

平衡增加的经验欧氏似然 *

申群海¹ 黄运生² 张军舰¹

(¹广西师范大学数学与统计学院, 桂林, 541004; ²开封文化艺术职业学院, 开封, 475004)

摘要

经验(欧氏)似然方法是近年来非常流行的一种非参数统计方法. 针对经验(欧氏)似然的凸包限制和计算复杂问题, 本文借助Emerson和Owen (2009)所提出的平衡增加思想对经验欧氏似然进行修正, 得到了平衡增加的经验欧氏似然. 随后论文从理论和模拟两个方面进行了研究. 理论上给出了该方法与经验欧氏似然检验函数之间的联系, 即在固定的样本量 n 下随着添加点位置的连续变化, 检验方法可以从简单的均值增加经验欧氏似然变化到经验欧氏似然检验; 模拟结果显示, 适当选取调整因子, 平衡增加的经验欧氏似然相对于(调整)经验欧氏似然而言, 在大多数情况下, 其分布更接近于对应的极限分布.

关键词: 经验欧氏似然, 凸包限制, 平衡增加.

学科分类号: O212.7.

§1. 引言

经验似然是在完全样本下提出的一种非参数统计推断方法, 它有类似于Bootstrap的抽样特性. 这一方法与经典的统计方法比较有较多突出的优点, 如用经验似然方法构造的置信区域有域保持性、变换不变性、经验似然区域的形状只与样本数据有关、Bartlett纠偏性及无需构造枢轴量等(见王启华, 2004). 因此, 该方法被广泛应用于如计量经济学、生物医学和市场调查等领域, 是目前统计理论和应用研究里比较热门的研究课题之一.

经验似然方法是由Owen (1988)系统提出并加以讨论, Qin和Lawless (1994)将其引入到较为一般的半参数模型中. 随后, 许多统计学家和应用学者不断把经验似然方法进行扩充并应用到各种各样的模型中. 比较有代表性的文献如下: Owen (2001)概括了上世纪有关经验似然的研究结果, 给出了一些当时还存在的问题及研究思路; 王启华(2004)对生存数据等方面的经验似然方法进行了综述; 张军舰(2006, 2008)给出经验似然的一种推广, 并介绍了经验似然在拟合优度检验方面的一些应用; Kitamura (2007)调查了经验似然在计量经济学方面的应用; Chen和Keilegom (2009)讨论了经验似然在回归方面的应用; Xue和Zhu (2010)给出了经验似然方法在非参和半参数模型中的应用. 尽管经验似然方法具有较多优势, 但也存在诸如计算量较大、在小样本时所构造的置信区域或估计方程可能

*国家自然科学基金(11261009)、国家社会科学基金(10CTJ004)、广西自然科学基金(2012GXNSFAA053004)资助.

本文2014年2月20日收到, 2014年7月8日收到修改稿.

doi: 10.3969/j.issn.1001-4268.2014.05.008

无解存在(即凸包限制)等问题. 因此, 寻找计算简单、小样本时无凸包限制, 且与经验似然又具有类似统计性质的方法显得非常有必要. 针对凸包限制问题, Chen等(2008)给出了调整的经验似然方法. 针对计算简单这一点, 比较典型的方法是经验欧氏似然方法, 详见罗旭(1994)所给出的系统讨论.

调整经验似然方法的基本思想是在原有数据点中人为地添加一个新的数据点, 把此点与原有数据合在一起组成新的样本, 然后重新利用经验似然方法. 这样做可以保证在小样本时所考虑的方程始终有解. 调整经验似然方法尽管可以解决凸包问题, 但在调整时由于人为因素, 改变了样本均值, 出现了样本不对称问题. 为了克服这一缺陷, Emerson和Owen(2009)提出对原始数据在参数真值的两侧同时进行修正, 即平衡增加经验似然, 并对其性质加以讨论.

另一方面, 经验欧氏似然是用欧氏距离代替经验似然中的似然距离而得到的一种非参数方法, 它具有与经验似然完全类似的性质. 并且其计算简单, 有较高的推广价值. 但经验欧氏似然仍然没有突破凸包限制问题. 为了克服凸包限制, 又使计算简单, 方便应用, 胡志明(2011)在Chen等(2008)的基础上, 对经验欧氏似然进行调整, 得到了调整经验欧氏似然方法. 类似地, 这种方法也有样本不对称问题.

