

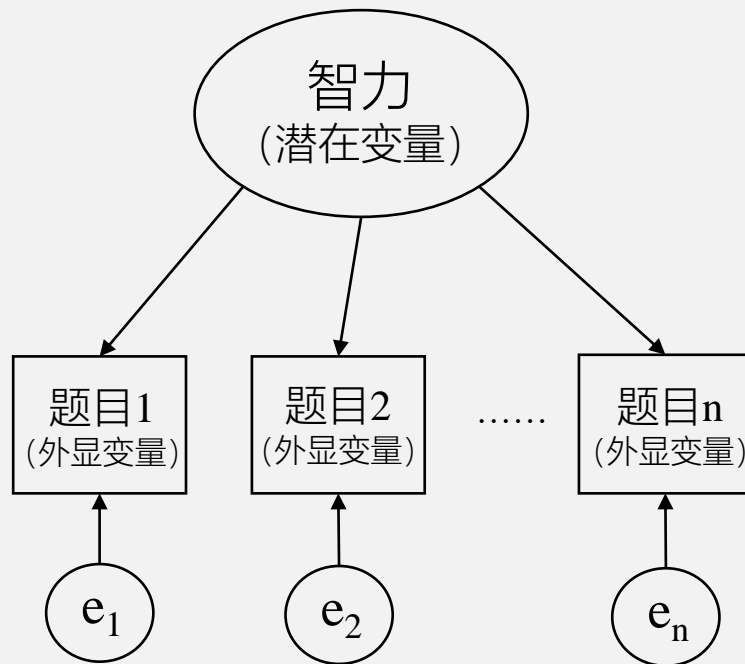
潜在变量模型

——潜在剖面分析 (Latent Profile Analysis, LPA)

报告人：黄颖诗

什么是潜在变量？

- 外显变量：可直接测量
 - e.g. 题目
- 潜在变量：无法直接测量
 - e.g. 智力



潜在变量模型

连续型潜变量

类别型潜变量

连续型指标

因素分析

潜在剖面分析

类别型指标

潜在特质分析
或项目反应理论

潜在类别分析

潜在剖面分析

- 分析模式
- 数学模型
- 分析过程

潜在剖面分析——分析模式

被试	题目				
	1	2	3	...	i
1	X_{11}	X_{12}	X_{13}	...	X_{1i}
2	X_{21}	X_{22}	X_{23}	...	X_{2i}
3	X_{31}	X_{32}	X_{33}	...	X_{3i}
...
p	X_{p1}	X_{p2}	X_{p3}	...	X_{pi}

潜在剖面分析——分析模式

	x_1	x_2	x_3	x_4	x_5	$P_y(Y)$
y_1	p_{11}	p_{21}	p_{31}	p_{41}	p_{51}	Σp_{i1}
y_2	p_{12}	p_{22}	p_{32}	p_{42}	p_{52}	Σp_{i2}
y_3	p_{13}	p_{23}	p_{33}	p_{43}	p_{53}	Σp_{i3}
y_4	p_{14}	p_{24}	p_{34}	p_{44}	p_{54}	Σp_{i4}
y_5	p_{15}	p_{25}	p_{35}	p_{45}	p_{55}	Σp_{i5}
$P_x(X)$	Σp_{1j}	Σp_{2j}	Σp_{3j}	Σp_{4j}	Σp_{5j}	1

- 联合概率 (分布)
- 边缘概率 (分布)
- 条件概率 (分布)

