第 41 卷 第 4 期 应 用 数 学 学 报 Vol. 41 No. 4
2018 年 7 月 ACTA MATHEMATICAE APPLICATAE SINICA July, 2018

对数线性模型下基于 ϕ- 散度的单边检验 *

金应华

(广东工业大学应用数学学院，广州 510006)

(E-mail: jyh@mail.ustc.edu.cn)

吴耀华

(中国科学技术大学统计与金融系，合肥 230026)

(E-mail: wuyh@ustc.edu.cn)

邵全喜

(CSIRO Data61, Leeuwin Centre, 65 Broockway Road, Floreat, WA 6014, Australia)

(E-mail: quanxi.shao@data61.csiro.au)

摘　要　单边检验是假设检验理论的重要组成部分之一．本文研究了在乘积多项抽样对数线性模型下的某种单边假设检验问题．基于 ϕ- 散度和约束最小 ϕ- 散度估计 (RMϕDE), 提出了三类检验统计量，证明了它们有相同的渐近分布，即类卡方 (Chi-bar-square) 分布．此三类统计量包含似然比 (likelihood ratio) 统计量和皮尔逊 (Pearson) 统计量等特例，推广了现有文献的研究结果．实例分析展示了此三类统计量的检验效果．模拟研究表明，在样本量较小时，功效散度 (power-divergence) 族中存在比似然比统计量和皮尔逊统计量表现更好的替代.

关键词　单边检验； ϕ- 散度；乘积多项抽样；对数线性模型；类卡方分布

MR(2000) 主题分类　62J12; 62F05

中图分类　O212.1

1　引言

假设由 k 个分量组成的计数向量 n 来自于乘积多项抽样，其中多项分布有 r(正整数) 个，且各自抽样是独立的．记此 r 个多项分布为

$$\mathcal{M}(N_1, \Pi_1), \cdots, \mathcal{M}(N_r, \Pi_r),$$

本文 2018 年 3 月 14 日收到．2018 年 5 月 24 日收到修改稿.

* 国家自然科学基金 (11401114), 广东省自然科学基金 (S2012040007622) 资助项目.

其中 N_i 和 $\Pi_i \equiv (\pi_{i1}, \cdots, \pi_{ik_i})^T$ 分别为第 i 个多项分布的试验次数和概率向量. 此处上标 T 表示向量或矩阵的转置. 进一步假设计数向量 n 由 r 部分组成且依次对应此 r 个多项分布. 记 $N = \Sigma_{i=1}^r N_i$ 和 $n \equiv (n_{11}, \cdots, n_{1k_1}; \cdots; n_{r1}, \cdots, n_{rk_r})^T$. 此时 $k = \sum\limits_{i=1}^r k_i$ 成立.

假设概率向量 Π_i 服从对数线性模型

$$\pi_{ij}(\theta) = \frac{\exp(w_{ij}^T\theta)}{\Sigma_{l=1}^{k_i} \exp(w_{il}^T\theta)}, \qquad \text{for } i = 1, \cdots, r, \quad j = 1, \cdots, k_i, \tag{1.1}$$

其中 w_{ij} 是 t 维列向量. 设 $k \times t$ 阶矩阵 $W = (w_{11}, \cdots, w_{1k_1}; \cdots; w_{r1}, \cdots, w_{rk_r})^T$ 列满秩且它的列向量组与 k 维列向量组 $\nu_1, \cdots, \nu_r$ 线性无关, 此处 $\nu_i = (\mathbf{0}_{k_1}^T, \cdots, \mathbf{0}_{k_{i-1}}^T, \mathbf{1}_{k_i}^T, \mathbf{0}_{k_{i+1}}^T, \cdots, \mathbf{0}_{k_r}^T)^T$, $\mathbf{1}_l$ 是元素全为 1 的 l 维列向量, $\mathbf{0}_l$ 是元素全为 0 的 l 维列向量. 设 $t < k - r$ 成立, 模型 (1.1) 的参数 θ 可在整个 t 维实数集 $\mathcal{R}^t$ 中取值.

设 $L(\theta)$ 是参数 θ 的似然函数, 可以证明

$$\log L(\theta) = C - N \sum_{i=1}^r \sum_{j=1}^{k_i} \frac{n_{ij}}{N} \log \frac{n_{ij}}{N p_{ij}(\theta)} = C - N \times D_{\text{kullback}}(\widehat{P}, P(\theta)),$$

其中 $\widehat{P} = n/N$, $P(\theta)^T = \left(\frac{N_1}{N}\Pi_1(\theta)^T, \cdots, \frac{N_r}{N}\Pi_r(\theta)^T\right)$, $p_{ij}(\theta)$ 为 $P(\theta)$ 中对应位置的分量, C 是与参数 θ 无关的常数. 通常称, 上式中的 $D_{\text{kullback}}(P, Q) = \sum\limits_{i=1}^k p_i \log\left(\frac{p_i}{q_i}\right)$ 为 Kullback-Leibler(KL) 散度, 可度量两概率向量 $P = (p_1, \cdots, p_k)^T$ 和 $Q = (q_1, \cdots, q_k)^T$ 之间的差异. 由上式可知, 极大似然估计 (MLE) 可如下定义

$$\widehat{\theta}_{\text{mle}} = \underset{\theta \in \mathcal{R}^t}{\operatorname{argmin}} D_{\text{kullback}}(\widehat{P}, P(\theta)).$$

Ali 和 Silvey[1] 在 1966 年引入了 ϕ- 散度

$$D_\phi(P, Q) = \sum_{i=1}^k q_i \phi\Big(\frac{p_i}{q_i}\Big), \qquad \phi \in \Phi^*,$$

其中 Φ^* 是凸函数 $\phi : [0, +\infty) \to R \cup \{\infty\}$ 组成的集合, 满足 $\phi(1) = 0$ 和 $\phi''(1) > 0$, 约定 $0\phi(0/0) = 0$ 和 $0\phi(p/0) = \lim\limits_{u\to\infty} \phi(u)/u$, 此处 p 为某个正数. 对任意在 1 处可导的函数 $\phi \in \Phi^*$, 可定义新函数 $\psi(x) = \phi(x) - \phi'(1)(x-1)$. 容易验证 $\psi(x) \in \Phi^*$. 可验证 $D_\phi(P, Q) = D_\psi(P, Q)$ 和 $\psi'(1) = 0$. 定义新的函数集 $\Phi \equiv \Phi^* \cap \{\phi : \phi'(1) = 0\}$. 因为对应的 ϕ- 散度相等, 可认为 Φ^* 和 Φ 是等价的. 一般地, 研究理论性质时考虑 $\phi \in \Phi$; 在实际应用中, 可选择 $\phi \in \Phi^*$.

