增维非光滑估计方程的刀切经验似然方法

韦阳, 李周平*, 杨帆

兰州大学数学与统计学院, 兰州 730000

E-mail: weiy17@lzu.edu.cn, lizhp@lzu.edu.cn, yangf10@lzu.edu.cn

收稿日期: 2018-03-29; 接受日期: 2018-07-10; 网络出版日期: 2019-04-22; * 通信作者

国家自然科学基金 (批准号: 11571154) 和中央高校基本科研业务费 (批准号: lzujbky-2018-110) 资助项目

摘要 本文考虑具有非光滑 U- 统计量结构的估计方程中增维参数的估计问题, 这一情形下, 本文利用刀切经验似然方法对参数进行统计推断, 并在一定正则条件下, 证明所构造刀切似然统计量的渐近性质. 最后, 通过 Monte Carlo 数值模拟和实际数据分析, 说明本文所提方法的优势.

关键词 高维数据 刀切经验似然 非光滑 估计方程 U- 统计量

MSC (2010) 主题分类 60F17, 60F15, 60A86

1 引言

具有 U- 统计量结构的非光滑估计方程在很多统计模型及其应用中是很常见的, 如分位数回归、Wilcoxon 秩回归和线性变换模型等. 特别地, 记 $X_1, \dots, X_n$ 是独立同分布的随机变量, 其对应的分布函数为 F , θ 为 p 维待估参数, 且其唯一地由以下 U- 估计方程决定:

$$U_n(\theta) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} h(X_{i_1}, \dots, X_{i_k}; \theta) = 0, \quad (1.1)$$

其中 $h(\cdot)$ 是一个 r ($r \geq p$) 维对称核函数 (symmetric kernel function), 且若记 θ_0 为 θ 的真值, 则其满足 $Eh(X_1, \dots, X_k; \theta_0) = 0$. 近几十年来, 涉及 U- 估计方程 (1.1) 中参数 θ 的估计被广泛研究, 统计学家已提出了许多估计方法, 例如, 在核函数 $h(\cdot)$ 是光滑的情形下, 估计量 $\hat{\theta}$ 的协方差矩阵较易得到; 而当 $h(\cdot)$ 非光滑时, 估计量 $\hat{\theta}$ 的协方差矩阵通常很难较好估计. 对于后一种情形, 现有的研究通常采用重采样方法 (resampling methods), 如自助法 (bootstrap)^[1]、随机扰动方法 (random perturbation)^[2]、重要性采样方法 (important sampling)^[3]、Monte Carlo 数值微分型重采样方法 (Monte Carlo numerical differentiation-type resampling method)^[4] 和基于重采样的高效收缩方法 (resampling-based efficient shrinkage method)^[5] 等.

英文引用格式: Wei Y, Li Z P, Yang F. Jackknife empirical likelihood for growing dimensional nonsmooth estimating equations (in Chinese). Sci Sin Math, 2019, 49: 1103–1118, doi: 10.1360/SCM-2018-0239

经验似然 (empirical likelihood, EL)^[6,7] 是 Owen 提出的一种构造未知参数的置信区间和假设检验的非参数方法. 较其他一些经典的统计方法, 该方法具有很多优点, 特别地, 经验似然无需构造轴统计量, 进而无需估计协方差阵, 而协方差阵的估计一般比较困难. 正因如此, 经验似然法被广泛研究并应用于回归、估计方程、高维数据和其他许多统计模型中. 有关经验似然法更详细的讨论, 推荐阅读 Owen 在 2001 年所撰写的专著 [8]. 在现代科学研究中, 高维数据变得越来越普遍, 例如, 在基因表达、成像和风险管理等领域中, 待估参数的维度非常大, 取值通常成百上千, 这引起了研究者们越来越多的关注, 该方向有最近的文献 [9–14] 等.

注意到, 在高维情形下, 现有文献中有关经验似然法的研究对象主要集中在具有光滑核函数结构以及线性约束条件下的 U- 统计量. 当约束条件是非线性时, 经验似然法的数值计算比较困难, 尤其对于涉及非光滑估计方程的高维参数的估计问题. 因此, 在非线性约束条件下, 找到 U- 估计方程 (1.1) 的解具有一定的挑战性. 另一方面, 尽管重采样方法在实践中有效且易于实现, 但是这类方法的一个缺点是, 估计量 $\hat{\theta}$ 的协方差矩阵的估计值, 比较依赖随机数发生器和整体重复次数的选取. 为保证估计值的准确性, 一般要求重复次数比较多, 这就导致了计算过程非常耗时. 因此, 关于 U- 估计方程 (1.1) 中参数的有效统计推断问题是一个值得研究的课题. 注意到, 针对 U- 统计量, 文献 [15] 建立了刀切经验似然方法 (jackknife empirical likelihood, JEL), 该方法能有效克服经验似然方法在非线性约束下计算复杂的问题. 随后, 文献 [16] 利用 JEL 方法进一步研究了非光滑 U- 估计方程中有限维参数的估计问题. 考虑到高维数据和 U- 估计方程的一般性, 我们将 JEL 方法应用到涉及增维非光滑 U- 估计方程的情形下, 并研究所构造的 JEL 统计量的渐近性质. 更多关于 JEL 方法的讨论和应用, 可以参见文献 [17–19] 等. 需要指出的是, 我们的工作不同于文献 [20] 之处在于, 其使用经验似然法处理非光滑估计方程, 且待估参数的维数是固定的.

本文结构如下: 第 1 节为本文引言; 第 2 节基于 JEL 方法, 构造出涉及增维非光滑 U- 估计方程中参数的刀切经验似然比统计量, 并给出其渐近性质; 第 3 和 4 节分别通过数值模拟和实际数据分析, 考察本文所提方法在有限样本和实际数据中的表现; 本文定理的详细证明过程在附录 A 中给出.

2 主要结论

假设 $X_1, \dots, X_n \in \mathbb{R}^d$ 是一组独立同分布随机变量, 待估参数 $\theta = (\theta_1, \dots, \theta_p)^T \in \Theta$, 且其唯一地由 r ($r \geq p$) 维 U- 估计方程 $U_n(\theta) := U(X_1, \dots, X_n; \theta) = 0$ 决定. 接下来, 构造待估参数 θ 的刀切经验似然比统计量 (参见文献 [16]).

首先, $U_n(\theta)$ 的刀切伪值的定义如下:

$$\hat{V}_i(\theta) = nU_n(\theta) - (n-1)U_{n-1}^{(-i)}(\theta), \quad i = 1, \dots, n, \quad (2.1)$$

其中 $U_{n-1}^{(-i)}(\theta) := U(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n; \theta)$. 其次, 我们对刀切伪值 $\{\hat{V}_i(\theta), i = 1, \dots, n\}$ 应用经验似然法, 由此得出参数 θ 的刀切经验似然函数为

$$\mathcal{L}(\theta) = \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i(\theta) = 0 \right\}, \quad (2.2)$$

相应的刀切经验似然比定义为

$$\mathcal{R}(\theta) = \left\{ \prod_{i=1}^n (np_i) : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i(\theta) = 0 \right\}. \quad (2.3)$$

不失一般性, 不妨取 $k = 2$. 对于 $k \geq 3$ 的情形, 可以类比 $k = 2$ 的情形得到.

