

多阶段混合增长模型的影响因素：距离与形态*

刘源^{1,2} 骆方¹ 刘红云¹

(¹北京师范大学心理学院应用实验心理北京市重点实验室, 北京 100875)

(²香港中文大学教育心理系, 香港)

摘要 通过模拟研究, 考察潜类别距离和发展形态等因素对多阶段混合增长模型的模型选择和参数估计的影响: (1)潜类别距离越大, 模型选择和分类效果越好。(2)混合模型的选择, 应以一定样本量(至少 200)为前提, 首先考虑 BIC 选出正确的分类模型, 再通过熵值、ARI 等选择分类确定性较高的模型。(3)多阶段的发展形态对正确模型的选择和分类的确定性均有一定程度影响。(4)潜类别距离和样本量越大, 参数估计精度越高。(5)在判断分类准确性的指标中, ARI 的选择更偏向于真实的模型。

关键词 多阶段混合增长模型(PGMM); 潜类别增长分析(LCGA); 潜类别距离(SMD); 发展形态
分类号 B841

1 引言

纵向研究(Longitudinal analysis)对于个体发展规律的描述、解释、预测都有重要的意义, 是对个体或群体特质长时间观察多次记录的研究范式。与横断研究相比, 追踪研究更耗时、费用不菲, 但是它可以对变量之间的因果关系做出合理的假设, 也可以就某一现象随时间的变化规律进行分析。

1.1 多阶段混合增长模型概要

在纵向研究中, 研究者可以用潜变量增长模型(Latent growth modeling, LGM)来描述群体的发展趋势以及个体差异。该模型假设观测群体同质和发展趋势连续。如果存在发展趋势不同质的多个群体, 也就是存在潜类别、未观测到的异质性(unobserved heterogeneous), 抑或发展趋势不连续(不能用一个连续函数表达), 即存在多个发展阶段, 这时一般的 LGM 不再适用。多阶段混合增长模型(Piecewise growth mixture modeling, PGMM) (Li, Duncan, Duncan, & Hops, 2001)是可以同时解决这两个问题的一种分析方法(图 1)。

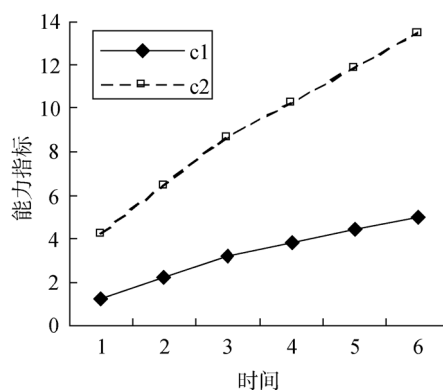


图 1 多阶段混合增长模型(PGMM)示意图

在图 1 中, 假设有 6 次观测, 存在两个潜在类别(c1 和 c2), 每个类别都有两个发展阶段, 转折点在时间点 t_3 处。一个 PGMM 模型可以如下定义:

$$\mathbf{y}_i^{(k)} = \mathbf{\Lambda}^{(k)} \boldsymbol{\eta}_i^{(k)} + \boldsymbol{\varepsilon}_i^{(k)} \quad (1)$$

$$\boldsymbol{\eta}_i^{(k)} = \boldsymbol{\mu}_{\boldsymbol{\eta}}^{(k)} + \boldsymbol{\zeta}_i^{(k)} \quad (2)$$

在方程 1 和方程 2 中, $\mathbf{y}_i^{(k)}$ 表示第 k 类被试 i 的观测值(6×1 向量), $\boldsymbol{\eta}_i^{(k)}$ 表示第 k 类被试 i 的潜在增长因子(这里潜在增长因子分别为截距、斜率 1 和斜率 2, 故为 3×1 向量), $\mathbf{\Lambda}^{(k)}$ 是描述第 k 类增长

收稿日期: 2013-08-07

* 国家自然科学基金(31100759)、全国教育科学“十二五”规划教育部重点课题(GFA111001)、教育部人文社会科学研究青年基金项目(11YJC190016)、北京市与中央在京高校共建项目(019-105812)资助。

通讯作者: 刘红云, E-mail: hylu@bnu.edu.cn

刘源, 骆方为共同第一作者。

模式的因子系数(6×3 矩阵), $\mu_{\eta}^{(k)}$ 表示第 k 类潜在增长因子的均值(3×1 向量)。随机部分 $\epsilon_i^{(k)}$ 表示第 k 类被试 i 的测量误差(6×1 向量), 对应的协方差矩阵为 $\Theta^{(k)}$, $\zeta_i^{(k)}$ 表示第 k 类被试 i 潜在增长因子的残差(3×1 向量), 即第 k 类别中个体间的差异, 对应的协方差矩阵为 $\Psi^{(k)}$ 。

特别地, 在 PGMM 中, 研究者通过 $\Lambda^{(k)}$ 来定义发展阶段。假设 t_3 是增长变化的转折点, 截距的载荷都固定为 1; t_1 到 t_3 的增长率为斜率 1, 故 t_3 之后的载荷设定为相等, 即 0, 1, 2, 2, 2, 2, 这时斜率 1 就作用于 t_1 到 t_3 的阶段; t_3 到 t_6 的增长率为斜率 2, 故 t_3 之前的载荷设定为相等, 即 2, 2, 2, 3, 4, 5, 这时斜率 2 就作用于 t_3 到 t_6 的阶段。二者的叠加就构成了 PGMM 模型特殊的定义时间的方式(公式 3)。

$$\Lambda^{(k)} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 2 & 5 \end{pmatrix} \quad (3)$$

在实际应用中, 研究者一般还会使用更为简单的潜类别增长分析(Latent class growth analysis, LCGA) (Nagin, 1999), 或者也叫做潜类别增长模型(Latent class growth modeling, LCGM)探讨存在不同发展潜在类别的问题。它同 PGMM 最主要的区别在于是否允许同一类别内的个体间存在变异。PGMM 允许同类内的个体间存在变异, 可以通过估计随机部分 $\zeta_i^{(k)}$ 来描述之; 而 LCGA 假设更为严格, 它要求类别内不存在个体差异, 即 $\zeta_i^{(k)}$ 被限定为零。

1.2 多阶段混合增长模型的参数估计和应用

PGMM 是混合增长模型(Growth mixture modeling, GMM) (B. Muthén & Asparouhov, 2007; B. Muthén et al., 2002)和多阶段增长模型(Piecewise growth modeling, PGM) (Duncan, Strycker, & Duncan, 1999; Raudenbush & Bryk, 2002)的综合模型。在参数估计上, PGMM 会受到两种模型参数估计各自问题的影响。在 PGM 中, 斜率的变化影响到模型的参数估计精度、检验力和第一类错误率, 斜率变化越大, 对于模型选择的影响也越大(刘源, 赵骞, 刘红云, 2013)。而 GMM 中, 潜在类别、测量信度等也会对模型估计精度产生影响, 潜类别距离影响分类的准确性: 距离越近, 其分类结果越差(Tolvanen, 2007)。此外, Steinley 和 Brusco (2011a, 2011b, 2011c)

