系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
27(1) (2007, 2), 59–67

混合 von Mises 模型的参数估计*

陈家骅 李鹏飞
(加拿大滑铁卢大学统计与精算科学系， 加拿大 N2L 3G1)
谭鲜明
(加拿大滑铁卢大学统计与精算科学系， 加拿大 N2L 3G1; 南开大学数学科学学院， 天津 300071)

摘要 有限混合 von Mises 模型在天文学、生物学、地理和医药等许多领域都有重要的应用. 可是，不论样本量有多大，此模型的似然函数都是无界的. 因此，参数的最大似然估计 (MLE) 是不相合的. 我们发现，与混合正态模型一样，上述困难可以通过引入关于分布浓度参数的一个惩罚函数或对参数空间添加适当的约束来克服. 在此文中，我们从理论上证明了这两种方法是可行的，相应的参数估计是强相合的，且是渐近有效的. 我们还通过计算机模拟来探讨这些新方法在有限样本情况下的统计性质，并与现有的矩估计作了比较. 结果发现，惩罚极大似然估计在均方误差方面表现最佳. 最后我们还分析了一组实际数据，以进一步介绍新的估计方法.

关键词 混合 von Mises 模型；约束最大似然；惩罚最大似然；强相合性.

MR(2000) 主题分类号 62H11, 62H12

1 引 言

与一元正态分布相对应， von Mises 分布常被称作是圆上的正态分布. 它是最主要的一种描述方向数据的模型. 有关此分布的许多性质可以在 [1] 中找到.

在许多实际问题中，方向数据常包含一些非同质性. 著名的海龟数据便是一个很好的例子 [2]. 文中指出，当把一批海龟在某个地点释放后，大多数海龟落地后便朝东北方向移动，但部分海龟却朝几乎完全相反的西南方向移动，其余的海龟的移动方向显得毫无规章. 这一随机现象很难用单一的 von Mises 分布来刻划，而 von Mises 的有限混合却显得非常合理，许多类似的实际例子可以从文 [1,3] 中找到.

单一 von Mises 分布的概率密度函数为

$$f(\theta;\mu,\kappa)=\frac{1}{2\pi I_0(\kappa)}\exp\{\kappa\cos(\theta-\mu)\},\quad |\theta|\le\pi, \tag{1}$$

其中 $|\mu|\le\pi$ 称作均值方向， $\kappa\ge 0$ 称作浓度参数， $I_0(\kappa)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\{\kappa\cos(\theta)\}\mathrm{d}\theta$. 混合 von Mises 分布的密度函数可表示为

$$f(\theta;G)=\sum_{j=1}^{p}\alpha_j f(\theta;\mu_j,\kappa_j).$$

* 加拿大 NSERC 基金，中国国家自然科学基金 (10601026) 资助课题.
收稿日期：2006-10-26.
谨以此文纪念成平研究员.

在这之中, $G=(\alpha_1,\alpha_2,\cdots,\alpha_p,\mu_1,\mu_2,\cdots,\mu_p,\kappa_1,\kappa_2,\cdots,\kappa_p)'$ 包含所有参数, $\alpha_j\ (j=1,2,\cdots,p)$ 是第 j 个成份的混合比例, μ_j, $\kappa_j\ (j=1,2,\cdots,p)$ 分别是第 j 个成份的位置及浓度参数, p 是有限混合 von Mises 分布的阶. 我们把参数空间记为

$$\Gamma=\Big\{G=(\alpha_1,\alpha_2,\cdots,\alpha_p,\mu_1,\mu_2,\cdots,\mu_p,\kappa_1,\kappa_2,\cdots,\kappa_p)':\\ \sum_{j=1}^{p}\alpha_j=1,\alpha_j\geq 0,\mu_j\in[-\pi,\pi],\kappa_j\geq 0,\ \ j=1,2,\cdots,p\Big\}.$$

假设我们获得了有限混合 von Mises 分布 $f(\theta;G)$ 的一组随机样本: $\theta_1,\theta_2,\cdots,\theta_n$. 在 p 已知的情况下, 我们应该如何估计 G 呢？首先我们注意到有限混合 von Mises 分布族是可识别的 [4,5], 因而估计 G 是有意义的. 众所周知, 在正规参数模型下, 极大似然估计具有许多优良的统计性质, 因而它常常是人们的首选 [6]. 然而由于混合 von Mises 模型的非正规性, 极大似然估计存在一个缺陷. 为说明此点, 令

$$l_n(G)=\sum_{i=1}^{n}\log\{f(\theta_i;G)\}$$

为对数似然函数. 如果让 $\mu_1=\theta_1$, $\kappa_1\to\infty$, 且其它参数取定值时, 我们发现 $l_n(G)\to\infty$. 显然 $\widehat{\mu}_1=\theta_1$ 不可能是 μ_1 的相合估计. 为此, Spurr and Koutbeiy 引进了矩方法 [7]. 直到现在, 估方法仍是估计 G 的最常用的方法. 然而, 矩估计通常不是一个渐近有效的估计.