对于 (2.3), 当 0 在点集 $\{\widehat{V}_1(\theta), \dots, \widehat{V}_n(\theta)\}$ 的凸包中时, 根据 Lagrange 乘数法, 可得

$$p_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda^T \widehat{V}_i(\theta)}, \quad i = 1, \dots, n, \quad (2.4)$$

其中 $\lambda = \lambda(\theta)$ 是 r 维的 Lagrange 乘子, 满足

$$0 = g(\lambda) := \frac{1}{n} \sum_{i=1}^n \frac{\widehat{V}_i(\theta)}{1 + \lambda^T \widehat{V}_i(\theta)}. \quad (2.5)$$

若将 (2.4) 代入 (2.3) 中, 并取 $\mathcal{R}(\theta)$ 的负对数形式, 则刀切经验对数似然比为

$$-\log \mathcal{R}(\theta) = \sum_{i=1}^n \log\{1 + \lambda^T \widehat{V}_i(\theta)\}. \quad (2.6)$$

记 $l(\theta) = -\log \mathcal{R}(\theta)$, $l(\theta, \lambda) = l(\theta)/n$. 由于 $h(X_1, X_2; \theta)$ 在 θ 处不可微, 因此使用一阶导数来求最优解是不可行的. 受文献 [21, 22] 的启发, 直接求解 (2.6) 是困难的, 我们转而寻找 $(\hat{\theta}, \hat{\lambda})$, 使得

$$l(\hat{\theta}, \hat{\lambda}) \leq \inf_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} l(\theta, \lambda) + o_p(n^{-\iota}), \quad (2.7)$$

其中 $\iota > 0$, $\hat{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} l(\hat{\theta}, \lambda)$, $\hat{\Lambda}_n(\theta) = \{\lambda : \lambda^T \widehat{V}_i(\theta) > -1, i = 1, \dots, n\}$. $\mathcal{D}_n = \{\theta : \|\theta - \theta_0\| \leq Ca_n\}$ 是 θ_0 的一个邻域, 其中 $C > 0$ 为常数, $a_n = (p/n)^{1/2}$, $\|\cdot\|$ 表示 Euclid 范数.

接下来给出本文的主要结果.

定理 2.1 当条件 (C1)–(C6) (在附录 A 中) 成立时, (2.7) 的最优解 $\hat{\theta}$ 满足:

(i) 当 $n \rightarrow \infty$ 时, $\hat{\theta} \xrightarrow{P} \theta_0$;

(ii) 当 $n \rightarrow \infty$ 时, $A_n \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, A\Omega A^T)$, $A_n A_n^T \rightarrow A A^T$, 其中 $A_n \in \mathbb{R}^{q \times p}$, q 为常数, $A A^T$ 是非负对称矩阵, Ω 的定义详见附录 A.

定理 2.2 当条件 (C1)–(C6) (在附录 A 中) 成立时,

$$\frac{2l(\theta_0) - p}{\sqrt{2p}} \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty.$$

根据定理 2.2, 在置信水平为 $1 - \alpha$ 下, 可以构造一个关于参数 θ 的渐近置信区间

$$I_{\alpha,1} = \{\check{\theta} : 2l(\check{\theta}) \leq p + z_\alpha \sqrt{2p}\},$$

其中 z_α 是 $N(0, 1)$ 分布的上侧 α 分位数.

在实际情况中, 人们有时只对参数 θ 的一部分感兴趣, 如 μ , 而剩余部分 ζ 则并不关心. 当参数的维数 p 比较大时, 通常考虑线性假设. 具体地, 记 $\theta = (\mu^T, \zeta^T)^T$, 其中 $\mu \in \mathbb{R}^d$, $\zeta \in \mathbb{R}^{p-d}$, $\theta_0 = (\mu_0^T, \zeta_0^T)^T$. 考虑以下假设检验:

$$H_0 : B_n \mu_0 = 0, \quad H_1 : B_n \mu_0 \neq 0,$$

其中 B_n 是一个 $q \times d$ 矩阵, 满足 $B_n B_n^T = I_q$, q 为取定的正整数. 由上可以看出, 此假设可以用于同时检验一组变量是否具有统计显著性, 或者通过选择特定的 B_n 来构建 μ 的置信域. 定义如下检验统计量:

$$\hat{l}(B_n) = 2 \min_{\mu, B_n \mu = 0} l(\theta) - 2l(\hat{\theta}). \quad (2.8)$$

以下定理将给出 $\hat{l}(B_n)$ 的渐近分布.

定理 2.3 在原假设 H_0 下, 当条件 (C1)–(C6) (在附录 A 中) 成立时,

$$\hat{l}(B_n) \xrightarrow{d} \chi_q^2, \quad n \rightarrow \infty.$$

基于定理 2.3, 可以构造参数 $B_n\mu$ 的一个水平为 $1 - \alpha$ 的渐近置信区间

$$I_{\alpha,2} = \left\{ \bar{u} : 2 \min_{\mu, B_n\mu=\bar{u}} l(\theta) - 2l(\hat{\theta}) \leq \chi_{q,1-\alpha}^2 \right\},$$

其中 $\chi_{q,1-\alpha}^2$ 是 χ_q^2 分布的 $1 - \alpha$ 分位数.

3 数值模拟

本节进行 Monte Carlo 数值模拟研究, 以评估本文所构建的刀切经验似然方法的有限样本性质. 模拟中我们考虑线性回归模型

$$Y_i = X_i^T \beta_0 + \epsilon_i, \quad (3.1)$$

其中

$$\beta_0 = (\beta_{01}, \beta_{02}, \dots, \beta_{0p})^T \in \mathbb{R}^p$$

是回归系数, 残差项 $\{\epsilon_1, \dots, \epsilon_n\}$ 是一组独立同分布的随机变量, 满足 $E(\epsilon_i) = 0$, 且与 X_i 独立, $i = 1, 2, \dots, n$. 接着, 基于 Wilcoxon 秩回归模型, 我们可得系数 β_0 的秩估计量 $\hat{\beta}^W$, 它是下面估计方程的解:

$$0 = W_n(\beta) := \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} H(Z_i, Z_j; \beta), \quad (3.2)$$

其中

$$H(Z_i, Z_j; \beta) = (X_i - X_j)[I\{e_i(\beta) < e_j(\beta)\} - I\{e_i(\beta) > e_j(\beta)\}],$$

这里

$$Z_i = (X_i, Y_i), \quad e_i(\beta) = Y_i - X_i^T \beta, \quad i, j = 1, \dots, n.$$

易见, Wilcoxon 秩回归估计方程 (3.2) 是估计方程 (1.1) 的一个特例.

首先, 选取模型 (3.1) 中的参数值. 对于任意 $k \in \{1, \dots, p\}$, 回归系数 β_{0k} 服从均匀分布 $U(-1, 1)$, 残差项 ϵ_i 服从正态分布 $N(0, 1)$, 特征变量 X_{ik} 的分布函数有以下两种情形: (i) 标准正态分布 $N(0, 1)$; (ii) Pareto(a) 分布, 其中 $a > 0$, 对应的分布函数记为 $G(x)$. 当 $x \geq 1$ 时, $G(x) = 1 - x^{-a}$, 而当 $x \neq 0$ 时, $G(x) = 0$, 这里, 取 $a = 4.5$. 注意到, 当 $X_{ik} \sim \text{Pareto}(4.5)$ 时, 其满足

$$EX_{ik}^4 < \infty, \quad EX_{ik}^5 = \infty.$$

若 X_{ik} 服从标准正态分布, 则其任意 K 阶距都是存在的.

其次, 根据条件 (C5) 以及特征变量 X_{ik} 的分布函数, 维数 p 分别取为 $p_1 = \lfloor c_1 n^{1/5-0.01} \rfloor$ 和 $p_2 = \lfloor c_2 n^{1/3-0.01} \rfloor$, 对应的窗宽参数为 $c_1 = 4, 6, 8$ 和 $c_2 = 2, 3, 4$, 选取样本量 $n = 200, n = 400$ 和 $n = 800$. 最后, 在每一种参数值组合情形下, 我们重复进行 1,000 次实验. 模拟结果见图 1 和 2. 当参数 (n, p) 选定, 图 1 和 2 分别表示在不同的参数维数 p_1 和 p_2 下统计量 $\frac{2l(\beta_0) - p}{\sqrt{2p}}$ 的 Q-Q 图.

图 1 当 $p_1 = \lfloor c_1 n^{1/5-0.01} \rfloor$ 时, $(2l(\beta_0) - p_1)/\sqrt{2p_1}$ 的 Q-Q 图

图 2 当 $p_2 = \lfloor c_2 n^{1/3-0.01} \rfloor$ 时, $(2l(\beta_0) - p_2)/\sqrt{2p_2}$ 的 Q-Q 图

同时, 为对比基于定理 2.1 (normal approximation, NA) 与 2.2 (jackknife empirical likelihood, JEL) 所构造的渐近置信区间, 取样本量 n 分别为 100、150 和 200, 维数 $p = \lfloor cn^{1/5-0.01} \rfloor$, 对应的窗宽参数 c 分别为 1、2 和 3. 然后, 计算在各种情形下两种方法构造的渐近置信区间、区间长度以及基于 1,000 次重复的覆盖率. 模拟结果见表 1.