还比较了聚类分析(K-means Clustering)和混合模型(Mixture Model)在潜在分类上的差异, 并探讨了潜类别间的重叠性对分类结果的影响。

在实际应用中, PGM 用于解决发展阶段不连续的问题, 对于发展速度的突变, 可以通过定义不同阶段的时间函数来描述; GMM 用于解决发展趋势不同质的问题, 对发展群体的潜在类别进行估计; 而 PGMM 可以同时解决发展趋势具有转折点和总体不同质的问题, 在能力发展、临床、社会行为等研究领域应用广泛(Kamata, Nese, Patarapichayatham, & Lai, 2012; Li et al., 2001; McAuley et al., 2011; B. Muthén & Brown, 2009; Uher et al., 2010)。Li 等人(2001)采用 PGMM 来研究青少年酗酒行为的发展形态, 他们定义了两个潜在类别, 两个类别的转折点不同(分别在三年级和四年级), 该研究为 PGMM 的适用性提供了实践证据。B. Muthén 和 Brown (2009)将 PGMM 应用到安慰剂的研究中。他们发现, 安慰剂组在用药之后, 患者的病情都出现了缓解(转折); 并且在安慰剂组内发现了两类群体: 反应组的趋势有明显的下降, 而非反应组这种趋势相对较弱。由此可以诊断出哪些被试在服用安慰剂后表现出了药物反应。

LCGA 是限定潜类别内增长因子方差为零的 PGMM, 故可以看成是 PGMM 的一个特例。一般地, LCGA 假设观测变量是类别变量, 但是由于统计软件(如 Mplus)的发展, LCGA 放宽了观测变量为类别变量的假设, 只需限定 PGMM 随机变异为零就可实现 LCGA 的定义(L. Muthén & Muthén, 2010)。此外, LCGA 的最大优点是估计简单, 通常将 LCGA 得到的估计值作为 PGMM 估计的初始值(Jung & Wickrama, 2008)。由于 LCGA 是一个简化模型, 不考虑类别内的个体差异, 分类精度比较高, 所以得到很多研究者的青睐, 尤其在青少年犯罪、酗酒等行为研究(Delucchi, Matzger, & Weisner, 2004; Nagin & Land, 1993)和发展领域(Luyckx, Schwartz, Goossens, Soenens, & Beyers, 2008)当中应用广泛。

具体应用中, 研究者一般都采用探索性的方法对实际数据进行潜在类的分析(Feldman, Masyn, & Conger, 2009; B. Muthén & Asparouhov, 2007; Reinecke & Seddig, 2011), 即同时使用 PGMM 和 LCGA 对数据进行拟合, 并且从 1 个类别开始, 逐步增加类别, 以“数据说话”代替繁琐的理论辩解, 通过鉴定指标来判断最终模型应该取多少个类别

(B. Muthén, Muthén, & Asparouhov, 2012)。大多数研究者偏向使用 BIC (Bayesian information criterion), LRT (Likelihood ratio test) 以及熵值 (entropy) 来判断类别的拟合情况和模型的分类情况 (Celeux & Soromenho, 1996; Feldman et al., 2009; Nagin, 1999; Nylund, Asparouhov, & Muthén, 2007)。然而在实际应用中, 强调类别的实际意义也是研究者应该重点考察和关心的问题 (B. Muthén et al., 2012; Vermunt, 2010)。

2 研究目的

在实际研究中, 研究者可以将 LCGA 看成混合模型的一种特殊的定义方式, 即不考虑 PGMM 中的个体差异。所以, 尽管二者可以看成是同一种模型, 但是限定的参数个数不同, 其估计结果也会有一定差异。其实, 类别判定的准确性与选择使用的模型有很大关系, LCGA 与 PGMM 的前提假设不同, 它们可能适用于不同类型的纵向数据。这两种定义方式在整体拟合、参数估计上是否相差甚远, 什么情况下可以简化模型, 类别距离和发展形态对这两种定义适用性的影响有多大, 目前尚缺乏相关研究的佐证。

另外, 由于 PGMM 的引入基于 GMM 和 PGM, 所以 GMM 与 PGM 的影响因素也会对 PGMM 的估计造成影响, 这一问题目前也没有进一步的研究。

最重要的, 由于多阶段模型的特殊性, 对于 PGMM 的定义会变得更为复杂。PGMM 中, 由于存在阶段性, 斜率会被叠加, 对于相同的潜在距离, 也会有不同的斜率组合。不同的发展形态 (即不同的斜率参数组合) 对 PGMM 估计结果的影响尚不清楚。所以, 为了讨论上述的问题, 本研究考虑采用模拟数据的方法, 使用软件 Mplus 6.0 (L. Muthén & Muthén, 2010), 探索在何种条件下, PGMM 和其简化模型可以得到相似的结果, 以及潜类别距离和发展形态等因素对模型参数估计的影响, 并尝试给出一般性的建议, 怎样选择适当的模型。

3 研究方法

3.1 模型定义

采用 Mplus 6.0 中 Monte Carlo 模块模拟生成数据。模型定义了两个潜在类别的两阶段模型, 共 6

次观测 (y_1 - y_6), 转折点在 y_3 处。本研究中限定两个类别的 Λ 、 Θ 和 Ψ 相等, 即 $\Lambda^{(1)} = \Lambda^{(2)} = \Lambda$, $\Theta^{(1)} = \Theta^{(2)} = \Theta$, $\Psi^{(1)} = \Psi^{(2)} = \Psi$, 但是在潜变量增长因子参数上存在差异, 即 $\mu_{\eta}^{(1)} \neq \mu_{\eta}^{(2)}$ 。

6 次测量误差矩阵 Θ 满足均值为 0, 方差都为 0.2 (公式 4, Θ 服从于均值为 0, 方差为对角矩阵的多元正态分布), 潜变量残差协方差矩阵 Ψ 设定为对角矩阵, 截距和两阶段斜率的方差分别为 1, 0.5, 0.2 (公式 5, Ψ 服从于均值为 0, 方差为对角矩阵的多元正态分布)。两个类别所占样本量的比率相同, 每种模拟条件下重复次数 (R) 为 500 次。

$$\Theta \sim N(0, \text{diag}[0.2, 0.2, 0.2, 0.2, 0.2, 0.2]) \quad (4)$$

$$\Psi \sim N(0, \text{diag}[1, 0.5, 0.2]) \quad (5)$$

3.2 模拟设计

(1) 潜在距离

一般地, 研究者使用马氏距离平方根 (Square root of the Mahalanobis distance, SMD) 作为潜类别距离的指标, 其计算公式如下:

$$\text{SMD} = \text{SMD}(\eta) = \sqrt{(\mu_{\eta}^{(1)} - \mu_{\eta}^{(2)})^T \Psi^{-1} (\mu_{\eta}^{(1)} - \mu_{\eta}^{(2)})} \quad (6)$$

其中 $\mu_{\eta}^{(k)}$ ($k = 1, 2$) 为潜变量向量的均值。本研究中分别取 SMD 的值为 1.5, 3 和 5, 代表类之间的距离小、中和大三个水平。

(2) 发展形态

针对多阶段模型 (本研究考虑两阶段), 不同类别在不同阶段的发展趋势会导致类别的发展形态不同。由于鲜有人涉足多阶段的形态问题, 这里参考 Tolvanen (2007) 在线性模型下的形态设定, 根据多阶段的特点, 定义如下 4 种发展形态 (见图 2):

① 平行形态: 初始状态存在差异, 但是两阶段的斜率均相同;

② 非平行形态 a: 初始状态存在差异, 第一阶段斜率相同, 第二阶段斜率不同;

③ 非平行形态 b: 初始状态存在差异, 第一阶段斜率不同, 第二阶段斜率相同;