如同一元正态混合分布, G 的渐近有效估计可以通过引入对浓度参数的惩罚项或者对其取值范围加以约束而得到. 具体的做法与 [8–10] 等文章的想法非常相似, 但那些文章的结论却无法直接在混合 von Mises 模型中应用. 因此, 我们在此文中介绍这两种方法, 并证明它们的强相合性及渐近正态性. 鉴于技术细节上与 [9] 和 [10] 很类似, 我们在第 2 节中只介绍理论结果, 而把十分简要的证明放在附录中. 在第 3 节里, 我们介绍模拟结果. 在我们的试验中, 带惩罚的极大似然估计的表现最佳, 这尤其体现在对浓度参数的估计上: 罚后的极大似然估计的均方误差明显小于其它方法. 在第 3 节中, 我们还用新方法对 [2] 中的实际数据进行了分析, 以进一步说明此方法的实际性质. 最后一节是全文的总结以及讨论.

2 惩罚最大似然估计与约束最大似然估计

令 C_n 为一在样本量 $n\to\infty$ 时也趋于无穷大的正常数. 我们把约束参数空间定义为

$$\Gamma_n=\Big\{G\in\Gamma:\max_{i,j}\frac{\kappa_i}{\kappa_j}\leq C_n\Big\}.$$

而 $l_n(G)$ 在 Γ_n 上的最大值点 $\widehat{G}_n$ 便是 G 的约束最大似然估计 (CMLE). 值得指出的是, 当 n 趋于无穷时, $\Gamma_n\to\Gamma$, 从而在 n 充分大时, 真实参数 $G_0\in\Gamma_n$.

令 $p_n(G)$ 为一取负值的函数, 定义

$$pl_n(G)=l_n(G)+p_n(G), \tag{2}$$

为带惩罚的似然函数. 而 $pl_n(G)$ 在 Γ 中的最大值点 $\widetilde{G}_n$ 便是 G 的惩罚最大似然估计 (PMLE). 为使 $\widetilde{G}_n$ 具有良好的统计性质, 我们必须选用合适的 $p_n(G)$. 令 $M=\max\{1,\max_\theta f(\theta;G_0)\}$. 我们要求

A1) $p_n(G)=\sum_{j=1}^{p}\widetilde{p}_n(\kappa_j)$.

A2) $p_n(G_0)=\sum_{j=1}^{p}\widetilde{p}_n(\kappa_{j0})=o(n)$.

A3) 当 n 充分大时，如果 $\kappa\geq n^2M^2$, 则有，$\widetilde{p}_n(\kappa)\leq -2\log^2 n\log(\kappa)$.

其中 $\widetilde{p}_n(\kappa)$ 是一个关于单个浓度参数 κ 的取负值的函数.

注 1　如仅为了保证 PMLE 的强相合性，$p_n(G)$ 的选择余地很大. 作为一个统计分析方法，满足 (A1) 的选择在使用 EM 算法时会十分方便. 条件 (A3) 仅仅要求惩罚函数足够大，系数 2 的采用仅为了方便证明的叙述. 另外，条件 (A3) 也指出，我们仅仅在 κ 过大时才需要加惩罚.

下面两个定理说明， PMLE 和 CMLE 都具有优良的渐近性质.

定理 2.1　1) 如果 $p_n(G)$ 符合条件 A1–A3, 那么惩罚最大似然估计 (PMLE) $\widetilde{G}_n$ 是 G_0 的强相合估计. 也就是说，当 $n\to\infty$ 时几乎必然地有 $\widetilde{G}_n\to G_0$.

2) 对任意给定正常数 m, 如果 $C_n\leq\exp(m\log^2 n)$, 则约束最大似然估计 (CMLE) $\widehat{G}_n$ 是 G_0 的强相合估计. 也就是说，当 $n\to\infty$ 时几乎必然地有 $\widehat{G}_n\to G_0$.

此定理的证明将在附录中给出. 在强相合性的基础上，我们很容易进一步获知这两种方法的渐近有效性. 为此，我们必须选用比较光滑的惩罚函数.

A4) 存在 G_0 的某个邻域 Γ_0, 使得 $p_n(G)$ 在其上关于 G 是三阶连续可微的，并且对于 $s=1,2,3$ 都有 $\|p_n^{(s)}(G)\|=o(\sqrt{n})$.

同时，令 $E_0(\cdot)$ 表示在真分布下的数学期望，及

$$I(G_0)=E_0\left(\left[\frac{\partial\log f(X;G_0)}{\partial G}\right]^{\tau}\left[\frac{\partial\log f(X;G_0)}{\partial G}\right]\right)$$

为 Fisher 信息阵. 下一定理给出 PMLE ($\widetilde{G}_n$) 和 CMLE ($\widehat{G}_n$) 的渐近分布.

定理 2.2　1)　假设 $p_n(G)$ 满足条件 A1–A4, 则有

$$\sqrt{n}(\widetilde{G}_n-G_0)\xrightarrow{\mathcal{D}}N(0,I^{-1}(G_0)),$$

其中 $\xrightarrow{\mathcal{D}}$ 表示依分布收敛.