本文计划借助Emerson和Owen(2009)的平衡增加经验似然思想, 讨论平衡增加经验欧氏似然函数的构造, 理论上给出所得平衡增加经验欧氏似然比检验与Hotelling T^2 检验之间的联系以及相应的极限分布. 然后再从多维数据方面, 对所得方法进行模拟比较.

本文第二节给出主要思想和结论, 第三节介绍模拟研究结果, 最后给出第二节主要结论的证明.

§2. 主要问题和结论

设 X_1, \dots, X_n 为来自总体 X 的一组独立同分布 (i.i.d.) d 维样本, X 的分布函数为 F (未知), $\theta = (\theta_1, \theta_2, \dots, \theta_p)^\top \in \Theta$ 为 p 维待估计参数, r ($r \geq p$) 维向量函数 $g(x, \theta) = (g_1(x, \theta), g_2(x, \theta), \dots, g_r(x, \theta))^\top$ 满足 $Eg(X, \theta) = 0$ 且函数 g 的形式已知, 即 g 为 Qin 与 Lawless (1994) 所定义的无偏估计函数. 我们对未知均值参数 θ 进行如下检验:

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0, \quad (2.1)$$

其中 θ_0 为已知参数真值. 由 Qin 与 Lawless (1994), 关于 (2.1) 的经验似然比检验函数为

$$R(\theta) = \sup \left\{ \prod_{i=1}^n n\omega_i \mid \sum_{i=1}^n \omega_i g(X_i, \theta) = 0, \omega_i \geq 0, \sum_{i=1}^n \omega_i = 1 \right\},$$

其中 $\omega_i = F\{X_i\}$ 为分布在 X_i 处的概率质量, -2 倍的经验对数似然比函数为

$$W_n(\theta) = -2 \log R(\theta).$$

由Qin与Lawless (1994)中定理2可得, 在适当的正则条件下, $W_n(\theta_0)$ 依分布收敛于自由度为 r 的卡方随机变量 χ_r^2 .

事实上, $\log \prod_{i=1}^n n\omega_i = \sum_{i=1}^n (\log \omega_i - \log 1/n)$ 可以看成是点 $(\omega_1, \omega_2, \dots, \omega_n)^\top$ 到点 $(1/n, 1/n, \dots, 1/n)^\top$ 的似然距离, 自然的一个推广是用这两点的欧氏距离 $\sum_{i=1}^n (\omega_i - 1/n)^2$ 来代替似然距离, Owen (1988) 依据这一思想, 给出了经验欧氏似然的定义, 即

$$R_E^*(\theta) = \inf \left\{ \sum_{i=1}^n \left(\omega_i - \frac{1}{n} \right)^2 \mid \sum_{i=1}^n \omega_i g(X_i, \theta) = 0, \omega_i \geq 0, \sum_{i=1}^n \omega_i = 1 \right\}.$$

经验(欧氏)似然方法的前提条件是上述集合中 ω_i 存在, 即0要在 $g(X_i, \theta)$ 的凸包内部, 否则求解过程可能无解. 当样本容量 $n \rightarrow \infty$ 时, μ 在 $g(X_i, \theta)$ 的凸包内部依概率1成立, 但在小样本高维数据下还是有较高的概率使得0不在 $g(X_i, \theta)$ 的凸包内部(详见Tsao (2004) 的进一步讨论). 针对凸包问题, 许多学者进行过讨论, 如Owen (2001) 的重复抽样(Bootstrap)、Bartlett调整、经验似然t方法、Bartolucci (2007) 的惩罚经验似然、以及Chen等(2008)和Emerson和Owen (2009) 的调整经验似然等. 本文主要集中在Emerson和Owen (2009) 的平衡增加调整思路对经验欧氏似然进行讨论.

