日本テスト学会講演
平成21年5月30日
早稲田大学36号館

テスト応答分析のための ベイズ的アプローチ

帝京大学
繁桝算男

OUTLINE

- 階層モデルとベイズ的アプローチの利点
- 項目応答へのベイズ的アプローチ：項目反応理論と古典的テスト理論
- テスト応答分析へのベイズ的アプローチの応用

二つの対立（？）軸

- 伝統的統計学（標本抽出理論） vs. ベイズ理論
- 古典的テスト理論 vs. 現代テスト理論（項目反応理論）
- あなたはどのように自分を位置付けますか？

1. ベイズ統計学とは？

確率法則を認知的に根拠のある体系として、
P(未知のもの｜既知のもの)
を得るために用いる。

- **未知の知りたいこと：θ**
- **既知の観測されたもの**：X(**データあるいは割り当てを示すダミー変数は、観測されることのあり、観測されないこともある。**e.g.**将来のデータ、調査観察研究の場合の割り当て。**)

$$p(\theta \mid X) \propto p(X \mid \theta)\, p(\theta)$$

- **主観的な設定が必要なもの**

データ(発生) モデル

事前分布　AND

データ取得メカニズム (eg. treatment assignment mechanism)

2. 階層モデル

- ベイズはもともと階層的

$$p(\theta \mid X) \propto p(X \mid \theta, \tau, \xi)\, p(\theta \mid \tau, \xi)\, p(\tau \mid \xi)$$

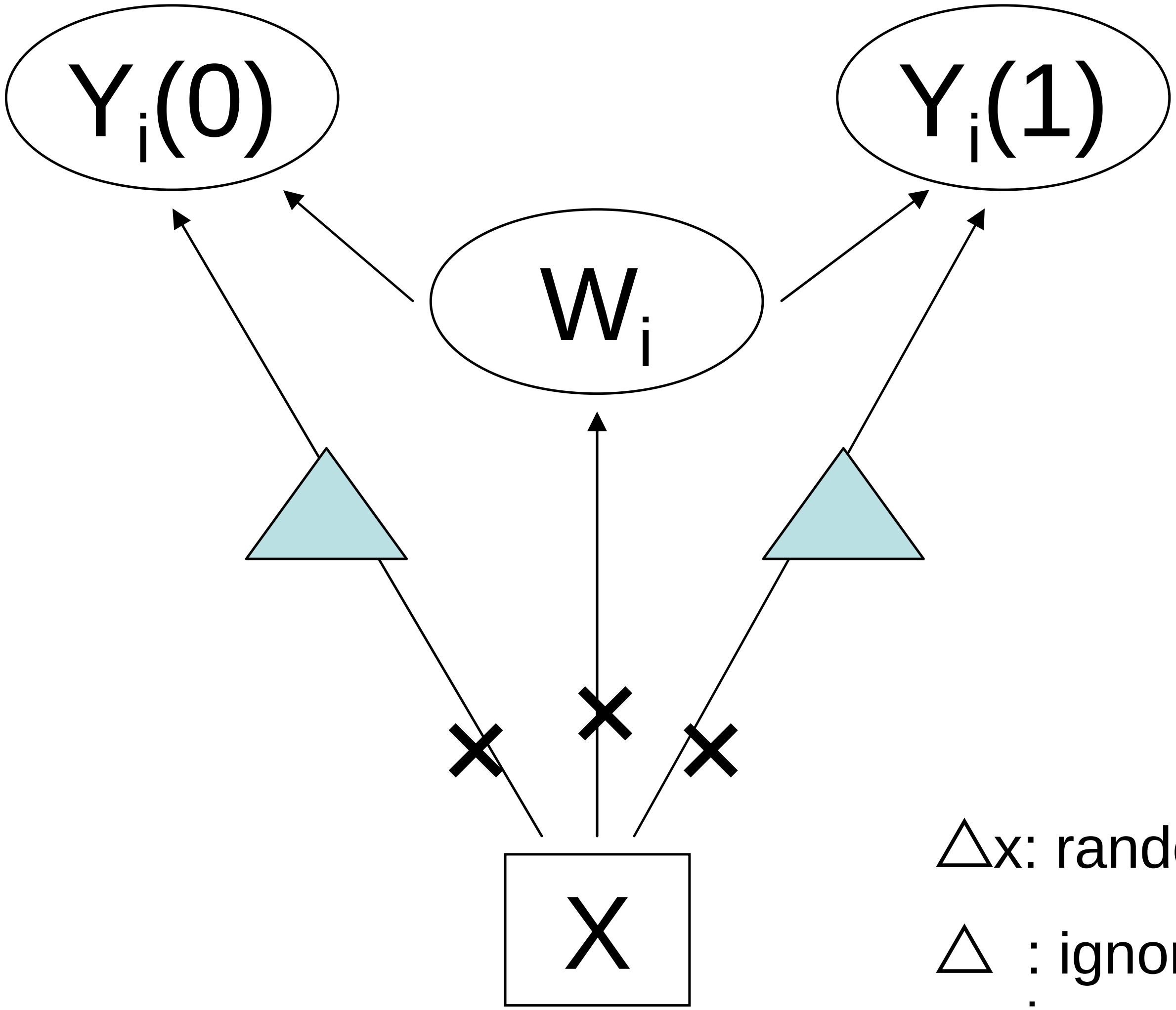
where θ : parameter, τ : hyper - parameter,

