亡率在横坐标为 20 处有一个噪声较大的训练样本点 (图 2(b) 中标记为 1 的点), 影响了 DGPR 和 DTPR 的预测. 凭借稳健性, DTPR 受到该点的影响较小, 拟合出的曲线更贴近实际的趋势.

表 5 记录了基于 100 次的重复实验下预测误差的平均值. 可以看到 DTPR 的表现明显优于 DGPR 和 eTPR, 而除了在 TAS 和 ACT 地区, DGPR 的表现大体好于 eTPR.

图 2 (a) 澳大利亚男性和女性随年龄变化的对数死亡率数据; (b) 采样得到的对数死亡率训练数据; (c) 女性对数死亡率测试数据, DTPR、DGPR 和 eTPR 的拟合曲线

表 5 依年龄划分的澳大利亚死亡率数据集上, DGPR、eTPR 和 DTPR 的预测误差

地区	DGPR	eTPR	DTPR
AUS	0.176	0.428	0.138
NSW	0.289	0.441	0.217
QLD	0.253	0.431	0.196
VIC	0.248	0.412	0.204
WA	0.288	0.378	0.224
SA	0.314	0.361	0.273
NT	0.418	0.435	0.393
TAS	0.328	0.317	0.280
ACT	0.366	0.302	0.278

6 结论

对于多元函数型响应数据而言, 常规的独立过程回归如 GPR 和 eTPR 会忽略响应变量之间的相关结构, 从而损失数据中的大量信息. 利用多个输出的 DGPR 模型可以解决这个问题, 相比诸如 GPR 的传统方法, 在相依数据上获得了更好的表现. 然而, DGPR 与 GPR 一样, 对异常点不具有稳健性. 本文提出了 DTPR 模型并证明了 DTPR 模型保留了 eTPR 模型和 DGPR 模型很多良好的性质.

我们在核函数中引入一个随机效应, 并通过卷积构造一个相依 t 过程. 我们证明了通过设计共享的随机函数, 模型便可以捕捉到函数型响应变量之间的相关性. DTPR 模型在考虑到响应变量之间的相依性的同时也具有良好的稳健性, 所以在有缺失数据和含异常点数据上表现得更好. 本文的数值实验也证明了这一点, DTPR 模型在几组实验中均给出了更准确的预测. 本文假设回归函数和误差函数为一个联合 DTP. 文献 [14] 证明了在联合 t 过程下, 回归函数与误差函数虽然不相关, 但却是不独立的. 当回归函数与误差函数为相互独立的 DTP 时, 估计过程将包含复杂的积分, 我们将在未来的工作中对其进行研究.

致谢 作者对主编、副主编以及三位审稿人的宝贵意见致以诚挚的感谢.

