

# Latent Class Analysis

McCutcheon, Allan L., 1987, *Latent Class Analysis*, Sage, Chap2

鎌田 健太郎

東京大学大学院教育学研究科

---

- カテゴリカルデータの分析方法
- 潜在変数モデル（cf. 因子分析）
- 潜在変数・顕在変数ともにカテゴリカル（離散変数）
  - 顕在変数（indicator, manifest variable）：データに入っている変数
  - 潜在変数（latent variable）：顕在変数の背後に想定される何らかの概念を表す変数
- 社会学での利用：似たような意識・選好などをもつグループの抽出など…
  - 性別役割態度（[山口 1999](#)）
  - 排外意識（[永吉 2014](#)）

- AとBは関連ありだが, Cの各層の内部 (局所) では関連なし (独立)

A	B	
	1	2
1	95	55
2	70	80

X-squared = 8.4175, df = 1,  
p-value = 0.003716,  
Odds ratio = 1.974026

C = 1

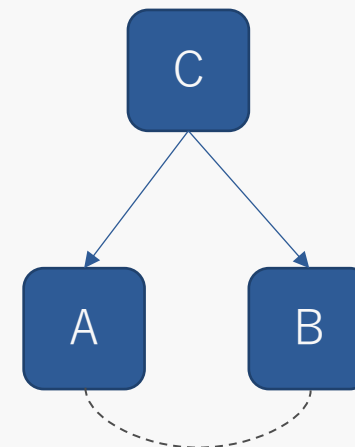
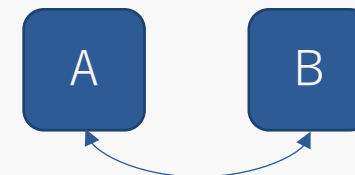
A	B	
	1	2
1	80	20
2	40	10

X-squared = 0, df = 1, p-value = 1  
odds ratio = 1

C = 2

A	B	
	1	2
1	15	35
2	30	70

X-squared = 0, df = 1, p-value = 1  
odds ratio = 1



•  $P_{ijk}^{ABC} = P_k^C \times P_{ik}^{\bar{A}C} \times P_{jk}^{\bar{B}C}$

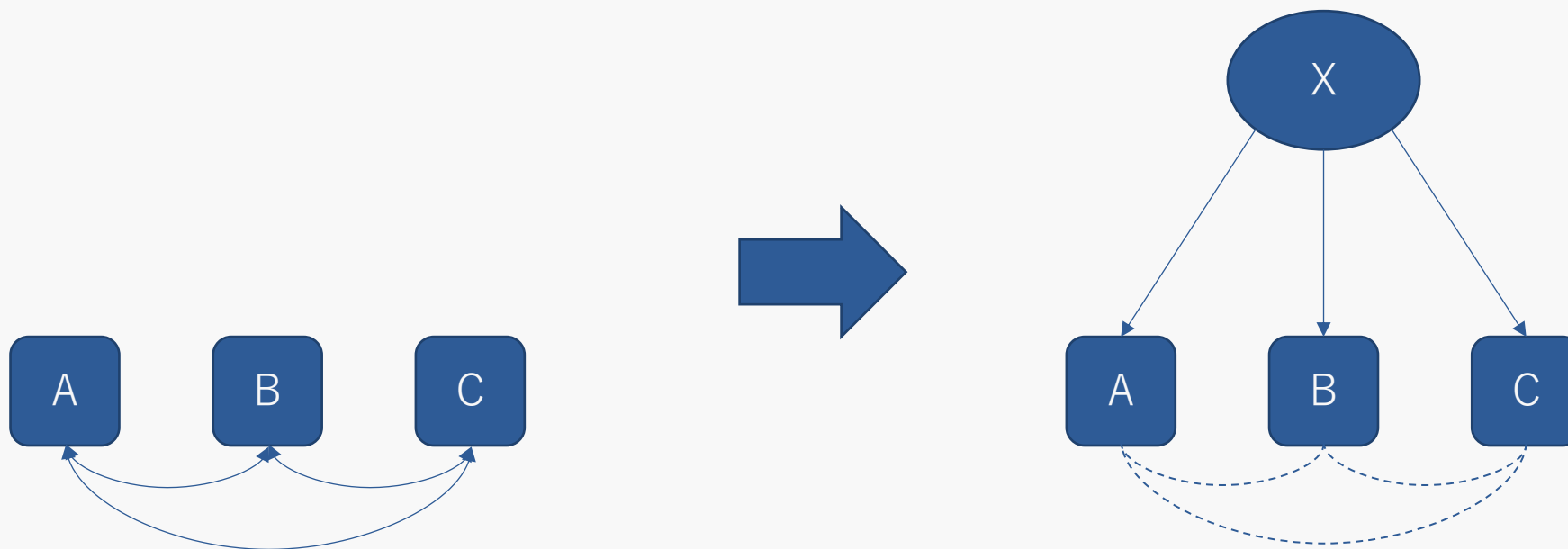
		C		
		1	2	
		0.500	0.500	← 変数Cの相対度数
A	1	0.667	0.333	← 変数Aの 条件付き確率
	2	0.333	0.667	
B	1	0.800	0.300	← 変数Bの 条件付き確率
	2	0.200	0.700	