and ξ : super - parameter

階層的に考える場合の注意

- 何を知りたいか？（e.g.将来の観測値か、パラメータか?）
- 何がわかっていて何がわかっていないか？(e.g.観測メカニズムは既知か？）
- 何が観測されるか？
- 未知の物を推論するために必要なものが抜けていないか？無駄なものが入っていないか？（e.g.観測メカニズムの知識は必要か？）
- モデルはなるべく“客観的”にする。即ち、だれしもが納得するようにしたい。

3．Rubin因果モデル



△×: randomized experiment

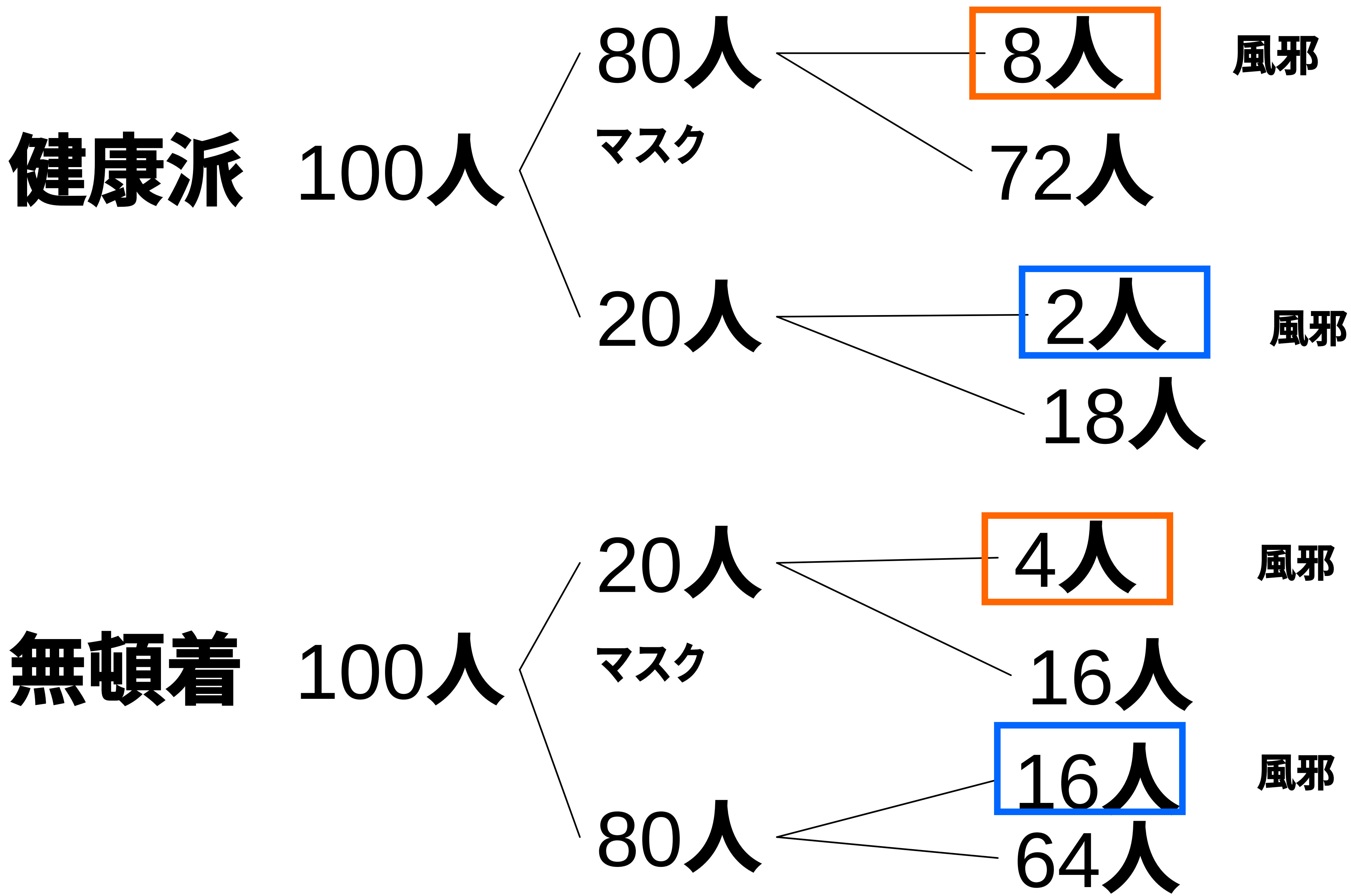
△ : ignorable treatment assignment

マスクは役に立つか？

- 同じ人がマスクをした場合とマスクをしない場合を比較する。→これは無理。
- ゆえに、同じ条件で、交換可能な観測ユニットを探す。理想的には、単純無作為化実験、a あるいは、共変数(健康意識) を所与として、確率化実験。たとえば、風邪にかかりたくないと強く思う人100人の8割にマスク、2割にマスクを着用させない。無頓着な人100人の2割にマスクを着用させ、8割にはそのまま。(無視可能な実験デザイン) 。

架空データ