参考文献

- 1 Rasmussen C E, Williams C K I. Gaussian Processes for Machine Learning. Cambridge: MIT Press, 2006
- 2 Mackay D J. Gaussian processes—A replacement for supervised neural networks? Lecture notes for a tutorial at Neural Information Processing Systems. <http://www.inference.org.uk/mackay/abstracts/gp.html>, 1997
- 3 Williams C K I, Rasmussen C E. Gaussian processes for regression. In: Advances in Neural Processing Systems 8. Cambridge: MIT Press, 1996, 514–520
- 4 Boyle P, Frean M. Dependent Gaussian processes. In: Advances in Neural Information Processing Systems 17. Cambridge: MIT Press, 2004, 217–224
- 5 Alvarez M A, Luengo D, Titsias M K, et al. Efficient multioutput Gaussian processes through variational inducing kernels. J Mach Learn Res, 2010, 9: 25–32
- 6 Nguyen V T, Bonilla E. Collaborative multi-output Gaussian processes. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence. Quebec: AUAI Press, 2014, 643–652
- 7 Liu H T, Cai J F, Ong Y. Remarks on multi-output Gaussian process regression. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence. Monterey: AUAI Press, 2014, 643–652
- 8 Neal R M. Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. Technical Report 9702. arXiv:physics/9701026, 1997
- 9 Vanhatalo J, Jylänki P, Vehtari A. Gaussian process regression with Student-t likelihood. In: Advances in Neural Information Processing Systems 22. Cambridge: MIT Press, 2009, 1910–1918
- 10 Malte K. Gaussian process models for robust regression, classification, and reinforcement learning. PhD Thesis. Darmstadt: Darmstadt University of Technology, 2006
- 11 Naishguzman A, Holden S B. Robust regression with twinned Gaussian processes. In: Neural Information Processing Systems. New York: Curran Associates, 2007, 1065–1072
- 12 Yu S P, Tresp V, Yu K. Robust multi-task learning with t -processes. In: Proceedings of the 24th International Conference on Machine Learning. New York: Association for Computing Machinery, 2007, 1103–1110
- 13 Zhang Y, Yeung D. Multi-task learning using generalized t process. J Mach Learn Res, 2010, 9: 964–971
- 14 Wang Z F, Shi J Q, Lee Y. Extended t -process regression models. J Statist Plann Inference, 2017, 189: 38–60
- 15 Wang Z F, Maengseok N, Lee Y, et al. A general robust t -process regression model. Comput Statist Data Anal, 2021, 154: 107093
- 16 Cao C Z, Shi J Q, Lee Y. Robust functional regression model for marginal mean and subject-specific inferences. Stat Methods Med Res, 2018, 27: 3236–3254
- 17 Shah A, Wilson A G, Ghahramani Z. Student- t processes as alternatives to Gaussian processes. J Mach Learn Res, 2014, 9: 877–885
- 18 Cui H J. On asymptotics of t -type regression estimation in multiple linear model. Sci China Ser A, 2004, 47: 628–639
- 19 He X M, Cui H J, Simpson D G. Longitudinal data analysis using t -type regression. J Statist Plann Inference, 2004,

- 122: 253–269
- 20 Maronna R A, Yohai V J. Robust functional linear regression based on splines. *Comput Statist Data Anal*, 2013, 65: 46–55
- 21 Shin H, Lee S. An RKHS approach to robust functional linear regression. *Statist Sinica*, 2016, 26: 255–272
- 22 Kalogridis I, Stefan V A. Robust functional regression based on principal components. *J Multivariate Anal*, 2019, 173: 393–415
- 23 Zhu H X, Brown P J, Morris J S. Robust, adaptive functional regression in functional mixed model framework. *J Amer Statist Assoc*, 2011, 106: 1167–1179
- 24 Qin G Y, Bai Y, Zhu Z Y. Robust empirical likelihood inference for generalized partial linear models with longitudinal data. *J Multivariate Anal*, 2012, 105: 32–44
- 25 Qin G Y, Bai Y, Zhu Z Y. Robust empirical likelihood inference for longitudinal data. *Statist Probab Lett*, 2009, 79: 2101–2108
- 26 Fan Y L, Qin G Y, Zhu Z Y. Variable selection in robust regression models for longitudinal data. *J Multivariate Anal*, 2012, 109: 156–167
- 27 Higdon D. Space and space-time modeling using process convolutions. In: *Quantitative Methods for Current Environmental Issues*. London: Springer, 2002, 37–56
- 28 Robinson G. That BLUP is a good thing: The estimation of random effects. *Statist Sci*, 1991, 6: 15–32
- 29 Lee Y, Nelder J A. Double hierarchical generalized linear models (with discussion). *J Roy Statist Soc Ser C*, 2006, 55: 139–185
- 30 Seeger M W, Kakade S M, Foster D P. Information consistency of nonparametric Gaussian process methods. *IEEE Trans Inform Theory*, 2008, 54: 2376–2382
- 31 Wang B, Shi J Q. Generalized Gaussian process regression model for non-Gaussian functional data. *J Amer Statist Assoc*, 2014, 109: 1123–1133
- 32 Frank R H, Elvezio R, Peter J R, et al. *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley & Sons, 1986

