系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
27(1) (2007, 2), 124–133

高维线性模型中的经验似然*

石 坚

(中国科学院数学与系统科学研究院, 北京 100080)

摘要 研究高维线性模型中的经验似然推断. 当协变量的维数随样本量增加时, 常规的经验似然推断失效. 在适当的正则条件下, 对修正的经验似然比统计量给出了渐近分布理论.

关键词 线性模型, 经验似然, 高维协变量

MR(2000) 主题分类号 62E20, 62F05

1 引 言

考虑如下线性模型

$$Y = X^{\mathrm{T}}\beta + \varepsilon, \tag{1.1}$$

其中 X 是维数为 k 的协变量, ε 是均值为 0 的误差变量, β 是回归参数, T 表示矩阵转置. 假设观测数据为 $(Y_1, X_1), (Y_2, X_2), \cdots, (Y_n, X_n)$.

对模型 (1.1), 实际中最感兴趣的参数是 β, 例如有关 β 的置信域或假设检验通常是人们所关注的统计问题. 当误差变量 ε 的分布完全未知时, 经典的方法是利用 β 的最小二乘估计的渐近正态性来构造 β 的置信域. 但是这种方法有两个缺陷. 第一个缺陷是基于渐近正态性导出的置信域是对称的, 而对称的置信域并不一定合适, 特别是当总体分布为非对称时. 第二个缺陷是要估计渐近方差或协方差, 这会给统计推断带来不便, 特别是对半参数非参数模型, 方差估计往往非常困难.

近二十年来, 另一种广受关注的方法是经验似然[1,2], 它是经典的参数似然比推断向非参数情形的自然推广. [2] 证明了对数经验似然比统计量仍然满足 Wilks 定理, 即渐近服从卡方分布. 此结论可以用于构造置信域和进行假设检验.

经验似然方法已得到统计界的广泛研究[3], 参见 Owen (2001) 的专著. 我们这里关心的问题是, 当模型 (1.1) 中的协变量 X 的维数 k 随样本量 n 增加时, 经验似然比统计量的渐近性质. 高维协变量在生物信息学的研究中经常出现, 例如微阵数据分析. 传统的统计方法在处理这类数据时要特别小心. 对模型 (1.1), Owen 指出[4], 当 k 为固定数时, 对数经验似然比统计量的渐近分布为 χ_k^2. 因此, 当 k 趋于无穷大时统计量将发散. 本文将研究如何修正经验似然比统计量以及在何种条件下它依分布收敛.

---

* 国家自然科学基金 (10231030,10071090) 资助课题.
收稿日期: 2006-10-20.
谨以此文纪念成平研究员.

2 主要结果

我们首先介绍 Owen 的经验似然. 对任意 $\beta \in R^k$ 以及 $1 \le i \le n$, 记 $\varepsilon_i(\beta) = Y_i - X_i^{\mathrm{T}}\beta$. 令 $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\}$ 为在参数真值 β_0 下 $\{\varepsilon_1(\beta_0), \varepsilon_2(\beta_0), \cdots, \varepsilon_n(\beta_0)\}$ 的简记，且假定它们是独立同分布的. 对任意 $\beta \in R^k$,Owen 的经验似然函数定义为[4]

$$L(\beta) = \sup_{\{\{p_i\}_{i=1}^n \in \mathcal{F}(\beta)\}} \prod_{j=1}^n p_j, \tag{2.1}$$

其中

$$\mathcal{F}(\beta) = \Big\{\{p_i \ge 0\}_{i=1}^n : \ \sum_{j=1}^n p_j = 1, \ \sum_{j=1}^n p_j X_j \varepsilon_j(\beta) = 0\Big\}.$$

类似于 [4], 由 Lagrange 乘子法有 $L(\beta) = \prod\limits_{i=1}^n p_i(\beta)$, 其中 $p_i(\beta) = (n(1 + \lambda^{\mathrm{T}} X_i \varepsilon_i(\beta)))^{-1}$, 且 $\lambda \in R^k$ 为如下方程的解

$$\sum_{i=1}^n \frac{X_i \varepsilon_i(\beta)}{(1 + \lambda^{\mathrm{T}} X_i \varepsilon_i(\beta))} = 0. \tag{2.2}$$

因此，相应的对数经验似然比统计量为

$$\widehat{T}_n(\beta) = -2 \log \Big(\prod_{i=1}^n (n p_i(\beta))\Big). \tag{2.3}$$

对参数真值 β_0, 记 $\widehat{T}_n = \widehat{T}_n(\beta_0)$. 对固定的维数 k, [4] 证明了 $\widehat{T}_n \xrightarrow{d} \chi_k^2$. 显然，当 k 随 n 同时趋于无穷时， $\widehat{T}_n \xrightarrow{p} \infty$. 因此，已有的统计推断对高维协变量情形失效. 针对此问题，我们下面提出一些适当的正则条件，对高协变量维数 k, 证明修正的统计量渐近服从正态分布.

令 $V_n = \sum\limits_{i=1}^n X_i X_i^{\mathrm{T}}$. 显然 V_n 是 $k \times k$ 的对称矩阵且 $V_n \ge 0$. 记 $\lambda_{1n} = \min eig(V_n)$ 和 $\lambda_{kn} = \max eig(V_n)$, 其中 $\min eig(\cdot)$ 和 $\max eig(\cdot)$ 分别表示矩阵的最小和最大特征根. 记 $M_n^2 = \max\limits_{1 \le i \le n} ||X_i||^2$, $||\cdot||$ 为通常的欧氏模. 由矩阵理论，有 $k\lambda_{1n} \le \sum\limits_{i=1}^n ||X_i||^2 \le nM_n^2$ 和 $\lambda_{kn} \le \sum\limits_{i=1}^n ||X_i||^2 \le nM_n^2$. 不妨假定 $||X_1|| \ne 0$. 我们提出如下的假设条件.