一类重要的 ϕ- 散度特例是功效散度 (power-divergence, Cressie 和 Read[2]) 族, 即 $I^\lambda(P, Q) \equiv D_{\phi_{(\lambda)}}(P, Q)$, 其中

$$\phi_{(\lambda)}(x) = \frac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x), \qquad \text{for } \lambda \neq 0, -1,$$

$$\phi_{(0)}(x)=\lim_{\lambda\to 0}\phi_{(\lambda)}(x)=x\log x-x+1, \tag{1.2}$$
$$\phi_{(-1)}(x)=\lim_{\lambda\to -1}\phi_{(\lambda)}(x)=-\log x+x-1.$$

可验证 $\phi_{(\lambda)}(x)$ 和 $\psi_{(\lambda)}(x)=\phi_{(\lambda)}(x)-(x-1)/(\lambda+1)$ 对应的散度相同．值得注意的是，$I^0(P,Q)$ 就是 KL 散度．

在多项模型下， Cressie 和 Read[2] 定义了最小功效散度估计 $\underset{\theta\in\mathcal{R}^t}{\operatorname{argmin}}\, I^{\lambda}(\widehat{P},P(\theta))$ 且研究了它的性质； $\lambda=0$ 对应 MLE, $\lambda=1$ 对应最小卡方 (minimum chi-squared) 估计．可认为，最小功效散度估计是 MLE 和最小卡方估计的推广． Pardo 和 Pardo[3] 研究了三向列联表下的最小功效散度估计． Morales, Pardo 和 Vajda[4] 定义了离散分布下的最小 ϕ- 散度估计 $\underset{\theta\in\mathcal{R}^t}{\operatorname{argmin}}\, D_{\phi}(\widehat{P},P(\theta))$ 并研究了它的性质．显然，最小 ϕ- 散度估计是最小功效散度估计的推广． Cressie 和 Pardo[5] 研究了最小 ϕ- 散度估计在模型选择方面的表现．可参考文献 Cressie, Pardo 和 Pardo[6] 等了解更多 ϕ- 散度在对数线性模型下的应用．在乘积多项抽样对数线性模型 (1.1) 下， Jin 和 Wu[7] 定义了最小 ϕ- 散度估计 $\underset{\theta\in\mathcal{R}^t}{\operatorname{argmin}}\, D_{\phi}(\widehat{P},P(\theta))$ 并研究了它的性质；在更广的模型下， Martin 和 Pardo[8] 用不同的方法研究了相同的问题．

本文将研究如下单边假设检验问题．设原假设为 $H_0: R\theta=\mathbf{0}_v$, 其中 R 是 $v\times t$ 阶行满秩矩阵，且 $v<t$. 设对应的单边对立假设为 $H_a: R\theta\le\mathbf{0}_v$ & $R\theta\ne\mathbf{0}_v$. 定义集合 $E\equiv\{1,2,\cdots,v\}$, 元素依次表示矩阵 R 的行指标．符号 $\mathcal{P}(E)$ 表示集合 E 的幂等集，即所有子集组成的集合．若 $S\in\mathcal{P}(E)$, $R(S)$ 表示 R 的子矩阵，集合 S 中的元素对应出现行的行指标．在对立假设 H_a 下，至少存在某个指标 $i\in E$ 使得 $R(\{i\})\theta<0$ 成立．本文将在 ϕ- 散度的基础上构造三类统计量用于研究下面的单边假设检验问题

$$H_0: R\theta=\mathbf{0}_v \quad\leftrightarrow\quad H_a: R\theta\le\mathbf{0}_v \ \&\ R\theta\ne\mathbf{0}_v. \tag{1.3}$$

检验 (1.3) 不同于 Jin 和 Wu[7] 和 Cressie 和 Pardo[9] 研究的嵌套 (hierarchical) 检验；嵌套检验是双边检验．一般地，与双边检验相比，单边检验研究难度较大，但它有更高的检验效率． Silvapulle 和 Sen[10] 系统地总结了单边检验领域的研究结果．在列联表数据的对数线性模型下， Martin, Mata 和 Pardo[11] 研究了似然比次序 (likelihood ratio ordering) 检验，它是检验 (1.3) 的特殊情形；考虑的模型是饱和对数线性模型，提出了两类基于 ϕ- 散度和 MLE 的统计量．本文所考虑的模型 (1.1) 比饱和模型要广得多，所使用的估计是约束最小 ϕ- 散度估计 (Jin 和 Wu[12]). Martin, Mata 和 Pardo[13] 用 ϕ- 散度研究了等比例二项 (isotonic binomial proportion) 分布的单边检验．

本文的其余章节安排如下：第 2 节介绍主要的理论结果；第 3 节给出第 2 节定理的理论证明；第 4 节列出实例分析和模拟研究的结果．

2 主要结果

定义参数空间 $\Theta_0=\{\theta: R\theta=\mathbf{0}_v\}$ 和 $\Theta_1=\{\theta: R\theta\le\mathbf{0}_v\}$, 显然有 $\Theta_0\subseteq\Theta_1$ 成立．设

$\theta_0 \in \Theta_0 \subseteq \Theta_1 \subseteq \mathcal{R}^t$ 是参数的真值. 假设下列条件成立

(A) $\alpha_i = \lim_{N\to\infty} \frac{N_i}{N} > 0, \text{for } i = 1, \cdots, r.$

Jin 和 Wu[12] 定义了约束最小 ϕ- 散度估计 (RMϕDE)

$$\widehat{\theta}_\phi = \mathop{\arg\min}_{\theta\in\Theta_0} D_\phi(\widehat{P}, P(\theta)). \tag{2.1}$$

有必要指明, 他们考虑的约束 (constraint) 是非线性的, 包含检验 (1.3) 中原假设的线性约束情形. 基于此事实, 可将 Jin 和 Wu[12] 得到的理论结果用于本文, 得到 $\widehat{\theta}_\phi$ 的渐近展开

$$\begin{pmatrix} \widehat{\theta}_\phi - \theta_0 \\ \widehat{\mu} \end{pmatrix} = \begin{pmatrix} H(\theta_0) \\ \phi''(1)B_{21}(\theta_0) \end{pmatrix} V(\widehat{P} - P(\theta_0)) + o_p(\mathbf{1}_{t+v}), \tag{2.2}$$

其中

$$\begin{aligned} &\Sigma_{\Pi_i(\theta_0)} = \operatorname{diag}(\Pi_i(\theta_0)) - \Pi_i(\theta_0)\Pi_i(\theta_0)^T, \\ &\Sigma = \operatorname{diag}\Big(\Big(\frac{N_i}{N}\Sigma_{\Pi_i(\theta_0)}\Big)_{i=1,\cdots,r}\Big), \qquad V \equiv W^T\Sigma\operatorname{diag}(P(\theta_0))^{-1}, \\ &A_{11} \equiv W^T\Sigma W, \qquad H(\theta_0) = A_{11}^{-1} - A_{11}^{-1}R^T(RA_{11}^{-1}R^T)^{-1}RA_{11}^{-1}, \\ &B_{21}(\theta_0) = (RA_{11}^{-1}R^T)^{-1}RA_{11}^{-1}. \end{aligned} \tag{2.3}$$

上式中的 $\widehat{\mu}$ 是拉格朗日乘子 (Lagrange multiplier) 估计.