2) 对任意给定正常数 m, 如果 $C_n\leq\exp(m\log^2 n)$, 则有

$$\sqrt{n}(\widehat{G}_n-G_0)\xrightarrow{\mathcal{D}}N(0,I^{-1}(G_0)).$$

很显然， $\widehat{G}_n$ 和 $\widetilde{G}_n$ 都是渐近有效估计. 但是，在有限样本下，它们的统计性质仍可能大不相同. 这一问题通常只能通过计算机模拟来探讨.

3　模拟结果与实际数据分析

在这一节中，我们通过计算机模拟来研究现有的 G 的矩估计和 PMLE, CMLE 的统计性质. 尽管矩估计通常不是渐近有效的，但我们并不能由此断定它的有限样本性质一定比不上渐近有效估计.

在模拟计算中，我们选用了 $p=2$ 的两个混合 von Mises 分布. 它们分别是

$$0.25VM\left(-\frac{\pi}{2},4\right)+0.75VM\left(\frac{\pi}{2},2\right),$$

和

$$0.25VM\Big(-\frac{\pi}{4},2\Big)+0.75VM\Big(\frac{\pi}{4},4\Big),$$

其中 $VM(\mu,\kappa)$ 表示均值为 μ, 浓度参数为 κ 的 von Mises 分布. 所选的这两个分布的特点是: 前一个混合分布的密度函数是双峰的, 而后一个的密度函数是单峰的.

我们为 PMLE 选择了 $p_n(G)=-c\kappa_1-c\kappa_2$, 及小, 中, 大 3 个 c 值, 即 $c=0.25,0.5,1$; 为 CMLE 选择了 $C_n=10$ 及 n 两个值, 其中 n 是样本量. 我们模拟了 $n=100$ 及 200 两种情况. 这样, 对每一组数据, 我们用数值方法获得 G 的六个点估计, 其中三个对应于 PMLE $(c=0.25,0.5,1)$, 两个对应于 CMLE $(C_n=10,n)$, 最后一个为矩估计.

最后, 对每一混合 von Mises 模型, 我们重复产生了 2000 组数据, 根据这 2000 个结果我们计算了这六种估计的偏差 (Bias), 标准差 (Std) 以及均方误差的平方根 (RMSE). 这些模拟结果被列入表 1 和表 2.

从这些模拟结果来看, 这六种方法对 μ_1, μ_2 及 α 的估计的精度是十分接近的. 但对于 κ_1 及 κ_2 的估计却差别很大. 总体而言, PMLE 在 $c=0.25,0.5,1$ 时的 κ_1 及 κ_2 的估计精度都比其它方法好. 特别是对 κ_1 的估计, 矩方法的精度在两个模型, 两种样本量下都差强人意. 约束最大似然估计尽管比矩估计有所改进, 但表现仍不尽如意. 因而, 我们初步断言, 惩罚极大似然估计是较优越的一种估计方法.

需要补充的是, 在计算 PMLE 和 CMLE 时我们采用了 EM 算法. 算法的初值选取为参数真值及文 [7] 在矩估计所采用的初值. 在计算矩估计时, 我们使用的是文 [7] 中采用的计算方法.

表 1 CMLE, PMLE 和矩估计的比较, 模型 I

		$n=100$					$n=200$				
		μ_1	μ_2	κ_1	κ_2	α	μ_1	μ_2	κ_1	κ_2	α
PMLE	Bias	0.00	0.00	−0.13	0.23	0.02	0.00	0.00	−0.05	0.11	0.01
$(c=0.25)$	Std	0.15	0.10	1.77	0.51	0.06	0.10	0.07	1.34	0.37	0.04
	RMSE	0.15	0.10	1.77	0.55	0.06	0.10	0.07	1.34	0.38	0.04
PMLE	Bias	0.00	0.00	−0.78	0.24	0.03	0.00	0.00	−0.43	0.13	0.01
$(c=0.5)$	Std	0.15	0.10	1.31	0.47	0.05	0.10	0.07	1.10	0.35	0.04
	RMSE	0.15	0.10	1.52	0.53	0.06	0.10	0.07	1.18	0.38	0.04
PMLE	Bias	0.00	0.00	−1.53	0.23	0.04	0.00	0.00	−0.97	0.15	0.02
$(c=1)$	Std	0.16	0.10	0.90	0.41	0.05	0.10	0.07	0.85	0.33	0.04
	RMSE	0.16	0.10	1.78	0.47	0.06	0.10	0.07	1.29	0.36	0.04
CMLE	Bias	0.00	0.00	1.13	0.19	0.01	0.00	0.00	0.52	0.08	0.00
$(C_n=10)$	Std	0.14	0.10	3.11	0.56	0.06	0.10	0.07	1.83	0.39	0.04
	RMSE	0.14	0.10	3.31	0.59	0.06	0.10	0.07	1.90	0.40	0.04
CMLE	Bias	0.00	0.00	1.34	0.19	0.01	0.00	0.00	0.53	0.08	0.00
$(C_n=n)$	Std	0.14	0.10	4.29	0.56	0.06	0.10	0.07	1.94	0.39	0.04
	RMSE	0.14	0.10	4.49	0.60	0.06	0.10	0.07	2.01	0.40	0.04
	Bias	0.00	0.00	11.92	0.19	0.01	0.00	0.00	2.63	0.08	0.00
矩估计	Std	0.15	0.11	42.59	0.59	0.07	0.10	0.08	17.31	0.42	0.05
	RMSE	0.15	0.11	44.21	0.62	0.07	0.10	0.08	17.51	0.43	0.05