④ 非平行形态 c: 初始状态存在差异, 两阶段斜率也均不同。

针对每种 SMD 的取值, 在上述 4 种形态下分别定义潜变量均值 (斜率和截距) 的真值 (如表 1), 再根据模型和参数真值生成数据。

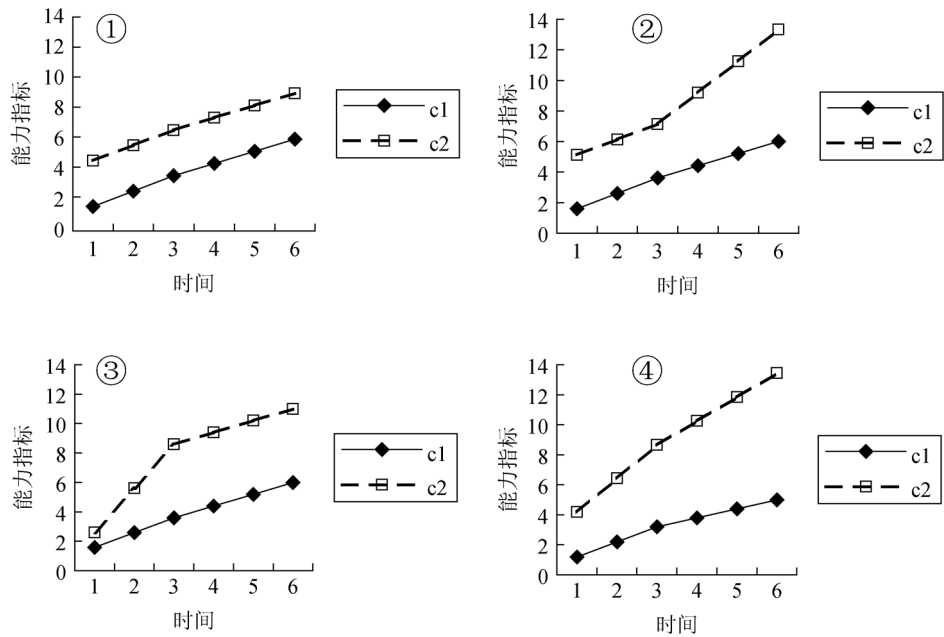


图 2 模拟研究中 4 种不同 PGMM 形态(示 SMD = 3)

表 1 不同形态下潜变量均值的真值

发展形态	参数	SMD = 5		SMD = 3		SMD = 1.5	
		c1	c2	c1	c2	c1	c2
①	截距	0.00	5.00	0.00	3.00	0.00	1.50
	斜率 1	1.00	1.00	1.00	1.00	1.00	1.00
	斜率 2	0.80	0.80	0.80	0.80	0.80	0.80
②	截距	0.00	1.00	0.00	1.00	0.00	1.00
	斜率 1	1.00	1.00	1.00	1.00	1.00	1.00
	斜率 2	0.80	2.99	0.80	2.07	0.80	1.30
③	截距	0.00	1.00	0.00	1.00	0.00	1.00
	斜率 1	1.00	4.46	1.00	3.00	1.00	1.79
	斜率 2	0.80	0.80	0.80	0.80	0.80	0.80
④	截距	0.00	1.00	0.00	1.00	0.00	1.00
	斜率 1	1.00	2.41	1.00	2.23	1.00	1.73
	斜率 2	0.60	2.60	0.60	1.60	0.60	0.80

根据上述条件，可以根据公式 7 计算出对应的观测变量的距离 SMD (y) (表 2 列出了给定 SMD 对应的 SMD (y)值)。

$$SMD(y)=\sqrt{\left(\mu_y^{(1)}-\mu_y^{(2)}\right)^T \Sigma^{-1}\left(\mu_y^{(1)}-\mu_y^{(2)}\right)} \quad (7)$$

其中， $\mu_y^{(k)}$ ($k = 1,2$)表示观测向量的均值， Σ 表示潜变量协方差矩阵 Ψ 对应的观测变量协方差矩阵，它和 Ψ 关系满足：

$$\begin{aligned} \Sigma &= \Lambda \Psi \Lambda^T + \Theta \\ \mu_y^{(k)} &= \Lambda \mu_{\eta}^{(k)} \end{aligned} \quad (8)$$

(3)样本量

根据前人的研究(Zhang & Willson, 2006)，建

议增长模型的最小样本量不低于 50。考虑到潜在类别，本研究的总样本量取 100、200 和 500 三个水平，分别作为小样本、中等样本和大样本，两个潜在类别的样本量相等。

表 2 四种形态下潜变量距离 SMD 对应的观测变量距离 SMD (y)

发展形态	SMD = 5	SMD = 3	SMD = 1.5
①	4.47	2.68	1.34
②	4.74	2.87	1.45
③	4.68	2.82	1.40
④	4.89	2.98	1.46

3.3 考察指标

(1) 潜类别模型比较

按照 Nylund 等人(2007)提出的关于混合模型的检验方法, 首先对模型选择进行比较。参考的指标包含以下几种:

a. 模型整体拟合。采用 BIC 指标作为整体拟合的检验。 L 为似然函数值, $\ln L$ 表示求似然函数值的自然对数, P 为参数个数, N 为样本量。BIC 的计算公式如下:

$$\text{BIC} = -2 \ln L + P \ln(N) \quad (9)$$

b. 熵值(entropy)。熵值是一个概率值, 与分到最可能的类别的后验概率相似, 它描述了正确分类概率的情况。其取值范围为 0~1, 0 表示分到每个类别的后验概率相同, 1 则表示分到某一个特点类别的后验概率为 1(Celeux & Soromenho, 1996; Feldman et al., 2009; Nagin, 1999)。熵值越大表示正确分类的确定性越高。公式 10 中, \hat{p}_{ik} 表示估计的个体 i 被分到类别 k 中的后验概率, K 为类别个数。

$$\text{entropy}_k = \sum_{k=1}^K \sum_{i=1}^n \hat{p}_{ik} \ln(\hat{p}_{ik}) \quad (10)$$

c. 总命中率。总命中率从分类个数出发, 通过频次统计来描述分类结果和实际数据之间的差异。根据信号检测论的原理, 计算出包含命中率和正确拒绝率的总命中率。

d. ARI (Adjusted Rand index)。根据许多学者的讨论(Steinley, 2004; Steinley & Brusco, 2011a, 2011b; Vermunt, 2011), ARI 可以考察分出的类别是否能返回到原有的类别, 定义为:

ARI =

$$\frac{\sum_{k \in T} \sum_{k \in U} \binom{N_{kh}}{2} - \left[\sum_{k \in T} \binom{N_{k+}}{2} \sum_{k \in U} \binom{N_{+h}}{2} \right] / \binom{N}{2}}{\frac{1}{2} \left[\sum_{k \in T} \binom{N_{k+}}{2} + \sum_{k \in U} \binom{N_{+h}}{2} \right] - \left[\sum_{k \in T} \binom{N_{k+}}{2} \sum_{k \in U} \binom{N_{+h}}{2} \right] / \binom{N}{2}} \quad (11)$$

其中, $\binom{N}{2}$ 表示在 N 里面选出 2 个的组合数量, 即 $\frac{N(N-1)}{2}$ 。 $k \in T$ 表示被分类的个体 k 原有的类别, $k \in U$ 表示被分类后个体 k 属于的类别。本研究中使用 R 软件中的 clues 程序包计算 ARI 值(使用 Rand 结果)。

(2) 参数估计精度

在选定了正确模型的基础上, 进一步比较参数估计精度, 分别从偏差(bias)、误差均方根(RMSE)

和 95%置信区间覆盖比率(CP)三个方面进行解释。其计算方法分别如下:

$$\text{bias} = \frac{1}{R} \sum_{r=1}^R (\hat{\tau}_{pr} - \tau_p) \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\tau}_{pr} - \tau_p)^2} \quad (13)$$

$$\text{CP} = \frac{1}{R} \sum_{r=1}^R \text{CI}_{95}(\hat{\tau}_{pr}) \quad (14)$$