附录 A 稳健性

命题 4.1 的证明与文献 [14] 中关于稳健性的证明相似, 主要的差别在于本文将成组数据 $\{y_{11}, \dots, y_{mn}\}$ 替换成了两条相依曲线数据下的情形 $\{y_{111}, \dots, y_{M2n}\}$. 考虑第 3.2 小节讨论的成组数据, 并沿用其记号. 令 $s_T(\beta; \mathbf{y}_1, \dots, \mathbf{y}_M) = (s_1(\beta; \mathbf{y}_1, \dots, \mathbf{y}_M), \dots, s_{M+1}(\beta; \mathbf{y}_1, \dots, \mathbf{y}_M))^T$ 为 DTPR 模型下 β 的得分函数, 其中

$$s_k(\beta; \mathbf{y}_1, \dots, \mathbf{y}_M) = \frac{\partial l(\beta; \nu)}{\partial \beta_k} = \frac{1}{2} \sum_{m=1}^M \text{Tr} \left((s_{1m} \Sigma_m^{-1} \mathbf{y}_m (\Sigma_m^{-1} \mathbf{y}_m)^T - \Sigma_m^{-1}) \frac{\partial \Sigma_m}{\partial \beta_k} \right), \quad (\text{A.1})$$

这里 β_k 为 $\beta = (\theta_1, \dots, \theta_M, \sigma)^T$ 的第 k 个分量.

在 DGPR 模型中, 得分函数的形式与 (A.1) 相同, 只不过其中的影响因子 $s_{1m} = 1$. 当对于某个 i 和 l 有 $y_{mil} \rightarrow \infty$ 时, DGPR 模型的得分函数趋于无穷. 而对于 DTPR 模型而言, 其影响因子为 $s_{1m} = (N + 2\nu)/(2\omega + \mathbf{y}_m^T \Sigma_m^{-1} \mathbf{y}_m)$. 故当某个观测值 $y_{mil} \rightarrow \infty$ 时, 影响因子分母上的 $\mathbf{y}_m^T \Sigma_m^{-1} \mathbf{y}_m$ 会影响得分函数, 使得它是有界的.

根据文献 [32], 对于给定的参数 ν , 参数 β 的影响函数为

$$\text{IF}(\mathbf{y}; \hat{\beta}, F) = - \left(\text{E} \left(\frac{\partial^2 l(\beta; \nu)}{\partial \beta \partial \beta^T} \right) \right)^{-1} s_T(\beta; \mathbf{y}),$$

其中 $l(\beta; \nu)$ 为在给定 ν 下 β 的边际对数似然, 形式为

$$l(\beta; \nu) = \sum_{m=1}^M \left\{ -\frac{N}{2} \log(2\pi(\nu - 1)) - \frac{1}{2} \log |\Sigma_m| - \left(\frac{N}{2} + \nu \right) \log \left(1 + \frac{\mathbf{y}_m^T \Sigma_m^{-1} \mathbf{y}_m}{2(\nu - 1)} \right) \right\}$$

$$+ \log \left(\Gamma \left(\frac{N}{2} + \nu \right) \right) - \log(\Gamma(\nu)) \Big\}. \quad (\text{A.2})$$

对 (A.2) 求关于 β_k 的二阶偏导, 有

$$\begin{aligned} \frac{\partial^2 l(\beta; \nu)}{\partial \beta_k \partial \beta_k} &= \frac{1}{2} \sum_{m=1}^M \text{Tr} \left((s_{1m} \Sigma_m^{-1} \mathbf{y}_m (\Sigma_m^{-1} \mathbf{y}_m)^\top - \Sigma_m^{-1}) \left(\frac{\partial^2 \Sigma_m}{\partial \beta_k \partial \beta_k} - \frac{\partial \Sigma_m}{\partial \beta_k} \Sigma_m^{-1} \frac{\partial \Sigma_m}{\partial \beta_k} \right) \right) \\ &\quad - \frac{1}{2} \text{Tr} \left(s_{1m} \Sigma_m^{-1} \mathbf{y}_m (\Sigma_m^{-1} \mathbf{y}_m)^\top \frac{\partial \Sigma_m}{\partial \beta_k} \Sigma_m^{-1} \frac{\partial \Sigma_m}{\partial \beta_k} \right) \\ &\quad + \frac{1}{2} \frac{s_{1m}^2}{N + 2\nu} \left\{ \text{Tr} \left(\Sigma_m^{-1} \mathbf{y}_m (\Sigma_m^{-1} \mathbf{y}_m)^\top \frac{\partial \Sigma_m}{\partial \beta_k} \right) \right\}^2. \end{aligned}$$

因为 $\partial^2 l(\beta; \nu) / \partial \beta \partial \beta^\top$ 是有界的, 故影响函数 $\text{IF}(\mathbf{y}; \hat{\beta}, F)$ 有界. 相反地, 在 DGPR 模型下, 由于得分函数无界, 故其影响函数也无界.