(A1) 对某 $r > 0$, 有 $E(|\varepsilon_1|^{2+r}) < \infty$.

(A2) 对某 $0 \le \delta < c(r)$, 有 $\frac{nM_n^2}{k\lambda_{1n}} = O(n^\delta)$, 其中 $c(r) = \frac{r}{(2(2+r))}$.

(A3) $k = k(n) \to \infty \ (n \to \infty)$, 且

$$k^{\frac{3}{2}} = \begin{cases} O(n^{c(r)-\delta}) & \text{若 } r < 1; \\ o(n^{\frac{1}{6}-\delta}) & \text{否则.} \end{cases}$$

我们有如下的定理.

定理 1 假设 (A1)–(A3) 成立，则在真值 $H_0: \beta = \beta_0$ 下，有

$$\frac{\widehat{T}_n - k}{\sqrt{2k}} \xrightarrow{d} N(0, 1). \tag{2.4}$$

因此，当协变量维数 k 以某种合理的速度趋于无穷大时，由上述定理我们仍可以利用经验似然方法构造 β 的置信域，不过此时有关临界值的确定依赖于正态分布而非卡方分布.

3 附录证明

不失一般性，我们以下假定 $E(\varepsilon_1^2)=1$. 由假设 (A.2), 我们有 $1\le \frac{nM_n^2}{k\lambda_{1n}}\le C_0n^\delta$ 对足够大的 n 和某常数 $C_0>0$ 成立. 进而有 $\lambda_{1n}\ge \frac{n^{1-\delta}M_n}{(C_0k)}\ge \frac{n^{1-\delta}\|X_1\|}{(C_0k)}$, 它至少以 $n^{1-\delta}k^{-1}$ 的阶趋于无穷大. 因此当 n 足够大后 V_n 正定.

我们先给出一些有关矩阵性质的预备引理. 令 $A=(a_{ij})_{k\times k}$ 为 $k\times k$ 的对称矩阵, 定义矩阵模 $\|\cdot\|_M$ 为 $\|A\|_M=\max\limits_{1\le i,j\le k}|a_{ij}|$, 即 A 的绝对值分量的最大值. 令 $B=A-I_k=(b_{ij})_{k\times k}$, 其中 I_k 为 $k\times k$ 的单位阵. 记 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 和 $\tau_1,\tau_2,\cdots,\tau_k$ 分别为 A 和 B 的特征根. 若 A 可逆, 则令 $\widetilde{B}=A^{-1}-I_k=(\widetilde{b}_{ij})_{k\times k}$. 显然有 $\min\limits_{1\le i\le k}|\tau_i+1|=\min\limits_{1\le i\le k}|\lambda_i|>0$.

引理 1 $\max\limits_{1\le i\le k}|\lambda_i|\le k\|A\|_M$.

引理 2 对任意 $a,b\in R^k,\|a\|=\|b\|=1$, 有 $|a^TAb|\le\max\limits_{1\le i\le k}|\lambda_i|$.

引理 3 若 A 可逆，则 $\|\widetilde{B}\|_M\le\max\limits_{1\le i\le k}\left|\frac{\tau_i}{(1+\tau_i)}\right|=\max\limits_{1\le i\le k}\left|\frac{(\lambda_i-1)}{\lambda_i}\right|$.

证 由于 B 为对称阵，则存在正交阵 C, 使得 $B=C^{\mathrm T}\mathrm{diag}(\tau_1,\tau_2,\cdots,\tau_k)C$. 那么有,

$$\begin{aligned}A^{-1}&=(I_k+B)^{-1}=(I_k+C^{\mathrm T}\mathrm{diag}(\tau_1,\cdots,\tau_k)C)^{-1}\\&=C^{\mathrm T}\mathrm{diag}((1+\tau_1)^{-1},(1+\tau_2)^{-1},\cdots,(1+\tau_k)^{-1})C\\&=I_k-C^{\mathrm T}\mathrm{diag}\Big(\frac{\tau_1}{(1+\tau_1)},\frac{\tau_2}{(1+\tau_2)},\cdots,\frac{\tau_k}{(1+\tau_k)}\Big)C.\end{aligned}$$

因此有 $\widetilde{B}=-C^{\mathrm T}\mathrm{diag}(\frac{\tau_1}{(1+\tau_1)},\frac{\tau_2}{(1+\tau_2)},\cdots,\frac{\tau_k}{(1+\tau_k)})C$. 令 e_i 为 R^k 中第 i 个分量为 1 的单位向量，其中 $1\le i\le k$. 那么对任意的 $1\le i,j\le k$, 有

$$|\widetilde{b}_{ij}|=|e_i^{\mathrm T}C^{\mathrm T}\mathrm{diag}\Big(\frac{\tau_1}{(1+\tau_1)},\frac{\tau_2}{(1+\tau_2)},\cdots,\frac{\tau_k}{(1+\tau_k)}\Big)Ce_j|.$$

由引理 2 知 $|\widetilde{b}_{ij}|\le\max\limits_{1\le i\le k}|\frac{\tau_i}{(1+\tau_i)}|$. 因此有 $\|\widetilde{B}\|_M\le\max\limits_{1\le i\le k}|\frac{\tau_i}{(1+\tau_i)}|$.