其中, R 表示重复总次数($R = 500$), r 为重复次数($r = 1, 2, \dots, R$)。 $\hat{\tau}_{pr}$ 为第 r 次运算参数 p 的估计值, τ_p 为参数 p 的真值(见表 1), $\hat{\sigma}_{pr}$ 为第 r 次运算参数 p 的估计标准误, 满足:

$$\begin{aligned} &\text{如果 } \hat{\tau}_{pr} \in (\tau_p \pm 1.96 \hat{\sigma}_{pr}), \text{ 那么 } \text{CI}_{95}(\hat{\tau}_{pr}) = 1 \\ &\text{否则 } \text{CI}_{95}(\hat{\tau}_{pr}) = 0 \end{aligned} \quad (15)$$

3.4 补充设计

当混合增长模型中存在三类及其以上的潜类别时, 在同一个模型中, 类别与类别之间的距离的大小存在差异(如第一类与第二类、第二类与第三类距离较小, 而第一类和第三类距离较大), 此时, 分析结果及其所受距离和发展模式的影响是否与两个潜在类别分析的结论一致? 为了验证结论的可推广性, 增加了一个补充设计。

对于有 3 个类别的情况, 设定 c1 和 c2 之间距离 $\text{SMD} = 3$, c2 和 c3 之间距离 $\text{SMD} = 3$, 故 c1 和 c3 之间的距离 $\text{SMD} = 6$ 。考虑发展形态完全平行形态和完全不平行形态两种情况。3 个类别的潜变量参数设定见表 3。

表 3 三类别潜变量均值的真值

参数	平行			不平行		
	c1	c2	c3	c1	c2	c3
截距	0.00	3.00	6.00	0.00	1.00	2.00
斜率 1	1.00	1.00	1.00	1.00	2.23	3.77
斜率 2	0.80	0.80	0.80	0.60	1.60	2.40

每个类别样本量为 100, 共计 $N = 300$, 其余条件与之前的研究相同。

3.5 分析方法

本研究针对上文定义生成的模拟数据, 采用 6 个观测变量、第三次测量为转折点的多阶段形式的模型, 分别使用含有方差部分的 PGMM 和简化的 LCGA (即公式 5 中设定为零矩阵)两种不同定义, 比较两种定义下模型的差异。用 Mplus 6.0 进行分析。

4 研究结果

4.1 潜类别模型比较和选择

4.1.1 距离对分类个数的影响

对于 PGMM 中潜类别个数的确定, 一般认为是一个探索性过程, 在此要求研究者逐步增加类别个数来观察整体拟合的变化, 以确定最佳类别的选择。所以, 基于模型整体拟合和分类结果(熵值), 分别对 1 个类别到 3 个类别的模型进行比较。同时, 因为 LCGA 简化了模型的复杂程度, 只关注于类别而不探究其类别内的变异, 实际应用中研究者常常将 LCGA 与 PGMM 进行比较, 以下也给出了 LCGA 得到的结果。

不同类别数目的 PGMM 结果(表 4)表明, 在 SMD 为 5 时, 2 类别模型的 BIC 选择概率最高。但是在 SMD 为 3 时, 样本量对 BIC 的选择有影响: 样本量在 200 或以上的 2 类别 BIC 被选择的概率最

高。但是在 SMD = 1.5 时, 一个类别的 BIC 显示为最佳, 说明该情况下模型倾向于不分类。

简化模型 LCGA 与 PGMM 的比较结果(表 5)表明, 在所有的模拟条件下, LCGA 比 PGMM 整体拟合(BIC)差。但是, 从熵值看出, 总的来讲, LCGA 比 PGMM 高。SMD = 5 时 PGMM 结果良好(>0.9), 但是在 SMD = 3 时, 其熵值均分布在 0.6~0.8; 当 SMD 减小到 1.5 时, 熵值下降到 0.4~0.7。说明潜在距离的大小对 PGMM 分类确定性的影响很大。而 LCGA 始终保持在 0.88 以上。一般认为, 正确分类需要达到 80% 才被认为良好分类(Celeux & Soromenho, 1996), 但从模拟研究来看, 该标准并没有达到。说明潜在类别距离对分类确定性的影响很大, 熵值在距离较小的时候偏向于 LCGA。

对于发展形态, 整体拟合几乎不受其影响(表 4), 而熵值受形态影响较大(表 5)。首先, LCGA 的

表 4 PGMM 不同类别模型的整体拟合 BIC 的选择概率(%)

潜在距离	样本量	①*			②			③			④		
		c = 1	c = 2	c = 3	c = 1	c = 2	c = 3	c = 1	c = 2	c = 3	c = 1	c = 2	c = 3
SMD = 5	N = 100	1	99	0	0	99	1	0	99	1	0	99	1
	N = 200	0	100	0	0	100	0	0	100	0	0	100	0
	N = 500	0	100	0	0	100	0	0	100	0	0	100	0
SMD = 3	N = 100	82	18	0	68	31	0	68	32	0	60	40	0
	N = 200	58	42	0	28	58	13	42	58	0	20	80	0
	N = 500	5	95	0	0	100	0	0	86	14	0	100	0
SMD = 1.5	N = 100	99	1	0	99	1	0	99	1	0	98	2	0
	N = 200	100	0	0	99	1	0	100	0	0	100	0	0
	N = 500	79	0	21	100	0	0	100	0	0	100	0	0

*注: ①、②、③和④分别表示 4 种形态: 平行形态、非平行形态 a~c, 下同。

表 5 PGMM 与 LCGA (C = 2)模型选择比较(BIC 和熵)

		SMD = 5				SMD = 3				SMD = 1.5			
		①	②	③	④	①	②	③	④	①	②	③	④
BIC													
PGMM	<i>N</i> = 100	1713	1715	1715	1716	1676	1683	1680	1686	1616	1622	1619	1621
	<i>N</i> = 200	3377	3381	3380	3382	3303	3316	3312	3323	3185	3195	3190	3196
	<i>N</i> = 500	8340	8350	8347	8354	8156	8188	8179	8205	7863	7888	7878	7890
LCGA	<i>N</i> = 100	2619	2695	2665	2719	2451	2569	2523	2625	2301	2376	2349	2389
	<i>N</i> = 200	5211	5359	5302	5408	4878	5112	5022	5222	4580	4728	4674	4756
	<i>N</i> = 500	12961	13331	13189	13452	12131	12714	12489	12987	11389	11757	11624	11827
熵													
PGMM	<i>N</i> = 100	0.94	0.97	0.99	0.98	0.64	0.82	0.77	0.77	0.58	0.61	0.61	0.62
	<i>N</i> = 200	0.92	0.95	0.97	0.96	0.70	0.63	0.81	0.66	0.54	0.71	0.58	0.55
	<i>N</i> = 500	0.93	0.96	0.97	0.97	0.71	0.76	0.77	0.76	0.49	0.48	0.47	0.48
LCGA	<i>N</i> = 100	0.95	0.99	0.98	1.00	0.94	0.94	0.94	0.97	0.96	0.93	0.94	0.93
	<i>N</i> = 200	0.94	0.99	0.97	1.00	0.90	0.93	0.91	0.97	0.90	0.88	0.89	0.89
	<i>N</i> = 500	0.96	0.98	0.97	0.99	0.94	0.95	0.95	0.97	0.92	0.93	0.92	0.94

熵值不受形态影响,保持在 0.88 以上;其次,PGMM 在距离较大的时候不受其影响,而在距离小时($SMD = 1.5$)受到形态影响较大,平行形态①的熵值普遍低于其他非平行形态②~④)。