附录 B 信息相合性

由于 DTPR 与文献 [14] 中的 eTPR 相似, 我们将给出关于命题 4.2 的一个总结性的证明, 其中的细节可参见文献 [14]. 为了证明命题 4.2, 我们先给出如下的引理.

引理 B.1 假定 DTP (3.3) 中的均值函数 $h_1(\mathbf{x}) = h_2(\mathbf{x}) = 0$, $\mathbf{y} = (y_1, \dots, y_{n_1+n_2})^\top$ 为 DTP 下的一组观测值. 记两条曲线的协变量为 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{n_1+n_2})^\top$. 假设协方差函数 k_{ij} 有界且关于其参数 θ 连续, 同时当 $n_1 \rightarrow \infty$ 和 $n_2 \rightarrow \infty$ 时, β 的估计量 $\hat{\beta}$ 几乎一致收敛于 β . 对于某个常数 c 和任意的正数 $\varepsilon > 0$, 有

$$\begin{aligned} &\frac{1}{n_1 + n_2} (-\log p_{\sigma_0, \hat{\theta}}(\mathbf{y} | \mathbf{X}) + \log p_{\sigma_0}(\mathbf{y} | f_0, \mathbf{X})) \\ &\leq \frac{1}{n_1 + n_2} \left\{ \frac{1}{2} \log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}| + \frac{q^2 + 2(\nu - 1)}{2(N + 2\nu - 2)(\|f_0\|_k^2 + c) + c} + \varepsilon \right\}, \end{aligned}$$

其中 $\mathbf{I}$ 为 $(n_1 + n_2) \times (n_1 + n_2)$ 的单位阵, $q^2 = (\mathbf{y} - f_0(\mathbf{X}))^\top (\mathbf{y} - f_0(\mathbf{X})) / \sigma_0^2$.

证明 记 r 为服从逆伽马分布 $\text{IG}(\nu, \nu - 1)$ 的一个随机效应. 给定 r , 我们有

$$\begin{pmatrix} f_1 \\ f_2 \\ \epsilon_1 \\ \epsilon_2 \end{pmatrix} \Bigg| r \sim \text{DGP} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, r \begin{pmatrix} k_{11} & k_{12} & 0 & 0 \\ k_{21} & k_{22} & 0 & 0 \\ 0 & 0 & k_{\epsilon_1} & 0 \\ 0 & 0 & 0 & k_{\epsilon_2} \end{pmatrix} \right),$$

其中 $\text{DGP}(h, k)$ 为均值函数为 h 、协方差函数为 k 的相依 Gauss 过程^[4].

令

$$\begin{aligned} \tilde{f} &= \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \Bigg| r \sim \text{DGP} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, r \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \right), \\ \tilde{\epsilon} &= \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \Bigg| r \sim \text{DGP} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, r \begin{pmatrix} k_{\epsilon_1} & 0 \\ 0 & k_{\epsilon_2} \end{pmatrix} \right). \end{aligned}$$

可以看出 $\tilde{f}$ 与 $\tilde{\epsilon}$ 相互独立. 因此得到 DGPR 模型

$$y(\mathbf{x}) = \tilde{f}(\mathbf{x}) + \tilde{\epsilon}(\mathbf{x}), \quad (\text{B.1})$$

其中

$$y(\mathbf{x}) = \begin{pmatrix} y_1(\mathbf{x}_1) \\ y_2(\mathbf{x}_2) \end{pmatrix}$$

且 $\mathbf{x}_i \in \mathbf{X}_i$. 基于 (B.1), 令

$$p_G(\mathbf{y} | r, \mathbf{X}) = \int_{\mathcal{F}} p_{\sigma_0}(\mathbf{y} | \tilde{f}, r, \mathbf{X}) d\tilde{p}_{\theta}(\tilde{f}), \quad (\text{B.2})$$

$$p_0(\mathbf{y} | r, \mathbf{X}) = p_{\sigma_0}(\mathbf{y} | f_0, r, \mathbf{X}), \quad (\text{B.3})$$

其中 p_0 代表真实模型和真实参数下的概率, p_{σ_0} 为给定噪声项真实参数下的概率, $\tilde{p}$ 为随机过程 $\tilde{f} \in \mathcal{F}$ 的测度.