表 1 在置信水平 0.95 下, 基于 JEL 方法和 NA 方法的置信区间、长度和覆盖率

n	p	$\{j: \beta_j\}$	JEL			NA		
			置信区间	区间长度	覆盖率	置信区间	区间长度	覆盖率
100	2	1	$[-1.1103, -0.7044]$	0.4059	0.950	$[-1.1099, -0.7017]$	0.4082	0.952
		2	$[-0.5899, -0.1804]$	0.4095	0.966	$[-0.5906, -0.1796]$	0.4110	0.961
	4	1	$[0.0968, 0.4867]$	0.3899	0.945	$[0.0874, 0.4932]$	0.4058	0.943
		2	$[0.5989, 0.9890]$	0.3901	0.953	$[0.5903, 0.9973]$	0.4069	0.944
		3	$[0.2669, 0.6561]$	0.3892	0.960	$[0.2608, 0.6667]$	0.4059	0.949
		4	$[0.0429, 0.4309]$	0.3881	0.960	$[0.0346, 0.4403]$	0.4057	0.947
	7	1	$[-0.8218, -0.4527]$	0.3692	0.950	$[-0.8373, -0.4348]$	0.4025	0.939
		2	$[-1.1763, -0.8065]$	0.3698	0.960	$[-1.1953, -0.7914]$	0.4038	0.948
		3	$[-0.1716, 0.1977]$	0.3694	0.949	$[-0.1902, 0.2131]$	0.4033	0.931
		4	$[0.2263, 0.5952]$	0.3689	0.957	$[0.2120, 0.6146]$	0.4026	0.941
		5	$[0.2228, 0.5883]$	0.3655	0.956	$[0.2053, 0.6060]$	0.4007	0.935
		6	$[-0.5700, -0.2045]$	0.3655	0.948	$[-0.5893, -0.1879]$	0.4013	0.931
		7	$[-0.5973, -0.2298]$	0.3676	0.961	$[-0.6138, -0.2115]$	0.4022	0.947
150	2	1	$[-0.2593, 0.0712]$	0.3306	0.946	$[-0.2595, 0.0721]$	0.3316	0.944
		2	$[-0.9948, -0.6628]$	0.3320	0.958	$[-0.9952, -0.6626]$	0.3326	0.956
	5	1	$[0.7640, 1.0809]$	0.3169	0.950	$[0.7576, 1.0876]$	0.3301	0.940
		2	$[0.7050, 1.0240]$	0.3190	0.956	$[0.6982, 1.0292]$	0.3311	0.948
		3	$[0.4568, 0.7743]$	0.3175	0.948	$[0.4497, 0.7793]$	0.3296	0.938
		4	$[-0.5934, -0.2763]$	0.3171	0.951	$[-0.5999, -0.2703]$	0.3296	0.940
		5	$[0.4038, 0.7196]$	0.3158	0.955	$[0.3966, 0.7256]$	0.3290	0.950
	7	1	$[-0.1337, 0.1749]$	0.3086	0.961	$[-0.1424, 0.1854]$	0.3278	0.943
		2	$[-0.9395, -0.6311]$	0.3084	0.949	$[-0.9483, -0.6211]$	0.3272	0.940
		3	$[0.3339, 0.6418]$	0.3079	0.958	$[0.3257, 0.6531]$	0.3274	0.949
		4	$[0.2852, 0.5943]$	0.3091	0.957	$[0.2773, 0.6055]$	0.3281	0.943
		5	$[0.4487, 0.7566]$	0.3080	0.945	$[0.4391, 0.7662]$	0.3272	0.933
		6	$[-0.5236, -0.2139]$	0.3097	0.948	$[-0.5317, -0.2030]$	0.3287	0.939
		7	$[0.6328, 0.9428]$	0.3099	0.948	$[0.6246, 0.9523]$	0.3277	0.938
200	2	1	$[0.5387, 0.8243]$	0.2856	0.952	$[0.5380, 0.8244]$	0.2864	0.944
		2	$[0.5834, 0.8702]$	0.2868	0.951	$[0.5826, 0.8699]$	0.2874	0.954
	5	1	$[-0.3448, -0.0667]$	0.2782	0.954	$[-0.3487, -0.0628]$	0.2859	0.953
		2	$[-0.8694, -0.5908]$	0.2786	0.966	$[-0.8732, -0.5874]$	0.2858	0.960
		3	$[-0.7504, -0.4742]$	0.2762	0.957	$[-0.7547, -0.4707]$	0.2840	0.951
		4	$[0.5863, 0.8637]$	0.2773	0.956	$[0.5831, 0.8686]$	0.2855	0.948
		5	$[-0.5814, -0.3030]$	0.2784	0.948	$[-0.5848, -0.2994]$	0.2855	0.941

续表

n	p	$\{j: \beta_j\}$	JEL			NA		
			置信区间	区间长度	覆盖率	置信区间	区间长度	覆盖率
8	1		[0.2609, 0.5286]	0.2677	0.966	[0.2539, 0.5363]	0.2824	0.960
	2		[-0.2113, 0.0578]	0.2691	0.953	[-0.2187, 0.0650]	0.2837	0.939
	3		[-1.0734, -0.8038]	0.2696	0.957	[-1.0809, -0.7970]	0.2839	0.949
	4		[0.4700, 0.7384]	0.2684	0.950	[0.4628, 0.7455]	0.2827	0.941
	5		[0.6195, 0.8887]	0.2692	0.959	[0.6127, 0.8965]	0.2838	0.946
	6		[-0.5172, -0.2481]	0.2691	0.961	[-0.5241, -0.2414]	0.2827	0.951
	7		[0.4470, 0.7164]	0.2695	0.950	[0.4399, 0.7239]	0.2839	0.946
	8		[-0.2417, 0.0279]	0.2696	0.969	[-0.2485, 0.0352]	0.2837	0.951

根据图 1、2 和表 1, 我们有如下结论:

(1) 从图 1 和 2 可以观察到, 统计量 $(2l(\beta_0) - p)/\sqrt{2p}$ 的样本分布非常接近标准正态分布, 特别是当 n 和 p 较大时, 这与理论结果一致.

(2) 根据图 1 和 2, 统计量 $(2l(\beta_0) - p)/\sqrt{2p}$ 在正态分布下的有限样本分布比在 Pareto 分布下更接近标准正态分布.

(3) 从图 1 和 2 中观察到, 基于 $N(0, 1)$ 的 Q-Q 图在较低和较高的分位数上欠拟合, 类似于文献 [10, 图 1]. 其可能是卡方分布与标准正态分布之间缺欠拟合造成的, 关于此现象的更多讨论参见文献 [10].

(4) 由表 1 可以看出, 基于定理 2.2 所构造的渐近置信区间明显要优于基于定理 2.1 所构造的渐近置信区间 (覆盖率更接近 0.95, 对应的置信区间长度更短).

4 实例分析

例 4.1 本例中, 我们利用 JEL 方法来分析 NO_2 数据集, 该数据来自于一项关于道路空气污染是否与交通流量和气象因素有关的研究, 该数据集可以从 <http://lib.stat.cmu.edu/datasets/NO2.dat> 中下载. 对于数据集中的某个观测值来说, 其响应变量 Y 表示 NO_2 浓度的对数值, 7 个特征变量分别是: 汽车数量的对数值 (cars)、离地 2 米的温度 (temperature 1)、风速 (wind speed)、离地 2 米与离地 25 米之间的温度差 (temperature 2)、风向 (wind direction)、一天中道路开放的时间 (hours) 和观察的天数 (days). 首先, 我们先将特征变量标准化, 再基于模型 (3.1) 来拟合数据.

基于 JEL 方法和 NA 方法, 我们分别构造待估参数在置信水平为 0.95 的渐近置信区间, 模拟结果见表 2. 从中可以看出, 在同一置信水平下, 基于 JEL 方法构造的置信区间长度大多短于基于 NA 法构造的置信区间长度.