表 2　CMLE, PMLE 和矩估计的比较，模型 II

		$n=100$					$n=200$				
		μ_1	μ_2	κ_1	κ_2	α	μ_1	μ_2	κ_1	κ_2	α
PMLE	Bias	–0.11	–0.01	0.68	0.11	–0.02	–0.04	–0.01	0.37	0.08	0.00
($c=0.25$)	Std	0.29	0.08	1.30	0.98	0.08	0.26	0.07	0.99	0.79	0.07
	RMSE	0.31	0.08	1.46	0.99	0.08	0.27	0.07	1.05	0.79	0.07
PMLE	Bias	–0.10	–0.01	0.28	-0.07	–0.02	–0.04	–0.01	0.19	–0.03	–0.01
($c=0.5$)	Std	0.26	0.08	0.84	0.82	0.07	0.24	0.06	0.75	0.69	0.06
	RMSE	0.28	0.08	0.89	0.82	0.07	0.24	0.07	0.77	0.69	0.06
PMLE	Bias	–0.08	–0.02	–0.16	–0.39	–0.02	–0.04	–0.02	–0.03	–0.21	–0.01
($c=1$)	Std	0.23	0.08	0.53	0.64	0.06	0.20	0.06	0.52	0.57	0.05
	RMSE	0.24	0.08	0.56	0.75	0.06	0.21	0.06	0.52	0.61	0.05
CMLE	Bias	–0.14	–0.01	2.21	0.28	–0.02	–0.05	-0.01	0.82	0.20	0.00
($C_n=10$)	Std	0.35	0.09	4.49	1.30	0.09	0.31	0.07	2.23	0.97	0.09
	RMSE	0.38	0.09	5.00	1.33	0.09	0.31	0.07	2.37	0.99	0.09
CMLE	Bias	–0.16	–0.02	4.06	0.26	–0.02	–0.05	–0.01	1.01	0.20	0.00
($C_n=n$)	Std	0.36	0.10	16.64	1.32	0.09	0.31	0.07	5.93	0.97	0.09
	RMSE	0.39	0.10	17.12	1.35	0.10	0.31	0.07	6.01	0.99	0.09
	Bias	–0.10	–0.01	7.52	0.22	–0.01	–0.03	0.00	1.71	0.16	0.00
矩估计	Std	0.37	0.09	32.84	1.24	0.10	0.32	0.07	12.98	0.96	0.09
	RMSE	0.38	0.09	33.68	1.25	0.10	0.32	0.07	13.09	0.97	0.09

最后我们把 PMLE 和 CMLE 应用于分析 Gould 的海龟数据. 这个数据来自一个研究某种海龟的试验，其目的是要考察这种海龟在陌生环境下寻找栖息地方向的能力. 在某个试验地释放 76 只海龟，然后记录它们最初的移动方向. 数据显示大多数海龟落地后即朝东北方向 (栖息地方向) 移动，但部分海龟朝向几乎完全相反的西南方向，剩下的极少部分的海龟的移动方向没有规律. 关于这个数据的一个解释是：这种海龟具有很强的定向能力，但是部分海龟混淆了正向和反向，因此向相反的方向移动 (参见 [11]). 我们用一个两成份的混合 von Mises 分布拟合这组数据. 矩方法得到的估计是

$$\bar{\mu}_1=0.351\pi,\ \bar{\mu}_2=-0.666\pi,\ \bar{\kappa}_1=2.913,\ \bar{\kappa}_2=4.799,\ \bar{\alpha}=0.819,$$

取 $c=0.5$ 时的 PMLE

$$\tilde{\mu}_1=0.353\pi,\ \tilde{\mu}_2=-0.658\pi,\ \tilde{\kappa}_1=2.639,\ \tilde{\kappa}_2=4.089,\ \tilde{\alpha}=0.828,$$

而关于 G 的 CMLE($C_n=10$) 是

$$\widehat{\mu}_1=0.353\pi,\ \widehat{\mu}_2=-0.660\pi,\ \widehat{\kappa}_1=2.619,\ \widehat{\kappa}_2=8.447,\ \widehat{\alpha}=0.837.$$

根据模拟计算的经验，我们认为用 PMLE 给出的 κ_1,κ_2 的估计更为可信.