4.1.2 分类准确性的判定

对于混合模型,分类的准确性是评价模型结果的很重要的指标,下面从总命中率和ARI的角度对比较模型所分类别与真实类别的一致性,表6呈现了正确的两类别PGMM和简化的两类别LCGA的总命中率和ARI的结果。

从表6的结果可知,总命中率和ARI有着较高的相关:对于PGMM,ARI与总命中率的相关为0.988,LCGA为0.790。从总命中率来看,其趋势和熵值较为接近。总命中率随着潜在距离的增大而增高, $SMD = 3$ 或更大时,其总命中率在80%以上;而在 $SMD = 1.5$ 的情况下,总命中率均在70%左右。

对ARI的考察发现,在 $SMD = 5$ 的时候,两种模型的ARI结果良好,并且可以看出PGMM优于LCGA; $SMD = 3$ 的时候,整体在0.8左右波动,但是当 $SMD = 1.5$ 的时候,其结果几乎都集中在0.6或以下。并且,在距离较小时,PGMM与LCGA的判断并没有给出明显的偏向,两种模型的选择概率相当。由此可知,当潜类别距离较大时,分类结果的返回效果很好,特别是 $SMD \geq 3$ 的时候,而且样本量越大,返回效果越好。但是整体而言,ARI的选择,特别是对于 $SMD = 1.5$ 的时候,并未像熵值或总命中率一样明显偏向LCGA。

对于发展形态对分类确定性的影响,和前文提到的熵值类似。不论PGMM或LCGA,平行形态下的总命中率和ARI均明显地低于其他非平行形态的趋势。

4.2 参数估计精度

4.2.1 PGMM 参数估计精度

参数估计偏差和RMSE的结果相类似,故在此仅列出RMSE的部分结果(图3)。对于PGMM,总体而言,不同潜类别距离(SMD)也影响RMSE, SMD 越大,RMSE越小,说明随着潜类别距离的增大,估计精度变高。在同一 SMD 的不同类别形态中,相差并不大。此外,随着样本量的增大,RMSE的值减小,说明估计精度随样本量的增加而增高。

95%CP结果如图4。总体而言, SMD 较大($SMD = 3$ 和5)时,CP覆盖比率都在90%以上。但是在 $SMD = 1.5$ 时,CP明显比其他情况差,且4种形态之间也有差别:平行形态的CP高于其他非平行形态,但是这种情况在 SMD 较大时并没有出现。此外,随着样本量的增大,CP也增大。

4.2.2 PGMM 与 LCGA 参数估计精度比较

特别地,鉴于LCGA分类结果良好,但研究者并不能判定其分类与真值之间的差别。故进一步对其进行参数精度的探讨。从表7可知,LCGA的RMSE结果都不如PGMM稳健,受到形态影响很大。特别地,在 $SMD = 3$ 或1.5的情况下,发展形态影响LCGA参数估计,而距离大的时候参数估计比较好。进一步地,当两类潜变量真值相同或相近

表6 PGMM 与 LCGA (C = 2)分类结果比较(总命中率和ARI)

		SMD = 5				SMD = 3				SMD = 1.5			
		①	②	③	④	①	②	③	④	①	②	③	④
总命中率													
PGMM	$N = 100$	0.99	0.99	0.99	0.99	0.85	0.89	0.83	0.92	0.59	0.65	0.67	0.68
	$N = 200$	0.99	0.99	0.99	0.99	0.89	0.91	0.88	0.93	0.59	0.67	0.68	0.70
	$N = 500$	0.99	0.99	0.99	0.99	0.91	0.92	0.92	0.93	0.60	0.73	0.72	0.73
LCGA	$N = 100$	0.91	0.97	0.95	0.99	0.78	0.88	0.84	0.93	0.65	0.73	0.71	0.75
	$N = 200$	0.91	0.97	0.95	0.99	0.78	0.89	0.85	0.93	0.65	0.73	0.71	0.75
	$N = 500$	0.91	0.99	0.95	0.99	0.78	0.89	0.85	0.89	0.65	0.73	0.71	0.75
ARI													
PGMM	$N = 100$	0.97	0.98	0.98	0.98	0.79	0.81	0.79	0.86	0.53	0.56	0.57	0.58
	$N = 200$	0.97	0.97	0.98	0.99	0.82	0.84	0.83	0.87	0.53	0.57	0.58	0.59
	$N = 500$	0.98	0.98	0.98	0.95	0.84	0.86	0.85	0.87	0.54	0.61	0.60	0.61
LCGA	$N = 100$	0.83	0.95	0.90	0.98	0.66	0.79	0.74	0.86	0.54	0.61	0.59	0.63
	$N = 200$	0.84	0.95	0.90	0.98	0.66	0.80	0.74	0.80	0.54	0.61	0.59	0.63
	$N = 500$	0.84	0.90	0.90	0.98	0.66	0.80	0.74	0.84	0.55	0.61	0.59	0.63

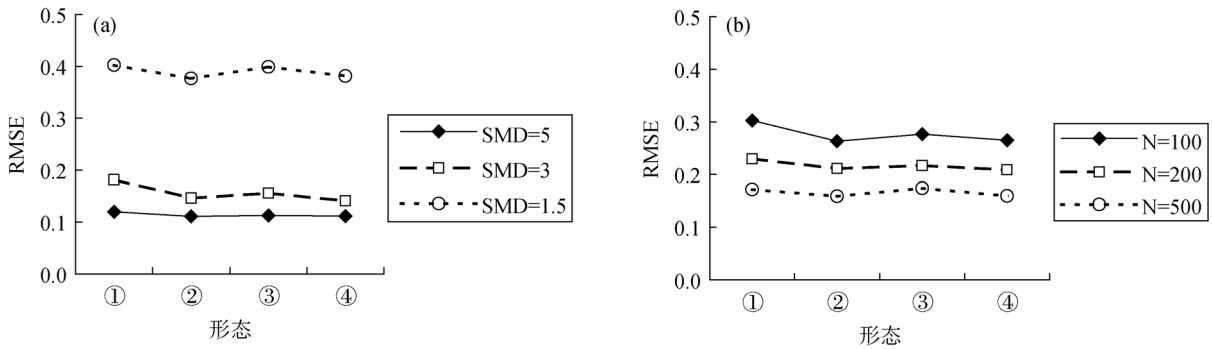


图 3 PGMM (C = 2)估计精度(RMSE)结果示意(a 表示 SMD, b 表示样本量)

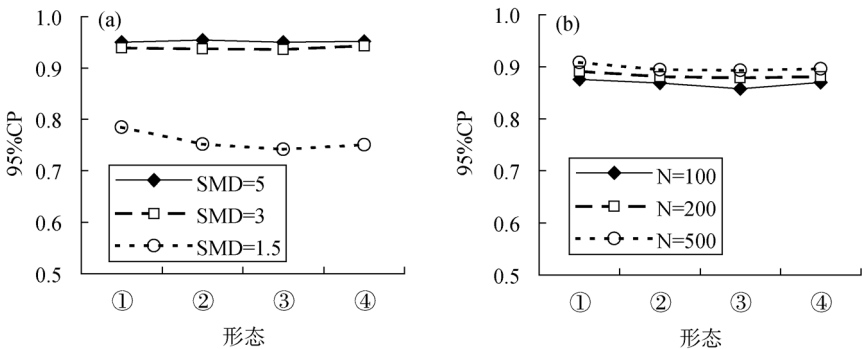


图 4 PGMM (C = 2)估计精度(95%CP)结果示意(a 表示 SMD, b 表示样本量)