记

$$\bar{g} = \bar{g}(\theta) = \frac{1}{n} \sum_{i=1}^n g(X_i, \theta), \quad S = \frac{1}{n-1} \sum_{i=1}^n (g(X_i, \theta) - \bar{g})(g(X_i, \theta) - \bar{g})^\top.$$

假定 $Eg(X, \theta_0)g(X, \theta_0)^\top$ 可逆, 则当 n 足够大时, S 以概率1可逆. 对检验问题(2.1), 可以得到经验欧氏似然检验函数具有类似Hotelling T^2 检验函数形式(详见罗旭, 1994), 即

$$T_E^2(\theta) = n\bar{g}(X, \theta)^\top S^{-1}\bar{g}(X, \theta).$$

如果 $g(X, \theta) = X - \theta$ 时, 此检验函数在线性变换下具有不变性等优点, 且在总体 X 分布为 d 元正态分布 $N_d(\mu, \Sigma)$ 时,

$$\frac{n-d}{(n-1)d} T_E^2(\theta) \sim F_{d, n-d}.$$

当 d, r 固定, $n \rightarrow \infty$ 时, 其极限分布是 χ_r^2 . 令

$$v^* = \bar{g}(X, \theta), \quad r^* = \|v^*\|, \quad u^* = \frac{v^*}{r^*},$$

根据平衡增加的思想(见Emerson和Owen, 2009), 向原始数据添加如下的两个数据点:

$$g_{n+1} = g_{n+1}(\theta) = -sc_{u^*}u^*, \quad g_{n+2} = g_{n+2}(\theta) = 2\bar{g}(\theta) + sc_{u^*}u^*,$$

其中 $c_{u^*} = ((u^*)^\top S^{-1}u^*)^{-1/2}$, s 为调整因子, 后面将进一步叙述. c_{u^*} 可以看作是一个Mahalanobis距离, 为讨论方便, 此处不妨假定 $P(\bar{g} = 0) = 0$, 这样就排除了 u^* 无意义的情况.

由此, 把 g_{n+1} 、 g_{n+2} 与原数据点放在一起组成新的数据集, 然后再利用经验欧氏似然方法, 可以得到平衡增加的经验欧氏似然为

$$R^*(\theta) = \inf \left\{ \sum_{i=1}^{n+2} \left(\omega_i - \frac{1}{n+2} \right)^2 \mid \sum_{i=1}^{n+2} \omega_i g(X_i, \theta) = 0, \omega_i \geq 0, \sum_{i=1}^{n+2} \omega_i = 1 \right\}. \quad (2.2)$$

为方便表达, 不妨记 $g_i = g(X_i, \theta)$, $i = 1, \dots, n$. 数据集 $\{g_1, \dots, g_n, g_{n+1}, g_{n+2}\}$ 中, $g_{n+1} - \bar{g}$ 与 $g_{n+2} - \bar{g}$ 方向相反, 且距离相等, 新的数据集没有改变均值. 当 $g(X, \theta) = X - \theta$ 时, 这种调整具有线性不变性(见Emerson和Owen, 2009的命题4.1), 可以避免检验函数有界的情况, 具体的优势参见Emerson和Owen(2009)第1172页的讨论. 下面我们讨论当调整因子 s 变化时, 平衡增加的经验欧氏似然函数的变化情况, 进一步发现平衡增加的经验欧氏似然函数与未调整的经验欧氏似然检验函数之间的关系.

为方便讨论, 我们对原始数据作如下标准化变换:

$$Z_i = A^{-1}(g_i - \bar{g}), \quad i = 1, \dots, n,$$

其中 A 为实空间上可逆的 $r \times r$ 阶可逆矩阵, 满足 $AA^T = S$. 由此易得 $\bar{Z} = (1/n) \sum_{i=1}^n Z_i = 0$. 令 $\eta = -A^{-1}\bar{g}$, $v = \bar{Z} - \eta = -\eta$, $r = \|v\| = \|\eta\|$, $u = v/r = -\eta/\|\eta\|$. 则 $-\eta = ru$, 参数 θ 的信息包含在 η 中. 由此记号, 经验欧氏似然函数可以简化为