4　总结与讨论

本文我们讨论了混合 von Mises 分布的参数估计问题. 特别地，我们研究了这种模型下的惩罚最大似然估计和约束最大似然估计的大样本性质，证明了 PMLE 和 CMLE 的强相合性和渐近正态性. 然后通过模拟比较了 CMLE, PMLE 和矩估计在有限样本情形下的表现. 模拟显示从均方误差的表现看， PMLE 是三种方法中表现最佳的.

另一个有趣的问题是如何对混合 von Mises 分布进行同质性检验. 亦即检验某组样本是来自一个混合的 von Mises 分布还是来自某个单一的 von Mises 分布. 一个可能的办法是要

求混合比例 α 不靠近 0 以及 1, 在这个限制下对数似然比检验统计量的极限分布往往比较简单 (参见 [12]. 另一种方法是在对数似然函数上添加一个关于混合比例的惩罚项. 在假设各子分布的浓度参数相等的条件下，这样得到的修正的对数似然比 (MLRT) 往往有简单的极限分布，并且具有局部最优势的性质[13]. 在各子分布的浓度参数不相等的情况下，同质性检验问题仍然没有得到充分研究. 我们下一步计划把 [12] 和 [13] 的方法结合起来，以求找到一种关于混合 von Mises 分布在上述情形下同质性的检验方法.

参 考 文 献

[1] Mardia K V and Jupp P E. Directional Statistics. John Wiley and Sons, 2000.

[2] Gould E. Studies on the orientation of turtles. Copeia, 1959, **2**: 174–176.

[3] Fisher N I. Statistical Analysis of Circular Data. Cambridge: Cambridge University Press, 1993.

[4] Fraser M D, Hsu Y S and Walker J J. Identifiability of finite mixtures of von Mises distributions. *Ann. Statist.*, 1981, **9**: 1130–1131.

[5] Holzmann H, Munk A and Stratmann B. Identifiability of finite mixtures - with applications to circular distributions. Sankhya, 2004, **66**: 440–450.

[6] Lindsay B G. Mixture models: theory, geometry and applications. 1995, Hayward: Institute for mathematical Statistics.

[7] Spurr B D and Koutbeiy M A. A comparison of various methods for estimating the parameters in mixture of von Mises distributions. *Commun. Statist. Simul. Comput.*, 1991, **20**: 725–741.

[8] Hathaway, R J. A constrained formulation of maximun-likelihood estimation for normal mixture distributions. *Ann. Statist.*, 1985, **13**: 795–800.

[9] Chen J, Tan X and Zhang R. Consistency of Penalized MLE for Normal Mixtures in Mean and Variance. Manuscript, 2006.

[10] Tan X, Chen J and Zhang R. Consistency of the constrained maximum likelihood estimator in finite normal mixture models. Manuscript 2006.

[11] Stephens M. Techniques for directional data. Technical Report 150, 1969, Department of Statistics, Stanford University.

[12] Chen J and Cheng P. The limit distribution of the restricted likelihood ratio statistic for finite mixture models. *Northeastern Mathematical Journal*, 1995, **11**: 365–374.

[13] Chen H, Chen J and Kalbfleisch J D. A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society*, Series B, 2001, **63**(1): 19–29.

[14] Serfling R J. Approximation Theorems of Mathematical Statistics. John Wiley and Sons, 1980.

[15] Wald A. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, 1949, **20**: 595–601.

附 录

尽管 [9] 和 [10] 讨论了混合正态模型下 PMLE 和 CMLE 的强相合性，但那些理论结果不能直接应用于混合 von Mises 模型. 不过，他们的技术手段及证明思路可直接应用于证明混合 von Mises 模型参数估计的强相合性. 因此，我们仅给出一个十分粗略的证明，感兴趣的读者可以参阅 [9] 和 [10].

为简略起见，我们只考虑 $p=2$ 的情形，一般情形的证明要复杂许多，但证明思路完全

雷同．为证明定理 2.1 中 PMLE 的强相合性，我们首先把参数空间划分成三个子空间：

$$\begin{aligned}\Gamma_1 &= \{G \in \Gamma : \kappa_1 \geq \kappa_2 \geq M_1\},\\ \Gamma_2 &= \{G \in \Gamma : \kappa_1 \geq M_2, \kappa_2 \leq M_1\},\\ \Gamma_3 &= \Gamma - (\Gamma_1 \cup \Gamma_2).\end{aligned}$$

确切的 M_1, M_2 的值略后会给出．上述三个子空间代表了三种情况．子空间 Γ_1 中各子分布的浓度参数都很大．这意味着相应的混合 von Mises 分布的密度在 μ_1, μ_2 的小邻域内很高．因而，落在那些邻域内的观察值对似然函数的贡献很大．幸而，这样的观察值的个数随着 κ_1, κ_2 的增加而减少．具体地，我们发现

引理 1　除去一个不依赖于 G 的零测集后，当 n 充分大以及在 G_0 下，

$$\begin{aligned}H_n(G) &= \sum_{i=1}^{n} I\Big\{\sin\Big[\frac{1}{2}(\theta_i - \mu)\Big] \leq \kappa^{-\frac{1}{2}}\log\kappa\Big\}\\ &\leq 4\log^2 n I(\kappa \geq n^2M^2) + (4nM\kappa^{-\frac{1}{2}}\log\kappa + 1)I(A \leq \kappa \leq n^2M^2),\end{aligned}$$