例 4.2 本例中, 我们将 JEL 方法应用于体脂肪数据集, 该数据集可以从 <http://lib.stat.cmu.edu/datasets/bodyfat> 下载. 对于数据集的每个观察, 我们收集响应变量 Y (体脂百分比), 以及 12 个解释变量, 即 Age(X_1)、BMI (X_2 、身体质量指数、 $\frac{\text{height}^2}{\text{weight}}$), 其余的 10 个变量分别记作 $X_i, i = 3, \dots, 12$. 借鉴文献 [14] 中的处理样本中的异常值的思路, 在我们删去样本中一个不合理的观测点 $Y = 0$ 之后, 还剩 251 个观察值. 接着, 我们将每个变量进行标准化, 最后基于模型 (3.1) 来拟合数据.

与表 2 类似, 表 3 记录了在置信水平为 0.95 下, 基于 JEL 方法和 NA 法构造待估参数的渐近置信区间. 由表 3 可知, 在相同的置信水平下, 基于 JEL 方法构造的置信区间长度大多短于基于 NA 构造的置信区间长度.

表 2 在置信水平为 0.95 下, 基于 JEL 方法和 NA 方法的置信区间 (例 4.1)

	JEL		NA	
	置信区间	区间长度	置信区间	区间长度
Cars	[0.2326, 0.5427]	0.3101	[0.2434, 0.5520]	0.3086
Temperature 1	[−0.1197, 0.0861]	0.2058	[−0.1449, 0.0894]	0.2343
Wind speed	[−0.3510, −0.1247]	0.2263	[−0.3627, −0.1078]	0.2549
Temperature 2	[−0.0641, 0.1618]	0.2259	[−0.0712, 0.1722]	0.2434
Wind direction	[−0.1386, 0.0786]	0.2172	[−0.1416, 0.1030]	0.2446
Hours	[−0.2113, 0.0723]	0.2836	[−0.2016, 0.0825]	0.3046
Days	[−0.0226, 0.1893]	0.2119	[−0.0453, 0.1993]	0.2446

表 3 在置信水平为 0.95 下, 基于 JEL 方法和 NA 方法的置信区间 (例 4.2)

	JEL		NA	
	置信区间	区间长度	置信区间	区间长度
Age	[0.0363, 0.0825]	0.0462	[−0.0254, 0.1847]	0.2101
BMI	[−0.5147, −0.2612]	0.2535	[−0.3286, −0.0797]	0.2489
Neck	[−0.2249, 0.0301]	0.2550	[−0.3899, −0.1177]	0.2722
Chest	[−0.2237, 0.0247]	0.2484	[−0.4145, −0.1451]	0.2694
Abdomen	[1.0687, 1.3380]	0.2693	[0.8759, 1.1611]	0.2852
Hip	[−0.3453, −0.0574]	0.2879	[−0.5023, −0.2163]	0.2860
Thigh	[0.0085, 0.0412]	0.0327	[−0.1639, 0.0964]	0.2603
Knee	[−0.0734, 0.1767]	0.2501	[−0.2529, 0.0176]	0.2705
Ankle	[−0.0288, 0.0155]	0.0443	[−0.1594, 0.0735]	0.2329
Biceps	[0.0083, 0.0408]	0.0325	[−0.1614, 0.0894]	0.2508
Forearm	[0.0240, 0.2243]	0.2003	[−0.0979, 0.1249]	0.2228
Wrist	[−0.2385, −0.0090]	0.2295	[−0.4018, −0.1549]	0.2469

致谢 稿件在审查过程中, 审稿人提出了重要的修改意见, 在此表示感谢.

参考文献

- 1 Hu F, Kalbfleisch J D. The estimating function bootstrap. *Canad J Statist*, 2000, 28: 449–481
- 2 Jin Z, Yang Z, Wei L. A simple resampling method by perturbing the minimand. *Biometrika*, 2001, 88: 381–390
- 3 Tian L, Liu J, Zhao Y, et al. Statistical inference based on non-smooth estimating functions. *Biometrika*, 2004, 91: 943–954
- 4 Zeng D, Lin D Y. Efficient resampling methods for nonsmooth estimating functions. *Biostatistics*, 2007, 9: 355–363
- 5 Xu J. Resampling-based efficient shrinkage method for non-smooth minimands. *J Nonparametr Stat*, 2013, 25: 731–743
- 6 Owen A B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 1988, 75: 237–249
- 7 Owen A B. Empirical likelihood ratio confidence regions. *Ann Statist*, 1990, 18: 90–120
- 8 Owen A B. *Empirical Likelihood*. London: Chapman and Hall/CRC, 2001
- 9 Hjort N L, McKeague I W, van Keilegom I. Extending the scope of empirical likelihood. *Ann Statist*, 2009, 37: 1079–1111
- 10 Chen S X, Peng L, Qin Y L. Effects of data dimension on empirical likelihood. *Biometrika*, 2009, 96: 711–722
- 11 Tang C Y, Leng C. Penalized high-dimensional empirical likelihood. *Biometrika*, 2010, 97: 905–920
- 12 Lahiri S N, Mukhopadhyay S. A penalized empirical likelihood method in high dimensions. *Ann Statist*, 2012, 40: 2511–2540

- 13 Leng C, Tang C Y. Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika*, 2012, 99: 703–716
- 14 Li Z, Liu Y, Liu Z. Empirical likelihood and general relative error criterion with divergent dimension. *Statistics*, 2017, 51: 1006–1022
- 15 Jing B Y, Yuan J, Zhou W. Jackknife empirical likelihood. *J Amer Statist Assoc*, 2009, 104: 1224–1232
- 16 Li Z, Xu J, Zhou W. On nonsmooth estimating functions via jackknife empirical likelihood. *Scand J Statist*, 2016, 43: 49–69
- 17 Li M, Peng L, Qi Y. Reduce computation in profile empirical likelihood method. *Canad J Statist*, 2011, 39: 370–384
- 18 Peng L. Approximate jackknife empirical likelihood method for estimating equations. *Canad J Statist*, 2012, 40: 110–123
- 19 Jing B Y, Li Z, Qin J, et al. Jackknife empirical likelihood method for case-control studies with gene-environment independence on controls. *Stat Interface*, 2012, 5: 293–302
- 20 Molanes Lopez E M, van Keilegom I, Veraverbeke N. Empirical likelihood for non-smooth criterion functions. *Scand J Statist*, 2009, 36: 413–432
- 21 Newey W K, Smith R J. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 2004, 72: 219–255
- 22 Parente P M D C, Smith R J. GEL methods for nonsmooth moment indicators. *Econom Theory*, 2011, 27: 74–113
- 23 Peng H, Fan J. Nonconcave penalized likelihood with a diverging number of parameters. *Ann Statist*, 2004, 32: 928–961
- 24 Lee A J. *U-Statistics: Theory and Practice*. New York: Marcel Dekker, 1990

附录 A

为行文方便, 不妨记

$$\left\{ \begin{array}{l} \tau(\theta) = (\tau_1(\theta), \dots, \tau_r(\theta))^T = Eh(X_1, X_2; \theta), \\ \widehat{V}_i(\theta) = (\widehat{V}_{i1}(\theta), \dots, \widehat{V}_{ir}(\theta))^T, \\ \bar{V}(\theta) = \frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\theta) = (\bar{V}_1(\theta), \dots, \bar{V}_r(\theta))^T, \\ \nu_n(\theta) = a_n^{-1}[\bar{V}(\theta) - \tau(\theta)], \\ S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\theta) \widehat{V}_i^T(\theta), \\ \phi(x, \theta) = (\phi_1(x, \theta), \dots, \phi_r(x, \theta))^T = Eh(x, X_1; \theta) - \tau(\theta), \\ \psi(x, y, \theta) = h(x, y; \theta) - \phi(x, \theta) - \phi(y, \theta) - \tau(\theta), \\ g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T = \tau(\theta) + 2\phi(x, \theta), \\ \sigma_{jk} = \text{Cov}(\phi_j(X_1, \theta_0), \phi_k(X_1, \theta_0)), \quad j, k = 1, \dots, r, \\ \Sigma(\theta) : \sqrt{n}(\bar{V}(\theta) - \tau(\theta)) \text{ 的渐近方差 - 协方差矩阵, 其元素为 } 4\sigma_{kl}(\theta), \quad k, l = 1, \dots, r, \\ \Sigma = \Sigma(\theta_0), \text{ 其元素为 } 4\sigma_{kl} = 4\sigma_{kl}(\theta_0), \quad k, l = 1, \dots, r, \\ G = \frac{\partial E g(X_1; \theta_0)}{\partial \theta^T}. \end{array} \right.$$