对随机效应积分, 我们可以得到 DTPR 下的条件概率密度

$$p_{\sigma_0, \hat{\theta}}(\mathbf{y} | \mathbf{X}) = \int p_G(\mathbf{y} | r, \mathbf{X}) g(r) dr, \quad (\text{B.4})$$

$$p_{\sigma_0}(\mathbf{y} | f_0, \mathbf{X}) = \int p_0(\mathbf{y} | r, \mathbf{X}) g(r) dr, \quad (\text{B.5})$$

其中 $g(r)$ 为逆伽马分布 $\text{IG}(\nu, \nu - 1)$ 的概率密度函数.

不妨先假设

$$-\log p_G(\mathbf{y} | r, \mathbf{X}) + \log p_0(\mathbf{y} | r, \mathbf{X}) \leq \frac{1}{2} \log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}| + \frac{r}{2} (\|f_0\|_k^2 + c) + c + (n_1 + n_2)\varepsilon \quad (\text{B.6})$$

成立. 如文献 [14] 所述, 不等式 (B.6) 的证明过程与文献 [30, 31] 中的过程类似, 因此这里忽略对其的证明, 证明的细节可参见文献 [14]. 对 (B.6) 稍作变换可得

$$\begin{aligned} -\log \int p_G(\mathbf{y} | r, \mathbf{X}) g(r) dr &\leq \frac{1}{2} \log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}| + c + (n_1 + n_2)\varepsilon \\ &\quad - \log \int p_0(\mathbf{y} | r, \mathbf{X}) \exp \left\{ - \left(\frac{r}{2} (\|f_0\|_k^2 + c) \right) \right\} g(r) dr. \end{aligned} \quad (\text{B.7})$$

记 $g^*(r)$ 为给定数据 $\mathcal{D}$ 下随机效应 r 的条件概率密度, 其分布为 $\text{IG}(\nu + (n_1 + n_2)/2, (\nu - 1) + q^2/2)$. 我们有

$$\begin{aligned} &\int p_0(\mathbf{y} | r, \mathbf{X}) \exp \left\{ - \left(\frac{r}{2} (\|f_0\|_k^2 + c) \right) \right\} g(r) dr \\ &= \int p_0(\mathbf{y} | r, \mathbf{X}) g(r) dr \int \exp \left\{ - \left(\frac{r}{2} (\|f_0\|_k^2 + c) \right) \right\} g^*(r) dr. \end{aligned} \quad (\text{B.8})$$

结合 (B.4)、(B.5)、(B.7) 和 (B.8), 可得

$$-\log p_{\sigma_0, \hat{\theta}}(\mathbf{y}, |\mathbf{X}) \leq \frac{1}{2} \log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}| + c - \log \int \exp \left\{ - \left(\frac{r}{2} (\|f_0\|_k^2 + c) \right) \right\} g^*(r) dr$$

$$\begin{aligned}
&\leq \frac{1}{2} \log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}| + c + \frac{\|f_0\|_k^2 + c}{2} \int r g^*(r) dr \\
&= \frac{1}{2} \log |\mathbf{I} + \sigma_0^{-2} \mathbf{K}| + \frac{q^2 + 2(\nu - 1)}{2(N + 2\nu - 2)} (\|f_0\|_k^2 + c) + c + (n_1 + n_2)\varepsilon.
\end{aligned}$$

至此完成了引理 B.1 的证明. $\square$

命题 4.2 的证明 容易证明 $q^2 = (\mathbf{y} - f_0(\mathbf{X}))^\top (\mathbf{y} - f_0(\mathbf{X})) / \sigma_0^2 = O(n_1 + n_2)$. 基于引理 B.1 和命题 4.2 中的假设, 当 $n_1 \rightarrow \infty$ 和 $n_2 \rightarrow \infty$ 时, 有

$$\frac{1}{n_1 + n_2} \mathbb{E}_{\mathbf{X}} (D[p_{\sigma_0}(\mathbf{y} | f_0, \mathbf{X}), p_{\sigma_0, \hat{\theta}}(\mathbf{y} | \mathbf{X})]) \rightarrow 0.$$