其中 A 为使得下列条件成立的某个正数：

1)　当 $\kappa \geq A$ 时，　$|\arcsin(\kappa^{-\frac{1}{2}}\log\kappa)| \leq 2\kappa^{-\frac{1}{2}}\log\kappa$;

2)　当 $\kappa \geq A$ 时，　$\kappa^{-\frac{1}{2}}\log\kappa$ 是关于 κ 的单调减函数．

证　对于任一给定的参数 G, H_n 是 n 个独立且具有相同二项分布的随机变量的和，且此二项分布的期望很小．因而我们可以使用 Bernstein 不等式 [14] 证明 $|H_n(G) - E\{H_n(G)\}| \geq n\epsilon$ 的概率很小，再利用 Borel-Cantalli 引理，此结论便可证明．而关于此结论对 G 的一致性的证明较繁锁，感兴趣的读者可以参阅 [9].

利用这一结论，我们将进一步证明，无论 $j = 1$ 或 $j = 2$, 都有

$$\sup_{G\in\Gamma_j} pl_n(G) - pl_n(G_0) \to -\infty \quad \text{a.s. 当 } n \to \infty. \tag{3}$$

我们首先考虑 Γ_1 的情形．为此我们需要先设定 M_1, 从而把 Γ_1 确定下来．令 $K_0 = E_0\{\log f(\theta; G_0)\}$, 易知 $K_0 < \infty$. 取 M_1 为满足如下条件的某正数．

1)　$M_1 > \max\{A, 64M^2, \exp(4)\}$,

2)　$2\log M_1 - \log^2 M_1 \leq 4(K_0 - 2)$,

3)　当 $\kappa \geq M_1$ 有，　$\exp\{\kappa\} \leq 2\pi I_0(\kappa)\sqrt{\kappa}$.

注意到当 $\kappa \to \infty$ 时，有 $\frac{\exp\{\kappa\}}{\{I_0(\kappa)\sqrt{\kappa}\}} \to \sqrt{2\pi}$ (参见文 [1]) 以及 $2\log\kappa - \log^2\kappa \to -\infty$. 因而 M_1 的存在性没有问题．

定义

$$A_1 = \Big\{i : \Big|\sin\Big(\frac{\theta_i - \mu_1}{2}\Big)\Big| < \kappa_1^{-\frac{1}{2}}\log\kappa_1\Big\},\quad A_2 = \Big\{i : \Big|\sin\Big(\frac{\theta_i - \mu_2}{2}\Big)\Big| < \kappa_2^{-\frac{1}{2}}\log\kappa_2\Big\}.$$

它们为观察值 $\theta_1, \theta_2, \cdots, \theta_n$ 中与 μ_1 或 μ_2 十分接近的那些观测值的指标集．对于任一指标集 S, 我们令

$$l_n(G; S) = \sum_{i\in S}\log f(\theta_i; G).$$

因此 $l_n(G)=l_n(G;A_1)+l_n(G;A_2A_1^c)+l_n(G;A_1^cA_2^c)$. 对于任一 $G\in\Gamma_1$, 我们有 $\kappa_1>\kappa_2$. 因而对于任一 $i\in A_1$, $f(\theta_i;G)\le I_0^{-1}(\kappa_1)\exp\{\kappa_1\}<\sqrt{\kappa_1}$, 及 $l_n(G;A_1)\le\frac{1}{2}n(A_1)\log(\kappa_1)$, 其中 $n(A_1)$ 为 A_1 中指标的个数. 同理, 对于任一 $i\in A_2-A_1=A_2A_1^c$, 我们可证, $f(\theta_i,G)\le\frac{1}{2}\log\kappa_2$. 因而 $l_n(G;A_2A_1^c)\le\frac{1}{2}n(A_2)\log\kappa_2$.

利用引理 1 的结论, 我们有

$$n(A_1)\le 4\log^2 nI(\kappa_1\ge n^2M^2)+(4nM\kappa_1^{-\frac{1}{2}}\log\kappa_1+1)I(A\le\kappa_1\le n^2M^2),$$

略去低阶项及常数系数我们发现,

$$\begin{aligned}l_n(G;A_1)\le&(\log\kappa_1)(\log^2 n)I(\kappa_1\ge n^2M^2)\\&+n(M\kappa_1^{-\frac{1}{2}}\log\kappa_1)I(A\le\kappa_1\le n^2M^2).\end{aligned}$$

类似地,

$$\begin{aligned}l_n(G;A_2A_1^c)\le&(\log\kappa_2)(\log^2 n)I(\kappa_2\ge n^2M^2)\\&+n(M\kappa_2^{-\frac{1}{2}}\log\kappa_2)I(M_1\le\kappa_2\le n^2M^2).\end{aligned}$$