注 1 若 $k\|B\|_M\overset{a.s.}{\to}0$, 我们可以用 $\|B\|_M$ 来给出 $\|\widetilde{B}\|_M$ 的界. 事实上, 若 $k\|B\|_M\overset{a.s.}{\to}0$, 由引理 1 知 $\max\limits_{1\le i\le k}|\tau_i|\le k\|B\|_M\overset{a.s.}{\to}0$, 而由引理 3 有 $\|\widetilde{B}\|_M\le\max\limits_{1\le i\le k}|\frac{\tau_i}{(1+\tau_i)}|$. 因此可知当 n 趋于无穷时有 $\|\widetilde{B}\|_M\le 2k\|B\|_M$ $(a.s.)$.

记 $B_n=V_n^{-\frac12}(\sum\limits_{i=1}^n X_iX_i^{\mathrm T}\varepsilon_i^2)V_n^{-\frac12}-I_k=\sum\limits_{i=1}^n(V_n^{-\frac12}X_iX_i^{\mathrm T}V_n^{-\frac12})(\varepsilon_i^2-1)=(b_{ijn})_{k\times k}$, 我们有如下关于 B_n 模收敛速度的引理.

引理 4 假设 (A1)-(A3) 成立，则有 $\|B_n\|_M\overset{a.s.}{=}O(kn^{-(b(r)-\delta)}\log n)$, 其中 $b(r)=\min(\frac12,\frac{r}{(2+r)})$, $b(r)>\delta$.

证 注意到对任意 $1\le s,l\le k$, 有

$$b_{sln}=e_s^{\mathrm T}B_ne_l=\sum_{i=1}^n(e_s^{\mathrm T}V_n^{-\frac12}X_i)(X_i^{\mathrm T}V_n^{-\frac12}e_l)(\varepsilon_i^2-1).$$

令 $a_{ni}^{(sl)}=(e_s^{\mathrm{T}}V_n^{-\frac{1}{2}}X_i)(X_i^{\mathrm{T}}V_n^{-\frac{1}{2}}e_l)$, 则有 $b_{sln}=\sum\limits_{i=1}^{n}a_{ni}^{(sl)}(\varepsilon_i^2-1)$. 我们先研究 $(kn^\delta)^{-1}b_{sln}=\sum\limits_{i=1}^{n}(kn^\delta)^{-1}a_{ni}^{(sl)}(\varepsilon_i^2-1)$ 的收敛速度. 对 $1\le i\le n$, 有 $(kn^\delta)^{-1}|a_{ni}^{(sl)}|\le(kn^\delta\lambda_{1n})^{-1}||X_i||^2\le\frac{M_n^2}{(kn^\delta\lambda_{1n})}$. 那么由 (A2) 有

$$\limsup_{n\to\infty}\max_{1\le s,l\le k}(kn^\delta)^{-1}\sum_{i=1}^{n}|a_{ni}^{(sl)}|\le\limsup_{n\to\infty}n^{1-\delta}\frac{M_n^2}{(\frac{k}{\lambda_{1n}})}\le C_0<\infty.$$

记 $d_{1n}=\max\limits_{1\le s,\,l\le k,\,1\le i\le n}(kn^\delta)^{-1}|a_{ni}^{(sl)}|$. 再由 (A2) 知 $d_{1n}\le\frac{M_n^2}{(kn^\delta\lambda_{1n})}=O(n^{-1})$. 由 (A1) 知存在某绝对常数 C_r 使得 $E(|\varepsilon_1^2-1|^{\frac{(2+r)}{2}})\le C_r(1+E(|\varepsilon_1|^{2+r})<\infty$. 因此, 由 Ghosh[5] 的引理 3 可得 $\max\limits_{1\le i\le n}|\varepsilon_i^2-1|\overset{a.s.}{=}o(n^{\frac{2}{(2+r)}})$. 这时, 再由 [6] 的引理 1, 不难推出 $\max\limits_{1\le s,\,l\le k}(kn^\delta)^{-1}|b_{sln}|\overset{a.s.}{=}O(\max(n^{\frac{2}{2+r}}d_{1n},d_{1n}^{\frac{1}{2}})\log n)$. 进而得

$$||B_n||_M=\max_{1\le s,\,l\le k}|b_{sln}|\overset{a.s.}{=}O(\max(n^{-\frac{r}{2+r}},n^{-\frac{1}{2}})n^\delta k\log n)\overset{a.s.}{=}O(kn^{-(b(r)-\delta)}\log n).$$

注 2　由假设条件 (A3) 可知 $||B_n||_M\overset{a.s.}{=}o(1)$.

对 $1\le i\le n$, 令 $Z_i=X_i\varepsilon_i$, $W_n=\sum\limits_{i=1}^{n}X_iX_i^{\mathrm{T}}\varepsilon_i^2=\sum\limits_{i=1}^{n}Z_iZ_i^{\mathrm{T}}$, $\bar{Z}_n=n^{-1}\sum\limits_{i=1}^{n}X_i\varepsilon_i$. 由引理 4 与 [6] 的引理 1, 不难推出

$$||V_n^{-\frac{1}{2}}\Big(\sum_{i=1}^{n}(Z_i-\bar{Z}_n)(Z_i-\bar{Z}_n)^{\mathrm{T}}\Big)V_n^{-\frac{1}{2}}-I_k||_M\overset{a.s.}{=}O(kn^{-(b(r)-\delta)}\log n).$$

记 $\widetilde{\lambda}_{1n}=\min eig(W_n)$. 我们有如下引理.