• 計算例

$$P_{111}^{ABC} = 0.5 \times 0.667 \times 0.8$$
$$= 0.2668 \simeq \frac{80}{300}$$

$$P_{22}^{AB} = P_{221}^{ABC} + P_{222}^{ABC}$$
$$= 0.5 \times 0.333 \times 0.2 + 0.5 \times 0.667 \times 0.7$$
$$= 0.26675 \simeq \frac{80}{300}$$

- 局所独立を達成できるCのような変数を推定する = 潜在クラス分析
- 定式化：A,B,Cの3つの顕在変数から潜在変数Xを推定
  - $\pi_{ijkt}^{ABCX} = \pi_t^X \times \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \pi_{kt}^{\bar{C}X}$



- Latent Class Probabilities ( $\pi_t^X$ ) : 各潜在クラスの相対的な大きさ
  - $\sum_t \pi_t^X = 1$  : 相対的な大きさ（確率）なので全てのクラスを足して1になる
- Conditional Probabilities ( $\pi_{it}^{\bar{A}X}, \pi_{jt}^{\bar{B}X}, \pi_{kt}^{\bar{C}X}$ ) : 各クラスの特徴を表す
  - $\pi_{11}^{\bar{A}X} = 0.7, \pi_{12}^{\bar{A}X} = 0.3$  : クラス1において,  $A = 1$ と回答する確率は0.7,  
 $A = 2$ と回答する確率は0.3なので, クラス1は $A = 1$ と回答する人が多めのクラス
  - $\sum_i \pi_{it}^{\bar{A}X} = \sum_j \pi_{jt}^{\bar{B}X} = \sum_k \pi_{kt}^{\bar{C}X} = 1$  : クラスtの内部で条件付き確率を足すと1になる

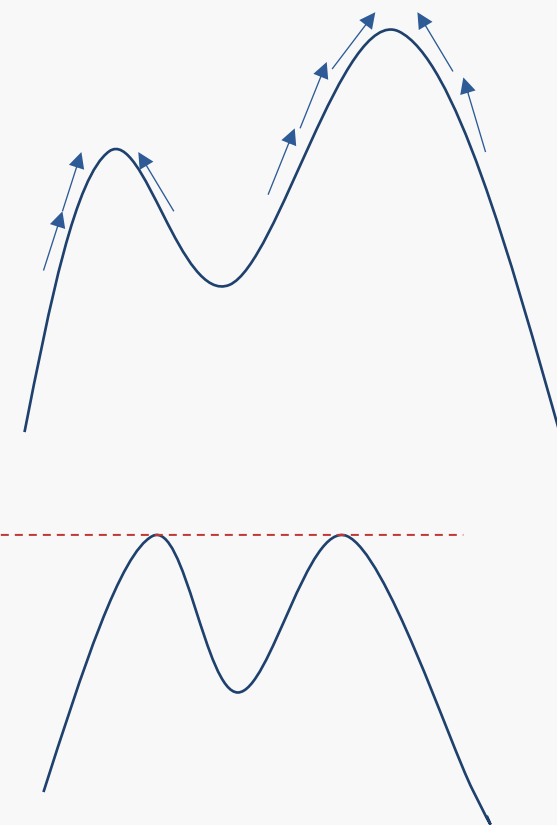
- 最終目標： $\hat{\pi}_t^X, \hat{\pi}_{it}^{\bar{A}X}, \hat{\pi}_{jt}^{\bar{B}X}, \hat{\pi}_{kt}^{\bar{C}X}$ を推定
- 以下が成り立つことがわかっている
  - $\hat{\pi}_{ijkt}^{ABCX} = \hat{\pi}_t^X \times \hat{\pi}_{it}^{\bar{A}X} \times \hat{\pi}_{jt}^{\bar{B}X} \times \hat{\pi}_{kt}^{\bar{C}X}$  : joint prob = LC prob  $\times$  Conditional prob
  - $\hat{\pi}_{ijk}^{ABC} = \sum_t \hat{\pi}_{ijkt}^{ABCX}$  : (1, 1, 1, t)の確率を全てのtで足していくと元のクロス表における(1, 1, 1)の確率が復元できる
  - $\hat{\pi}_{ijkt}^{ABC\bar{X}} = \hat{\pi}_{ijkt}^{ABCX} / \hat{\pi}_{ijk}^{ABC}$  : (i, j, k)において、クラスtに割り当てられる条件付き確率
    - 個人を潜在クラスに割り当てて、潜在クラスと他の変数の関連を見るなど
    - 各個人においてクラス1からクラスtまでこの確率を比べて、最も確率の高いクラスにその人を割り当てたりする (modal assignment)