$$T_E^2(\eta) = nr^2 u^T u.$$

添加点 g_{n+1} 、 g_{n+2} 变为

$$Z_{n+1} = \eta - su, \quad Z_{n+2} = -\eta + su = (2r + s)u.$$

从而, $Z_{n+1} - \eta = -su$, $Z_{n+2} - \eta = -2\eta + su = (2r + s)u$. 此时平衡增加的经验欧氏似然为

$$\tilde{R}(\eta) = \inf \left\{ \sum_{i=1}^{n+2} \left(\omega_i - \frac{1}{n+2} \right)^2 \mid \sum_{i=1}^{n+2} \omega_i Z_i = \theta, \omega_i \geq 0, \sum_{i=1}^{n+2} \omega_i = 1 \right\}.$$

由拉格朗日乘子法, 令

$$G = -\frac{1}{2} \sum_{i=1}^{n+2} ((n+2)\omega_i - 1)^2 - (n+2)t^T \left(\sum_{i=1}^{n+2} \omega_i (Z_i - \eta) \right) - \lambda \left(\sum_{i=1}^{n+2} \omega_i - 1 \right).$$

可得

$$\frac{\partial G}{\partial \omega_i} = -(n+2)((n+2)\omega_i - 1) - (n+2)t^T(Z_i - \eta) - \lambda = 0. \quad (2.3)$$

对(2.3)式最后一等号两边求和得

$$(n+2)t^T \sum_{i=1}^{n+2} (Z_i - \eta) + \lambda(n+2) = 0.$$

从而

$$\lambda = -t^T \sum_{i=1}^{n+2} (Z_i - \eta) = (n+2)t^T \eta.$$

代入(2.3)式最后一等式得

$$\omega_i = \frac{1}{n+2} (1 - t^T Z_i).$$

此时

$$\tilde{R}(\eta) = \frac{1}{(n+2)^2} \sum_{i=1}^{n+2} (t^T Z_i)^2 = \frac{1}{(n+2)^2} \sum_{i=1}^{n+2} t^T Z_i Z_i^T t, \quad (2.4)$$

其中 t 满足

$$\frac{1}{n+2} \sum_{i=1}^{n+2} (1 - t^T Z_i)(Z_i - \eta) = 0. \quad (2.5)$$

我们发现当调整因子 s 变化时有如下结论:

定理 2.1 设 X_1, \dots, X_n 为来自总体 X 的一组独立同分布(i.i.d.) d 维样本, 对固定的样本容量 n , 当 $s \rightarrow \infty$ 时,

$$2ns^2 \tilde{R}(\eta) \rightarrow T_E^2(\eta),$$

其中, $\tilde{R}(\eta)$ 见(2.4)所示, $T_E^2(\eta)$ 为关于参数 η 的经验欧氏似然检验函数.

这里调整因子 s 至关重要, 由添加点定义, $s = 0$ 时, 相当于只利用均值进行平衡调整, 对固定的样本容量 n , s 决定了平衡调整后的样本凸包大小. 随着 s 从0变化到 ∞ , 检验函数的调整力度也在不断加大, 这说明在小样本情况下, 平衡增加的经验欧氏似然改变了检验函数的结构. 由于 η 是 θ 的函数, 上述结论也可以看作是关于 θ 的函数.

为便于比较, 我们需要知道当 s 固定时, 在零假设成立下检验统计量的极限分布. 事实上, 类似于胡志明(2011)的证明思路, 容易发现这个极限分布为 χ_r^2 . 也即有如下结论:

定理 2.2 设 X_1, \dots, X_n 为来自总体 X 的一组独立同分布(i.i.d.) d 维样本, 假定 $Eg(X, \theta_0)g(X, \theta_0)^T$ 可逆, 对固定的调整因子 s , 当 $n \rightarrow \infty$ 时,

$$n[n-1+2(r+s)^2]R^*(\theta_0) \rightarrow \chi_r^2,$$

其中 $R^*(\theta)$ 见(2.2)所示. 也即在零假设下平衡增加的经验欧氏似然经验函数的极限分布为 χ_r^2 .

由上述两个结论, 我们可以借助模拟或其极限分布构造相应的估计或检验. 在此就不一一细述. 上述两个定理的较为详细的证明过程将留在第四节给出.

