

KMM 算法在星系团研究中的应用

闫鹏飞¹, 丁蓉², 袁启荣³

(1. 青岛科技大学数理系 266042 山东 青岛)

(2. 南京师范大学学报编辑部 210097 江苏 南京)

(3. 南京师范大学物理科学与技术学院 210097 江苏 南京)

[摘要] 介绍了 KMM 算法基本原理, 并将此算法运用于近邻双星系团系统 A399/A401. KMM 算法对该系统中的 215 颗星系进行了成员判定, 其中 127 颗星系属于 A401, 88 颗星系属于 A399. 实践表明, KMM 算法在研究星系团动力学演化方面很有效.

[关键词] KMM 算法, 星系团, 成员星系

[中图分类号] P14, [文献标识码] A, [文章编号] 1001-4616(2004)04-0039-04

0 引言

作为宇宙中最大的引力束缚体系, 星系团的形成和演化引起了人们广泛的研究兴趣. 对星系团结构及其动力学的研究就是其中的一个重要方面, 它有助于理解星系团的形成和演化过程. N 体模型模拟表明, 星系团内部的子结构可能形成于星系维里化之前^[1, 2]. 对于具有子结构的星系团, 区分星系究竟属于哪一个子星系团就显的尤为重要.

一种被称为 KMM 的混合模型算法近期被引入星系团的子结构研究中^[3], 该算法在星系团子结构成员判定方面很有效. 本文将简要介绍 KMM 算法, 并首次将此算法运用到一个正在进行相互作用的双星系团系统中, 对该区域中的星系成员进行判定.

1 KMM 算法简介

作为混合模型算法之一, KMM 算法被广泛应用于天体物理研究(对银河系不同星族的恒星计数研究^[4, 5]). KMM 算法用于一维的天文学数据, 研究河外球状星团体系的金属丰度、颜色的分布情况等^[6, 7]; 用于多维的天文数据, 根据星系团的速度和位置数据来确定子结构, 作为进一步动力学分析的依据^[3].

KMM 算法尝试以最大似然法将数据分配到指定的 g 个 K 维子系统中^[7]. 这 g 个子系统被假定是 K 维高斯分布的, 每一维可能有不同的标度长度. 这种算法还对采用多子系统拟合相对于单系统拟合的优化程度作出了估计. 同时, 它也对 g 个子系统的中心和尺度估计进行了优化.

KMM 算法假定 N 个观测数据 $x_j (j = 1, 2, \dots, N)$ 都是从概率密度函数为高斯函数的子系统中抽取出来的. 混合密度可以表示为

$$f(x; \phi) = \sum_{i=1}^g \pi_i f_i(x; \theta_i),$$

其中 g 是需要拟合的子系统的总数目, π_i 是各子系统的混合比, θ_i 是代表 g 个子系统中所有未知参数的矢量, 且 $\phi = (\pi', \theta')$ 是未知模型参数和混合比的矢量转置矩阵. KMM 算法先得到最佳的单高斯拟合(它可以明确的得到数据的平均值 μ 和方差 σ^2). 单高斯分布的概率定义为

$$L_c(1) = \sum_{j=1}^N \ln f(x_j; \theta),$$

即在 x_j 处求得的单变量高斯值的自然对数的总和.

收稿日期: 2004-09-09.

基金项目: 国家自然科学基金资助项目(10273007).

作者简介: 闫鹏飞, 1979 - , 青岛科技大学数理系, 主要从事天体物理的教学与研究. E-mail: ypfyan0822@sina.com

通讯联系人: 袁启荣, 1968 - , 南京师范大学物理科学与技术学院副教授, 主要从事天体物理的教学与研究.