首先, 根据 Hoeffding 分解定理, 有

$$\bar{V}_l(\theta) = \tau_l(\theta) + \frac{2}{n} \sum_{i=1}^n \phi_l(X_i, \theta) + \binom{n}{2}^{-1} \sum_{i < j}^n \psi_l(X_i, X_j, \theta),$$

由此可得

$$\begin{aligned}\widehat{V}_{il}(\theta) &= \tau_l(\theta) + 2\phi_l(X_i, \theta) + \frac{2}{n-2} \sum_{\substack{j=1 \\ j \neq i}}^n \psi_l(X_i, X_j, \theta) - \binom{n-1}{2}^{-1} \sum_{i < j}^n \psi_l(X_i, X_j, \theta) \\ &=: g_l(X_i, \theta) + R_{ni,l}(\theta).\end{aligned}\quad (\text{A.1})$$

对于 $\theta \in \mathcal{D}_n$, 易得

$$\mathbb{E}R_{ni,l}^2(\theta) \leq Cn^{-1}\mathbb{E}\psi_l^2(X_1, X_2, \theta) + Cn^{-2}\mathbb{E}\psi_l^2(X_1, X_2, \theta) \rightarrow 0. \quad (\text{A.2})$$

故 $R_{ni,l}(\theta) = O_p(n^{-1/2}) \xrightarrow{P} 0$ 和 $\widehat{V}_{il}(\theta) \xrightarrow{P} g_l(X_i, \theta)$, $i = 1, \dots, n$, $l = 1, \dots, r$.

在证明主要结果之前, 我们先列出如下一些正则条件:

- (C1) 参数 θ 的支撑集 Θ 是 $\mathbb{R}^p$ 中一个紧集, 且 $\mathbb{E}h(X_1, X_2; \theta) = 0$ 在 Θ 中存在唯一解, 记为 θ_0 ;
- (C2) 当 $\alpha > 2$ 且 n 足够大时, $\mathbb{E}[\sup_{\theta \in \Theta} (\|h(X_1, X_2; \theta)\| r^{-1/2})^\alpha] < \infty$;
- (C3) 当 $\theta \in \Theta$, 核函数 $h_k(X_1, X_2; \theta)$ 连续, $\tau(\theta) = \mathbb{E}h(X_1, X_2; \theta)$ 对 θ 在 θ_0 二阶可导, $\text{rank}(G) = p$, $\|\frac{\partial \sigma_{ij}(\theta)}{\partial \theta_k}\|$ 有界, $k, i, j = 1, \dots, r$;
- (C4) 当 n 足够大时, 对于任意 $\theta \in \mathcal{D}_n$, 存在 b 和 B , 使得 $\Sigma(\theta)$ 的特征值满足 $0 < b \leq \gamma_1\{\Sigma(\theta)\} \leq \dots \leq \gamma_r\{\Sigma(\theta)\} \leq B < \infty$;
- (C5) 当 $n \rightarrow \infty$ 时, $p^t/n \rightarrow 0$, 其中 $t = \max\{\frac{5}{2-\alpha}, 3\}$, $p/r \rightarrow r_0$, $0 < r_0 < 1$;
- (C6) 当 $\delta_n \rightarrow 0$ 时,

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} \frac{\|\nu_n(\theta) - \nu_n(\theta_0)\|}{1 + a_n^{-1}\|\theta - \theta_0\|} \xrightarrow{P} 0.$$

注 A.1 条件 (C1) 和 (C2) 用于保证 (2.6) 极小值点的存在性和一致性, 以及控制核函数 $h(X_1, X_2; \theta)$ 的尾部形状. 条件 (C3) 表示 $\mathbb{E}h(X_1, X_2; \theta)$ 在 θ_0 附近的光滑性. 条件 (C4) 假设似然函数的信息矩阵是正定的, 且其特征值是一致有界的. 条件 (C5) 是对 p 和 n 的收敛速率的要求. 条件 (C6) 是 Bahadur- 型连续模量. 条件 (C1)–(C6) 并不是限制很强的假设条件, 对这些条件所表达的实际含义以及其合理性的更多讨论, 可以参见文献 [13, 16, 23] 等.

为证明定理 2.1–2.3, 我们需要以下引理.

引理 A.1 当条件 (C2)–(C5) 成立时, 对任意 $\theta \in \Theta$, 有

$$\|S_n(\theta) - \Sigma(\theta)\| = O_p\left(\frac{p}{\sqrt{n}}\right).$$

证明 对于任意 $\theta \in \Theta$, $k = 1, \dots, r$, 记 $\sigma_{h,k}^2(\theta) = \text{Var}(h_k(X_1, X_2; \theta))$, $\sigma_{h,k}^2(\theta) < \infty$, 因为 $\mathbb{E}h_k^2(X_1, X_2; \theta) < \infty$, 所以,

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\theta) \widehat{V}_i^T(\theta) = \frac{1}{n} \sum_{i=1}^n [\widehat{V}_i(\theta) - \bar{V}(\theta)][\widehat{V}_i(\theta) - \bar{V}(\theta)]^T + \bar{V}(\theta) \bar{V}^T(\theta). \quad (\text{A.3})$$

下面考虑 $\Sigma(\theta)$ 中的每个元素, 根据文献 [24, 定理 2], 可得

$$\text{Cov}(\bar{V}_j(\theta), \bar{V}_k(\theta)) = \frac{4(n-2)}{n(n-1)} \sigma_{jk}(\theta) + \frac{2}{n(n-1)} \text{Cov}(h_j(X_1, X_2; \theta), h_k(X_1, X_2; \theta)).$$

由于 $\text{Cov}(\bar{V}_j(\theta), \bar{V}_k(\theta))$ 的刀切估计值为

$$\widehat{\text{Cov}}_{jk}(\text{Jack}) = \frac{1}{n(n-1)} \sum_{i=1}^n [\hat{V}_{ij}(\theta) - \bar{V}_j(\theta)][\hat{V}_{ik}(\theta) - \bar{V}_k(\theta)],$$

那么, 当 $n \rightarrow \infty$ 时, 以概率 1 成立,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n [\hat{V}_{ij}(\theta) - \bar{V}_j(\theta)][\hat{V}_{ik}(\theta) - \bar{V}_k(\theta)] &= (n-1) \widehat{\text{Cov}}_{jk}(\text{Jack}) \\ &= (n-1) \{ \text{Cov}(\bar{V}_j(\theta), \bar{V}_k(\theta)) + o(n^{-1}) \} \\ &= 4\sigma_{jk}(\theta) + o(1). \end{aligned} \quad (\text{A.4})$$

结合 (A.3) 和 (A.4), 对任意 $M_\epsilon > 0$, 根据 Chebyshev 不等式, 易得

$$\begin{aligned} P\left(\|S_n(\theta) - \Sigma(\theta)\| \geq \frac{p}{\sqrt{n}} M_\epsilon\right) &\leq \frac{n}{n^2 p^2 M_\epsilon^2} E \sum_{j,k=1}^r \left[\sum_{i=1}^n \{ \hat{V}_{ij}(\theta) \hat{V}_{ik}(\theta) - 4\sigma_{jk}(\theta) \} \right]^2 \\ &= O\left(\frac{1}{M_\epsilon^2}\right). \end{aligned}$$

证毕. □

引理 A.2 当条件 (C2) 成立时, 对于 $\theta \in \Theta$, 有

$$\frac{n(\bar{V}(\theta) - \tau(\theta))^T \Sigma(\theta)^{-1} (\bar{V}(\theta) - \tau(\theta)) - p}{\sqrt{2p}} \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty.$$