将参数限制在 Θ_1 中, 定义约束最小 ϕ- 散度估计

$$\widetilde{\theta}_\phi = \mathop{\arg\min}_{\theta\in\Theta_1} D_\phi(\widehat{P}, P(\theta)). \tag{2.4}$$

据我们所知, 到目前为止, 还没有关于 $\widetilde{\theta}_\phi$ 的理论结果. 受 Martin, Mata 和 Pardo[11] 和 Jin, Ming 和 Wu[14] 的启发, 为研究单边检验 (1.3), 构造三类统计量, 如下

$$\begin{aligned} &\mathbf{T}_{\phi_1,\phi_2}(\widehat{P}, P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})) = \frac{2N}{\phi_1''(1)}(D_{\phi_1}(\widehat{P}, P(\widehat{\theta}_{\phi_2})) - D_{\phi_1}(\widehat{P}, P(\widetilde{\theta}_{\phi_2}))), \\ &\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})) = \frac{2N}{\phi_1''(1)} D_{\phi_1}(P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})), \\ &\widetilde{\mathbf{S}}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})) = \frac{2N}{\phi_1''(1)} D_{\phi_1}(P(\widehat{\theta}_{\phi_2}), P(\widetilde{\theta}_{\phi_2})). \end{aligned} \tag{2.5}$$

上式中, $\phi_1, \phi_2 \in \Phi$ 且可不相同. 设 $\phi_1(x) \in \Phi$, 定义函数 $\widetilde{\phi}_1(x) \equiv x\phi_1(1/x)$, 有 $\widetilde{\phi}_1(x) \in \Phi$ 成立, 且 $\widetilde{\mathbf{S}}_{\phi_1,\phi_2} = \mathbf{S}_{\widetilde{\phi}_1,\phi_2}$. 当 $\phi_1 = \phi_2 = \phi_{(0)}$ 时, $\mathbf{T}_{\phi_1,\phi_2}$ 和 $\mathbf{S}_{\phi_1,\phi_2}$ 是似然比统计量; 当 $\phi_1 = \phi_{(1)}$ 和 $\phi_2 = \phi_{(0)}$ 时, $\mathbf{S}_{\phi_1,\phi_2}$ 是皮尔逊统计量. 可参看 [11], [15, 338 页] 和 [16], 得到更多的细节.

定理 1 设 $\phi_1, \phi_2 \in \Phi$ 且二阶连续可导. 在检验 (1.3) 的原假设和条件 **(A)** 下, 有

$$\begin{aligned} &\lim_{N\to\infty} \mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})) \le x\big) = \lim_{N\to\infty} \mathrm{P}\big(\widetilde{\mathbf{S}}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})) \le x\big) \\ = &\lim_{N\to\infty} \mathrm{P}\big(\mathbf{T}_{\phi_1,\phi_2}(\widehat{P}, P(\widetilde{\theta}_{\phi_2}), P(\widehat{\theta}_{\phi_2})) \le x\big) = \sum_{h=0}^{v} \varpi_h(\theta_0)\mathrm{P}(\chi^2_{v-h} \le x). \end{aligned} \tag{2.6}$$

上式中， $\chi_0^2 \equiv 0$， $\varpi_h(\theta_0) = \sum\limits_{S\in\mathcal{P}(E),c(S)=h} \mathrm{P}(Z_1(S) \geq \mathbf{0}_h)\mathrm{P}(Z_2(S) \geq \mathbf{0}_{v-h})$, $Z_1(S) \sim N(\mathbf{0}_{c(S)}, \Sigma_1(\theta_0, S))$ 和 $Z_2(S) \sim N(\mathbf{0}_{v-c(S)}, \Sigma_2(\theta_0, S))$, 其中

$$\begin{aligned}&\Sigma_1(\theta_0,S) = (\phi_2^{''}(1))^2(R(S)A_{11}^{*-1}R^T(S))^{-1}, \Sigma_2(\theta_0,S) = R(S^C)H^{S*}(\theta_0)R^T(S^C),\\ &H^{S*}(\theta_0) = A_{11}^{*-1} - A_{11}^{*-1}R^T(S)(R(S)A_{11}^{*-1}R^T(S))^{-1}R(S)A_{11}^{*-1}.\end{aligned} \tag{2.7}$$

此处， $c(\cdot)$ 表示集合的势 (cardinality) 函数， S^C 表示集合 S 的补集． A_{11}^* 是 A_{11} 的极限，即将 A_{11} 中 $\frac{N_i}{N}$ 替换成 α_i 而得到的矩阵．本文中，所有带上标 * 的符号意义相同．

由定理 1 知， (2.5) 式定义的统计量的渐近分布就是 Silvapulle 和 Sen[10] 提到的类卡方 (Chi-bar-square) 分布 $\overline{\chi}^2$. 要计算 $\overline{\chi}^2$ 的上尾概率，就必须先计算权值 $\varpi_h(\theta_0)$. 在模拟计算权值时，会用到以下定理.

定理 2 在 定理1 的条件下，权值 $\varpi_h(\theta_0)$ 有如下的概率表示

$$\begin{aligned}\varpi_{v-h}(\theta_0) =&\varpi_h(\theta_0;(RA_{11}^{*-1}R^T)^{-1},\mathcal{R}_+^v)\\ =&\mathrm{P}\big(\arg\min_{\zeta\in\mathcal{R}_+^v}(Q-\zeta)^T(RA_{11}^{*-1}R^T)(Q-\zeta)\in\mathcal{R}_+^v(h)\big),\end{aligned} \tag{2.8}$$

其中 $Q \sim N(\mathbf{0}_v, (RA_{11}^{*-1}R^T)^{-1})$, $\mathcal{R}_+^v(h)$ 是 $\mathcal{R}_+^v = \{\zeta \in \mathcal{R}^v : v \geq \mathbf{0}_v\}$ 的子集，其 v 维向量元素恰好有 h 个正分量，余下 $v-h$ 个分量为 0.

证 可参见 [10] 中命题 3.6.1(6-7) 和 79 页模拟 3. 证毕.

因参数真值 θ_0 未知，故不能直接利用定理 2 计算权值．解决的办法是，用在检验 (1.3) 的原假设 H_0 下的极大似然估计 $\widehat{\theta}_{\phi_{(0)}}$ 替换未知真值 θ_0. 一旦完成替换，就可用蒙特卡罗模拟计算权值 $\varpi_h(\theta_0)$. 在 [10] 和 [11] 中，称这类用极大似然估计替换未知真值的检验为局部 (local) 检验；一般认为，局部检验是对应理论检验的较好逼近.

3　理论证明

引理 1 设 k 维随机变量 X 服从标准正态分布 $N(\mathbf{0}_k, I_k)$, Q 是 k 阶幂等对称方阵，秩为 v. 设 $d_1,\cdots,d_v,\cdots,d_k$ 是线性无关的 k 维列向量组，满足 $Qd_i = d_i$, $i=1,\cdots,v$ 和 $Qd_i = \mathbf{0}_k$, $i=v+1,\cdots,k$, 则在给定 $d_i^TX \geq 0$, $i=1,\cdots,v$ 下，二次型 X^TQX 的条件分布为 χ_v^2.

引理 1 的类似结论及证明，可参见 [11] 和 [17, 414 页].