- $p_{ijk}$  : ABCの3元クロス表における (i, j, k) の相対度数
- 最終目標は以下の式で推定できる
  - $\hat{\pi}_t^X = \sum_{ijk} p_{ijk} \hat{\pi}_{ijkt}^{ABC\bar{X}}$
  - $\hat{\pi}_{it}^{\bar{A}X} = \sum_{jk} p_{ijk} \hat{\pi}_{ijkt}^{ABC\bar{X}} / \hat{\pi}_t^X$
  - $\hat{\pi}_{jt}^{\bar{B}X} = \sum_{ik} p_{ijk} \hat{\pi}_{ijkt}^{ABC\bar{X}} / \hat{\pi}_t^X$
  - $\hat{\pi}_{kt}^{\bar{C}X} = \sum_{ij} p_{ijk} \hat{\pi}_{ijkt}^{ABC\bar{X}} / \hat{\pi}_t^X$



- EMアルゴリズム：不完全データに対する最尤推定アルゴリズムの一つ
1. 初期値を定める ( $\bar{\pi}_t^X, \bar{\pi}_{it}^{\bar{A}X}, \bar{\pi}_{jt}^{\bar{B}X}, \bar{\pi}_{kt}^{\bar{C}X}$ )
  2. これまで見てきた式に代入していき,  $\bar{\pi}_{ijkt}^{ABC\bar{X}}$  を計算 (スライド7)
  3. データの観測度数 ( $p_{ijk}$ ) を利用して初期値を更新 (スライド8)
  4. 新たに得られた ( $\bar{\pi}_t^X, \bar{\pi}_{it}^{\bar{A}X}, \bar{\pi}_{jt}^{\bar{B}X}, \bar{\pi}_{kt}^{\bar{C}X}$ ) を元に  $\bar{\pi}_{ijkt}^{ABC\bar{X}}$  を計算
  5. 一定回数の繰り返し or 推定値の変化幅が一定以下になる (収束する) まで繰り返し...

- 局所解の問題：極大値が複数あることが普通
  - 初期値を変えながら何十回か推定を行い、  
globalな最大値を求める
- 自由度がマイナスにならないように
  - $$df = IJK - (1 + \underset{\substack{\uparrow \\ \text{Grand} \\ \text{Effect}}}{(T - 1)} + \underset{\substack{\uparrow \\ \text{LC} \\ \text{prob}}}{T(I - 1)} + \underset{\substack{\uparrow \\ \text{AのCond} \\ \text{Prob}}}{T(J - 1)} + \underset{\substack{\uparrow \\ \text{BのCond} \\ \text{Prob}}}{T(K - 1)})$$
- Identification Problem（パラメタの識別可能性）
  - 尤度を最大化するパラメタの組が複数ある可能性（＝解が一意に定まらない）
  - 確率を固定する（cf. Chap3）ことなどで対処



- [三輪哲, 2009, 「潜在クラスモデル入門」『理論と方法』 24\(2\): 345–56.](#)
- [藤原翔・伊藤理史・谷岡謙, 2012, 「潜在クラス分析を用いた計量社会学的アプローチ——地位の非一貫性, 格差意識, 権威主義的伝統主義を例に」『年報人間科学』 33: 43–68.](#)
- [井出草平の研究ノート \(はてなブログ\)](#)
  - さまざまなソフトウェア (R, Stata, Mplusなど) による分析例が載っている
- [Vermunt, Jeroen, 1997, LEM: A General Program for the Analysis of Categorical Data.](#)
  - 確率に制約をかけた潜在クラスや多集団分析といった複雑なモデルの関係性などがわかる
- [Collins, Linda M. & Stephanie T. Lanza, 2010, \*Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences\*, Hoboken, New Jersey: John Wiley & Sons, Inc.](#)