潜在剖面分析——分析模式

	x_1	x_2	x_3	x_4	x_5	$P_y(Y)$
y_1	p_{11}	p_{21}	p_{31}	p_{41}	p_{51}	Σp_{i1}
y_2	p_{12}	p_{22}	p_{32}	p_{42}	p_{52}	Σp_{i2}
y_3	p_{13}	p_{23}	p_{33}	p_{43}	p_{53}	Σp_{i3}
y_4	p_{14}	p_{24}	p_{34}	p_{44}	p_{54}	Σp_{i4}
y_5	p_{15}	p_{25}	p_{35}	p_{45}	p_{55}	Σp_{i5}
$P_x(X)$	Σp_{1j}	Σp_{2j}	Σp_{3j}	Σp_{4j}	Σp_{5j}	1

xy 的联合概率 $P(X=x_i, Y=y_j)$
 多个条件且所有条件同时成立

Y 的边缘概率 $P(Y=y_j)$
 $P(X=x_i)$ 或 $P(Y=y_j)$,
 仅与单个随机变量有关

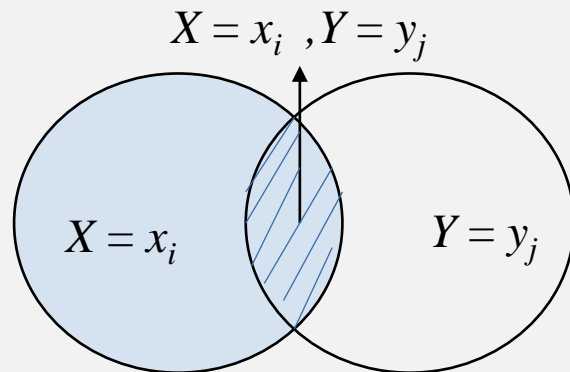
X 的边缘概率 $P(X=x_i)$

$$\forall x \in X, P(X = x) = \sum_y P(X = x, Y = y) \quad p(x) = \int p(x, y) dy$$

潜在剖面分析——分析模式

- 条件概率（分布）

条件概率表示在条件 $X = x_i$ 成立的情况下， $Y = y_j$ 的概率，记作 $P(Y = y_j | X = x_i)$ 或 $P(y_j | x_i)$ 。



$$p(Y = y_j | X = x_i) = \frac{p(X = x_i, Y = y_j)}{p(X = x_i)}$$

潜在剖面分析——分析模式

- 贝叶斯后验概率

先验概率：知道原因推结果

后验概率：根据结果推原因

$$p(Y = y_j | X = x_i) = \frac{p(X = x_i, Y = y_j)}{p(X = x_i)}$$

$$p(Y|X) = \frac{p(XY)}{p(X)}$$

$$p(XY) = p(Y|X)p(X)$$

$$p(XY) = p(YX), p(YX) = p(X|Y)p(Y).$$

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

潜在剖面分析——数学模型

- 若 X 与 Y 两个变量独立无关，则：

$$P_{ij} = P_i^x P_j^y$$

- X 与 Y 关联的实质：
 - 变量 X 与 Y 的关联能够被一个潜在类别变量解释。

潜在剖面分析——数学模型

- 概率参数化

- 潜在类别概率 π_t^c (方差贡献率) $P(C = t)$

被试属于潜在类别 t 的概率，相当于各个潜在类别的类别大小 (Class size)，数值越大的类别具有越重要的地位。

- 条件概率 $\pi_{it}^{\bar{x}^c} \pi_{jt}^{\bar{y}^c}$ (因子负荷) $P(X = i | C = t); P(Y = j | C = t)$

在被试属于潜在类别 t 的条件下，随机抽取一个人，在外显变量上作答的概率，数值越大说明潜在变量对该外显变量影响越强。

$$\sum_i \pi_{it}^{\bar{x}^c} = \sum_j \pi_{jt}^{\bar{y}^c} = 1$$

潜在剖面分析——数学模型

- 基本公式：

$$\text{联合概率} \leftarrow \pi_{ij}^{XY} = \sum_{t=1}^T \pi_t^C \pi_{it}^{\bar{X}^C} \pi_{jt}^{\bar{Y}^C}$$