- 健康に気をつける人の内、マスクをした人の1割が風邪を引き、同様に、マスクをしていない人も１割が風邪を引いた。また、無頓着な人の内、マスクをした人もしない人も2割が風邪を引いた。もし母集団において健康に注意する人無頓着な人が5分5分だとすると、母集団では、マスクをつける人もつけない人も１５パーセントが風邪を引くことになり、マスクの効果がないことは明らかである。


健康派
100人
80人
マスク
20人
8人
風邪
72人
2人
風邪
18人
無頓着
100人
20人
マスク
80人
4人
風邪
16人
16人
風邪
64人

確率化が行われない場合

ここで単純にデータ発生メカニズムを無視すれば、マスクをした人100人の内12人(健康派80人の1割、無頓着20人の2割) が風邪を引き、マスクをしない人100人の内18人 (健康派20人の1割無頓着80人の2割) 、マスクをする効果があると誤って結論するかも知れない。

research model

assignment mechanism

$$p(Y_{mis} \mid X, Y_{obs}, W)$$

$$= \frac{\boxed{p(X, Y(0), Y(1))}\boxed{P(W \mid X, Y(0), Y(1))}}{\int p(X, Y(0), Y(1)) P(W \mid X, Y(0), Y(1)) dY_{mis}},$$

ignorable treatment assignmentが成立するならば、

$$= \frac{P(X, Y(0), Y(1))}{\int P(X, Y(0), Y(1)) dY_{mis}}$$

(cf. ignorable assignment :