引理 5　假设 (A1)-(A3) 成立, 则有 $\liminf\limits_{n\to\infty}\big(\frac{\widetilde{\lambda}_{1n}}{\lambda_n}\big)\ge\frac{1}{2}$ $(a.s.)$.

证　记 $\widetilde{B}_n=\sum\limits_{i=1}^{n}X_iX_i^{\mathrm{T}}(\varepsilon_i^2-1)=(\widetilde{b}_{ijn})_{k\times k}$. 类似于引理 4 的推导, 不难推出 $||\widetilde{B}_n||_M\overset{a.s.}{=}O(k\lambda_{1n}n^{-(b(r)-\delta)}\log n)$. 对任意 $a\in R^k$, $||a||=1$, 我们有 $a^{\mathrm{T}}\widetilde{B}_na=a^{\mathrm{T}}W_na-a^{\mathrm{T}}V_na$. 由引理 1 知 $|a^{\mathrm{T}}W_na-a^{\mathrm{T}}V_na|\le k||\widetilde{B}_n||_M$ 以及 $a^{\mathrm{T}}V_na-k||\widetilde{B}_n||_M\le a^{\mathrm{T}}W_na$. 因而有 $\lambda_{1n}-k||\widetilde{B}_n||_M\le a^{\mathrm{T}}W_na$, 进而知 $\widetilde{\lambda}_{1n}\ge\lambda_{1n}-k||\widetilde{B}_n||_M$. 这样不难推出

$$\frac{\widetilde{\lambda}_{1n}}{\lambda_{1n}}\ge1-O(kn^{-(b(r)-\delta)}\log n).\quad(a.s.)$$

因此引理得证.

我们下面来研究如下 U 统计量

$$T_n=\Big(\sum_{i=1}^{n}V_n^{-\frac{1}{2}}X_i\varepsilon_i\Big)^{\mathrm{T}}\Big(\sum_{i=1}^{n}V_n^{-\frac{1}{2}}X_i\varepsilon_i\Big)=\sum_{i=1}^{n}\sum_{j=1}^{n}X_i^{\mathrm{T}}V_n^{-1}X_j\varepsilon_i\varepsilon_j$$

的渐近分布. 为避免与已有记号的混淆, 对 $1\le i,j\le n$, 我们下面记 $a_{ij}=X_i^{\mathrm{T}}V_n^{-1}X_j$, $A=(a_{ij})_{n\times n}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}V_n^{-1}(X_1,X_2,\cdots,X_n)$, $B=\mathrm{diag}(a_{11},a_{22},\cdots,a_{nn})$. 那么有

$T_n=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}a_{ij}\varepsilon_i\varepsilon_j$. 为给出 T_n 的渐近分布，我们先研究统计量 $\widetilde{T}_n$ 的渐近性质，其中

$$\widetilde{T}_n=T_n-\sum_{i=1}^{n}a_{ii}\varepsilon_i^2=\sum_{1\le i\ne j\le n}a_{ij}\varepsilon_i\varepsilon_j.$$

由 [7] 可知，$\widetilde{T}_n$ 是具有一般形式的洁净 (clean) 的 U 统计量，即其系数跟样本量 n 有关. 我们有如下引理.

引理 6 假设 (A1)-(A3) 成立，则有

$$\frac{\widetilde{T}_n-E(\widetilde{T}_n)}{\sqrt{Var(\widetilde{T}_n)}}\xrightarrow{d}N(0,1).$$

证 主要是验证 [7] 关于洁净 U 统计量渐近正态性的条件. 首先，注意到

$$\sum_{i=1}^{n}a_{ii}=\sum_{i=1}^{n}X_i^{\mathrm{T}}V_n^{-1}X_i=\mathrm{tr}\Big(\sum_{i=1}^{n}V_n^{-1}X_iX_i^{\mathrm{T}}\Big)=\mathrm{tr}(V_n^{-1}V_n)=k,$$

其中 $\mathrm{tr}(\cdot)$ 表示矩阵的迹. 又 $E(\widetilde{T}_n)=0$, 且

$$\begin{aligned}E(\widetilde{T}_n^2)&=2\sum_{1\le i\ne j\le n}a_{ij}^2=2\sum_{1\le i\ne j\le n}X_i^{\mathrm{T}}V_n^{-1}X_jX_j^{\mathrm{T}}V_n^{-1}X_i\\&=2\sum_{1\le i,j\le n}X_i^{\mathrm{T}}V_n^{-1}X_jX_j^{\mathrm{T}}V_n^{-1}X_i-2\sum_{i=1}^{n}(X_i^{\mathrm{T}}V_n^{-1}X_i)^2\\&=2\sum_{i=1}^{n}X_i^{\mathrm{T}}V_n^{-1}X_i-2\sum_{i=1}^{n}a_{ii}^2=2k-2\sum_{i=1}^{n}a_{ii}^2.\end{aligned}$$

注意到，$\sum\limits_{i=1}^{n}a_{ii}^2\le\lambda_{1n}^{-2}\sum\limits_{i=1}^{n}||X_i||^4\le\frac{nM_n^4}{\lambda_{1n}^2}=O(k^2n^{-(1-2\delta)})$, 再由 (A1)-(A3) 可知 $k^{-1}\sum\limits_{i=1}^{n}a_{ii}^2=O(kn^{-(1-2\delta)})=o(1)$. 因此有 $\frac{E(\widetilde{T}_n^2)}{(2k)}\to1$ 和 $Var(\widetilde{T}_n)=2k(1+o(1))$.