属于类别 t 的被试在 Y 上的反应为 j 的条件概率 $P(Y=j | C=t)$

属于某一个特定类别的概率, $P(C=t)$

- 假设为连续分布，总模型服从多元正态分布：

$$f(M) = \sum_{c=1}^T p(C) f(M | \mu_c, \Sigma_c)$$

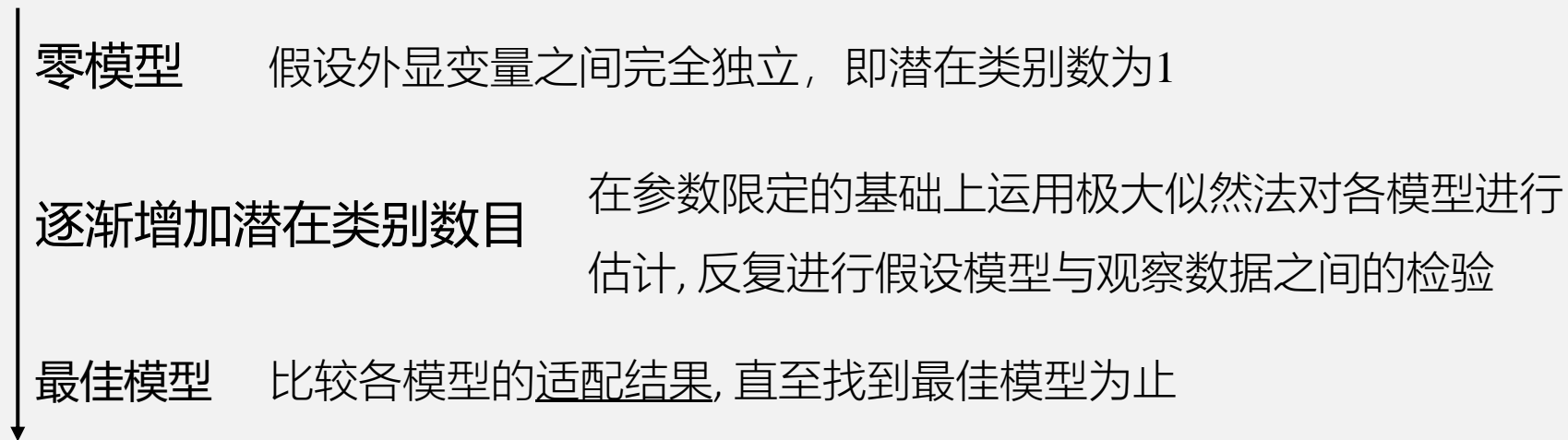
m 个外显变量向量的联合概率密度， C 指潜在变量， t 指潜在类别数，每个潜在类别有自己的均值向量 μ_c 和协方差矩阵 Σ_c ， $p(C)$ 指属于某类别的先验概率。

潜在剖面分析——分析过程

- 模型选择——确定潜变量水平数（有多少个潜在类别变量）
↓
- 模型解释——对各类别进行归纳和命名
↓
- 个体归类——对每位被试进行分类

潜在剖面分析——分析过程

- 模型选择



潜在剖面分析——分析过程

- 三大类拟合指数

- 信息评价准则：越低越好

赤池信息准则（AIC）、贝叶斯信息准则（BIC），以及它们的修正版，如AIC3（赤池信息准则3）、CAIC、SABIC

- 似然比检验：假设增加类别前更匹配数据($p < 0.05$)