$$P(W \mid X, Y(0), Y(1)) = P(W \mid X, Y_{obs}))$$

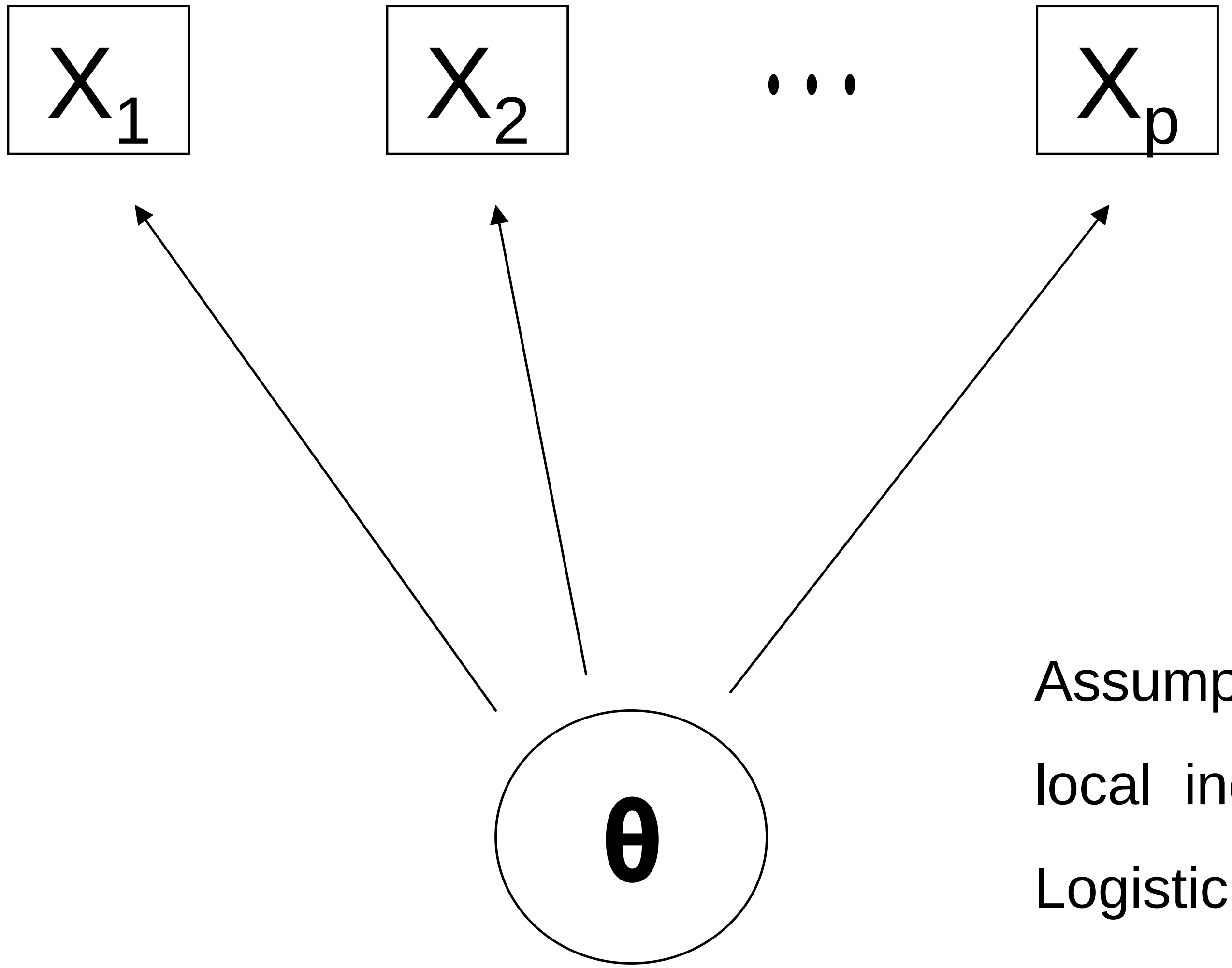
一般的な図式

$$p(\theta \mid X, W) = \frac{p(\theta, X)P(W \mid \theta, X)}{\int p(\theta, X)P(W \mid \theta, X)d\theta}$$

$$= \frac{p(\theta)p(X \mid \theta)P(W \mid \theta, X)}{\int p(\theta)p(X \mid \theta)P(W \mid \theta, X)d\theta}$$