表 7 PGMM 与 LCGA (C = 2)参数估计精度比较(RMSE 和 95%CP, N = 500)

参数		SMD = 5				SMD = 3				SMD = 1.5			
		①	②	③	④	①	②	③	④	①	②	③	④
RMSE													
PGMM	截距	0.08	0.07	0.07	0.07	0.10	0.09	0.09	0.08	0.34	0.32	0.37	0.33
	斜率 1	0.05	0.05	0.05	0.05	0.06	0.06	0.07	0.06	0.23	0.21	0.22	0.23
	斜率 2	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.14	0.14	0.14	0.14
LCGA	截距	0.30	0.08	0.09	0.07	0.36	0.12	0.12	0.10	0.16	0.12	0.12	0.13
	斜率 1	0.13	0.07	0.11	0.05	0.23	0.15	0.14	0.05	0.29	0.25	0.07	0.08
	斜率 2	0.10	0.04	0.07	0.03	0.18	0.05	0.11	0.04	0.22	0.07	0.20	0.14
95%CP													
PGMM	截距	0.95	0.96	0.96	0.96	0.95	0.95	0.96	0.96	0.83	0.78	0.76	0.78
	斜率 1	0.96	0.95	0.96	0.96	0.96	0.96	0.95	0.96	0.72	0.73	0.79	0.78
	斜率 2	0.95	0.96	0.95	0.95	0.95	0.95	0.93	0.95	0.75	0.79	0.76	0.75
LCGA	截距	0.48	0.92	0.90	0.94	0.26	0.75	0.78	0.83	0.71	0.82	0.85	0.81
	斜率 1	0.32	0.85	0.78	0.95	0.00	0.16	0.66	0.96	0.00	0.00	0.92	0.84
	斜率 2	0.14	0.91	0.53	0.94	0.00	0.90	0.25	0.94	0.00	0.67	0.00	0.03

时, CP 不足 50%。由此可见, 对于潜类别距离较小的模型, 尽管 LCGA 能正确分类, 但其参数估计精度较差, 对于不同类别形态的估计有偏。特别是两个类别的真值相同或相近的参数, 其 95%CP 均在 0 附近, 说明其参数估计效果不高。

4.3 三类别模型结果

对于三类别多阶段混合增长模型的分类准确性, 平行发展模式得到的总命中率和 ARI 的值分别为 0.772 和 0.761, 非平行发展模式得到的总命中率和 ARI 的值分别为 0.824 和 0.802。与前面两个类

别的结论一致,平行形态下的总命中率和 ARI 均低于非平行形态的趋势。表 8 给出了三类别情况下模型分类的结果。

表 8 三类别模型两两比较的分类概率

实际组	判别组					
	平行			不平行		
	c1	c2	c3	c1	c2	c3
c1	0.86	0.12	0.02	0.84	0.14	0.03
c2	0.15	0.68	0.17	0.15	0.75	0.09
c3	0.00	0.08	0.77	0.01	0.04	0.88

注:对角线元素(加粗)为同一个类别的正确分类概率,非对角线元素为误判概率;左下角与右上角的元素(斜体)为距离较远的两类(SMD = 6)误判概率,其余单元格为距离较近的两类(SMD = 3)误判概率。

从表 8 的结果可以看出,类别 c1 和 c3 的命中率比类别 c2 高,这是因为类别 c2 在中间,与类别 1 和类别 3 的距离均为 3,有近似相同的机会分到两边的类别里去。类别 c1 误判为类别 c2 的概率比其误判为类别 c3 的概率高,类别 c3 误判为类别 c2 的概率比其误判为类别 c1 的概率高,说明类别间距离越近,误判的概率越高,距离越远,误判的概率越低。这与两类别情境下得到的判断准确性与距离关系的结论一致。

5 讨论

本文基于 PGMM 模型,重点讨论了潜在类别间的距离对模型选择和参数估计的影响;同时,结合多阶段模型的特点,首次探讨了不同发展形态对模型参数估计的影响;另外,尝试使用 ARI 统计指标验证了分类模型的效果。本研究的结果可以为实际应用者提供一些模型选择和应用上的建议。

5.1 两难问题:如何选择模型

混合模型分析中,模型选择是一个非常复杂的问题,很多研究者建议 BIC 是选择模型首要考虑的指标(B. Muthén & Asparouhov, 2007; Nylund et al., 2007; Vermunt & Magidson, 2004),但是从本研究结果看,即使在模拟情景下,仅根据 BIC 指标并不一定能够选出正确的模型。BIC 这一指标受潜类别距离的影响较大,当潜类别距离比较小时,根据 BIC 的结果可能并不能有效判断出总体中存在不同发展类别的结果。本研究的结果表明,BIC 会容易受到样本量的影响,这一结果与 Nylund 等人(2007)的研究结论一致。这就告诉实际应用者,采用 BIC

指标选择模型,应保证比较大的样本量($N \geq 200$)。另外,在 PGMM 模型中,对于中等的类别间距离(SMD = 3)平行发展的模式类别,其 BIC 指标更容易受到小样本的影响。

本研究的结果表明,在类别距离很小的时候,PGMM 的分类效果较差(熵值 <0.6),其鉴别力的效度也存在争议(Lubke, 2012)。这表明,在允许组内存在变异的情况下,如果类别间距离较小,对个体做出类别判断的确定性很低,根据已有的信息并不能很好判断某个个体属于哪个类别,这种情况在平行形态的时候更为明显。与此相对应,对于 PGMM 模型,样本量越大,熵值提示的分类结果越差,从这个角度而言,熵值并不是一个十分可靠的判断标准。但是值得注意的是,对于简化的 LCGA 模型,熵值受潜类别距离和样本量的影响较小,如果仅仅根据熵值的大小选择模型,那么则会更加倾向于选择简化的模型。

关于分类正确性的指标,本研究除了常用的命中率,还对 ARI 做了一些探索。从 ARI 与总命中率的相关系数可以看出,对于真实模型(PGMM),二者之间达到了高相关;而对于 LCGA,尽管相关也达到了 0.790,但相比于真实模型,二者的选择偏好略有不同。结果显示 ARI 比较偏向于真实的模型,特别是在 SMD = 1.5 的情况下,ARI 并不像熵值或者总命中率那样明显偏向 LCGA。其对于真实模型的还原程度比其他的指数更高,在 Steinley 和 Brusco (2011b)的研究中就使用了 ARI 去判断采用 BIC 分类结果和 CH 指数分类结果的好坏。Chiang 和 Mirkin (2007)也采用 ARI 来判断各种分类指数分类结果返还到真实类别的程度。由此可知在研究者面临多个选择指标的时候,ARI 也是对分类指标的考察,而非仅仅对数据的考察。此外,根据对 LCGA 参数估计精度的探讨发现,LCGA 尽管有较高的熵值和总命中率,但是偏差、RMSE 和覆盖区间都较差,这也从另一个侧面为 ARI 指数选择 PGMM 而非 LCGA 提供了证据。总体来看,ARI 是一个较为优良的指标,其可以为最后模型的选择提供更有说服力的依据。

值得一提的是,本研究结果似乎单方面支持 PGMM 而非 LCGA,这主要是由于数据生成的“真”模型导致的,即本研究中假设潜在类别内个体之间存在随机变异,而 LCGA 模型限定了潜类别内个体间不存在变异。由于这种限制,使得 LCGA 更加简单,在实际研究中,如果潜类别内个体间的随机变