对于 $A_1\cup A_2$ 之外的观察值 θ_i, 利用公式 $\cos(2x)=1-2\sin^2x$, 以及关于 M_1 的第 3 个条件, 我们可以证明

$$\log f(\theta_i;G)\le\frac{1}{2}[\log\kappa_2-\log^2\kappa_2]\le\frac{1}{2}[\log M_1-\log^2 M_1].$$

即当 M_1 充分大时, 它们对似然函数的贡献趋于负无穷. 同时, 由于 $n((A_1\cup A_2)^c)\ge n-(n(A_1)+n(A_2))\ge\frac{1}{2}n$, 我们得到 $l_n(G;(A_1\cup A_2)^c)\le\frac{n}{4}[\log M_1-\log^2 M_1]$. 综合起来, 几乎必然地有

$$\begin{aligned}pl_n(G)&=[l_n(G;A_1)+\widetilde{p}_n(\kappa_1)]+[l_n(G;A_1^cA_2)+\widetilde{p}_n(\kappa_2)]+l_n(G;A_1^cA_2^c)\\&\le Mn\kappa_1^{-\frac{1}{2}}\log\kappa_1+Mn\kappa_1^{-\frac{1}{2}}\log\kappa_1+\frac{n}{4}[\log M_1-\log^2 M_1]\\&\le 2MnM_1^{-\frac{1}{2}}\log M_1+\frac{n}{4}[\log M_1-\log^2 M_1]\\&\le n(K_0-1).\end{aligned}$$

在上面的表达式中, 最后一步用到了关于 M_1 的第 2 个条件. 由大数定律有 $\frac{pl_n(G_0)}{n}=K_0+o(1)$. 从而, 当 n 足够大时, 几乎必然有

$$\sup pl_n(G)-pl_n(G_0)\le -n.$$

因此, (3) 在 Γ_1 情形下为真. 现在我们考虑 Γ_2 的情形. 在这种情况下, 当 κ_1 很大时, 少数几个与 μ_1 十分接近的观察值对似然函数的贡献可以很大. 但我们选取的惩罚函数可以使这些点在 κ_1 很大时的贡献变得得不偿失. 具体的证明如下. 令

$$g(\theta_i;G)=\alpha\exp\Big\{-\kappa_1\sin^2\Big(\frac{\theta_i-\mu_1}{2}\Big)\Big\}+(1-\alpha)I_0^{-1}(\kappa_2)\exp\{\kappa_2\cos(\theta_i-\mu_1)\}.$$

因 $I_0^{-1}(\kappa_1)\exp\{\kappa_1\}\le\sqrt{\kappa_1}$, 以及当 $\theta_i\in A_1^c$ 时

$$\begin{aligned}I_0^{-1}(\kappa_1)\exp\{\kappa_1\cos(\theta_i-\mu_1)\}&\le\exp\Big\{\frac{1}{2}\log\kappa_1-2\kappa_1\sin^2\Big(\frac{\theta_i-\mu_1}{2}\Big)\Big\}\\&\le\exp\Big\{-\kappa_1\sin^2\Big(\frac{\theta_i-\mu_1}{2}\Big)\Big\}\end{aligned}$$

我们有 $l_n(G)=l_n(G;A_1)+l_n(G;A_1^c)\le \frac{1}{2}n(A_1)\log\kappa_1+\sum\limits_{i=1}^{n}\log g(\theta_i;G)$.

对于任意给定的 G, 由强大数律及 Jensen 不等式，必存在 $\Delta>0$ 使得

$$n^{-1}\Big\{\sum_{i=1}^{n}\log g(\theta_i;G)-l_n(G_0)\Big\}\xrightarrow{a.s.}-\Delta. \tag{4}$$

由于 $\sup\{g(\theta;G):G\in\Gamma_2\}\le I_0^{-1}(M_1)\exp\{M_1\}$, 我们可以将 g 的定义延拓到紧致化后的 Γ_2. 从而可以证明，几乎必然地有

$$\sup\Big\{n^{-1}\sum_{i=1}^{n}\log g(\theta_i;G)-n^{-1}l_n(G_0);\ \ G\in\Gamma_2\Big\}\to-\Delta(M_2),$$

且 $\Delta(M_2)$ 是一 M_2 的单调增函数. 这样一来，我们有，在 M_2 充分大时，

$$\sup_{G\in\Gamma_2}pl_n(G)-pl_n(G_0)\le 2MnM_2^{-\frac{1}{2}}\log M_2-n\Delta(M_2)\le-\frac{1}{2}n\Delta(M_2)\to-\infty\quad \text{a.s.},$$

当 $n\to\infty$, 这就证明了 (3) 在 $j=2$ 时成立.

由于 (3) 成立，我们必然有，当 $n\to\infty$ 时， PMLE 以概率 1 属于 Γ_3. 而在 Γ_3 上的 PMLE 的强相合性可以直接应用 [15] 的结论. 这样，我们的证明便完成了.