通常の調査観察研究では、観測メカニズムは無視される。

4. 項目反応(応答)理論

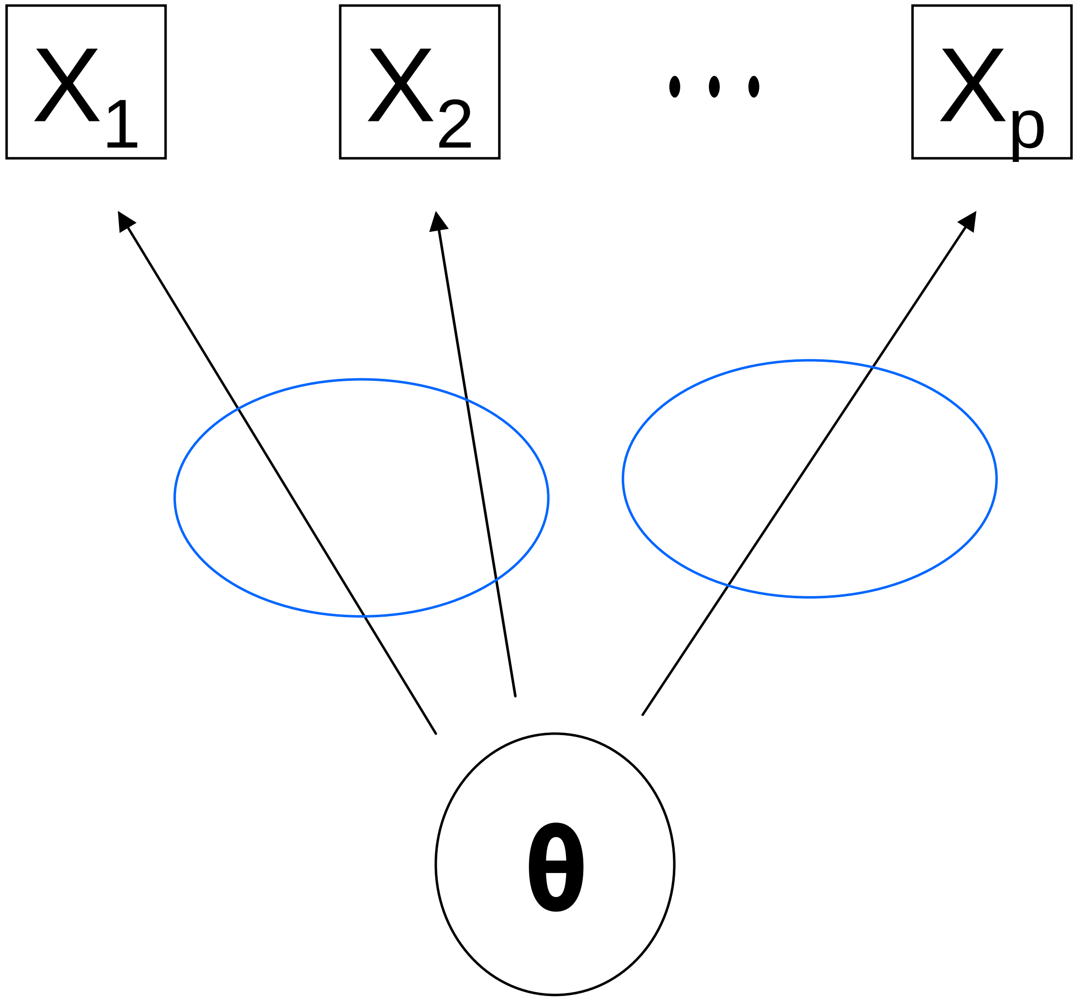


Assumptions involved