E-mail: yuanyiqirong@njnu.edu.cn
万方数据

使用者需要首先输入要拟合的子系统数目以及各高斯分布的矢量平均值及方差的初值. 接着, KMM 算法运用最初给定的或者在计算中产生的 ϕ 值进入迭代步骤. 在每一轮计算中, KMM 算法将得到每一个数据点属于各个子系统的概率, 这些概率将形成该数据被分配到任一子系统的置信度估计. 数据点将被分配到成员概率最高的那个子系统. 分配结束后, KMM 算法通过计算每一个子系统的 π_i , μ_i 和 σ_i^2 的值从而得到新的矢量 ϕ . 这样, g 个子系统拟合的概率就是:

$$L_c(\phi) = \sum_{i=1}^g \sum_{j=1}^N z_{ij} [\ln \pi_i + \ln f_i(x_j; \theta)].$$

这里, z_{ij} 是一个指示变量, 如果对象 j 属于 i 子系统, $z_{ij} = 1$, 否则 $z_{ij} = 0$. 根据新的矢量 ϕ , KMM 算法继续计算每一个数据点属于各个子系统的概率, 从而得到新的分配方案. 这一迭代程序直到 $L_c(\phi)$ 已经收敛 (即对象分配趋于稳定) 方才结束. 假设所有的子系统都具有相同的协方差矩阵, 上面得到的概率序列通常是有界的, 这种算法在 ϕ 参数空间会收敛到一个局部的最大值. 但在异方差情况下, KMM 算法得到的概率序列可能就不是有界的.

$L_c(\phi)$ 的最终数值可用来确定数据是否与零假设 (即单系统假设) 相一致. 引入概率比值检验的统计量

$$\lambda = \frac{L_c(\phi)}{L_c(1)}$$

来对单系统模型到多子系统模型优化程度进行估计. 统计量 λ 的显著性水平可以通过比较 $-2\ln\lambda$ 和 χ^2 分布 (自由度数是两种比较模型中不包括混合比在内的参数数目差值的两倍) 得到, 但这只是统计量显著性的一个近似估计.

KMM 算法运用上述分析方法, 得到 g 子系统拟合优于单系统情况的统计显著性估计, 该项显著性水平是用 $P(\lambda \geq \lambda_{\text{obs}})$ (即从单高斯分布中抽取的一个样本分布中对应的 λ 值大于或等于 λ 观测值的概率) 的形式来表示的. 如果零假定 (单系统假设) 是正确的, 那么 P 值至少应该大于 0.1. P 值越小, 表明零假定越不能够很好地描述数据. 习惯上, $P < 0.05$ 表示数据分布与零假设基本不符, 而 $0.05 < P < 0.1$ 则表明数据与零假设有微小的不一致性.

另外, KMM 算法还给出了每个子系统的平均值 μ_i , 混合比 π_i 和共同的协方差 σ^2 的估计值 (在异方差的情况下, 每个子系统的协方差是分别计算给出的). 该算法还将每个数据点分配到成员概率最高的子系统中, 并给出了分配到各个子系统的置信度. 选择最佳的拟合模型是以拟合的显著性水平、正确分配率以及拟合所得到的高斯分布的平均值及其离散的稳定性为标准的. 这样, KMM 算法不仅提供了一种强有力的假设检验工具, 而且在指定的 g 子系统拟合中, 客观地将数据点分配到各个子系统中.

2 KMM 算法应用实例

KMM 算法最适用于从理论上或经验上证实了高斯模型合理性的情形. 对于一个维里化了的星系团系统, 引力相互作用将使其空间分布和视向速度分布趋于高斯分布. 所以, KMM 算法可以引入星系团的动力学研究中.