异很小, 则可以考虑使用之。另外, LCGA 模型在实际中还有一些特殊的应用。例如对于类别观测变量(二分变量、三分变量)或计数数据, 则应使用 LCGA。针对非连续型数据, LCGA 可以推广到二项分布、泊松分布、零膨胀泊松分布(Zero-Inflated Poisson distribution)等情形, 其应用价值不容忽视。有关 LCGA 的更多应用可以参考 L. Muthén 和 Muthén (2010)的操作手册。

综上所述, 很难根据某个单一的指标做出混合模型类别数的正确判断。在实际应用中, 研究者需要慎重考虑模型的选择问题。因为涉及到取舍的问题: 如果选择较好的分类结果, 研究者可能需要在模型整体拟合上做出让步; 但是如果研究者严格按照模型拟合来选定模型, 可能会出现分类不确定的情况。结合模拟研究分类准确性的指标, LCGA 模型高的熵值并不一定带来了正确的分类结果, 如潜类别距离为 1.5 时, LCGA 的熵值在 0.9 左右, 但是其对应的 ARI 却仅在 0.6 左右; PGMM 低的熵值对应于差的分类结果, 高的熵值对应于好的分类结果, 从这一角度, 正确模型的熵值与模型分类结果的准确性更为一致。在实际应用中, 对于混合模型类别数的选择, 首先还是应该考虑 BIC 的选择; 其次, 对于差异不大的 BIC 值, 要更加关注不同类别理论上的意义和解释, 应该在同时考虑模型整体拟合和理论上不同类别意义可解释的前提下, 选择熵值高的模型。

5.2 潜类别距离和发展形态对模型估计结果的影响

不难理解, 潜类别间的距离对混合模型的类别数和参数估计的精度都有很大的影响, 潜类别间的距离越小, 混合模型分类结果变差, 进而模型中参数估计的精度也变差。与前人研究一致的是, Tolvanen 建议一般在 $SMD = 3$ 的时候, 模型就能正确分类(Tolvanen, 2007)。从本研究的结果看来, 如果需要正确分类, $SMD \geq 3$ 是一个必要条件。同时参数估计精度的结果也表明, 潜类别间的距离不小于 3 时, 才有可能得到相对精确的参数估计结果。

关于形态对结果的影响, 本研究发现 PGMM 的结果不受其影响; 然而 LCGA 的一部分结果却大相径庭, 主要体现在参数估计精度, 特别是 95%CP 极大地受到形态的影响。在 LCGA 中, 对于非平行形态, 参数估计精度和 PGMM 相差无几; 但是对于平行形态, 参数估计精度明显不如 PGMM。特别地, 从 95%CP 的结果部分可知, 平行形态部分的

值都在 50%以下, 甚至有 0 出现。这说明, 对于平行形态的 LCGA, 估计出的模型几乎完全不能找到真实的位置在什么地方。对于这种结果的出现, 推测和 LCGA 模型本身的限定有关。由于 LCGA 并不考虑组内变异, 其假设一个组内的个体不存在个体差异, 所以在 $SMD = 1.5$ 的时候, 其图像本身就很难辨识出两个不同的组别。特别是对于平行形态而言, 两个组的个体交织在一起, 而且发展趋势相同, 在这种情况下 LCGA 就很难再对其进行准确的分类, 所以参数精度比较低。而对于非平行的发展形态, 尽管在发展初期其交织在一起, 但是随着时间变化, 其各自的趋势得以彰显。故对于非平行形态, 参数可以被正确的估计。此外, 由于 LCGA 没有考虑变异部分, 这就使得置信区间的范围也变得特别窄, 一旦有些许分类不准确, 其覆盖区间也会接近于 0。综上所述, 如果研究者偏向于使用 LCGA, 那么数据本身则需要有较高的辨识度, 非平行形态的结果较好, 但是整体而言, LCGA 的结果并没有 PGMM 的稳健。

5.3 研究展望

本研究首次探讨了潜类别间的距离和发展形态对 PGMM 模型选择和参数估计的影响, 但是尚有一些问题值得进一步探讨。

本研究发现, 潜在距离是影响模型选择和参数估计的重要因素, 但是潜在距离本身并不是一个单一的指标。 SMD 只是从马氏距离的角度综合考察潜类别距离的一个指标。不难看出, 对于潜类别间的距离, 除了潜变量因素的真值, 随机部分的方差 Ψ 也是 SMD 所考虑的部分。回归到原始数据上, 可以用类别的重叠(overlap)来描述之。本研究中只针对公式 5 的情况做了模拟。为了进一步探讨随机部分方差的影响, 作为初步尝试, 我们将随机部分设定为 $\Psi \sim N(0, \text{diag}[0.5, 0.25, 0.1])$, 并在 $SMD = 5$, 样本量为 500 的条件下进行模拟。从结果来看, 如果随机部分方差变小, 模型拟合和分类结果均会变好; 但是随机部分的变化并不影响选择正确的模型。由此推断随机部分的大小会影响模型的整体拟合, 但是并没有影响正确模型的选择。当然, 这只是一个初步的探索。对于重叠的问题, 可能还有其他的一些研究思路(McLachlan, 2011; Steinley & Brusco, 2011a, 2011b, 2011c; Vermunt, 2011), 有待更多的研究进行讨论。

本研究虽然探讨了潜类别间距离大小对分类准确性和参数估计结果的影响, 但是值得注意的是

本研究的结论在实际研究中很难直接使用。虽然根据公式 6~8, 研究者可以计算出潜变量的距离和观测变量的距离, 但是距离的计算更多是一个事后的过程, 对于一组实际的数据, 我们并无法“事先”知道这些潜在类。因此, 在实际应用中保守的方法是先对数据进行预估计, 从预估计的结果中检查得到的类与类之间距离的大小, 再根据距离的大小评判是否该方法合适。由于马氏距离不受到量尺变化的影响, 所以本研究的结论可以应用于实际数据的分析, 帮助研究者判断类与类之间距离的大小是否适合采用混合模型, 分类的结果可靠性如何。

此外, 模型中待估参数的个数也可能会影响模型的选择和模型估计的结果, 本研究中没有考虑模型中待估参数个数多少的影响, 在后续研究中有必要进一步考察模型的复杂性对混合模型分类结果和参数估计的影响。

6 总结与建议

本研究基于模拟研究的范式, 对 PGMM 中潜类别距离和发展形态对模型选择和参数估计的影响进行探讨, 提出以下几点结论和建议。

(1) 潜类别距离影响模型选择和分类效果。潜类别间距离较大时, BIC、熵值表现出一致性, 均能选出正确的模型, 得到正确的分类结果; 但当潜类别间的距离很小($SMD = 1.5$), 不分类的模型能够比分类模型更好拟合数据, 即便使用 PGMM 强行分类, 分出来的类别也可能是错误的。

(2) 关于混合模型的选择, 研究者应充分考虑模型的拟合与分类结果确定性之间可能存在的相悖关系, 在满足一定样本量(至少 200)的前提下, 首先考虑 BIC 指标选出正确的分类模型, 再通过熵值、ARI 等选择分类确定性较高的模型。

(3) 发展形态对正确模型的选择和分类的确定性均有一定程度的影响, 潜类别间距离中等($SMD = 3$), 样本量中等时($N = 200$), 平行发展形态正确模型选择的比例和熵值均低于非平行形态。形态的差异对参数估计结果的影响相对较小。