证明 由 (A.1)、(A.2) 和中心极限定理, 易得上述结论. □

引理 A.3 当条件 (C2) 成立时, 对于 $\theta \in \Theta$, 有

$$\max_{1 \leq i \neq j \leq n} \sup_{\theta \in \Theta} \|h(X_i, X_j; \theta)\| = o_p(n^{1/\alpha} r^{1/2}), \quad n \rightarrow \infty.$$

证明 根据条件 (C2) 和文献 [15, 引理 A.4], 易得上述结论. □

根据引理 A.3 可知, 对于任意 ξ 且 ξ 满足 $\frac{3}{5\alpha} + \frac{1}{5} \leq \xi < \frac{3}{10} + \frac{2}{5\alpha}$, 有

$$\max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|\hat{V}_i(\theta)\| = o_p(n^{1/\alpha} r^{1/2}), \quad \max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|\lambda^T \hat{V}_i(\theta)\| = o_p(n^{-\xi+1/\alpha} r^{1/2}), \quad (\text{A.5})$$

其中 $\lambda \in \Lambda_n = \{\lambda : \|\lambda\| \leq n^{-\xi}\}$, $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$, $\theta \in \Theta$.

由于引理 A.4 的证明过程类似于文献 [13, 引理 2 和 3] 的证明过程, 故省略其证明过程.

引理 A.4 当条件 (C1)–(C4) 成立时, 存在 $\lambda_{\theta_0} = \arg \max_{\lambda \in \hat{\Lambda}(\theta_0)} l(\lambda, \theta_0)$, $\|\lambda_{\theta_0}\| = O_p(a_n)$, 且 $\sup_{\lambda \in \hat{\Lambda}(\theta_0)} l(\lambda, \theta_0) \leq O_p(a_n^2)$ 以概率 1 成立.

引理 A.5 当条件 (C1)–(C5) 成立时, $\|\lambda_{\hat{\theta}}\| = O_p(a_n)$.

证明 令 $\lambda_{\hat{\theta}} = \rho u$, 其中 $\|u\| = 1$. 由 (2.5), 可得

$$\begin{aligned} 0 &= \|g(\rho u)\| \\ &\geq \frac{1}{n} \left| u^T \left\{ \sum_{i=1}^n \hat{V}_i(\hat{\theta}) - \rho \sum_{i=1}^n \frac{\hat{V}_i(\hat{\theta}) u^T \hat{V}_i(\hat{\theta})}{1 + \rho u^T \hat{V}_i(\hat{\theta})} \right\} \right| \end{aligned}$$

$$\geq \rho u^T \frac{S_n(\hat{\theta})}{1 + \rho \max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|\hat{V}_i(\theta)\|} u - \frac{1}{n} \left| u^T \sum_{i=1}^n \hat{V}_i(\hat{\theta}) \right|.$$

根据引理 A.2, 可得 $\|\frac{1}{n} \sum_{i=1}^n \hat{V}_i(\hat{\theta})\| = O_p(a_n)$. 结合 $u^T S_n(\hat{\theta})u \geq b + o_p(1)$, 可得 $\|\lambda_{\hat{\theta}}\| = O_p(a_n)$. $\square$

定理 2.1 的证明 (i) 类似于文献 [21, 定理 3.1] 的证明过程, 易证其具有一致性; (ii) 渐近正态性的证明可以通过几个步骤完成.

步骤 1 首先, 根据三角不等式, 可得

$$\|\tau(\hat{\theta})\| \leq \|\bar{V}(\hat{\theta}) - \bar{V}(\theta_0) - \tau(\hat{\theta})\| + \|\bar{V}(\hat{\theta})\| + \|\bar{V}(\theta_0)\|.$$

结合引理 A.3 以及估计 $\hat{\theta}$ 的一致性, 可知 $a_n^{-1}\|\bar{V}(\hat{\theta})\| = O_p(1)$, 特别地, $a_n^{-1}\|\bar{V}(\theta_0)\| = O_p(1)$. 另外, 根据条件 (C6), 可得

$$a_n^{-1}\|\bar{V}(\hat{\theta}) - \bar{V}(\theta_0) - \tau(\hat{\theta})\| \leq (1 + a_n^{-1}\|\hat{\theta} - \theta_0\|)o_p(1).$$

进一步, 有

$$a_n^{-1}\|\tau(\hat{\theta})\| \leq (1 + a_n^{-1}\|\hat{\theta} - \theta_0\|)o_p(1) + O_p(1).$$

注意到 $\tau(\theta)$ 在 θ_0 处可微, 即 $\|\tau(\hat{\theta})\| \geq C\|\hat{\theta} - \theta_0\|$, 则有

$$a_n^{-1}\|\hat{\theta} - \theta_0\| \leq (1 + a_n^{-1}\|\hat{\theta} - \theta_0\|)o_p(1) + O_p(1).$$

因此, $\|\hat{\theta} - \theta_0\| = O_p(a_n)$.

步骤 2 因为 $l(\theta, \lambda)$ 在其最优处是不可微的, 所以接下来构造一个光滑函数来近似求解它的最优值. 记 $\hat{\eta} = (\hat{\theta}^T, \hat{\lambda}^T)^T$ 和 $\eta_0 = (\theta_0^T, 0^T)^T$, 且构造的光滑函数定义为

$$L_n(\theta, \lambda) = [G(\theta - \theta_0)]^T \lambda + \bar{V}(\theta_0)^T \lambda - \frac{1}{2} \lambda^T \Sigma \lambda. \quad (\text{A.6})$$

为验证我们所构造函数 $L_n(\theta, \lambda)$ 的合理性, 需要证明

$$|l(\hat{\theta}, \hat{\lambda}) - L_n(\hat{\theta}, \hat{\lambda})| = o_p(a_n^2).$$

我们先将 $l(\hat{\theta}, \hat{\lambda})$ 在 $\lambda = 0$ 处进行 Taylor 展开, 即

$$l(\hat{\theta}, \hat{\lambda}) = \hat{\lambda}^T \bar{V}(\hat{\theta}) - \frac{1}{2} \hat{\lambda}^T \left[\frac{1}{n} \sum_{i=1}^n \frac{\hat{V}_i(\hat{\theta}) \hat{V}_i^T(\hat{\theta})}{(1 + \hat{\lambda}^T \hat{V}_i(\hat{\theta}))^2} \right] \hat{\lambda},$$

其中 $\dot{\lambda}$ 的每个分量都处在 0 与 $\hat{\lambda}$ 之间. 因此,

$$\begin{aligned} |l(\hat{\theta}, \hat{\lambda}) - L_n(\hat{\theta}, \hat{\lambda})| &\leq |(\bar{V}(\hat{\theta}) - \bar{V}(\theta_0) - G(\hat{\theta} - \theta_0))^T \hat{\lambda}| + \left| \frac{1}{2} \hat{\lambda}^T \left[\frac{1}{n} \sum_{i=1}^n \frac{\hat{V}_i(\hat{\theta}) \hat{V}_i^T(\hat{\theta})}{(1 + \dot{\lambda}^T \hat{V}_i(\hat{\theta}))^2} - \Sigma \right] \hat{\lambda} \right| \\ &=: I_1 + I_2. \end{aligned}$$

注意到

$$\begin{aligned} I_1 &\leq \|\bar{V}(\hat{\theta}) - \bar{V}(\theta_0) - G(\hat{\theta} - \theta_0)\| \|\hat{\lambda}\| \\ &\leq \{\|\bar{V}(\hat{\theta}) - \bar{V}(\theta_0) - \tau(\hat{\theta})\| + \|G(\hat{\theta} - \theta_0) - \tau(\hat{\theta})\|\} \|\lambda\| \\ &\leq \{(1 + a_n^{-1}\|\hat{\theta} - \theta_0\|)o_p(a_n) + o_p(\|\hat{\theta} - \theta_0\|)\} \|\lambda\| \\ &= o_p(a_n^2), \end{aligned}$$

