轮廓似然函数及其应用*

韩栋　陈征Δ

摘　要　目的　介绍轮廓似然（Profile Likelihood）方法及其两个应用。**方法**　拟合 logistic 回归模型，计算轮廓似然的置信区间，并与 Wald 置信区间进行比较；采用轮廓似然方法解决障碍参数过多的模型拟合问题。**结果**　轮廓似然可以解决偏态分布下 Wald 置信区间失效的问题；最大轮廓似然估计与最大似然估计的结果是一致的，而轮廓似然仅需要较弱的假设条件。**结论**　参数呈非正态分布时，轮廓似然置信区间要优于 Wald 置信区间；轮廓似然作为最大似然估计的替代方法，可以解决最大似然无法计算或计算困难的问题，也可以提高模型的适用性。

关键词　轮廓似然　最大似然估计　置信区间　障碍参数

在回归分析中，似然函数通常会含有多个参数，但有时只有其中一个或几个是欲研究的参数，称为目标参数（Parameter of Interest），其它参数就被称作障碍参数（Nuisance Parameter），这些障碍参数对模型的求解有时会有阻碍作用。当存在多个障碍参数时，标准的似然方法无法消除或减少它们，所以变得不可靠或完全无效，而轮廓似然（Profile Likelihood，PL）作为一种处理障碍参数的方法能够解决障碍参数过多的问题。

另外，在目标参数呈非正态分布时，如果计算基于正态分布的 Wald 型置信区间（Wald CI）将会产生偏差[1]，尤其在无法计算目标参数的标准误时，Wald CI 也无法计算。而轮廓似然置信区间（PL CI）是基于 χ^2 分布且无需计算标准误，因此，PL CI 能够解决参数不服从正态分布和标准误无法计算时置信区间的计算问题。

本文将描述轮廓似然的定义及其两个应用，模拟比较 PL CI 与 Wald CI 的优劣并运用 PL 方法解决障碍参数过多和参数呈非正态分布时的问题。

原理与方法

轮廓似然定义

轮廓似然函数是固定目标参数时，对障碍参数求最大化后的函数，因而不是真正的似然函数。设 θ 表示目标参数或目标参数向量，γ 表示障碍参数或障碍参数向量，假设 $X_1,\ldots,X_n$ 为独立同分布且密度函数为 $p_{\theta,\gamma}(x)$，则 $l(\theta,\gamma)=\prod_{i=1}^{n} p_{\theta,\gamma}(X_i)$，然后轮廓似然函数被定义为 $pl(\theta)=l[\theta,\hat{\gamma}(\theta)]$，其中，$\hat{\gamma}(\theta)$ 为固定 θ 时，γ 的最大似然估计值（MLE），即：$pl(\theta)=\max_{\gamma} l(\theta,\gamma)$。

轮廓似然置信区间

Wald CI 是根据一个预先给定的置信水平和参考分布（在线性回归分析中选用 t 分布，其它为标准正态分布）选定分位数，采用“估计值±分位数×估计值的标准误”来计算模型中某个参数的置信区间。如果目标参数的分布呈偏态分布或无法计算其标准误时，Wald CI 的结果不可靠，而 PL CI 对以上特殊情况并不敏感，是一种更加稳健的方法。PL 方法可应用于所有基于似然理论的统计分析。

目标参数 θ 的 95%PL CI 是由检验水准为 0.05 时似然比检验无统计学意义的所有 θ 构成，即所

*：广东省科技计划项目(2010B031600100)、广东省“211 工程”三期重点学科建设项目(GW201005)资助.
作者地址：广州市同和南方医科大学，公共卫生与热带医学学院，生物统计学系（510515）
韩栋：南方医科大学生物统计专业硕士一年级。Δ通讯作者：陈征，zchen@smu.edu.cn。