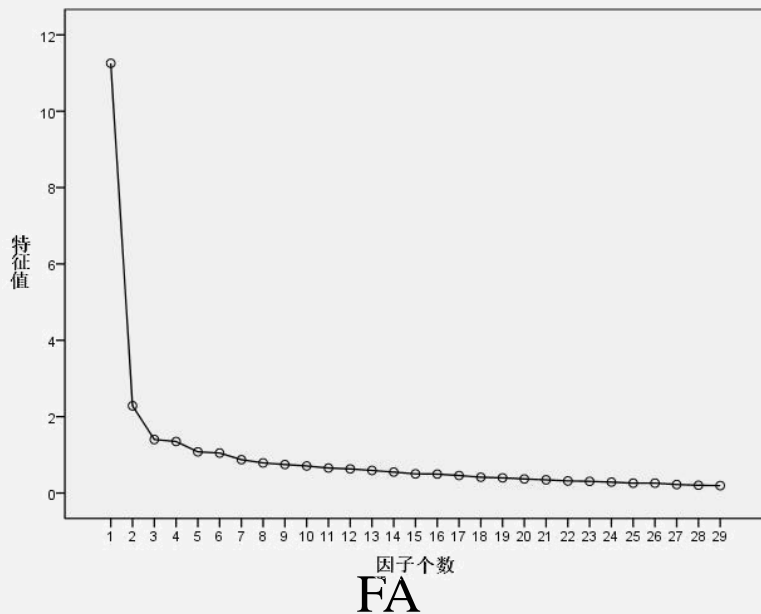
LMR、aLMR

- 分类不确定性：越接近1越好

熵值（Entropy）

潜在剖面分析——分析过程

- 模型选择

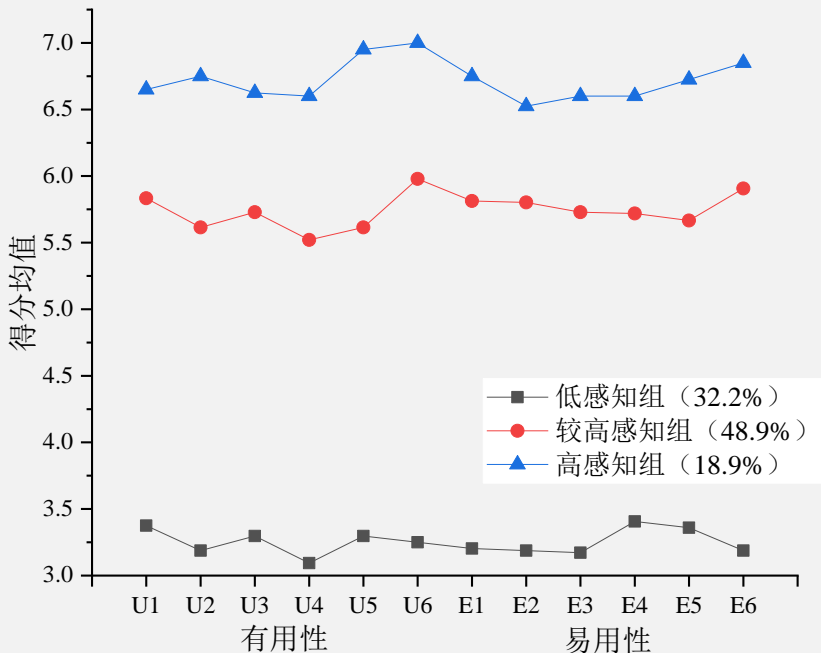


Model	AIC	BIC	Entropy
1	7063.259	7105.274	
2	6508.669	6596.201	0.928
3	6157.773	6290.821	0.942
4	6014.392	6192.956	0.934
5	5973.560	6197.641	0.888

LCA

潜在剖面分析——分析过程

- 模型解释
 - 右图为感知用户体验分类，根据三类别模型中各个题目的得分均值，得分均值从高到低，可将三种类别依次命名为高感知组，较高感知组，和低感知组。



潜在剖面分析——分析过程

- 个体归类

- 在一定的概率模型之下 (model-based), 利用概率的估计与比较进行分类。

作答模式 (XY)	Class1	Class2	Class3
11	0.001	0.343	0.013
12	0.042	0.001	0.050
...
55	0.057	0.016	0.008

$$P(\text{属于类别1}|\text{作答模式为11}) = 0.001/(0.001+0.343+0.013) = 0.003$$

$$P(\text{属于类别2}|\text{作答模式为11}) = 0.343/(0.001+0.343+0.013) = \mathbf{0.961}$$

$$P(\text{属于类别3}|\text{作答模式为11}) = 0.013/(0.001+0.343+0.013) = 0.036$$

Thank you for your attention!

2019-03-02