§3. 模拟研究

在用经验(欧氏)似然方法进行检验时, 零假设下检验统计量的真实 χ^2 分布与其极限分布的差距比较大, 我们用模拟数据简单说明这一点. 我们取总体为 $d = 4$ 维的标准正态分布,

样本容量为 $n = 10$, 分别用未调整经验似然和经验欧氏似然方法模拟在零假设下检验统计量的分布, 然后与其极限分布 χ_4^2 进行比较, 模拟重复次数为1万次, 其结果如图1所示. 图中横坐标表示 χ_4^2 的分位数, 纵坐标表示表示模拟得到的经验分位数, 直线表示用其极限分布得到的结果(即以 χ_4^2 的分位数为横坐标, 以 χ_4^2 的分位数为纵坐标); EL线表示用经验似然得到的结果(即以 χ_4^2 的分位数为横坐标, 以模拟得到的经验似然分位数为纵坐标); EEL线表示用经验欧氏似然得到的结果(即以 χ_4^2 的分位数为横坐标, 以模拟得到的经验欧氏似然分位数为纵坐标). 图中直线上的黑色菱形表示90%分位点. 我们知道, 理论上经验似然和经验欧氏似然的极限分布是 χ_4^2 , 所以图中的EL线和EEL线应该有向图中的直线逐渐收拢的趋势. 但由图1可以看出, 在零假设下, 经验(欧氏)似然检验统计量分布与其极限分布差距较大.

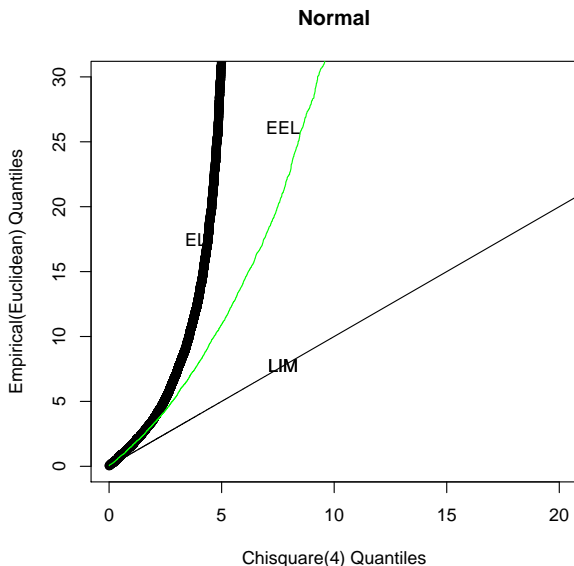


图1 经验(欧氏)分位数与极限分位数的比较

借助同样的方法, 我们比较调整方法和平衡增加调整方法的情况. 为便于比较, 我们分别选取四个 $d = 4$ 维分布总体, 每个总体具有独立且相同的边际分布, 所选取的边际分布在偏度和峰度方面都有所区别, 以便更有代表性. 这些边际分布分别为标准正态分布 $N(0, 1)$, 自由度为7的t分布 $t(7)$, 卡方分布 χ_1^2 和第一自由度为4、第二自由度为10的F分布 $F(4, 10)$. 当边际分布为 $N(0, 1)$ 时, 其偏度和峰度均为0; 边际分布为 $t(7)$ 时, 其偏度为0, 峰度为2; 边际分布为 χ_1^2 分布时, 其偏度为 $2\sqrt{2}$, 峰度为12; 边际分布为 $F(4, 10)$ 分布时, 其偏度为4, 峰度为54. 所考虑的样本容量均为 $n = 20$. 为使图形更直观, 我们对同一分布模拟数据的经验欧氏似然(EEL)、调整经验欧氏似然(AEEL)、平衡增加的经验欧氏似然(BAEEL)以及其对应的极限分布分别画在一张图上进行比较, 详见图2所示. 在利用调整经验欧氏似然(AEEL)时, 调整因子 a_n 取为 $\log(\log(n))/4$; 在利用平衡增加的经验欧氏似然(BAEEL)时,

调整因子 s 取为 $\log(n)/2$. 之所以选择以上两种调整因子, 是因为这两种调整因子不仅满足论文中的有关要求, 而且模拟经验比较也发现, 在收敛到其极限分布方面, 这两种调整因子也有较好的表现.