其中最后一个不等式由条件 (C3) 和 (C6) 可得. 另外, 根据引理 A.1、条件 (C3) 和 (C5), 有

$$\begin{aligned} I_2 &\leq \|\hat{\lambda}\|^2 \|S_n(\hat{\theta}) - \Sigma\| \leq \|\hat{\lambda}\|^2 \{\|S_n(\hat{\theta}) - \Sigma(\hat{\theta})\| + \|\Sigma(\hat{\theta}) - \Sigma\|\} \\ &= O_p(a_n^2) \left(O_p\left(\frac{p}{\sqrt{n}}\right) + O_p\left(\frac{p^2}{\sqrt{n}}\right) \right) \\ &= O_p(a_n^2) O_p\left(\frac{p^2}{\sqrt{n}}\right) = o_p(a_n^2). \end{aligned}$$

综上所述,

$$|l(\hat{\theta}, \hat{\lambda}) - L_n(\hat{\theta}, \hat{\lambda})| = o_p(a_n^2).$$

步骤 3 由步骤 1 和 2 可知, 现在只需考虑 $\min_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}^r} L_n(\theta, \lambda)$. 由于 $L_n(\theta, \lambda)$ 是光滑函数, 且 Θ 是紧集, 所以其全局最优解 $\tilde{\eta} = (\tilde{\theta}^T, \tilde{\lambda}^T)^T$ 存在, 且满足

$$G^T \tilde{\lambda} = 0, G(\tilde{\theta} - \theta_0) + \bar{V}(\theta_0) - \Sigma \tilde{\lambda} = 0, \quad (\text{A.7})$$

等价地,

$$\begin{pmatrix} 0 \\ \bar{V}(\theta_0) \end{pmatrix} + M(\tilde{\eta} - \eta_0) = 0,$$

其中

$$M = \begin{pmatrix} 0 & G^T \\ G & -\Sigma \end{pmatrix}.$$

进一步, 有

$$\tilde{\theta} - \theta_0 = -\Omega G^T \Sigma^{-1} \bar{V}(\theta_0), \quad \tilde{\lambda} = \Sigma^{-1} [I - G \Omega G^T \Sigma^{-1}] \bar{V}(\theta_0), \quad (\text{A.8})$$

其中 $\Omega = (G^T \Sigma^{-1} G)^{-1}$.

根据引理 A.2, 对于 $A_n \in \mathbb{R}^{q \times p}$, 其满足 $A_n A_n^T \rightarrow A A^T$, 易得 $A_n \sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, A \Omega A^T)$.

步骤 4 最后, 我们需要证明 $\tilde{\eta}$ 渐近等价于 $\hat{\eta}$, 即证明 $\|\hat{\eta} - \tilde{\eta}\| = o_p(a_n)$. 类似于文献 [22, 定理 2.2] 的证明过程, 定理 2.1 容易得证. $\square$

定理 2.2 的证明 令 $w_i(\theta) = \lambda_\theta^T \hat{V}_i(\theta)$, 由 (A.5), 可得 $\max_{1 \leq i \leq n} \|w_i(\theta)\| = o_p(1)$. 对 (2.5) 进行 Taylor 展开, 即

$$0 = \frac{1}{n} \sum_{i=1}^n \hat{V}_i(\theta_0) - S_n(\theta_0) \lambda_{\theta_0} + R_n,$$

其中

$$R_n = \frac{1}{n} \sum_{i=1}^n \hat{V}_i(\theta_0) \frac{w_i^2(\theta_0)}{(1 + \varpi_i)^3},$$

这里 $|\varpi_i| \leq |w_i(\theta_0)|$. 因此, $\max_{1 \leq i \leq n} |\varpi_i| = o_p(1)$,

$$R_n = \frac{1}{n} \sum_{i=1}^n \hat{V}_i(\theta_0) w_i^2(\theta_0) \{1 + o_p(1)\} := R_n^* \{1 + o_p(1)\}.$$

因为

$$\begin{aligned}\|R_n^*\| &\leq \frac{1}{n} \sum_{i=1}^n \|\widehat{V}_i(\theta_0)\| w_i^2(\theta_0) \leq \max_{1 \leq i \leq n} \|\widehat{V}_i(\theta_0)\| \left\| \frac{1}{n} \sum_{i=1}^n \lambda_{\theta_0}^T \widehat{V}_i(\theta_0) \widehat{V}_i(\theta_0)^T \lambda_{\theta_0} \right\| \\ &= o_p(n^{1/\alpha} r^{1/2}) \|\lambda_{\theta_0}\|^2 O_p(\gamma_p(S_n(\theta_0))) = o_p(n^{1/\alpha-1} p^{3/2}),\end{aligned}$$

所以,

$$\lambda_{\theta_0} = S_n^{-1}(\theta_0) \bar{V}(\theta_0) + S_n^{-1}(\theta_0) R_n.$$

根据 Taylor 公式, 可得

$$\log(1 + w_i(\theta_0)) = w_i(\theta_0) - \frac{w_i^2(\theta_0)}{2} + \frac{w_i^3(\theta_0)}{3(1 + \gamma_i)^4},$$

其中 $|\gamma_i| \leq |w_i(\theta_0)|$. 因此,

$$\begin{aligned}2l(\theta_0) &= 2 \sum_{i=1}^n \log\{1 + \lambda_{\theta_0}^T \widehat{V}_i(\theta_0)\} \\ &= n \bar{V}^T(\theta_0) S_n^{-1}(\theta_0) \bar{V}(\theta_0) - n R_n S_n^{-1}(\theta_0) R_n + \frac{2}{3} \sum_{i=1}^n \frac{(\lambda_{\theta_0}^T \widehat{V}_i(\theta_0))^3}{(1 + \gamma_i)^4} \\ &= n \bar{V}^T(\theta_0) \Sigma^{-1} \bar{V}(\theta_0) + n \bar{V}^T(\theta_0) (S_n^{-1}(\theta_0) - \Sigma^{-1}) \bar{V}(\theta_0) - n R_n^T S_n^{-1}(\theta_0) R_n + R_n^*,\end{aligned}\quad (\text{A.9})$$

其中 $R_n^* = \frac{2}{3} \sum_{i=1}^n w_i^3(\theta_0)(1 + o_p(1))$. 对于 (A.9) 的第二项, 结合引理 A.1、A.2、条件 (C4) 和文献 [10], 若 $p^3/n \rightarrow 0$, 则

$$n \bar{V}^T(\theta_0) (S_n^{-1}(\theta_0) - \Sigma^{-1}) \bar{V}(\theta_0) = o_p(\sqrt{p}). \quad (\text{A.10})$$

另外, 容易验证

$$|n R_n^T S_n^{-1}(\theta_0) R_n| \leq \frac{n \|R_n\|^2}{\gamma_1(S_n(\theta_0))} = o_p(\sqrt{p}) \quad (\text{A.11})$$

和

$$\begin{aligned}|R_n^*| &\leq C \sum_{i=1}^n |w_i(\theta_0)|^3 \leq C \left\{ \sum_{i=1}^n w_i^2(\theta_0) \sum_{i=1}^n w_i^4(\theta_0) \right\}^{1/2} \\ &\leq C \sqrt{n \lambda_{\theta_0}^T S_n(\theta_0) \lambda_{\theta_0}} \left\{ \sum_{i=1}^n \|\lambda_{\theta_0}\|^4 \widehat{V}_i^4(\theta_0) \right\}^{1/2} \\ &= o_p(\sqrt{p}).\end{aligned}\quad (\text{A.12})$$

结合 (A.9)–(A.12), 可得

$$l(\theta_0) = n \bar{V}^T(\theta_0) \Sigma^{-1} \bar{V}(\theta_0) + o_p(\sqrt{p}),$$

由上式和引理 A.2, 引理得证. □

定理 2.3 的证明 首先, 记 B_n^* 是一个 $d \times p$ 矩阵, 其满足 $B_n^* \theta = \mu$, B_n 是一个 $q \times d$ 矩阵, 且满足 $B_n B_n^T = I_q$, 记 $H = B_n B_n^*$, 是一个 $q \times p$ 矩阵. 在此假设下, 我们有 $H \theta_0 = B_n B_n^* \theta_0 = B_n \mu_0 = 0$. 根据 Lagrange 乘子法, 可得

$$l_1(\theta) := \frac{1}{n} \sum_{i=1}^n \log\{1 + \lambda^T \widehat{V}_i(\theta)\} + \nu^T H \theta, \quad (\text{A.13})$$