为研究单边检验 (1.3), 先研究如下的双边检验

$$H_0 : R\theta = \mathbf{0}_v \quad \longleftrightarrow \quad H_1 : R(S)\theta = \mathbf{0}_{c(S)}, \qquad R\theta \neq \mathbf{0}_v, \tag{3.1}$$

其中 $S \in \mathcal{P}(E)$. 定义 $\Theta_0(S) = \{\theta : R(S)\theta = \mathbf{0}_{c(S)}\}$. 定义约束最小 ϕ- 散度估计

$$\widehat{\theta}_\phi^S = \operatorname*{argmin}_{\theta\in\Theta_0(S)} D_\phi(\widehat{P}, P(\theta)). \tag{3.2}$$

用 $\widehat{\theta}^S_{\phi_2}$ 替换 (2.5) 式中参数估计 $\widetilde{\theta}_{\phi_2}$, 得到如下统计量

$$\begin{aligned}
&\mathbf{T}_{\phi_1,\phi_2}(\widehat{P},P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))=\frac{2N}{\phi_1^{''}(1)}(D_{\phi_1}(\widehat{P},P(\widehat{\theta}_{\phi_2}))-D_{\phi_1}(\widehat{P},P(\widehat{\theta}^S_{\phi_2})),\\
&\mathbf{S}_{\phi_1,\phi_2}(P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))=\frac{2N}{\phi_1^{''}(1)}D_{\phi_1,\phi_2}(P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2})),\\
&\widetilde{\mathbf{S}}_{\phi_1,\phi_2}(P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))=\frac{2N}{\phi_1^{''}(1)}D_{\phi_1,\phi_2}(P(\widehat{\theta}_{\phi_2}),P(\widehat{\theta}^S_{\phi_2})).
\end{aligned}\tag{3.3}$$

引理 2 在 定理1 的条件和检验 (3.1) 的原假设下，有

$$\mathbf{S}_{\phi_1,\phi_2}(P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))=\mathbf{T}_{\phi_1,\phi_2}(\widehat{P},P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))+o_p(1),\tag{3.4}$$

且 (3.3) 式中的三类统计量有相同的渐近分布，即卡方分布 $\chi^2_{v-c(S)}$.

证 可参见 [12,14] 和 [16]. 证毕.

定理 1 的证明 先列出统计量 $\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))$ 的证明. 对于 (2.4) 式定义的约束最小 ϕ- 散度估计 $\widetilde{\theta}_{\phi_2}$, 一定存在一个子集 $S\in\mathcal{P}(E)$ 使得 $\widetilde{\theta}_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$ 成立，此处 $\widetilde{\theta}^S_{\phi_2}$ 定义如下

$$\widetilde{\theta}^S_{\phi_2}=\underset{\theta\in\Theta_1(S)}{\operatorname{argmin}}\,D_{\phi_2}(\widehat{P},P(\theta)),\tag{3.5}$$

其中 $\Theta_1(S)=\{\theta:R(S)\theta=\mathbf{0}_{c(S)},\ R(S^C)\theta<\mathbf{0}_{v-c(S)}\}$.

显然，对于任意给定样本， $\widetilde{\theta}_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$ 仅对某个唯一的子集 $S\in\mathcal{P}(E)$ 成立. 由全概率公式，可得

$$\mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x\big)=\sum_{S\in\mathcal{P}(E)}\mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x,\widetilde{\theta}_{\phi_2}=\widetilde{\theta}^S_{\phi_2}\big).\tag{3.6}$$

[18] 中定理 4.2.13 给出了求解 (2.4) 式所对应优化问题的 KKT(Karush-Kuhn-Tucker) 充分条件，如下

$$\begin{aligned}
&\frac{\partial}{\partial\theta}D_{\phi_2}(\widehat{P},P(\theta))+\sum_{i=1}^{v}\mu_iR^T(\{i\})=0,\\
&\mu_iR(\{i\})\theta=0,\qquad \text{for}\ \ i=1,\cdots,v,\\
&\mu_i\ge 0,\qquad \text{for}\ i=1,\cdots,v.
\end{aligned}\tag{3.7}$$

上式中的 μ_i 是 KKT 乘子. 一旦已知 KKT 充分条件对所有可能的子集 $S\in\mathcal{P}(E)$ 皆成立，则互补松弛 (complementary slackness) 条件 $\mu_i>0$, $i\in S$ 和 $R(\{i\})\theta<0$, $i\in S^C$ 就能够确定特定子集 $S\in\mathcal{P}(E)$ 满足等式 $\widetilde{\theta}_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$, 此时条件 $R(\{i\})\theta=0$, $i\in S$ 和 $\mu_i=0,i\in S^C$ 是多余的. 基于以上事实，可得

$$\begin{aligned}
&\mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x,\widetilde{\theta}_{\phi_2}=\widetilde{\theta}^S_{\phi_2}\big)\\
=&\mathrm{P}\{\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x,\widetilde{\mu}(S)>\mathbf{0}_{c(S)},\ R(S^C)\widetilde{\theta}^S_{\phi_2}<\mathbf{0}_{v-c(S)}\},
\end{aligned}\tag{3.8}$$

其中 $\widetilde{\mu}(S)$ 是当 $\widetilde{\theta}_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$ 成立时非零 KKT 乘子依次排列组成的列向量．在检验 (1.3) 的原假设 H_0 下，可得

$$\begin{aligned}&\mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x\big)\\&=\sum_{S\in\mathcal{P}(E)}\mathrm{P}\{\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x,\widetilde{\mu}(S)>\mathbf{0}_{c(S)},R(S^C)(\theta_0-\widetilde{\theta}^S_{\phi_2})>\mathbf{0}_{v-c(S)}\}.\end{aligned} \tag{3.9}$$

另一方面，在考虑 (3.2) 式对应的拉格朗日优化问题时，所得的估计 $\widehat{\theta}^S_{\phi_2}$ 和 $\widehat{\mu}^S$ 也满足条件 (3.7) 中的前两个条件；也就是说，有 $\widehat{\theta}^S_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$ 和 $\widehat{\mu}^S=\widetilde{\mu}(S)$ 成立．

当 $\widehat{\theta}^S_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$ 时，有 $\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))=\mathbf{S}_{\phi_1,\phi_2}(P(\widehat{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))$ 成立．由 (2.2) 式和 引理2 的证明思路，可得

$$\begin{aligned}\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}^S_{\phi_2}),P(\widehat{\theta}_{\phi_2}))=&Z_0^T(L_S-L)^2Z_0+o_p(1)\\\equiv&(M_3(\theta_0,S)Z_0)^T(M_3(\theta_0,S)Z_0)+o_p(1),\end{aligned} \tag{3.10}$$

其中

$$\begin{aligned}&Z_0=\sqrt{N}\mathrm{diag}(P(\theta_0))^{-\frac{1}{2}}(\widehat{P}-P(\theta_0)),\qquad M_3(\theta_0,S)=L_S-L,\\&L\equiv\mathrm{diag}\,(P(\theta_0))^{-\frac{1}{2}}\Sigma WH(\theta_0)W^T\Sigma\mathrm{diag}(P(\theta_0))^{-\frac{1}{2}}\end{aligned}$$

和

$$L_S\equiv\mathrm{diag}\,(P(\theta_0))^{-\frac{1}{2}}\Sigma WH^S(\theta_0)W^T\Sigma\mathrm{diag}\,(P(\theta_0))^{-\frac{1}{2}}.$$