记 $E=A-B$, 并令 $\mu_1,\mu_2,\cdots,\mu_n$ 为 E 的特征根. 我们下面要证明 $\max\limits_{1\le i\le n}\frac{\mu_i^2}{k}\to0$, 它在 de Jong[7] 的理论中对验证 $\widetilde{T}_n$ 的渐近正态性起关键作用. 易知，存在正交阵 C 使得 $A-B=E=C^{\mathrm{T}}\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)C$. 注意到，$A^2=A$ 且 $E^2=C^{\mathrm{T}}\mathrm{diag}(\mu_1^2,\mu_2^2,\cdots,\mu_n^2)C=A^2-AB-BA+B^2=A+B^2-AB-BA$. 我们有

$$\begin{aligned}\sum_{i=1}^{n}\mu_i^2&=\mathrm{tr}(E^2)=\mathrm{tr}(A+B^2-AB-BA)=\mathrm{tr}(A)+tr(B^2)-2\mathrm{tr}(AB)\\&=\sum_{i=1}^{n}a_{ii}+\sum_{i=1}^{n}a_{ii}^2-2\sum_{i=1}^{n}a_{ii}^2=\sum_{i=1}^{n}a_{ii}-\sum_{i=1}^{n}a_{ii}^2\\&=k-\sum_{i=1}^{n}a_{ii}^2.\end{aligned}$$

所以 $\sum_{i=1}^{n}\frac{\mu_i^2}{k}\le 1$. 因为 $\max_{1\le i\le n}\frac{\mu_i^4}{k^2}\le\sum_{i=1}^{n}\frac{\mu_i^4}{k^2}\le\max_{1\le i\le n}\frac{\mu_i^2}{k}$, 有

$$\max_{1\le i\le n}\frac{\mu_i^2}{k}\to 0\Longleftrightarrow\sum_{i=1}^{n}\frac{\mu_i^4}{k^2}\to 0.$$

对 $1\le i\le n$, 记 $\widetilde{\mu}_i=\frac{\mu_i}{k^{\frac{1}{2}}}$. 则有

$$\frac{(A-B)}{k^{\frac{1}{2}}}=C^{\mathrm{T}}\mathrm{diag}(\widetilde{\mu}_1,\widetilde{\mu}_1,\cdots,\widetilde{\mu}_n)C=\frac{E}{k^{\frac{1}{2}}}.$$
$$\sum_{i=1}^{n}\widetilde{\mu}_i^4=\mathrm{tr}(\frac{E}{k^{\frac{1}{2}}})^4=\frac{\mathrm{tr}(E^4)}{k^2}=\frac{\mathrm{tr}((A-B)^4)}{k^2}.$$

又

$$\begin{aligned}\mathrm{tr}((A-B)^4)&=\mathrm{tr}((A+B^2-AB-BA)^2)\\&=\mathrm{tr}(A-4AB+4AB^2-4AB^3+2ABAB)\\&=\sum_{i=1}^{n}a_{ii}-4\sum_{i=1}^{n}a_{ii}^2+4\sum_{i=1}^{n}a_{ii}^3-4\sum_{i=1}^{n}a_{ii}^4+2\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ii}a_{jj}a_{ij}^2.\end{aligned}$$

注意到 $k^{-2}\sum_{i=1}^{n}a_{ii}^2=O(n^{-(1-2\delta)})=o(1)$ 且 $\max_{1\le i\le n}|a_{ii}|\le\frac{M_n^2}{\lambda_{1n}}=O(kn^{-(1-\delta)})=o(1)$. 那么，我们有

$$k^{-2}\sum_{i=1}^{n}a_{ii}^3\le k^{-2}\Big(\sum_{i=1}^{n}a_{ii}^2\Big)\max_{1\le i\le n}|a_{ii}|\to 0,$$
$$k^{-2}\sum_{i=1}^{n}a_{ii}^4\le k^{-2}\Big(\sum_{i=1}^{n}a_{ii}^2\Big)\max_{1\le i\le n}a_{ii}^2\to 0.$$

同时，$k^{-2}\sum_{i=1}^{n}a_{ii}=\frac{1}{k}=o(1)$, 且

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ii}a_{jj}a_{ij}^2\le\Big(\sum_{i=1}^{n}a_{ii}\Big)\Big(\sum_{j=1}^{n}\frac{||X_i||^2||X_j||^4}{\lambda_{1n}^3}\Big)\le\frac{knM_n^6}{\lambda_{1n}^3}.$$

因而有 $k^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ii}a_{jj}a_{ij}^2\le\frac{nM_n^6}{k\lambda_{1n}^3}=O(k^2n^{-(2-3\delta)})=o(1)$.

最后，我们有 $\sum_{i=1}^{n}\frac{\mu_i^4}{k^2}=\sum_{i=1}^{n}\widetilde{\mu}_i^4=\frac{tr((A-B)^4)}{k^2}\to 0$, 因此 $\max_{1\le i\le n}\frac{\mu_i^2}{k}\to 0$. 因为 $E\frac{(\widetilde{T}_n^2)}{(2k)}\to 1$, 有

$$\frac{\max_{1\le i\le n}\mu_i^2}{E(\widetilde{T}_n^2)}=\frac{k}{E(\widetilde{T}_n^2)}\cdot\frac{\max_{1\le i\le n}\mu_i^2}{k}\to 0.$$

由 [7] 的定理 5.2, 可知

$$\frac{\widetilde{T}_n-E(\widetilde{T}_n)}{\sqrt{E(\widetilde{T}_n^2)-(E(\widetilde{T}_n))^2}}\xrightarrow{d}N(0,1).$$

引理得证.