2.1 数据

邻近星系团系统 A401 和 A399 很长一段时间以来就被认为是一对星系团, 并且可能已经有了相互作用^[8]. 我们的研究基于从美国的河外星系数据库 (NASA/IPAC Extragalactic Database) 中得到这一天区的星系观测数据. 在该星系团系统中心 (赤经 $2^{\text{h}}58.5^{\text{m}}$, 赤纬 $+13^{\circ}20'$) 附近 $100'$ 的范围内共有 1 331 颗星系, 其中仅 240 颗星系具有已知的光谱红移值, 其视向速度处于 $18\,000\text{ km s}^{-1} < cz < 25\,000\text{ km s}^{-1}$ 的有 215 颗. 这 215 颗星系应该属于这个双星系团系统. 图 1 是这 215 颗星系的

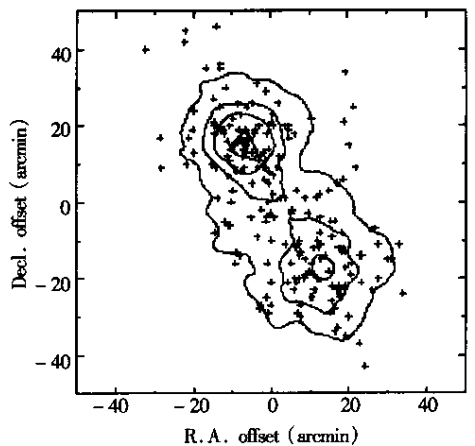


图 1 215 颗星系的空间等强度分布图

空间分布和面密度等强度图. 从图中, 我们可以看出, 星系整体上很明显的分成了两
— 40 — 万方数据

团.由于 A401 和 A399 的视向速度很接近,因此从 215 颗星系中区分属于两团的星系要困难些.下面我们将 KMM 算法用于三维数据——星系在天空中的投影位置(x,y)和视向速度(v),来判断星系究竟属于哪一个星系团.

我们首先输入各星系团的起始星系分配表和起始参数,通过迭代运算,得到最终的星系分配表和拟合参数.表 1 列出了我们所用的起始参数和星系分配以及通过 KMM 算法相应得到的最终结果(x_1,y_1,v_1)和(x_2,y_2,v_2)分别是 A401 和 A399 星系团整体的平均位置和速度, $\sigma(x_1,y_1,v_1)$ 和 $\sigma(x_2,y_2,v_2)$ 分别是相应的标准偏差, f_1 和 f_2 是两团统计样本所占的百分比,表 1 同时还给出了 KMM 算法对整体星系分配方案的置信度估计.

表 1 应用于 KMM 算法的起始输入参数及得到的结果

编号	(x_1,y_1,v_1)	$\sigma(x_1,y_1,v_1)$	(x_2,y_2,v_2)	$\sigma(x_2,y_2,v_2)$	(f_1,f_2)	置信度/%
输入参数						
1	(5.9,16.9,22133)	(9.8,9.9,1208)	(-11.4,-15.7,21536)	(9.8,9.5,1227)	(0.521,0.479)	
2	(4.9,14.9,22099)	(10.2,11.0,1189)	(-12.7,-17.9,21491)	(9.2,8.0,1254)	(0.586,0.414)	
3	(4.9,14.7,22098)	(10.1,11.2,1185)	(-12.8,-18.0,21486)	(9.1,8.0,1261)	(0.591,0.409)	
4	(5.0,16.8,22126)	(10.5,9.7,1204)	(-10.7,-16.2,21532)	(10.4,8.9,1232)	(0.53,0.47)	
5	(5.8,18.6,22164)	(10.3,8.7,1233)	(-9.8,-14.3,21561)	(10.6,10.1,1201)	(0.474,0.526)	
6	(4.0,14.6,22080)	(10.7,11.1,1212)	(-11.7,-18.3,21505)	(10.3,7.7,1234)	(0.595,0.405)	
输出参数						
1	(5.0,14.8,22107)	(10.1,11.5,1188)	(-12.4,-17.0,21495)	(9.2,9.0,1239)	(0.575,0.425)	95.1
2	(4.9,14.5,22107)	(10.2,11.7,1185)	(-12.6,-17.3,21479)	(9.1,8.8,1241)	(0.586,0.414)	95.2
3	(4.8,14.4,22107)	(10.2,11.7,1185)	(-12.6,-17.4,21477)	(9.1,8.8,1241)	(0.587,0.413)	95.2
4	(5.0,14.8,22107)	(10.1,11.5,1188)	(-12.4,-17.0,21494)	(9.2,9.0,1239)	(0.576,0.424)	95.1
5	(5.0,14.8,22107)	(10.1,11.5,1188)	(-12.4,-17.0,21495)	(9.2,9.0,1239)	(0.575,0.425)	95.1
6	(4.9,14.5,22107)	(10.2,11.7,1186)	(-12.6,-17.3,21481)	(9.2,8.9,1240)	(0.585,0.415)	95.2