《应用概率统计》版权所有

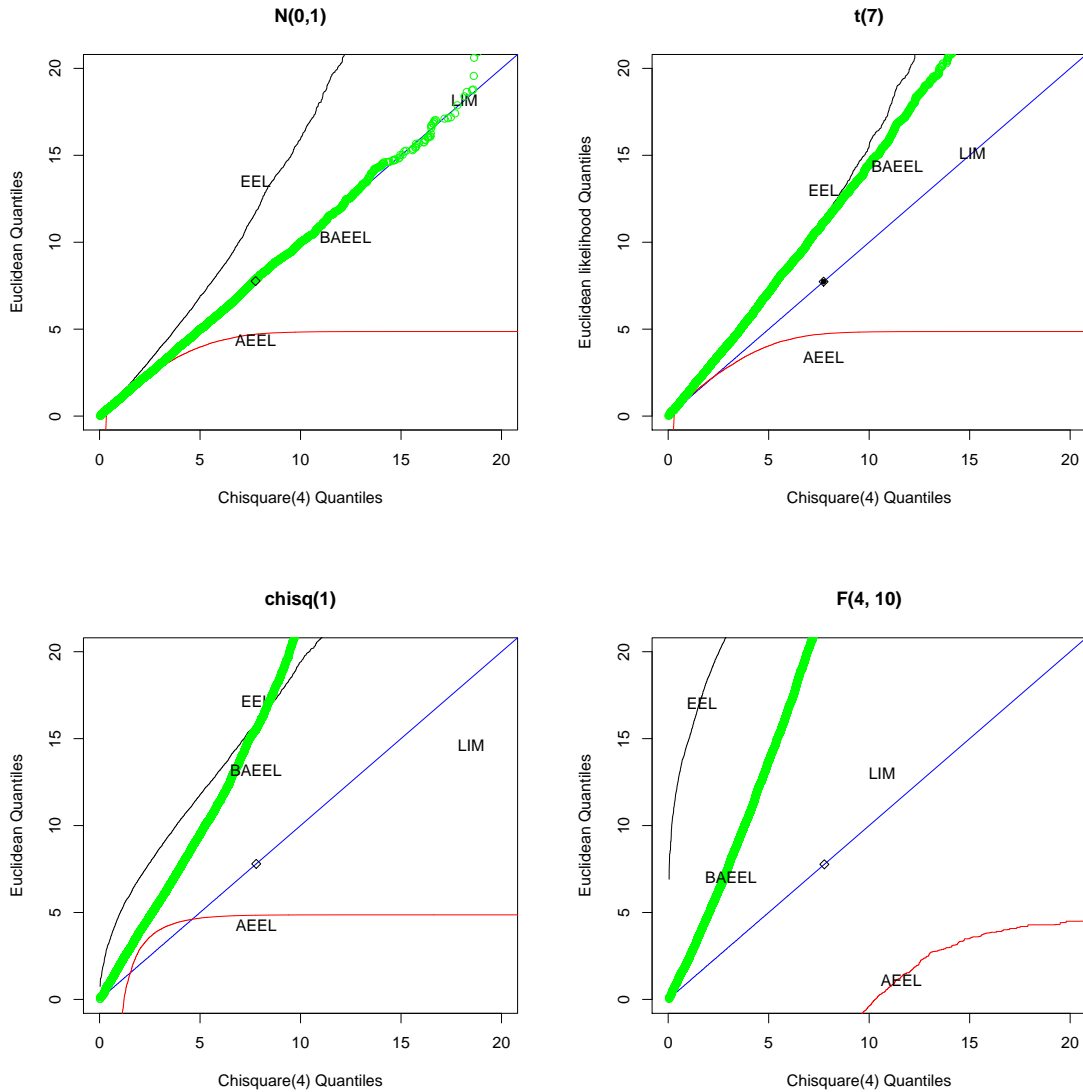


图2 分位数与分位数的比较图

以 $N(0, 1)$ 为例, 直线表示用其极限分布得到的结果(即以 χ_4^2 的分位数为横坐标, 以 χ_4^2 的分位数为纵坐标); EEL线表示用经验欧氏似然得到的结果(即以 χ_4^2 的分位数为横坐标, 以模拟得到的经验欧氏似然分位数为纵坐标); AEEL线表示用调整经验欧氏似然得到的结果(即以 χ_4^2 的分位数为横坐标, 以模拟得到的调整经验欧氏似然分位数为纵坐标); BAEEL线表示用平衡增加的经验欧氏似然得到的结果(即以 χ_4^2 的分位数为横坐标, 以模拟

得到的平衡增加的经验欧氏似然分位数为纵坐标). 由图2可以看出, 调整经验欧氏似然与其极限分布 χ_4^2 的起初优于欧氏似然, 但在90%分位点之前优势不大, 而平衡增加的经验欧氏似然则在90分位点之后仍保持与极限分布有较高的一致性. 因此我们可以说而平衡增加的经验欧氏似然整体上优于经验欧氏似然和调整经验欧氏似然.

由前面几节的结果可以发现, 平衡增加的经验欧氏似然计算简单, 有效突破了传统经验似然的凸包限制, 性质较好, 值得进一步推广和应用.

§4. 定理的证明

定理2.1的证明: 由(2.5)式, 并利用序列 Z_i 的性质, 易得

$$\begin{aligned} 0 &= \sum_{i=1}^{n+2} (Z_i - \eta) - \sum_{i=1}^{n+2} t^\top Z_i (Z_i - \eta) \\ &= (n+2)ru - \sum_{i=1}^{n+2} (Z_i - \eta) Z_i^\top t \\ &= (n+2)ru - \sum_{i=1}^{n+2} Z_i Z_i^\top t. \end{aligned}$$

进而

$$t = \left(\frac{1}{n+2} \sum_{i=1}^{n+2} Z_i Z_i^\top \right)^{-1} ru \triangleq B^{-1} ru,$$

其中

$$B = \frac{1}{n+2} \sum_{i=1}^{n+2} Z_i Z_i^\top.$$

把 t 的表达式代入(2.4)式可得