推论 7 假设 (A1)-(A3) 成立, 则有

$$\frac{T_n-k}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

证 注意到 $T_n-k=\widetilde{T}_n+\sum\limits_{i=1}^n a_{ii}(\varepsilon_i^2-1)$. 因此只要证明 $\sum\limits_{i=1}^n \frac{a_{ii}(\varepsilon_i^2-1)}{k^{\frac{1}{2}}}=o_p(1)$ 即可. 为此, 再次由 [6] 的引理 1, 不难推出 $\sum\limits_{i=1}^n a_{ii}(\varepsilon_i^2-1)\overset{a.s.}{=}O(kn^{-(b(r)-\delta)}\log n)\overset{a.s.}{=}o(1)$. 由引理 6 知

$$\frac{T_n-k}{\sqrt{2k}}=\frac{\widetilde{T}_n}{\sqrt{2k}}+o_p(1)=\frac{\widetilde{T}_n-E(\widetilde{T}_n)}{\mathrm{Var}(\widetilde{T}_n)}\cdot\frac{\mathrm{Var}(\widetilde{T}_n)}{\sqrt{2k}}+o_p(1)\xrightarrow{d}N(0,1).$$

我们以下主要考虑经验似然比统计量的渐近分布. 对 $1\le i\le n$, 记 $\widehat{p}_i=p_i(\beta_0)=(n(1+\lambda^{\mathrm T}X_i\varepsilon_i))^{-1}=(n(1+\lambda^{\mathrm T}Z_i))^{-1}$, 那么有

$$\widehat{T}_n=\widehat{T}_n(\beta_0)=-2\log\Big(\prod_{i=1}^n(n\widehat{p}_i)\Big)=2\sum_{i=1}^n\log(1+\lambda^{\mathrm T}Z_i),$$

其中 λ 满足方程 $\sum\limits_{i=1}^n Z_i(1+\lambda^{\mathrm T}Z_i)^{-1}=0$. 可另记 $\lambda=\rho\theta$, 其中 $\rho\ge 0$, $\theta\in R^k$, $||\theta||=1$.

注意: 下面的推导都是在真值 $\beta=\beta_0$ 下进行的.

引理 8 假设 (A1)-(A3) 成立, 则有 $||\lambda||=O_p((kn)^{\frac{1}{2}}\frac{M_n}{\lambda_{1n}})$.

证 首先, 记 $Z_n^*=\max\limits_{1\le i\le n}||Z_i||$. 则由 [5] 的引理 3, 有

$$Z_n^*\le M_n\max_{1\le i\le n}|\varepsilon_i|\overset{a.s.}{=}o(M_nn^{\frac{1}{2+r}}).$$

因为 $\sum\limits_{i=1}^n Z_i(1+\lambda^{\mathrm T}Z_i)^{-1}=0$, 那么有

$$0=\sum_{i=1}^n\frac{\lambda^{\mathrm T}Z_i}{(1+\lambda^{\mathrm T}Z_i)}=\lambda^{\mathrm T}\sum_{i=1}^n Z_i-\lambda^{\mathrm T}\sum_{i=1}^n\frac{Z_iZ_i^{\mathrm T}}{(1+\lambda^{\mathrm T}Z_i)}\lambda.$$

可得 $\lambda^{\mathrm T}\sum\limits_{i=1}^n Z_i=\lambda^{\mathrm T}\sum\limits_{i=1}^n Z_iZ_i^{\mathrm T}(1+\lambda^{\mathrm T}Z_i)^{-1}\lambda$, 亦

$$\rho\theta^{\mathrm T}\sum_{i=1}^n Z_i=\rho^2\sum_{i=1}^n\frac{\theta^{\mathrm T}Z_iZ_i^{\mathrm T}\theta}{(1+\rho\theta^{\mathrm T}Z_i)}.$$

因此有

$$|\theta^{\mathrm T}\sum_{i=1}^n Z_i|\ge\frac{\widetilde{\lambda}_{1n}\rho}{1+\rho Z_n^*}\Longrightarrow\rho\le\frac{||\sum\limits_{i=1}^n Z_i||}{(\widetilde{\lambda}_{1n}-Z_n^*||\sum\limits_{i=1}^n Z_i||)}.$$

还需要知道 $\|\sum\limits_{i=1}^{n} Z_i\|$ 的阶. 由推论 7, 我们有

$$\frac{(V_n^{-\frac{1}{2}}\sum\limits_{i=1}^{n} Z_i)^{\mathrm{T}}(V_n^{-\frac{1}{2}}\sum\limits_{i=1}^{n} Z_i)-k}{\sqrt{2k}}\xrightarrow{d} N(0,1),$$

即知 $(\sum\limits_{i=1}^{n} Z_i)^{\mathrm{T}}V_n^{-1}(\sum\limits_{i=1}^{n} Z_i)=O_p(k)$. 由 $\lambda_{kn}^{-1}\|\sum\limits_{i=1}^{n} Z_i\|^2\le(\sum\limits_{i=1}^{n} Z_i)^{\mathrm{T}}V_n^{-1}(\sum\limits_{i=1}^{n} Z_i)$ 以及 $\lambda_{kn}\le nM_n^2$, 可得 $\|\sum\limits_{i=1}^{n} Z_i\|=O_p((kn)^{\frac{1}{2}}M_n)$. 那么 $Z_n^*\|\sum\limits_{i=1}^{n} Z_i\|=o_p(k^{\frac{1}{2}}n^{\frac{4+r}{2(2+r)}}M_n^2)$. 由引理 5 与假设条件 (A1)–(A3), 不难推出

$$\widetilde{\lambda}_{1n}^{-1}Z_n^*\|\sum_{i=1}^{n} Z_i\|=o_p(k^{\frac{3}{2}}n^{-(\frac{r}{2(2+r)}-\delta)})=o_p(1).$$

因此可得 $\rho=O_p((kn)^{\frac{1}{2}}\frac{M_n}{\lambda_{1n}})$, 即 $\|\lambda\|=O_p((kn)^{\frac{1}{2}}\frac{M_n}{\lambda_{1n}})$.