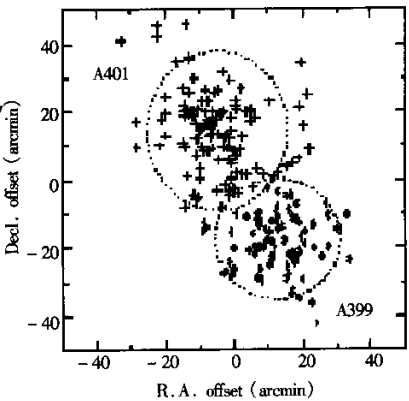
我们粗略估计出每个星系团的中心位置、视向速度及方差矩阵(两个团的方差矩阵相同),并作为输入参数,由 KMM 算法自动判定星系的归属.第 2 轮计算的输入参数是以第 1 轮计算所得到的星系分配表作为输入参数,其输出的星系分配表再作为第 3 轮计算的输入参数.我们发现,第 3 轮计算所输出的星系分配表和第 2 轮的完全相同,这表明 KMM 算法收敛的较好,同时该分配是较合理的.

然后,我们以星系团系统中心位置的赤纬为分界线,即把 $y > 0$ arcmin 的星系作为星系团 A401 的成员,其余的属于 A399,将这种星系分配表作为第四轮计算的输入参数.另外,我们还尝试了其它的星系分配方案,第 5、6 轮计算是分别把 $y = 5.0、-5.0$ arcmin 作为星系团的分界线,将所对应的星系分配表作为 KMM 程序的输入参数.随后,我们将以上 3 轮计算的输出结果进行迭代运算,所得的星系分配表和第 3 轮计算的结果是完全一致的,并且收敛得较快.图 2 是对应于第 3 轮计算所输入的星系分配表的星系空间分布图.

在上述列出的 6 轮计算中,只有第 1 轮计算输入同方差参数,其余 5 轮均是输入异方差参数.

3 结论

从以上的应用实例可以看出,KMM 算法在 A399/A401 系统的星系成员判定方面是非常有效的,它是区分星系团体系中子团星系的强有力的工具.我们发现,在选择同方差的情况下,虽然起始输入的未知参数不同,但是最后得到的结果都是完全一致的.另外,我们对异方差输入的情况下 KMM 算法的收敛性质进行了研究,发现输入不合理的初始参数,可能得不到收敛的结果.总体上说,KMM 算法的收敛速度和初始星系分配方案的合理性密切相关,也和子结构的位置和视向速度的弥散度初始估计有关.



点线是 2σ 误差椭圆

点线是 2σ 误差椭圆

图 2 由 KMM 算法得到的两团的空间分布图

[参考文献]