此处， $H^S(\theta_0)$ 为将 $H(\theta_0)$ 中的 R 替换成 $R(S)$ 而得到的矩阵．由 $\widehat{\mu}^S=\widetilde{\mu}(S)$ 和 (2.2) 式，可得

$$\begin{aligned}\sqrt{N}\widetilde{\mu}(S)=&\big(\phi_2''(1)(R(S)A_{11}^{-1}R(S)^T)^{-1}R(S)A_{11}^{-1}W^T\Sigma\mathrm{diag}\,(P(\theta_0))^{-\frac{1}{2}}\big)Z_0+o_p(1)\\\equiv&M_1(\theta_0,S)Z_0+o_p(1).\end{aligned} \tag{3.11}$$

由 $\widehat{\theta}^S_{\phi_2}=\widetilde{\theta}^S_{\phi_2}$, 可得

$$\begin{aligned}\sqrt{N}R(S^C)(\theta_0-\widetilde{\theta}^S_{\phi_2})=&\big(-R(S^C)H^S(\theta_0)W^T\Sigma\mathrm{diag}\,(P(\theta_0))^{-\frac{1}{2}}\big)Z_0+o_p(1)\\\equiv&M_2(\theta_0,S)Z_0+o_p(1).\end{aligned} \tag{3.12}$$

由 (3.9)–(3.12), 有

$$\begin{aligned}&\lim_{N\to\infty}\mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x\big)\\&=\sum_{S\in\mathcal{P}(E)}\mathrm{P}\{Z_3^T(S)Z_3(S)\le x,Z_1(S)>\mathbf{0}_{c(S)},Z_2(S)>\mathbf{0}_{v-c(S)}\}\\&=\sum_{S\in\mathcal{P}(E)}\mathrm{P}\{Z_3^T(S)Z_3(S)\\&\qquad\le x|(Z_1^T(S),Z_2^T(S))^T>\mathbf{0}_v\}\mathrm{P}\{Z_1(S)>\mathbf{0}_{c(S)},Z_2(S)>\mathbf{0}_{v-c(S)}\},\end{aligned} \tag{3.13}$$

其中

$$Z_0 \xrightarrow{\mathcal{L}} Z_0^* \sim N(\mathbf{0}_k, \mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^*\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}})$$
$$Z_3(S) = M_3^*(S,\theta_0)Z_0^*, \quad Z_1(S) = M_1^*(S,\theta_0)Z_0^*, \quad Z_2(S) = M_2^*(S,\theta_0)Z_0^*. \tag{3.14}$$

进一步，可验证

$$\begin{aligned}&\Sigma^{*\frac{1}{2}}\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}M_3^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}}(M_2^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}})^T\\ =&(M_2^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}})^T,\\ &\Sigma^{*\frac{1}{2}}\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}M_3^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}}(M_1^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}})^T\\ =&\mathbf{0}_{k\times c(S)}.\end{aligned} \tag{3.15}$$

可证明，矩阵 $\Sigma^{*\frac{1}{2}}\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}M_3^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}}$ 的秩为 $v-c(S)$, 矩阵 $M_2^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^{*\frac{1}{2}}$ 是行满秩的，秩为 $v-c(S)$. 由 (3.15) 式和 引理1, 可得

$$\mathrm{P}\{Z_3^T(S)Z_3(S)\le x|(Z_1^T(S),\ Z_2^T(S))^T \ge \mathbf{0}_v\} = \mathrm{P}\{\chi^2_{v-c(S)}\le x\}. \tag{3.16}$$

另外，可验证

$$M_1^*(S,\theta_0)\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}\Sigma^*\mathrm{diag}\,(P^*(\theta_0))^{-\frac{1}{2}}M_2^{*T}(S,\theta_0) = \mathbf{0}_{c(S)\times c(S^C)}. \tag{3.17}$$

(3.17) 式表明 $Z_1(S)$ 和 $Z_2(S)$ 是独立的. 由上面的一系列结论，可得

$$\begin{aligned}&\lim_{N\to\infty}\mathrm{P}\big(\mathbf{S}_{\phi_1,\phi_2}(P(\widetilde{\theta}_{\phi_2}),P(\widehat{\theta}_{\phi_2}))\le x\big)\\ =&\sum_{S\in\mathcal{P}(E)}\mathrm{P}\{\chi^2_{v-c(S)}\le x\}\mathrm{P}\{Z_1(S)\ge\mathbf{0}_{c(S)}\}\mathrm{P}\{Z_2(S)\ge\mathbf{0}_{v-c(S)}\}\\ =&\sum_{h=0}^{v}\varpi_h(\theta_0)\mathrm{P}\{\chi^2_{v-h}\le x\}.\end{aligned} \tag{3.18}$$

上式即定理的结论. 其余两类统计量的证明类似，在此省略.

4 实例分析和模拟研究

4.1 实例分析

例 表 1 为列联表，收集了 1991 年缅因州在汽车和轻型卡车交通事故中对 68,694 位乘客进行观察的数据，按照受伤 (I), 安全带使用 (S), 事故位置 (L) 和性别 (G) 进行交叉分类，参看 [19, 第 8.4.2 节].

表 1　交通数据的交叉分类计数表

Gender	Location	Seat Belt	Injury No	Injury Yes
Female	Urban	No	7,287	996
		Yes	11,587	759
	Rural	No	3,246	973
		Yes	6134	757
Male	Urban	No	10,381	812
		Yes	10,969	380
	Rural	No	6,123	1,084
		Yes	6,693	513

经过分析， Agresti[19] 得到对数线性模型 (GLS,GI,IL,IS) 能非常好地拟合表 1 数据. 注意该模型仅有一个三因子项 λ_{gls}^{GLS}, 每对变量是条件相关的，在变量 I 的每个水平下剩余三个变量中的任意两个变量之间的关系会受到第三个变量的影响. 因为变量 I 没有出现在三因子项中，则它与任意某个变量之间的优势比 (odds ratio) 不受剩余两个变量的影响. 基于以上结论，可根据变量 I 的两个水平把模型 (GLS,GI,IL,IS) 分解成两个不独立的局部子模型 (GLS). 在变量 I 取水平 "否" 时，子模型为

$$\log(\pi_{1gls}) = \lambda + \lambda_g^G + \lambda_l^L + \lambda_s^S + \lambda_{gl}^{GL} + \lambda_{gs}^{GS} + \lambda_{ls}^{LS} + \lambda_{gls}^{GLS}. \tag{4.1}$$

在变量 I 取水平 "是" 时，子模型为

$$\log(\pi_{2gls}) = \lambda + \lambda_g^G + \lambda_l^L + \lambda_s^S + \lambda_{gl}^{GL} + \lambda_{gs}^{GS} + \lambda_{ls}^{LS} + \lambda_{gls}^{GLS}. \tag{4.2}$$

模型 (4.1)–(4.2) 都是饱和模型. 由 [19] 的结论, 可以假设, 两模型中的参数 $\lambda_{gl}^{GL}, \lambda_{gs}^{GS}, \lambda_{ls}^{LS}$ 和 λ_{gls}^{GLS} 相同，其余参数不同. 此时，模型 (4.1)–(4.2) 就是模型 (1.1) 的特例，对应参数向量 θ 为 $(\lambda_{(g)1}^{1G}, \lambda_{(g)1}^{2G}, \lambda_{(l)1}^{1L}, \lambda_{(l)1}^{2L}, \lambda_{(s)1}^{1S}, \lambda_{(s)1}^{2S}, \lambda_{(gl)11}^{GL}, \lambda_{(gs)11}^{GS}, \lambda_{(ls)11}^{LS}, \lambda_{(gls)111}^{GLS})^T$.