有使似然比统计量小于等于 3.84($\chi^2(1)$ 分布的 95%分位数)的 θ 值。用公式表示为满足 $\ln[pl(\theta)] \cong \ln\left[pl(\hat{\theta})\right] - 3.84/2 = \ln\left[pl(\hat{\theta})\right] - 1.92$ 的所有 θ 值构成了 95%PL CI，其中 $\hat{\theta}$ 是 θ 的最大轮廓似然估计值。用 $\chi^2(1)$ 分布的 $(1-\alpha)$%分位数代替 3.84 可以计算其他置信水平为 $(1-\alpha)$% 的置信区间。

实例

多个障碍参数出现的问题

在对 2003 年 SARS 病死率估计的研究中，陈征等[2]基于竞争风险理论建立模型：令 n_i、d_i、c_i 和 a_i 分别指代在第 i 点的新增患者、死亡人数、治愈康复人数和观察人数（at risk），h_{1i}、h_{2i} 分别表示死亡与治愈的危险率，其中 i=1...s，表示不同时间点。根据实际数据观察可假设治愈-死亡危险率比 $R_i = h_{2i}/h_{1i} \equiv R$ 是一个常数，则病死率估计值为$(1+R)^{-1}$。关于 R 和 h_{1i} 的对数似然函数为：

$$l = \sum_{i=1}^{s}\{d_i \log(h_{1i}) + c_i \log(Rh_{1i}) + (a_i - d_i - c_i)\log(1 - h_{1i} - Rh_{1i})\} \tag{1}$$

因为病死率估计公式只与 R 有关，因此上式中 R 为目标参数，其它参数(h_{1i}，i=1,2,3,...,s)为障碍参数，此时似然函数中有(s+1)个参数，而且随着观察时间点增多（s 增大），障碍参数个数在不断增加，因此不能直接使用标准最大似然估计求解参数。基于实际数据研究[3]及 Lam[4]研究，模型又假设 $h_{1i} \equiv h_1$ 为常数，从而将对数似然函数(1)中的参数个数减至可求解的两个（R 和 h_1）。然而，将每个时间点的两个危险率均设为常数，此条件过于苛刻，但无此假设无法使用 MLE 估计参数。

使用轮廓似然方法解决上述问题：

此处仅假设 R_i 为常数,即 $R_i = h_{2i}/h_{1i} \equiv R$，基于似然函数公式(1)，解方程组 $\partial l/\partial h_{1i} = 0$,得出 $\hat{h}_{1i} = (d_i + c_i)/\left[a_i(1+R)\right]$，然后将 $\hat{h}_{1i}$ 代替 h_{1i} 代入公式(1)得出对数轮廓似然函数：

$$pl = \sum_{i=1}^{s}\left\{d_i \log\frac{d_i + c_i}{a_i(1+R)} + c_i \log\frac{R(d_i + c_i)}{a_i(1+R)} + (a_i - d_i - c_i)\log\left(1 - \frac{d_i + c_i}{a_i}\right)\right\}$$

此时轮廓似然函数中只包含一个目标参数 R，解方程 $\frac{dpl}{dR} = -\frac{\sum_{i=1}^{s}(d_i + c_i)}{(1+R)} + \frac{\sum_{i=1}^{s} c_i}{R} = 0$，得 R 的最大轮廓似然估计值为：

$$\hat{R} = \sum_{i=1}^{s} c_i \Big/ \sum_{i=1}^{s} d_i \tag{2}$$

然后轮廓观察信息量(Observed Profile Information)[5]为：

$$I(\hat{R}) = -\frac{\partial^2 pl}{\partial R^2} = -\frac{\sum_{i=1}^{s}(d_i + c_i)}{(1+R)^2} + \frac{\sum_{i=1}^{s} c_i}{R^2}$$

则 R 的近似方差估计是：

$$\text{avar}\left(\hat{R}\right) = \sum_{i=1}^{s} c_i \sum_{i=1}^{s}(d_i + c_i) \Big/ \left(\sum_{i=1}^{s} d_i\right)^3 \tag{3}$$