$$\tilde{R}(\eta) = \frac{1}{(n+2)^2} \sum_{i=1}^{n+2} t^\top Z_i Z_i^\top t = \frac{r^2}{n+2} u^\top B^{-1} B B^{-1} u = \frac{r^2}{n+2} u^\top B^{-1} u.$$

而

$$\begin{aligned} B &= \frac{1}{n+2} \sum_{i=1}^{n+2} Z_i Z_i^\top \\ &= \frac{1}{n+2} \left(\sum_{i=1}^n Z_i Z_i^\top + Z_{n+1} Z_{n+1}^\top + Z_{n+2} Z_{n+2}^\top \right) \\ &= \frac{1}{n+2} \left(\sum_{i=1}^n A^{-1} (g_i - \bar{g})(g_i - \bar{g})^\top A^{-1} + 2(r+s)^2 uu^\top \right) \\ &= \frac{1}{n+2} \left((n-1)I_d + 2(r+s)^2 uu^\top \right) = \frac{n-1}{n+2} \left(I_d + \frac{2(r+s)^2}{n-1} uu^\top \right), \end{aligned}$$

其中 I_d 表示 d 维单位阵. 容易推得

$$\left(I_d + \frac{2(r+s)^2}{n-1} uu^\top \right)^{-1} = I_d - \frac{2(r+s)^2}{n-1+2(r+s)^2} uu^\top.$$

所以

$$B^{-1} = \frac{n+2}{n-1} \left[I_d - \frac{2(r+s)^2}{n-1+2(r+s)^2} uu^\top \right].$$

从而

$$\begin{aligned} \tilde{R}(\eta) &= \frac{r^2}{n+2} u^\top B^{-1} u \\ &= \frac{r^2}{n-1} u^\top u - \frac{r^2}{n-1} \frac{2(r+s)^2}{n-1+2(r+s)^2} u^\top u \\ &= \frac{1}{n-1+2(r+s)^2} r^2 u^\top u. \end{aligned}$$