定理 1 的证明　由 $\sum\limits_{i=1}^{n} Z_i(1+\lambda^{\mathrm{T}}Z_i)^{-1}=0$, 不难推出

$$\sum_{i=1}^{n} Z_i=\sum_{i=1}^{n}\frac{Z_iZ_i^{\mathrm{T}}}{(1+\lambda^{\mathrm{T}}Z_i)}\lambda=W_n\lambda-\sum_{i=1}^{n}\frac{(\lambda^{\mathrm{T}}Z_i)^2Z_i}{(1+\lambda^{\mathrm{T}}Z_i)}.$$

若记 $R_n=\sum\limits_{i=1}^{n}(\lambda^{\mathrm{T}}Z_i)^2Z_i(1+\lambda^{\mathrm{T}}Z_i)^{-1}$, 则有 $\lambda=W_n^{-1}\sum\limits_{i=1}^{n} Z_i+W_n^{-1}R_n$. 我们要证明 $W_n^{-1}R_n$ 以概率趋于零. 由于

$$\max_{1\le i\le n}|\lambda^{\mathrm{T}}Z_i|\le\|\lambda\|Z_n^*=o_p(k^{\frac{3}{2}}n^{-(\frac{r}{2(2+r)}-\delta)})=o_p(1),$$

显然有

$$\|R_n\|\le\sum_{i=1}^{n}\left\|\frac{(\lambda^{\mathrm{T}}Z_i)^2Z_i}{(1+\lambda^{\mathrm{T}}Z_i)}\right\|\le\sum_{i=1}^{n}\frac{\|\lambda\|^2\|Z_i\|^3}{|(1+\lambda^{\mathrm{T}}Z_i)|}\le\frac{\|\lambda\|^2M_n^3}{(1-\|\lambda\|Z_n^*)}\sum_{i=1}^{n}|\varepsilon_i|^3$$

对足够大的 n 成立. 那么，由大数律和 [5] 的引理 3, 知

$$\sum_{i=1}^{n}|\varepsilon_i|^3\overset{a.s.}{=}\begin{cases}O(n) & r\ge 1;\\ o(n^{\frac{3}{(2+r)}}) & r<1.\end{cases}$$

因此，由引理 5 和引理 8, 有

$$\|W_n^{-1}R_n\|\le\widetilde{\lambda}_{1n}^{-1}\|R_n\|\le\widetilde{\lambda}_{1n}^{-1}\frac{\|\lambda\|^2M_n^3}{(1-\|\lambda\|Z_n^*)}\sum_{i=1}^{n}|\varepsilon_i|^3=\begin{cases}O_p(k^4n^{-1+3\delta}) & r\ge 1;\\ o_p(k^4n^{-(\frac{1+2r}{2+r}-3\delta)}) & r<1.\end{cases}$$

由假设条件 (A3) 中的 $k^{\frac{3}{2}}n^{-(c(r)-\delta)}=O(1)$, 不难导出 $\|W_n^{-1}R_n\|=o_p(1)$.

由不等式 $|\log(1+x)-x+\frac{x^2}{2}|\le|x|^3$ 对 $x>-1$ 成立，可知

$$|\log(1+\lambda^{\mathrm{T}}Z_i)-\lambda^{\mathrm{T}}Z_i+\frac{(\lambda^{\mathrm{T}}Z_i)^2}{2}|\le|\lambda^{\mathrm{T}}Z_i|^3$$

依概率对 $1\le i\le n$ 一致成立，因为 $\max\limits_{1\le i\le n}|\lambda^{\mathrm T}Z_i|=o_p(1)$. 而

$$\widetilde{R}_n\triangleq\sum_{i=1}^n|\lambda^{\mathrm T}Z_i|^3\le||\lambda||^3M_n^3\sum_{i=1}^n|\varepsilon_i|^3=\begin{cases}O_p(k^{\frac{9}{2}}n^{-(\frac12-3\delta)}) & r\ge 1;\\ o_p(k^{\frac92}n^{-(\frac{3r}{2(2+r)}-3\delta)}) & r<1.\end{cases}$$

显然有 $\widetilde{R}_n=o_p(1)$.

那么，由上述推导与简单运算可得

$$\begin{aligned}\widehat{T}_n&=2\sum_{i=1}^n\log(1+\lambda^{\mathrm T}Z_i)=2\sum_{i=1}^n\Big(\lambda^{\mathrm T}Z_i-\lambda^{\mathrm T}Z_iZ_i^{\mathrm T}\frac{\lambda}{2}\Big)+\widetilde{R}_n\\&=\Big(\sum_{i=1}^nZ_i\Big)^{\mathrm T}W_n^{-1}\Big(\sum_{i=1}^nZ_i\Big)-R_n^{\mathrm T}W_n^{-1}R_n+o_p(1).\end{aligned}$$

注意到

$$|R_n^{\mathrm T}W_n^{-1}R_n|\le\widetilde{\lambda}_{1n}^{-1}||R_n||^2\le\widetilde{\lambda}_{1n}^{-1}||\lambda||^4M_n^6\Big(\sum_{i=1}^n|\varepsilon_i|^3\Big)^2=\begin{cases}O_p(k^7n^{-(1-5\delta)}) & r\ge1;\\ o_p(k^7n^{-(\frac{3r}{2+r}-5\delta)}) & r<1.\end{cases}$$

那么，在假设 $k^{\frac32}n^{-(c(r)-\delta)}=O(1)$ 下，相似地我们可推知 $|R_n^{\mathrm T}W_n^{-1}R_n|=o_p(1)$. 因而有 $\widehat{T}_n=(\sum\limits_{i=1}^nZ_i)^{\mathrm T}W_n^{-1}(\sum\limits_{i=1}^nZ_i)+o_p(1)\triangleq\widehat{T}_n^*+o_p(1)$.