Muphy[6]等证明了在一般情况下，轮廓似然方法与最大似然方法是等价的，本例也验证了轮廓似然估计(式(2)和(3))与 MLE 结果[2]一致。由于轮廓似然方法的假设相比 MLE 方法[2]的假设弱化了很多，因此当存在障碍参数时，使用轮廓似然方法可以提高方法的适用性。

偏态分布的轮廓似然置信区间

2.1 数值模拟

此节对不同偏态分布情况下 PL CI 和 Wald CI 的置信水平进行检测。为了模拟非正态分布参数，选取 logistic 模型 $\log(p_i/1-p_i)=\beta_1+\beta_2 x_i$，并设定 x_i 分别为（60，65，75，90），$\beta_1=-6.5,\beta_2=0.1$。采用二项分布，每个 x 下的试验次数分别设定为 3、8、20，以每一个 p_i 为发生率，模拟出每个试验次数下的事件发生次数与失败次数，拟合 logistic 回归模型并计算 PL CI 和 Wald CI 界值在 $\chi^2(1)$ 分布下的置信水平。相对轮廓似然值(relative PL，RPL)定义为：轮廓似然值/最大轮廓似然值。根据似然理论，RPL 表示数据对两个参数估计值支持程度的比值，取值为(0,1]，因此可采用 RPL 比较不同数据情况下的置信限处的似然。轮廓似然不对称性指标的计算公式[7]为：

$$\text{不对称性}=\frac{\left|(PL\,CI\text{上限}-\text{估计值})-(\text{估计值}-PL\,CI\text{下限})\right|}{(PL\,CI\text{上限}-PL\,CI\text{下限})}\times 100\%.$$

表示置信限到估计值距离之差占置信区间长度的百分比，不对称性越趋近于 0，表示 PL CI 越趋于对称。模拟结果反映在表 1 和图 1 上。

表 1 轮廓似然置信区间与 Wald 置信区间的置信水平

试验次数	PL CI 不对称性	β_2 的 PL CI	PL CI 的置信水平	β_2 的 Wald CI	Wald CI 的置信水平
3	28.9%	-0.012~0.362	95.0%	-0.050~0.293	93.0%
8	13.6%	0.013~0.182	95.0%	0.004~0.168	94.3%
20	8.3%	0.041~0.150	95.0%	0.037~0.145	94.7%