由此, 显然可得对固定的样本容量 n , 当 $s \rightarrow \infty$ 时,

$$2ns^2 \tilde{R}(\eta) \rightarrow T_E^2(\eta).$$

定理证毕. \square

定理2.2的证明: s 固定, 当 $n \rightarrow \infty$ 时, 由定理2.1的证明过程, 显然有

$$n[n-1+2(r+s)^2] R^*(\theta_0) = n[n-1+2(r+s)^2] \tilde{R}(\eta) = nr^2 u^\top u \rightarrow \chi_r^2.$$

定理证毕. \square

参 考 文 献

- [1] 王启华, 经验似然统计推断方法发展综述, 数学进展, **33(2)**(2004), 141–151.
- [2] Owen, A.B., Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75(2)** (1988), 237–249.
- [3] Qin, J. and Lawless, J., Empirical likelihood and general estimating equations, *The Annals of Statistics*, **22(1)**(1994), 300–325.
- [4] Owen, A.B., *Empirical Likelihood*, New York: Chapman and Hall, 2001.
- [5] 张军舰, 广义非参似然拟合优度检验, 中国科学院数学与系统科学研究院博士学位论文, 2006.
- [6] 张军舰, 非参似然拟合优度检验, 北京工业大学博士后出站报告, 2008.
- [7] Kitamura, Y., Empirical likelihood methods in econometrics: theory and practice, in Blundell, R., Newey, W.K. and Persson, T. (Editors), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Vol. 3 (Econometric Society Monographs)*, 2007, 174–237.
- [8] Chen, S.X. and Keilegom, I.V., A review on empirical likelihood methods for regression, *Test*, **18(3)** (2009), 415–447.
- [9] Xue, L.G. and Zhu, L.X., *Empirical Likelihood in Nonparametric and Semiparametric Models*, Beijing: Science Press, 2010.
- [10] Chen, J., Variyath, A.M. and Abraham, B., Adjusted empirical likelihood and its properties, *Journal of Computational and Graphical Statistics*, **17(2)**(2008), 426–443.

- [11] 罗旭, 半参数模型的经验欧氏似然估计的大样本性质, *应用概率统计*, **10(4)**(1994), 344–352.
- [12] Emerson, S.C. and Owen, A.B., Calibration of the empirical likelihood method for a vector mean, *Electronic Journal of Statistics*, **3**(2009), 1161–1192.
- [13] 胡志明, 调整的经验欧氏似然及其性质, 广西师范大学硕士学位论文, 2011.
- [14] Tsao, M., Bounds on coverage probabilities of the empirical likelihood ratio confidence regions, *The Annals of Statistics*, **32(3)**(2004), 1215–1221.
- [15] Bartolucci, F., A penalized version of the empirical likelihood ratio for the population mean, *Statistics & Probability Letters*, **77(1)**(2007), 104–110.

Balanced Augmented Empirical Euclidean Likelihood

SHEN QUNHAI¹ HUANG YUNSHENG² ZHANG JUNJIAN¹

(¹*School of Mathematics and Statistics, Guangxi Normal University, Guilin, 541004*)

(²*Kaifeng Vocational College of Culture and Arts, Kaifeng, 475004*)

Empirical (Euclidean) likelihood is a very popular nonparametric statistical method in recent years. In view of the convex hull restrictions and complex calculation of empirical (Euclidean) likelihood, the balanced augmented empirical Euclidean likelihood (BAEEL) is proposed by using the idea of Emerson and Owen (2009). Then the BAEEL method is investigated from two aspects of theory and simulation. In theory, the connection between BAEEL method and the empirical Euclidean likelihood method is deduced. That is, with fixed sample size and the continuous varied location of augmented points, the test can be varied from the simple mean augmented empirical Euclidean likelihood to empirical Euclidean likelihood test. Simulation results show that the distribution of the BAEEL converges its limit distribution more rapidly than that of the (adjusted) empirical Euclidean likelihood in most cases.

Keywords: Empirical euclidean likelihood, convex hull restrictions, balanced augmentation.

AMS Subject Classification: 62G10, 62H15.