为导出 $\widehat{T}_n$ 的渐近分布，我们先研究 $\widehat{T}_n^*$ 的渐近性质. 注意到

$$\widehat{T}_n^*=\widehat{T}_n^*-T_n+T_n=\Big(\sum_{i=1}^nZ_i\Big)^{\mathrm T}(W_n^{-1}-V_n^{-1})\Big(\sum_{i=1}^nZ_i\Big)+T_n=:R_n^*+T_n.$$

而由推论 7 知 $\frac{(T_n-k)}{(2k)^{\frac12}}\xrightarrow{d}N(0,1)$. 因此只要证明 $\frac{R_n^*}{k^{\frac12}}\xrightarrow{p}0$ 即可. 令 λ_k^* 为 $V_n^{\frac12}W_n^{-1}V_n^{\frac12}-I_k$ 的特征根绝对值的最大值. 经简单计算，可得

$$\frac{R_n^*}{(2k)^{\frac12}}=\frac{\Big(V_n^{-\frac12}\sum\limits_{i=1}^nZ_i\Big)^{\mathrm T}(V_n^{\frac12}W_n^{-1}V_n^{\frac12}-I_k)\Big(V_n^{-\frac12}\sum\limits_{i=1}^nZ_i\Big)}{(2k)^{\frac12}},$$

$$\frac{|R_n^*|}{(2k)^{\frac12}}\le\frac{\lambda_k^*\Big(\sum\limits_{i=1}^nZ_i\Big)^{\mathrm T}V_n^{-1}\Big(\sum\limits_{i=1}^nZ_i\Big)}{(2k)^{\frac12}}=\frac{\lambda_k^*(T_n-k+k)}{(2k)^{\frac12}}\le\lambda_k^*\Big(\frac{k}{2}\Big)^{\frac12}+\lambda_k^*O_p(1).$$

因此只要证明 $\lambda_k^*k^{\frac12}\xrightarrow{p}0$. 由引理 1 有 $\lambda_k^*\le k||V^{\frac12}W_n^{-1}V_n^{\frac12}-I_k||_M$, 而由引理 4 有 $||V_n^{-\frac12}W_nV_n^{-\frac12}-I_k||_M\overset{a.s.}{=}O(kn^{-(b(r)-\delta)}\log n)\overset{a.s.}{=}o(1)$. 因此，当 n 足够大时，由引理 1 和引理 3, 我们有

$$||V_n^{\frac12}W_n^{-1}V_n^{\frac12}-I_k||_M\le2k||V_n^{-\frac12}W_nV_n^{-\frac12}-I_k||_M\overset{a.s.}{=}O(k^2n^{-(b(r)-\delta)}\log n).$$

最后，对足够大的 n, 有

$$\lambda_k^*k^{\frac12}\le2k^{\frac32}||V_n^{\frac12}W_n^{-1}V_n^{\frac12}-I_k||_M\overset{a.s.}{=}O(k^{\frac52}n^{-(b(r)-\delta)}\log n)\overset{a.s.}{=}o(1).$$

因此 $\frac{R_n^*}{k^{\frac{1}{2}}} = o_p(1)$. 由推论 7 知

$$\frac{(\widehat{T}_n^* - k)}{(2k)^{\frac{1}{2}}} \xrightarrow{d} N(0,1),$$

进而

$$\frac{(\widehat{T}_n - k)}{(2k)^{\frac{1}{2}}} \xrightarrow{d} N(0,1).$$

定理得证.

参 考 文 献

[1] Owen A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 1988, **75**: 237–249.
[2] Owen A. Empirical likelihood confidence regions. *Ann. Statist.*, 1990, **18**: 90–120.
[3] Owen A. Empirical likelihood. *Chapman & Hall*, New York, 2001.
[4] Owen A. Empirical likelihood for linear models. *Ann. Statist.*, 1991, **19**: 1725–1747.
[5] Ghosh M. A note on bootstrapping the sample median. *Ann. Statist.*, 1984, **12**: 1130–1135.
[6] Shi J and Lau T S. Empirical likelihood for partially linear models. *J. Multivariate Analysis*, 2000, **72**: 132–148.
[7] de Jong P. A central limit theorem for generalized quadratic forms. *Probab. Theory & Related Fields*, **75**: 261–277.

EMPIRICAL LIKELIHOOD FOR HIGHER DIMENSIONAL LINEAR MODELS

Shi Jian

(*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100080)

Abstract Empirical likelihood for higher dimensional linear model is studied in this paper. When the dimension number of covariates increases with the sample size, conventional empirical likelihood inference becomes invalid. To this end, some mild regularity conditions are imposed and an asymptotic distribution theory is established for modified statistic.

Key words Linear models, empirical likelihood, higher dimensional convariates.