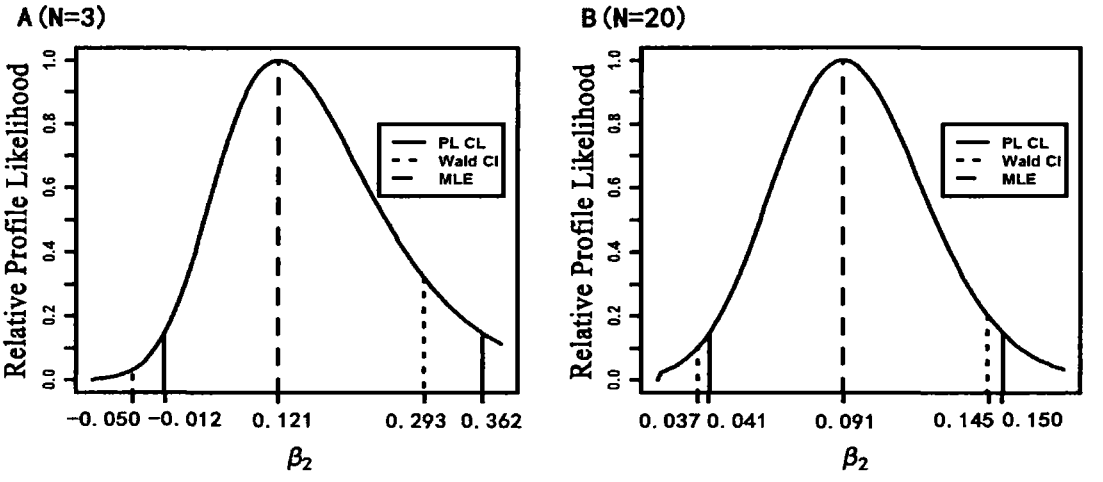

图 1 不同试验次数下的相对轮廓似然值

由表 1 和图 1 可以看出，随着试验次数增大，Wald CI 与 PL CI 趋于一致，PL CI 也逐渐趋于对称。试验次数较小时（N=3），95%Wald CI 的置信水平仅为 93.0%，由于采用 PL 方法，95%PL CI 的置信水平被控制在 95.0%，此时 PL CI 不对称为 28.9%。

图 1-A 中，Wald CI 下限至 PL CI 下限间的 RPL 值在 0.03~0.15 之间，而 Wald CI 上限至 PL CI

上限间 RPL 值的区间为 0.15~0.36，由于两个 CI 上限间的 RPL 值均大于两个 CI 下限间的 RPL 值，根据似然理论以及似然比检验的原理，Wald CI 下限至 PL CI 下限间包括真实值的可能性均比 Wald CI 上限至 PL CI 上限间包括真实值的可能性要低。图 1-B 的结论与此类似，因此 PL CI 置信区间更可信。

2.2 白鼠毒性实验

利用 PL 来分析白鼠毒性实验[8]，n_i 表示总的白鼠数，r_i 表示死亡鼠数，x_i 表示毒药剂量,数据如下表：

表 2 白鼠毒性实验数据

x_i	422	744	948	2069
r_i	0	1	3	5
n_i	5	5	5	5

对以上数据拟合 logistic 回归模型：$\log(p_i/1-p_i)=\beta_1+\beta_2\log x_i \quad (i=1\ldots4)$。结果见表 3，经 Wald 检验，毒药剂量的对数值对白鼠的死亡率没有影响（P=0.119），但由图 2 可以看出，β_2 的轮廓似然函数值呈正偏态，因此采用 Wald 法不可靠。如果采用似然比检验，由表 3 的结果显示，毒药剂量的对数值对白鼠的死亡率的影响有统计学意义（P<0.001），毒药剂量对数值的 PL CI 为(2.283,21.491)，不对称性达到 41.2%。

表 3 似然比检验与 Wald 检验

估计值±SE	检验方法	统计量	P 值	CI	置信水平	PL CI 不对称性
7.930±5.081	似然比	Chisq=15.745	<0.001	PL CI = (2.283, 21.491)	94.9%	41.2%
	Wald 法	z=1.561	0.119	Wald CI = (-2.029, 17.889)	93.7%	

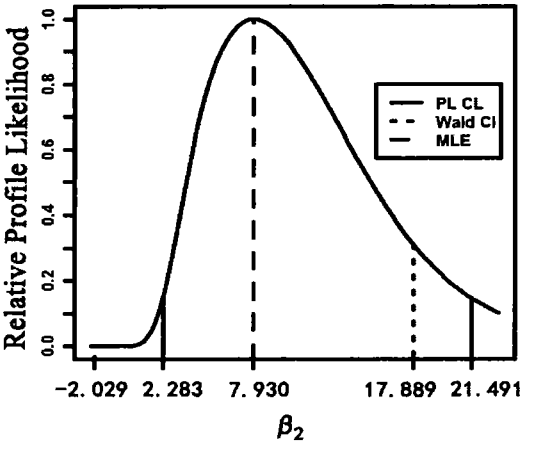

图 2 白鼠毒性实验中系数值的相对轮廓似然值

讨论

本文就轮廓似然方法及其应用进行了阐述，并用模拟与实例说明轮廓似然在估计参数值和计算置信区间等方面都有较强的实用性。除了文中所述的一些性质外，在参数模型中，对数轮廓似然函数的二阶导函数是观察信息量的估计值，甚至是在轮廓似然函数不能写成外显函数的情况下，数值计算方法也可以计算出信息矩阵的估计值。轮廓似然方法还有其他特殊的性质，如利用轮廓似然方法消去普通似然函数中的基准危险率，从而推导出拟合 Cox 回归