故命题 4.2 得证. $\square$

附录 C 多个输出的情形

我们很容易便可以将第 3 节讨论的两个输出的 DTPR 模型拓展至多个输出的情形. 假设观测数据来自 S 个总体, 则可以得到 S 条有相依性的曲线, 即 $i = 1, \dots, S$. 令 $W_1(\mathbf{x}), \dots, W_T(\mathbf{x})$ 为 T ($T \geq S$) 个独立的 Gauss 白噪声过程并有若干个光滑的核函数 g . 记核函数的个数为 R . 注意到构造 S 个 t 过程的方法是十分灵活的, 因此可以根据数据相关性的不同而使用不同相依结构的 DTPR 模型. 我们将以 $S = 3$ 为例重点阐述如何构造不同相依结构的 DTP. 考虑过程回归模型

$$\begin{aligned}
y_i(\mathbf{x}) &= f_i(\mathbf{x}) + \epsilon_i(\mathbf{x}), \quad i = 1, 2, 3, \\
f_i(\mathbf{x}) &= h_i(\mathbf{x}) + U_i(\mathbf{x}) + V_i(\mathbf{x}),
\end{aligned}$$

其中 h_i 为均值函数, U_i 为刻画相依性的 t 过程, V_i 为描述第 i 条曲线个体特性的相互独立的 t 过程. 为叙述方便, 令均值函数 $h_i(\cdot) = 0$.

假设 $T = 3$ 和 $R = 3$. 对三条曲线 $i = 1, 2, 3$, 令 $V_i(\mathbf{x}) = \sqrt{r}W_i(\mathbf{x}) \star g_i(\mathbf{x})$ 而 $U_i(\mathbf{x}) = 0$, 可以看出 $\{f_i\}_{i=1}^S$ 仅包含独立的部分 V_i , 过程之间是相互独立的. 此时回归模型退化为独立的回归模型.

假设三条曲线共同拥有同一个效应, 记作 W_0 . 对于 $i = 1, 2, 3$, 取 $V_i(\mathbf{x}) = \sqrt{r}W_i(\mathbf{x}) \star g_i(\mathbf{x})$, $U_i(\mathbf{x}) = \sqrt{r}W_0(\mathbf{x}) \star g_{i+3}(\mathbf{x})$. 在这个情形下, $T = 4$, $R = 6$. 此时模型分别考虑了每个曲线各自的特性, 同时也考虑到了三者共有的特性, 不过并没有考虑曲线两两之间的效应.

如果仅考虑曲线两两之间的相依性, 而不考虑三者共同的效应, 我们可以按如下方法构造:

$$\begin{aligned}
V_i(\mathbf{x}) &= \sqrt{r}W_i(\mathbf{x}) \star g_i(\mathbf{x}), \\
U_1(\mathbf{x}) &= \sqrt{r}W_4(\mathbf{x}) \star g_4(\mathbf{x}) + \sqrt{r}W_5(\mathbf{x}) \star g_5(\mathbf{x}), \\
U_2(\mathbf{x}) &= \sqrt{r}W_5(\mathbf{x}) \star g_6(\mathbf{x}) + \sqrt{r}W_6(\mathbf{x}) \star g_7(\mathbf{x}), \\
U_3(\mathbf{x}) &= \sqrt{r}W_4(\mathbf{x}) \star g_8(\mathbf{x}) + \sqrt{r}W_6(\mathbf{x}) \star g_9(\mathbf{x}),
\end{aligned}$$

这里 W_4 使得 f_1 和 f_3 之间具有相依性, W_5 刻画了 f_1 和 f_2 的相依性, 而 W_6 则负责对 f_2 和 f_3 的相依性进行建模. 此时, 有 $T = 6$ 和 $R = 9$. 当然, 如果真实的曲线是相互独立的, 这个模型也仍然有效. 因为模型包含了独立的效应, 而相依效应 $U_i(\cdot)$ 中的超参数会根据数据进行估计, 模型中的过程 $U_i(\cdot)$ 会接近于常数 0, 从而使得模型不再考虑曲线之间的相依性.