从表 1 的实际发生背景出发，可以猜测未系安全带受伤事故的发生概率要比系安全带情形大，即有 $\log(\pi_{2gl2}) \le \log(\pi_{2gl1})$, $g=1,2$, $l=1,2$ 成立；考虑到市区交通的拥挤度和低流动性，也可猜测市区未受伤事故的发生概率要比郊区大，即有 $\log(\pi_{1g2s}) \le \log(\pi_{1g1s})$, $g=1,2$, $s=1,2$ 成立；考虑到男性有更好的运动协调能力和预判能力，可以猜测在给定位置和安全带使用两变量的条件下性别和受伤两变量之间的条件优势比会大于 1, 即有 $\log\left(\frac{\pi_{11ls}\pi_{22ls}}{\pi_{21ls}\pi_{12ls}}\right) \le 1$, $l=1,2$, $s=1,2$ 成立. 为分析以上三个猜测，可以构造形如检验 (1.3) 的单边检验，依次记为

$$H_0: R_1\theta = \mathbf{0}_4 \ \leftrightarrow \ H_a: R_1\theta \le \mathbf{0}_4 \ \& \ R_1\theta \ne \mathbf{0}_4, \tag{4.3}$$

$$H_0: R_2\theta = \mathbf{0}_4 \ \leftrightarrow \ H_a: R_2\theta \le \mathbf{0}_4 \ \& \ R_2\theta \ne \mathbf{0}_4 \tag{4.4}$$

和

$$H_0: R_3\theta = 0 \ \leftrightarrow \ H_a: R_3\theta \le 0 \ \& \ R_3\theta \ne 0. \tag{4.5}$$

表 2 检验 (4.5) 的统计量的值

λ_1	λ_2 0	2/3	1	1.5	2
-1.5	406.3071	407.4233	408.1595	409.5074	411.1860
-1	404.4551	405.3943	406.0361	407.2357	408.7559
-0.5	403.5381	404.3087	404.8596	405.9149	407.2795
0	403.5466	404.1548	404.6171	405.5305	406.7404
2/3	404.9922	405.3902	405.7363	406.4615	407.4638
1	406.3319	406.6261	406.9143	407.5448	408.4417
1.5	409.1199	409.2591	409.4598	409.9463	410.6819
2	412.8556	412.8387	412.9505	413.2894	413.8585

由第 2 节定理 1 知，可用 (2.5) 式定义的统计量 $\mathbf{T}_{\phi_1,\phi_2}, \mathbf{S}_{\phi_1,\phi_2}$ 和 $\widetilde{\mathbf{S}}_{\phi_1,\phi_2}$ 研究检验 (4.3)–(4.5). 本文对功效散度族进行了实例分析和模拟研究，选择功效指标 $\lambda_1 \in \{-1.5,-1,-0.5,0,2/3,1,1.5,2\}$ 和 $\lambda_2 \in \{0,2/3,1,1.5,2\}$, 分别对应 ϕ_1 和 ϕ_2. 用一百万次重复来估计 定理2 中的权值，再用权值的估计去计算 定理1 中极限分布的上 0.05 分位数. 表 2 列出了研究检验 (4.5) 时统计量 $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 的值，皆远远大于估计的上 0.05 分位数 2.613896. 此结果表明，有充分的理由拒绝检验 (4.5) 的原假设 H_0 而接受对立假设 H_a. 用统计量 $\mathbf{T}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$, $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 和 $\widetilde{\mathbf{S}}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 研究检验 (4.3)–(4.5) 时，可得到很多类似结果，统计量的值皆远远大于相应的临界值估计，在此不一一列出. 这些结果表明，未系安全带受伤事故的发生概率要比系安全带情形大，市区未受伤事故的发生概率要比郊区大，女性事故受伤的发生概率比男性大.

4.2 模拟研究

考虑模型 (4.1)–(4.2) 及检验 (4.3) 和 (4.5). 设参数真值为 $\theta_0 = \mathbf{0}_{10} + \delta_{1,i}\mathbf{c_1}$, $i = 1,2,\cdots,7$, 其中 $\delta_1^T = (0,0.025,0.05,0.075,0.1,0.125,0.150)$ 和 $\mathbf{c_1}^T = (0,0,0,0,0,1,0,1,1,1)$. 考虑假设检验问题

$$H_0:\theta_0 = \mathbf{0}_{10} \ \leftrightarrow\ H_a:\theta_0 = \mathbf{0}_{10} + \delta_{1,i}\mathbf{c_1}, \qquad \text{for} \ \ i = 2,\cdots,7. \tag{4.6}$$

设参数真值为 $\theta_0 = \mathbf{0}_{10} + \delta_{2,i}\mathbf{c_2}$, $i = 1,2,\cdots,7$, 其中 $\delta_2^T = (0,0.075,0.15,0.225,0.3,0.375,0.45)$ 和 $\mathbf{c_2}^T = (0,1,0,0,0,0,0,0,0,0)$. 考虑假设检验问题

$$H_0:\theta_0 = \mathbf{0}_{10} \ \leftrightarrow\ H_a:\theta_0 = \mathbf{0}_{10} + \delta_{2,i}\mathbf{c_2}, \qquad \text{for} \ \ i = 2,\cdots,7. \tag{4.7}$$

检验 (4.6)–(4.7) 分别满足检验 (4.3) 和 (4.5), 其中指标 i 可度量检验 (4.3) 和 (4.5) 中真值 θ_0 和原假设 H_0 之间的差距. 当 $i = 1$ 时原假设为真，否则对立假设为真；指标 i 值越大，差距越大.

在检验 (4.6)–(4.7) 的原假设下，真值参数 θ_0 已知. 由 定理1 和 定理2 可知，用蒙特卡罗模拟方法可计算统计量 $\mathbf{T}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$, $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 和 $\widetilde{\mathbf{S}}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 的临界值，得到在检

验水平 0.05 下，检验 (4.6)–(4.7) 的临界值分别为 6.354925 和 2.70487. 依次定义用统计量 $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 研究检验 (4.6)–(4.7) 的检验水平和检验功效

$$\begin{aligned}&\mathrm{P}\{\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}>6.354925|\theta_0=\mathbf{0_{10}}\},\mathrm{P}\{\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}>6.354925|\theta_0=\mathbf{0_{10}}+\delta_{\mathbf{1,i}}\mathbf{c_1}\},\\&\mathrm{P}\{\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}>2.70487|\theta_0=\mathbf{0_{10}}\},\mathrm{P}\{\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}>2.70487|\theta_0=\mathbf{0_{10}}+\delta_{\mathbf{2,i}}\mathbf{c_2}\},\end{aligned}\tag{4.8}$$

其中 $i=2,3,4,5,6,7$. 以下用 1000 次重复来估计 (4.8) 式中的概率.