定理 2.1 的第二部分考虑对参数空间直接加以约束，其证明的方法与第一部分十分接近. 为简略起见，我们就不重复了. 详细的证明将在第二作者的论文中给出，感兴趣的读者可以直接与第二作者联系.

INFERENCE FOR VON MISES MIXTURE IN MEAN DIRECTION AND CONCENTRATION PARAMETERS

Chen Jiahua　Li Pengfei

(*Department of Statistical and Acturial Sciences, University of Waterloo, ON, Canada* N2L 3G1)

Tan Xianming

(*LPMC and School of Mathematical Sciences, Nankai University, Tianjin* 300071)

Abstract　The finite mixtures of von Mises distributions in both mean direction and concentration parameters are widely used in many disciplines, including astronomy, biology, ecology, geology and medicine. It is well known that the likelihood function is unbounded for any sample size. Hence, the ordinary maximum likelihood estimator (MLE) is not consistent. Similar to normal mixtures in both mean and variance parameters, this drawback of MLE will disappear by introducing a penalty function to the log-likelihood function or putting constraints on component concentration parameters (Chen et al., 2006 and Tan et al., 2006). In this paper, we prove that both of the penalized maximum likelihood estimator and the constrained maximum likelihood estimator are asymptotically consistent and efficient. The finite sample performance of penalized MLE and constrained MLE are compared with the moment estimator (Spur and Koutbeiy, 1991) through simulations. The PMLE is found to have the best performance in term of mean square error. A real data example is used to illustrate the proposed methods.

Key words　Constrained maximum likelihood, mixture of von Mises distributions, penalized maximum likelihood, strong consistency.

混合von Mises模型的参数估计

作者： 陈家骅， 李鹏飞， 谭鲜明， Chen Jiahua， Li Pengfei， Tan Xianming
作者单位： 陈家骅,李鹏飞,Chen Jiahua,Li Pengfei(加拿大滑铁卢大学统计与精算科学系,加拿大,N2L 3G1)， 谭鲜明,Tan Xianming(加拿大滑铁卢大学统计与精算科学系,加拿大,N2L 3G1;南开大学数学科学学院,天津,300071)
刊名： 系统科学与数学ISTIC PKU
英文刊名： JOURNAL OF SYSTEMS SCIENCE AND MATHEMATICAL SCIENCES
年，卷(期)： 2007,27(1)

参考文献(15条)

1. Mardia K V;Jupp P E Directional Statistics 2000

2. Gould E Studies on the orientation of turtles 1959

3. Fisher N I Statistical Analysis of Circular Data 1993

4. Fraser M D;Hsu Y S;Walker J J Identifiability of finite mixtures of von Mises distributions 1981

5. Holzmann H;Munk A;Stratmann B Identifiability of finite mixtures-with applications to circular distributions 2004

6. Lindsay B G Mixture models:theory,geometry and applications 1995

7. Spurr B D;Koutbeiy M A A comparison of various methods for estimating the parameters in mixture of von Mises distributions 1991

8. Hathaway,R J A constrained formulation of maximun-likelihood estimation for normal mixture distributions 1985

9. Chen J;Tan X;Zhang R Consistency of Penalized MLE for Normal Mixtures in Mean and Variance 2006

10. Tan X;Chen J;Zhang R Consistency of the constrained maximum likelihood estimator in finite normal mixture models 2006

11. Stephens M Techniques for directional data[Technical Report 150] 1969

12. Chen J;Cheng P The limit distribution of the restricted likelihood ratio statistic for finite mixture models 1995(11)

13. Chen H;Chen J;Kalbfleisch J D A modified likelihood ratio test for homogeneity in finite mixture models 2001(01)

14. Serfling R J Approximation Theorems of Mathematical Statistics 1980

15. Wald A Note on the consistency of the maximum likelihood estimate 1949

本文读者也读过(10条)

6. 刘国光.许世刚 基于Copula方法深圳A股、B股投资组合风险值实证分析[期刊论文]-淮海工学院学报（自然科学版）2004,13(4)

7. 孙祝岭.SUN Zhu-ling 求MLE的间接方法[期刊论文]-数学的实践与认识2008,38(17)

8. 颜宁生.YAN Ning-sheng 样条指函数[期刊论文]-数学的实践与认识2005,35(11)

9. 李华民 基于copula技术的电力市场金融风险价值计算方法研究[学位论文]2007

10. 郭三刚.曹吉利.张琳.GUO San-gang.CAO Ji-li.ZHANG Lin 一种基于非均匀惩罚因子的序列无约束最优化外点新算法[期刊论文]-陕西理工学院学报（自然科学版）2008,24(3)

引证文献(4条)

1.闫宝伟,郭生练,余维 长江和清江洪水过程遭遇风险分析[期刊论文]-水力发电学报 2013(01)

2.闫宝伟,郭生练,余维 长江和清江洪水过程遭遇风险分析[期刊论文]-水力发电学报 2013(01)

3.李伟 Copula函数在多变量洪水联合分布中的应用研究[学位论文]硕士 2013

4.郑志安 县域农业产业链发展中政府行为研究[学位论文]博士 2007