local independence

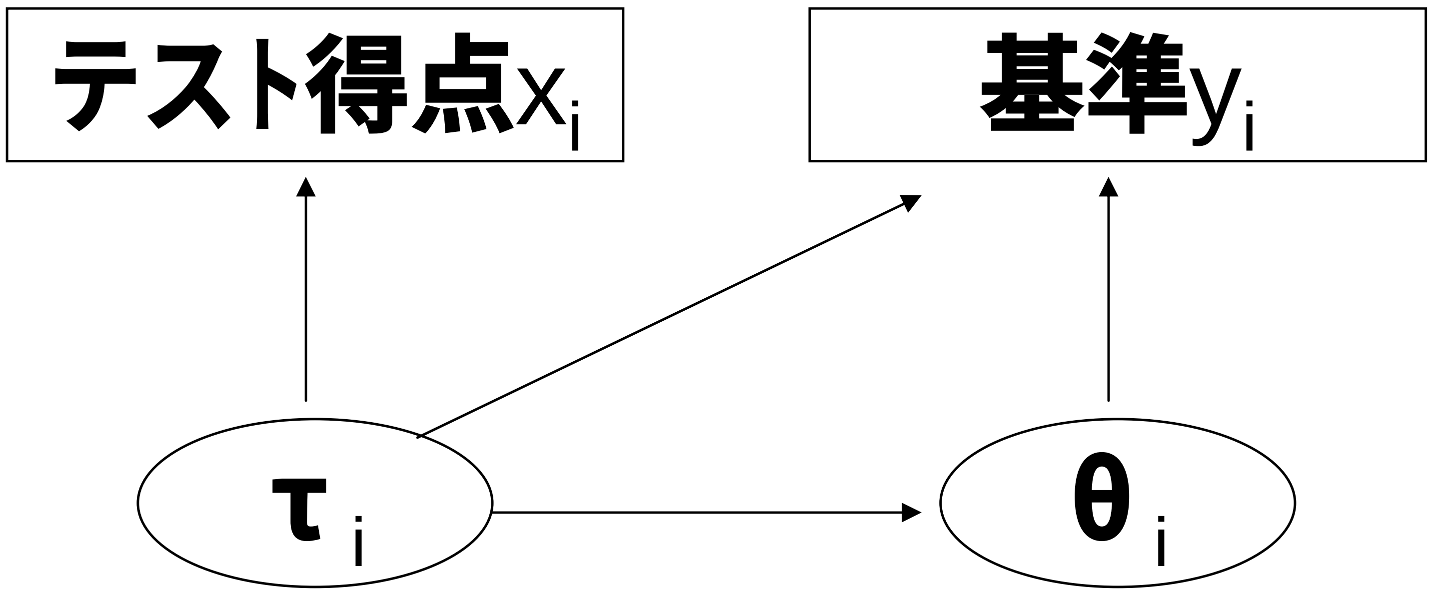
Logistic regression

大問形式(ブックレットモデル)