接下来对多个输出情形下的回归进行讨论, 此时考虑 S 为任意大于 1 的常数. 由可加性, 有

$$\begin{pmatrix} f_1 \\ \vdots \\ f_S \end{pmatrix} \sim \text{DTP} \left(\nu, \omega, \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{11} & \dots & k_{1S} \\ \vdots & \ddots & \vdots \\ k_{S1} & \dots & k_{SS} \end{pmatrix} \right).$$

不失一般性, 假设每条曲线的观测点个数都相同, 记为 n . 所有曲线上的样本大小记为 N , 自然地, 有 $N = Sn$. 令观测数据为 $\mathcal{D} = \{\mathbf{X}_i, \mathbf{y}_i, i = 1, \dots, S\}$, 其中

$$\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in})^\top, \quad \mathbf{y}_i = (y_{i1}, \dots, y_{in})^\top.$$

过程的一个实现 $(f_1(\mathbf{X}_1), \dots, f_S(\mathbf{X}_S))^\top$ 的协方差矩阵 $\mathbf{K}$ 为一个 $N \times N$ 的矩阵

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \dots & \mathbf{K}_{1S} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{S1} & \dots & \mathbf{K}_{SS} \end{pmatrix},$$

其中每个 $\mathbf{K}_{ij}$ 为一个 $n \times n$ 的矩阵

$$\mathbf{K}_{ij} = \begin{pmatrix} k_{ij}(\mathbf{x}_{i1} - \mathbf{x}_{j1}) & \dots & k_{ij}(\mathbf{x}_{i1} - \mathbf{x}_{jn}) \\ \vdots & \ddots & \vdots \\ k_{ij}(\mathbf{x}_{in} - \mathbf{x}_{j1}) & \dots & k_{ij}(\mathbf{x}_{in} - \mathbf{x}_{jn}) \end{pmatrix}.$$

将所有协变量和观测值按照顺序分别记作向量的形式

$$\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_S^\top)^\top, \quad \mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_S^\top)^\top.$$

对于给定的协变量 $\mathbf{X}$, 观测值 $\mathbf{y}$ 服从 $\text{DMTD}(\nu, \omega, \mathbf{0}, \Sigma)$, 其中 $\Sigma = \mathbf{K} + \sigma^2 \mathbf{I}$, σ^2 为噪声项的方差.

第 i 条曲线上一个新的样本点 $\mathbf{u}$ 的预测可以表示为

$$\hat{y}_i(\mathbf{u}) = \text{E}(y_i(\mathbf{u}) \mid \mathcal{D}) = \boldsymbol{\zeta}_\mathbf{u}^\top \Sigma^{-1} \mathbf{y},$$

其预测的方差为 $s_0(K_{ii}(\mathbf{u}) - \boldsymbol{\zeta}_\mathbf{u}^\top(\mathbf{u}) \Sigma^{-1} \boldsymbol{\zeta}_\mathbf{u} + \sigma^2)$, 其中影响因子

$$s_0 = \text{E}(r \mid \mathcal{D}) = \frac{\mathbf{y}^\top \Sigma \mathbf{y} + 2\omega}{N + 2(\nu - 1)},$$

而新样本点 $\mathbf{u}$ 与训练数据 $\mathbf{X}$ 的 $N \times 1$ 协方差矩阵为

$$\boldsymbol{\zeta}_\mathbf{u} = (k_{i1}(\mathbf{u}, \mathbf{x}_{11}), \dots, k_{i1}(\mathbf{u}, \mathbf{x}_{1n}), \dots, k_{iS}(\mathbf{u}, \mathbf{x}_{S1}), \dots, k_{iS}(\mathbf{u}, \mathbf{x}_{Sn}))^\top.$$

Dependent t-process regression model

Zhanfeng Wang, Huarong Luo, Ruixing Ming & Yaohua Wu

Abstract Dependent t-process regression (DTPR) is introduced to model functional data with multivariate functional responses, where a dependent t-process (DTP) via bringing in a random effect and a convolution integration is constructed to build the correlated structure of multivariate responses. It shows that DTPR has more robustness against outliers compared with Gaussian process regression (GPR), while it possesses the interpretability as GPR. This paper develops an estimation procedure to predict functional responses. Robustness and information consistency of the proposed prediction are studied. Numerical studies including simulation and real examples show that the proposed DTPR performs well.

Keywords dependent functional responses, robustness, information consistency

MSC(2020) 62G35, 62G08

doi: 10.1360/SSM-2020-0263