其中 ν 是一个 q 维 Lagrange 乘子. 注意到 $l_1(\theta)$ 在其最优解处不可微, 类似于定理 2.1, 我们构造一个光滑函数来近似 $l_1(\theta)$, 定义如下:

$$\tilde{L}_n(\theta, \lambda) = [G(\theta - \theta_0)]^T \lambda + \bar{V}(\theta_0)^T \lambda - \frac{1}{2} \lambda^T \Sigma \lambda + \nu^T H \theta. \quad (\text{A.14})$$

令

$$\begin{cases} \tilde{Q}_{1n}(\theta, \lambda, \nu) = G(\theta - \theta_0) + \bar{V}(\theta_0) - \Sigma \lambda, \\ \tilde{Q}_{2n}(\theta, \lambda, \nu) = G^T \lambda + H^T \nu, \\ \tilde{Q}_{3n}(\theta, \lambda, \nu) = H \theta. \end{cases}$$

所以, (A.14) 的最优解 $(\check{\theta}, \check{\lambda}, \check{\nu})$ 也是 $\tilde{Q}_{jn}(\theta, \lambda, \nu) = 0$ ($j = 1, 2, 3$) 的解. 对于 $\theta \in \mathcal{D}_n$, 有 $\|\lambda\| = O_p(a_n)$ 和 $\tilde{Q}_{2n}(\check{\theta}, \check{\lambda}, \check{\nu}) = 0$, 所以, $\|\nu\| = O_p(a_n)$, 且可以验证

$$\begin{aligned} \frac{\partial \tilde{Q}_{1n}(\theta_0, 0, 0)}{\partial \lambda^T} &= -\Sigma, & \frac{\partial \tilde{Q}_{1n}(\theta_0, 0, 0)}{\partial \theta^T} &= G, & \frac{\partial \tilde{Q}_{1n}(\theta_0, 0, 0)}{\partial \nu^T} &= 0, \\ \frac{\partial \tilde{Q}_{2n}(\theta_0, 0, 0)}{\partial \lambda^T} &= G^T, & \frac{\partial \tilde{Q}_{2n}(\theta_0, 0, 0)}{\partial \theta^T} &= 0, & \frac{\partial \tilde{Q}_{2n}(\theta_0, 0, 0)}{\partial \nu^T} &= H^T, \\ \frac{\partial \tilde{Q}_{3n}(\theta_0, 0, 0)}{\partial \lambda^T} &= 0, & \frac{\partial \tilde{Q}_{3n}(\theta_0, 0, 0)}{\partial \theta^T} &= H, & \frac{\partial \tilde{Q}_{3n}(\theta_0, 0, 0)}{\partial \nu^T} &= 0. \end{aligned}$$

将 $\tilde{Q}_{jn}(\check{\theta}, \check{\lambda}, \check{\nu}) = 0$ 在 $(\theta_0, 0, 0)$ 附近展开, 可得

$$\begin{aligned} 0 &= \tilde{Q}_{jn}(\check{\theta}, \check{\lambda}, \check{\nu}) \\ &= \tilde{Q}_{jn}(\theta_0, 0, 0) + \frac{\tilde{Q}_{jn}(\theta_0, 0, 0)}{\partial \lambda^T} \check{\lambda} + \frac{\tilde{Q}_{jn}(\theta_0, 0, 0)}{\partial \theta^T} (\check{\theta} - \theta_0) + \frac{\tilde{Q}_{jn}(\theta_0, 0, 0)}{\partial \nu^T} \check{\nu} + \tilde{R}_{jn}. \end{aligned}$$

因此,

$$\begin{pmatrix} -\bar{V}(\theta_0) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\Sigma & G & 0 \\ G^T & 0 & H^T \\ 0 & H & 0 \end{pmatrix} \begin{pmatrix} \check{\lambda} \\ \check{\theta} - \theta_0 \\ \check{\nu} \end{pmatrix} + \tilde{R}_n. \quad (\text{A.15})$$

根据文献 [13], 有 $\|\tilde{R}_n\| = o_p(n^{-1/2})$. 记 $\tilde{\eta} = (\theta^T, \nu^T)^T$,

$$D_{11} = -\Sigma, \quad D_{12} = (G, 0), \quad D_{21} = D_{12}^T, \quad D_{22} = \begin{pmatrix} 0 & H^T \\ H & 0 \end{pmatrix}.$$

进一步, 有

$$\check{\mu} - \mu_0 = D_{22.1}^{-1} D_{21} D_{11}^{-1} \bar{V}(\theta_0) + \tilde{R}_{1n}, \quad \|\tilde{R}_{1n}\| = o_p(n^{-1/2}),$$

其中 $D_{22.1} = D_{22} - D_{21}D_{11}^{-1}D_{12}$. 因此,

$$\tilde{\lambda} = \{\Sigma^{-1} + \Sigma^{-1}G\hat{Z}G^T(-\Sigma)^{-1}\}\bar{V}(\theta_0).$$

其中 $\hat{Z} = \Omega\{I - H^T(H\Omega H^T)^{-1}H\Omega\}$.

记 $(\bar{\theta}_2, \bar{\lambda}, \bar{\nu})$ 和 $(\hat{\theta}, \hat{\lambda})$ 分别是 $l_1(\theta)$ 和 $l(\theta)$ 的最优解. 类似于定理 2.1 的证明, 可得

$$\bar{\lambda} = \tilde{\lambda} + o_p(n^{-1/2}), \quad \hat{\lambda} = \Sigma^{-1}[I - G\Omega G^T\Sigma^{-1}]\bar{V}(\theta_0) + o_p(n^{-1/2}).$$

接着对 $l((\theta_{01}, \hat{\theta}_{02}))$ 和 $l(\hat{\theta})$ 分别应用 Taylor 公式, 根据之前证明的引理 A.1–A.5, 可得

$$\hat{l}(B_n) = n\bar{V}^T(\theta_0)\Sigma^{-1/2}D\Sigma^{-1/2}\bar{V}(\theta_0) + o_p(1), \quad (\text{A.16})$$

其中 D 是一个幂等矩阵, 且有

$$D = \Sigma^{-1/2}G\Omega H^T(H\Omega H^T)^{-1}H\Omega G^T\Sigma^{-1/2}.$$

由上式可以看出 $\text{tr}(D) = q$. 根据幂等矩阵的性质, 可得 $D = Z_n^T Z_n$, 其中 Z_n 是一个 $q \times r$ 矩阵, 且满足 $Z_n Z_n^T = I_q$. 由定理 2.2, 有 $\sqrt{n}Z_n\Sigma^{-1/2}\bar{V}(\theta_0) \xrightarrow{d} N(0, I_q)$. 因此,

$$\begin{aligned} \hat{l}(B_n) &= n\bar{V}(\theta_0)\Sigma^{-1/2}D\Sigma^{-1/2}\bar{V}(\theta_0) + o_p(1) \\ &= n\{Z_n\Sigma^{-1/2}\bar{V}(\theta_0)\}^T\{Z_n\Sigma^{-1/2}\bar{V}(\theta_0)\} + o_p(1) \\ &\xrightarrow{d} \chi_q^2. \end{aligned}$$

证毕. □

Jackknife empirical likelihood for growing dimensional nonsmooth estimating equations

Yang Wei, Zhouping Li & Fan Yang

Abstract We consider the estimation problem of parameters that is determined by nonsmooth U-statistic structured (U-type) estimating equations with growing dimensional parameters. Under the high dimensional setup, we extend the jackknife empirical likelihood (JEL) method to do inference for the growing number of parameters, and establish the asymptotic properties of the estimators under certain mild conditions. We carry out Monte Carlo simulation studies and real data analysis to illustrate the performance of our new method.

Keywords high-dimensional data, jackknife empirical likelihood, nonsmooth, estimating equations, U-statistics

MSC(2010) 60F17, 60F15, 60A86

doi: 10.1360/SCM-2018-0239