表 3–4 分别列出了 (4.8) 式中检验水平和指标 $i=4$ 时检验功效的估计. λ_1 取负值时的检验功效估计要比 λ_1 取正值时的大 (好); 同时， λ_1 取负值时的检验水平估计要比取正值时的差，因为此时的检验水平估计远离理论的检验水平 0.05 . 此现象就是检验水平和检验功效之间的互补性 (trade-off)[2−6,14,16] . 另外，从表 3–4 可以看出，随着样本量由小变大, 对应的检验水平估计变得更加靠近理论的检验水平; 对应的检验功效估计变得更大 (好) . [20] 定义了下列不等式

$$|\mathrm{logit}(1-\alpha_n)-\mathrm{logit}(1-\alpha)|\le d,\tag{4.9}$$

其中 $\mathrm{logit}(p)=\log(p/(1-p))$. 如果 $d=0.35$ 时不等式 (4.9) 成立，就认为两概率值 α_n,α 非常接近；如果 $d=0.7$ 时不等式 (4.9) 成立，就认为两概率值 α_n,α 比较接近. 当 $\alpha=0.05$, $d=0.35$ 时，满足不等式 (4.9) 的 α_n 的取值范围为 $[0.0357,0.0695]$; 当 $\alpha=0.05$, $d=0.7$ 时，满足不等式 (4.9) 的 α_n 的取值范围为 $[0.0254,0.0959]$.

表 3 检验 (4.6) 的检验水平 (左) 和检验功效 (右) 估计，样本量：上 (50, 50), 下 (75, 75)

	λ_2					λ_2				
λ_1	0	2/3	1	1.5	2	0	2/3	1	1.5	2
−1.5	0.087	0.086	0.089	0.089	0.094	0.299	0.289	0.290	0.287	0.294
−1	0.082	0.078	0.079	0.082	0.083	0.281	0.269	0.268	0.271	0.274
−0.5	0.077	0.071	0.070	0.070	0.076	0.269	0.258	0.256	0.256	0.257
0	0.069	0.068	0.068	0.068	0.068	0.261	0.246	0.245	0.246	0.245
2/3	0.068	0.063	0.062	0.063	0.062	0.254	0.234	0.228	0.222	0.221
1	0.068	0.061	0.060	0.060	0.061	0.251	0.224	0.220	0.215	0.213
1.5	0.067	0.057	0.057	0.055	0.056	0.251	0.219	0.210	0.201	0.199
2	0.068	0.057	0.056	0.056	0.054	0.253	0.216	0.207	0.198	0.194
−1.5	0.091	0.082	0.082	0.080	0.084	0.362	0.365	0.366	0.375	0.383
−1	0.080	0.077	0.073	0.071	0.073	0.355	0.347	0.354	0.353	0.357
−0.5	0.077	0.069	0.068	0.069	0.069	0.346	0.339	0.336	0.337	0.341
0	0.071	0.067	0.063	0.062	0.065	0.342	0.331	0.330	0.325	0.326
2/3	0.069	0.060	0.059	0.057	0.057	0.337	0.316	0.313	0.313	0.313
1	0.068	0.057	0.056	0.056	0.056	0.337	0.309	0.306	0.302	0.301
1.5	0.069	0.056	0.054	0.052	0.052	0.337	0.306	0.300	0.296	0.288
2	0.069	0.056	0.052	0.048	0.048	0.340	0.304	0.295	0.286	0.274

在表 3 中, 样本量为 $(50,50)$, λ_1,λ_2 取 $0,2/3,1,1.5,2$ 时, 检验水平估计满足 $d=0.35$ 时的不等式 (4.9), 此时统计量 $\mathbf{S}_{\phi_{(0)},\phi_{(0)}}$ 有最大的检验功效估计；检验水平估计全部满

足 $d=0.7$ 时的不等式 (4.9), 此时统计量 $\mathbf{S}_{\phi_{(-1.5)},\phi_{(0)}}$ 有最大的检验功效估计. 样本量为 $(75,75)$, λ_1 取 $-0.5,0,2/3,1,1.5,2$ 且 λ_2 取 $0,2/3,1,1.5,2$(除 $\mathbf{S}_{\phi_{(-0.5)},\phi_{(0)}}$ 和 $\mathbf{S}_{\phi_{(0)},\phi_{(0)}}$) 时, 检验水平估计满足 $d=0.35$ 时的不等式 (4.9), 此时统计量 $\mathbf{S}_{\phi_{(-0.5)},\phi_{(0)}}$ 有最大的检验功效估计; 检验水平估计全部满足 $d=0.7$ 时的不等式 (4.9), 此时统计量 $\mathbf{S}_{\phi_{(-1.5)},\phi_{(2)}}$ 有最大的检验功效估计. 在表 4 中, 样本量为 $(15,15)$ 时, 除 λ_1 取 $-1.5,-1,-0.5,0$ 且 λ_2 取 0 以外, 检验水平估计皆满足 $d=0.35$ 时的不等式 (4.9); 除 λ_1 取 -1.5 且 λ_2 取 0 以外, 检验水平估计皆满足 $d=0.7$ 时的不等式 (4.9). 样本量为 $(20,20)$ 时, 除 λ_1 取 -1.5 且 λ_2 取 0 以外, 检验水平估计皆满足 $d=0.35$ 时的不等式 (4.9); 检验水平估计皆满足 $d=0.7$ 时的不等式 (4.9). 样本量为 $(50,50)$ 时, 统计量 $\mathbf{S}_{\phi_{(-1.5)},\phi_{(0)}}$ 有最大的检验功效估计; 样本量为 $(75,75)$ 时, 统计量 $\mathbf{S}_{\phi_{(-1.5)},\phi_{(0)}}$ 和 $\mathbf{S}_{\phi_{(-1)},\phi_{(0)}}$ 有最大的检验功效估计. 用统计量 $\mathbf{T}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$, $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 和 $\widetilde{\mathbf{S}}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 研究检验 (4.6)-(4.7) 时, 可得到类似结果, 在此省略.

综合以上几个方面, 可得以下两条结论: (1) 只要样本量足够大, 无论 λ_1 和 λ_2 取何值, 统计量的检验水平会无限接近理论检验水平, 检验功效会无限接近 1, 这与定理1 的结论吻合; (2) 样本量总是有限的, 当样本量小时, 不同统计量之间有区别. 在表 3–4 中, $\lambda_1,\lambda_2=2/3,1$ 时的统计量 $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 就要比 $\mathbf{S}_{\phi_{(1)},\phi_{(0)}}$(皮尔逊统计量) 和 $\mathbf{S}_{\phi_{(0)},\phi_{(0)}}$ (似然比统计量) 表现好, 因为它们的检验水平估计更接近理论的检验水平 0.05, 此时的检验功效估计表现尚可. 因此在实际数据分析中, 尤其在样本量小时, 推荐用 $\lambda_1,\lambda_2=2/3,1$ 时的统计量 $\mathbf{T}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$, $\mathbf{S}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 和 $\widetilde{\mathbf{S}}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ 作为皮尔逊统计量和似然比统计量的有力替代, 来进行单边检验分析. 在 [5,6,8,11,14,16] 中, 有类似的结论.