(4) 模型的参数估计精度受到类别间距离和样本量的影响, 类别间距离越大, 参数估计精度越高; 样本量越大, 参数估计精度越高。

(5) ARI 是一个较为优良的指标, 不仅和总命中率高度相关, 且采用 ARI 的模型选择更偏向于真实的模型。

参 考 文 献

- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2), 195–212.
- Chiang, M. M.-T., & Mirkin, B. (2007). Experiments for the number of clusters in K-means. In *Progress in Artificial Intelligence* (pp. 395–405). Springer Berlin Heidelberg.
- Delucchi, K. L., Matzger, H., & Weisner, C. (2004). Dependent and problem drinking over 5 years: A latent class growth analysis. *Drug and Alcohol Dependence*, 74(3), 235–244.
- Duncan, S. C., Strycker, L. A., & Duncan, T. E. (1999). Exploring associations in developmental trends of adolescent substance use and risky sexual behavior in a high-risk population. *Journal of Behavioral Medicine*, 22(1), 21–34.
- Feldman, B. J., Masyn, K. E., & Conger, R. D. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology*, 45(3), 652–676.
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302–317.
- Kamata, A., Nese, J. F. T., Patarapichayatham, C., & Lai, C.-F. (2012). Modeling nonlinear growth with three data points: Illustration with benchmarking data. *Assessment for Effective Intervention*, 38(2), 105–116.
- Li, F., Duncan, T. E., Duncan, S. C., & Hops, H. (2001). Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), 175–204.
- Liu, Y., Zhao, Q., & Liu, H. (2013). Methods comparison of piecewise growth modeling. *Psychological Exploration*, 33(5), 415–422.
- [刘源, 赵骞, 刘红云. (2013). 多阶段增长模型的方法比较. *心理学探新*, 33(5), 415–422.]
- Lubke, G. (2012). Old issues in a new jacket: Power and validation in the context of mixture modeling. *Measurement: Interdisciplinary Research and Perspectives*, 10(4), 212–216.
- Luyckx, K., Schwartz, S. J., Goossens, L., Soenens, B., & Beyers, W. (2008). Developmental typologies of identity formation and adjustment in female emerging adults: A latent class growth analysis approach. *Journal of Research on Adolescence*, 18(4), 595–619.
- McAuley, E., Mailey, E. L., Mullen, S. P., Szabo, A. N., Wójcicki, T. R., White, S. M., ... Kramer, A. F. (2011). Growth trajectories of exercise self-efficacy in older adults: Influence of measures and initial status. *Health Psychology*, 30(1), 75–83.
- McLachlan, G. J. (2011). Commentary on Steinley and Brusco (2011): Recommendations and cautions. *Psychological Methods*, 16(1), 80–81.
- Muthén, B., & Asparouhov, T. (2007). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Advances in Longitudinal Data Analysis* (pp. 143–165). Raton, FL: Chapman & Hall/CRC Press.
- Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., ... Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3(4), 459–475.
- Muthén, B., & Brown, H. C. (2009). Estimating drug effects in

- the presence of placebo response: Causal inference using growth mixture modeling. *Statistics in Medicine*, 28(27), 3363–3385.
- Muthén, B., Muthén, L., & Asparouhov, T. (2012). *Latent variable modeling using Mplus*. Beijing: National Survey Research Center at Renmin University of China.
- Muthén, L., & Muthén, B. (2010). Mplus (Version 6.0) [Computer software]. Los Angeles, CA: Muthén & Muthén.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2), 139–157.
- Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model. *Criminology*, 31(3), 327–362.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed). Thousand Oaks, California: Sage Publications.
- Reinecke, J., & Seddig, D. (2011). Growth mixture models in longitudinal research. *AStA Advances in Statistical Analysis*, 95, 415–434.
- Steinley, D. (2004). Properties of the Hubert-Arabie Adjusted Rand index. *Psychological Methods*, 9(3), 386–396.
- Steinley, D., & Brusco, M. J. (2011a). Choosing the number of clusters in K-means clustering. *Psychological Methods*, 16(3), 285–297.
- Steinley, D., & Brusco, M. J. (2011b). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, 16(1), 63–79.
- Steinley, D., & Brusco, M. J. (2011c). K-means clustering and mixture model clustering: Reply to McLachlan (2011) and Vermunt (2011). *Psychological Methods*, 16(1), 89–92.
- Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study* (Unpublished doctoral dissertation), University of Jyväskylä.
- Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., ... McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine*, 40(8), 1367–1377.
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469.
- Vermunt, J. K. (2011). K-means may perform as well as mixture model clustering but may also be much worse: Comment on Steinley and Brusco (2011). *Psychological Methods*, 16(1), 82–88.
- Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. In A. Bryman, T. Liao, & M. Lewis-Beck (Eds.), *The Sage encyclopedia of social sciences research methods* (pp. 549–553). California: Sage Publications.
- Zhang, D., & Willson, V. L. (2006). Comparing empirical power of multilevel structural equation models and hierarchical linear models: Understanding cross-level interactions. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(4), 615–630.

Factors of Piecewise Growth Mixture Model: Distance and Pattern

LIU Yuan^{1,2}; LUO Fang¹; LIU Hongyun¹

(¹ Beijing Key Laboratory of Applied Experimental Psychology; School of Psychology, Beijing Normal University, Beijing 100875, China)

(² Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong, China)

Abstract

The piecewise growth mixture model (PGMM) has been a very popular analytical approach in recent studies of longitudinal data. PGMM builds on the piecewise growth model (PGM) and the growth mixture model (GMM). It is used to locate the turning point of growth trajectory as well as to identify the latent class of the population. It is particularly useful in detecting the non-continuous growing trend in a heterogeneous population. A simplified version of the model, the latent class growth analysis (LCGA), has also been often used with a restriction on the variance of PGMM.

Understandably, factors affecting PGM and GMM will affect the estimates and performance of PGMM. These factors may include the change of the slope, the distance of latent classes, and the sample size. PGMM being developed from the two growth-related models (PGM, GMM) also attempts to analyze the growth pattern in latent growth trajectory as a special and newly emerged issue. Even for models with the same distance, their different slopes can be combined to form different patterns. This issue has not been fully explored in previous literature. Yet in empirical studies, factors such as the distance of the latent classes, the growth pattern, the existing criteria of model fit indices, and the precision of parameter estimates are well worth examining issues.

In the present simulation study, a two-class-two-period model was adopted. The three simulation conditions

being considered were: the sample size, the distance of latent class, and the pattern of the growth trajectory. The sample size was set to be 100, 200 and 500. The distance of the latent classes was defined as the squared Mahalanobis distance (SMD), with 1.5, 3 and 5 being used to represent the small, medium and large distance of latent classes respectively. Four different types of growth pattern were selected to represent one parallel and three non-parallel patterns. Finally, the LCGA was selected as the reference model to see whether PGMM could be further simplified or not.

The results showed that: (1) the distance between the latent classes (SMD) was a crucial factor that influenced the model selection and parameter estimation. Large distance would lead to consistent BIC and entropy when the right models were selected; while small distance ($SMD = 1.5$) would not. (2) When mixture modeling was taken into consideration, it was suggested that a sample size of at least 200 should be used. BIC index should be the preferred choice to be used for model selection; the entropy, ARI and other indices were also recommended to further reference. (3) The pattern of the growth trajectory would affect model selection; specifically, non-parallel patterns of the trajectory would help model selection (higher entropy and higher total hit ratio) for medium distance ($SMD = 3$) and medium sample size ($N = 200$) conditions. However, as compared to LCGA, the pattern of the growth trajectory had little influence on PGMM. (4) Parameter estimation was affected by the sample size and distance of latent classes. Parameter estimates would become more precise as the sample size and the distance increased. (5) ARI was a reasonably good index belonging to the recovery indices family. ARI was highly correlated with the total hit ratio and thus would lead to recommendations of models closer to the true model.

Key words piecewise growth mixture modeling (PGMM); latent class growth analysis (LCGA); distance of latent classes (SMD); pattern of the growth trajectory