[1] White S D M. The dynamics of rich clusters of galaxies[J]. Monthly Notes of Royal Astron Society ,1976 ,177 :717—733.
[2] Burns J O ,Roettiger K ,Ledlow M , *et al.* The coma cluster after lunch : Has a galaxy group passed through the cluster core[J]. Astrophys Journal ,1994 ,427 :187.
[3] Colless Matthew , Dunn Andrew M. Structure and Dynamics of the Coma Cluste[J]. Astrophys Journal ,1996 ,458 :435—454.
[4] Ashman Keith A , Bird Christina M , Zepf Stephen E. Detecting bimodality in astronomical dataset[J]. Astronomical Journal ,1994 ,108(6) :2348—2361.
[5] Nemec J M , Nemec A F L. Mixture models for studying stellar populations. I-Univariate mixture models , parameter estimation , and the number of discrete population components[J]. Publications of Astronomical Society of Pacific ,1991 ,103 :95—121.
[6] Ostrov Pablo , Geisler Doug , Forte Juan C. The metallicity gradient and distribution function of globular clusters around NGC 1399 [J]. Astronomical Journal ,1993 ,105(5) :1762—1778.
[7] Lee Myung G , Geisler Doug. Metal abundances for a large sample of globular clusters in M87[J]. Astronomical Journal ,1993 ,106(2) :493—506.
[8] Fabian A C , Nulsen P E J , Canizares C R. Cooling flows in clusters of galaxies[J]. Nature ,1984 ,310 :733—740.

Application of the KMM Algorithm on the Studies of Galaxy Clusters

Yan Pengfei¹ ,Ding Rong² ,Yuan Qirong³

(1.Department of Mathematics and Physics , Qingdao University of Science and Technology , 266042 , Qingdao , China)
(2.Editorial Board of Journal of Nanjing Normal University , 210097 , Nanjing , China)
(3.School of Physical Science and Technology , Nanjing Normal University , 210097 , Nanjing , China)

Abstract :The principal of the KMM algorithm was presented and applied to a nearby pair of galaxy clusters A399/A401 . The membership for 215 galaxies in this system has been determined by the KMM algorithm . As a result , 127 galaxies are found to belong to A401 , and 88 galaxies are in A399 . It shows that the KMM algorithm is effective in the investigation of the dynamical evolution of galaxy clusters .

Key words :KMM algorithm , cluster of galaxies , member galaxies

[责任编辑 :丁蓉]

(上接第 35 页)

[参考文献]

[1] 徐承绪 ,关虹. 自正交拉丁方存在性的一个简短证明[J]. 应用数学学报 ,1993 (2) :185—190.
[2] 徐承绪 ,卢准炜. Pandiagonal magic squares[J]. Lecture Notes in Computer Science ,1995 (959) :388—391.
[3] Hudson C B. On Pandiagonal magic squares of order $6t \pm 1$ [J]. Math Mag ,1972 (45) :94—96.
[4] Maurice Kraitichik. Mathematical Recreations[M]. New York : Dover Publications , INC ,175.
[5] Hedayat A. A comlete solution to the existence and nonexistence of Knut Vik designs and orthogonal Knut Vik designs[J]. J Combinatorial Theory(A) ,1977 (22) :331—337.

The Existence of Pandiagonal Magic Squares

Xu Chengxu , Lu Zhunwei

(Taiyuan University of Technology , 030024 , Taiyuan , China)

Abstract :This paper also gives the definition of quadratic integer squares alone with the product between them and replace of proving the existence of pandiagonal magic squares of orders mn by proving the existence of pandiagonal magic squares of orders m and n respectively . Then constructs the family of pandiagonal magic squares of all orders $n = 4 \times 2^k$ and $n = 12 \times 2^k$ ($k \geq 0$). Again , using the product of quadratic integer squares , constructs the family of pandiagonal magic squares of all orders $n \neq 2, 3, 4t + 2$. It is well known that the magic square of order $n = 2$ does not exist and magic square of order $n = 3$ is unique and not a pandiagonal magic square . Note that Mr. Raynor has shown nonexistence of pandiagonal magic square of orders $n = 4t + 2$. Therefore , the existence problem of pandiagonal magic squares is completely solved .

Key words :integer square , quadratic integer square , product of quadratic integer square , self orthogonal pandiagonal integer square , pandiagonal magic square

[责任编辑 :陆炳新]