表 4 检验 (4.7) 的检验水平 (左) 和检验功效 (右) 估计,
样本量: 左上 $(15,15)$, 左下 $(20,20)$, 右上 $(50,50)$, 右下 $(75,75)$

	λ_2					λ_2				
λ_1	0	2/3	1	1.5	2	0	2/3	1	1.5	2
−1.5	0.131	0.063	0.058	0.060	0.060	0.308	0.292	0.287	0.287	0.289
−1	0.073	0.060	0.056	0.053	0.055	0.305	0.288	0.284	0.284	0.286
−0.5	0.071	0.057	0.053	0.052	0.053	0.304	0.287	0.284	0.282	0.286
0	0.070	0.053	0.052	0.050	0.050	0.304	0.287	0.283	0.281	0.286
2/3	0.067	0.052	0.050	0.048	0.048	0.302	0.284	0.281	0.281	0.285
1	0.066	0.053	0.052	0.048	0.048	0.301	0.284	0.281	0.280	0.285
1.5	0.065	0.054	0.052	0.049	0.052	0.302	0.283	0.280	0.281	0.286
2	0.065	0.056	0.055	0.052	0.053	0.302	0.283	0.282	0.281	0.287
−1.5	0.091	0.056	0.053	0.051	0.050	0.386	0.378	0.378	0.376	0.381
−1	0.065	0.053	0.051	0.049	0.049	0.386	0.378	0.377	0.375	0.378
−0.5	0.063	0.052	0.049	0.048	0.048	0.384	0.377	0.376	0.375	0.378
0	0.062	0.049	0.048	0.048	0.047	0.383	0.376	0.374	0.373	0.375
2/3	0.061	0.048	0.047	0.045	0.045	0.381	0.374	0.372	0.372	0.373
1	0.060	0.048	0.047	0.043	0.043	0.381	0.374	0.371	0.372	0.373
1.5	0.059	0.048	0.047	0.043	0.041	0.380	0.373	0.371	0.371	0.373
2	0.059	0.049	0.047	0.043	0.041	0.380	0.373	0.371	0.372	0.373

参 考 文 献

[1] Ali S, Silvey S. A general class of coefficients of divergence of one distribution from another. *J. of Royal Statistical Society, Series B (Methodological)*, 1966, 28(1): 131–142

[2] Cressie N, Read T. Multinomial goodness-of-fit tests. *J. of the Royal Statistical Society, Series B (Methodological)*, 1984, 46(3): 440–464

[3] Pardo L, Pardo M. Minimum power-divergence estimator in three-way contingency tables. *J. of Statistical Computation and Simulation*, 2003, 73(11): 819–831

[4] Morales D, Pardo L, Vajda I. Asymptotic divergence of estimates of discrete distribution. *J. of statistical Planning and Inference*, 1995, 48(3): 347–369

[5] Cressie N, Pardo L. Model checking in loglinear models using ϕ- divergences and mles. *J. of Statistical planning and inference*, 2002, 103(1-2): 437–453

[6] Cressie N, Pardo L, Pardo M. Size and power consideration for testing log-linear models using ϕ-divergence test statistics. *Statistica Sinica*, 2003, 13(2): 555–570

[7] Jin Y, Wu Y. Minimum ϕ-divergence estimator and hierarchical testing in loglinear models under product-multinomial sampling. *J. of Statistical planning and inference*, 2009, 139(10): 3488–3500

[8] Martin N, Pardo L. New families of estimators and test statistics in log- linear models. *Journal of Multivariate Analysis*, 2008, 99(8): 1590–1609

[9] Cressie N, Pardo L. Minimum ϕ-divergence estimator and hierarchical testing in loglinear models. *Statistica Sinica*, 2000, 10(3): 867–884

[10] Silvapulle M, Sen P. Constrained statistical inference: Inequality, order and shape restrictions. John Wiley & Sons Inc., 2005

[11] Martin N, Mata R, Pardo L. Phi-divergence statistics for the likelihood ratio order: An approach based on the loglinear models. *J. of Multivariate Analysis*, 2014, 130(3): 387–408

[12] Jin Y, Wu Y. Analysis of ϕ-divergence for loglinear models with constraints under product-multinomial sampling. *Acta Mathematicae Applicatae Sinica*, 2013, 29(1): 93–104

[13] Martin N, Mata R, Pardo L. Wald type and phi-divergence based test- statistics for isotonic binomial proportions. *Mathematics and Computers in Simulation*, 2016, 120: 31–49

[14] Jin Y, Ming R, Wu Y. Nonadditivity in loglinear models using ϕ- divergences and mϕes under product-multinomial sampling. *J. of Statistical Planning and Inference*, 2013, 143(2): 356–367

[15] Christensen R. Loglinear model and logistic regression. Springer-Verlag, New York, 1997

[16] Pardo L, Pardo M. Nonadditivity in loglinear models using ϕ-divergences and mles. *J. of Statistical Planning and Inference*, 2005, 127(1-2): 237–252

[17] Kudo A. A multivariate analogue of the one-sided test. *Biometrika*, 1963, 50(3/4): 403–418

[18] Bazaraa M, Sherali H, Shetty C. Nonlinear programming: Theory and algorithms. John Wiley & Sons Inc., 2006

[19] Agresti A. Categorical data analysis. John Wiley & Sons Inc., 2002

[20] Dale J. Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. *J. of Royal Statistical Society, Series B (Methodological)*, 1986, 48(1): 48–59

One-sided Test Based on ϕ-divergence under Log-linear Models

JIN Yinghua

(*School of Applied Mathematics, Guangdong University of Technology, Guangzhou* 510006*, China*)

(*E-mail: jyh@mail.ustc.edu.cn*)

WU Yaohua

(*Department of Statistics and Finance,*

University of Science and Technology of China, Hefei 230026*, China*)

(*E-mail: wuyh@ustc.edu.cn*)

SHAO Quanxi

(*CSIRO Data61, Leeuwin Centre, 65 Broockway Road, Floreat, WA 6014, Australia*)

(*E-mail: quanxi.shao@data61.csiro.au*)

Abstract One-sided test is an important part of hypothesis test theory. Some kind of one-sided test for log-linear models under product-multinomial sampling is studied in this paper. Based on ϕ-divergence and the restricted minimum ϕ-divergence estimator (RMϕDE), three families of statistic are proposed. All the three proposed statistics have identical asymptotic distribution as Chi-bar-square distribution, and include the likelihood ratio statistic and the Pearson statistic as special cases. These proposed statistics can be treated as a generalization of the results in existing literature with three improvements: the model in this paper is much wider than the saturated log-linear model they considered; the form of one-sided test is much more general than the likelihood ratio ordering test they constructed; and RMϕDE used to construct these statistic is more general than the maximum likelihood estimator (MLE) they used. Real data analysis demonstrates the effect of these statistic. A simulation study shows that some members of the power-divergence family could act as superior alternatives to the likelihood ratio statistic and the Pearson statistic under finite sample sizes.

Key words one-sided test; ϕ-divergence; product-multinomial sampling; log-linear model; Chi-bar-square distribution

MR(2000) Subject Classification 62J12; 62F05

Chinese Library Classification O212.1