时使用的偏似然函数[6]；也可以利用轮廓似然方法消去基准危险率后，构造全轮廓似然函数(full-profile likelihood function)[9]，在中小样本情况下，最大全轮廓似然估计值比最大偏似然估计值更有用；与标准的似然方法相比，利用轮廓似然方法处理有删失的生存时间数据时，无需对删失类

型进行假设[10]。除了轮廓似然方法外，处理障碍参数的方法还有边际似然、条件似然、联合似然等。由于以上三种似然方法的使用都需要依赖一定的特殊结构，而本文所述的轮廓似然没有这种限制，甚至在轮廓似然函数不能被写成显性函数的形式时，轮廓似然方法依然适用。因此轮廓似然作为一种处理障碍参数的方法更可行[11]。

现在，轮廓似然方法仍在发展。因为轮廓似然函数不是直接基于密度函数的乘积而获得的似然函数，从而导致了PL存在一些不能符合似然函数要求的性质(如偏性、方差估计过度最优化等)，因此一些改进的轮廓似然方法又被提出，如conditional profile likelihood、modified profile likelihood、adjusted profile likelihood等。

Profile Likelihood and Its Application

Han Dong, Chen Zheng
Department of Biostatistics, School of Public Health and Tropical Medicine,
Southern Medical University (510515), Guangzhou

Abstract **Objective** To introduce the method and application of profile likelihood. **Method** Comparing the confidence interval of profile likelihood with that of Wald under the logistic model; Using profile likelihood to fit models with overmany parameters. **Results** Confidence interval of profile likelihood can solve the problem that confidence interval of Wald fails; The maximum profile likelihood estimator is equivalent to the maximum likelihood estimator with a weaker assumption. **Conclusions** Profile likelihood confidence interval is superior to Wald confidence interval for the parameter from a skewed distribution; Profile likelihood estimator can be worked out when the MLE is hardly computed, so that the applicability is increased.

Key words Profile Likelihood; Maximum likelihood estimator; Confidence interval;Nuisance parameter

参考文献

[1] Venzon D J, Moolgavkar S H. A method for computing profile-likelihood-based confidence intervals[J]. Applied Statistics,1988,37(1):87-94.

[2] 陈征，Tsuyoshi N. 基于竞争风险理论和概要型数据的病死率估计模型[J]. 中国卫生统计,2010,27(003):249-252.

[3] Chen Z, Nakamura T. Statistical evidence for the usefulness of Chinese medicine in the treatment of SARS[J]. Phytotherapy Research,2004,18(7):592-594.

[4] Lam K F, Deshpande J V, Lau E, et al. A test for constant fatality rate of an emerging epidemic: with applications to severe acute respiratory syndrome in Hong Kong and Beijing[J]. Biometrics,2008,64(3):869-876.

[5] Tsodikov A, Garibotti G. Profile information matrix for nonlinear transformation models[J]. Lifetime data analysis,2007,13(1):139-159.

[6] Murphy S A, Van der Vaart A W. On profile likelihood[J]. Journal of the American Statistical Association,2000,95(450):449-465.

[7] Royston P. Profile likelihood for estimation and confidence intervals[J]. Stata Journal,2007,7(3):376-387.

[8] Aitkin M. Statistical modelling: the likelihood approach[J]. The Statistician,1986,35(2):103-113.

[9] Ren J, Zhou M. Full likelihood inferences in the Cox model: an empirical likelihood approach[J]. Annals of the Institute of Statistical Mathematics, Published online: 18 February 2010, doi:10.1007/s10463-010-0272-y.

[10] Zhang Z. Profile Likelihood and Incomplete Data[J]. International Statistical Review,2010,78(1):102-116.

[11] Montoya J, Díaz-Francés E, Sprott D. On a criticism of the profile likelihood function[J]. Statistical Papers,2009,50(1):195-202.