Assumptions involved;
Multi-Normality within
a group of items

5. 古典的テスト理論（真の得点理論）



- 期待値τ_iは一様には定義できない
  →一般化可能性理論
- 期待値τ_iは知りたい値θ_iとは異なる
  →階層的理論へ

5-1 一般化可能性理論

$$x_{ijk} = \tau_i + \gamma_k + \varepsilon_{ijk}$$

$$\text{where } i : i\text{-th unit,}$$

$$j : j\text{-th rater, and } k :$$

$$k\text{-th repetition}$$

ベイズ的アプローチの利点:いろいろな分散成分の関数、たとえば、各種のの信頼性係数の分布を導くことが出来る。

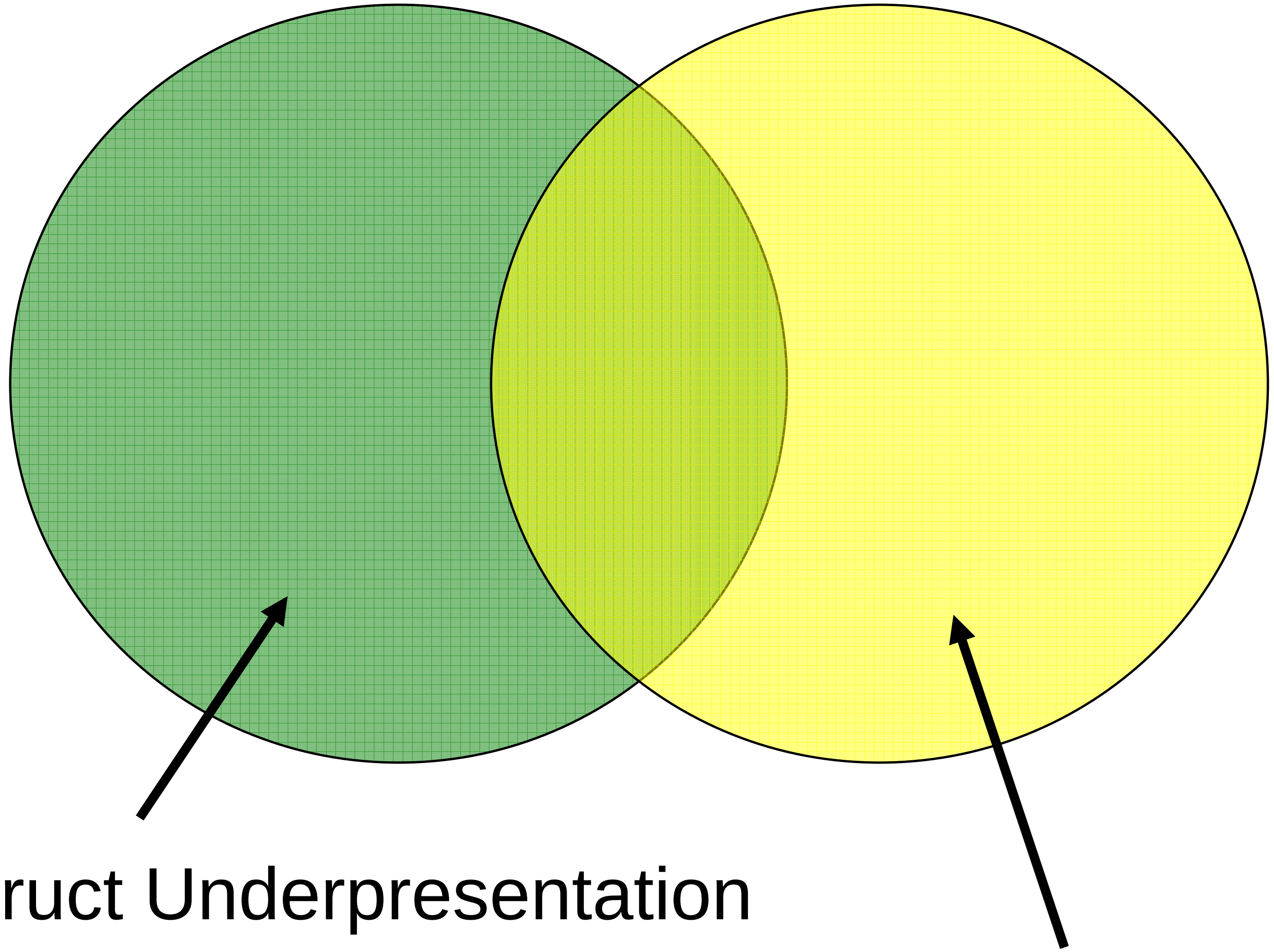
5-2　妥当性の問題

- いかにxがθと関係しているか
- 相関係数

$$x = G_x(\tau) + \varepsilon_x$$

$$y = G_y(\theta) + \varepsilon_y$$

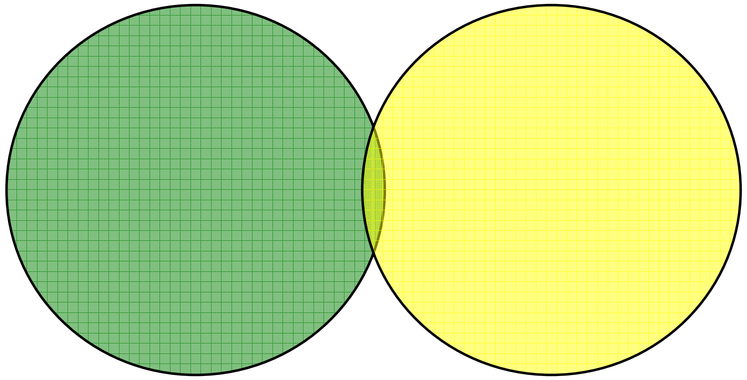
- 構成概念（θ）



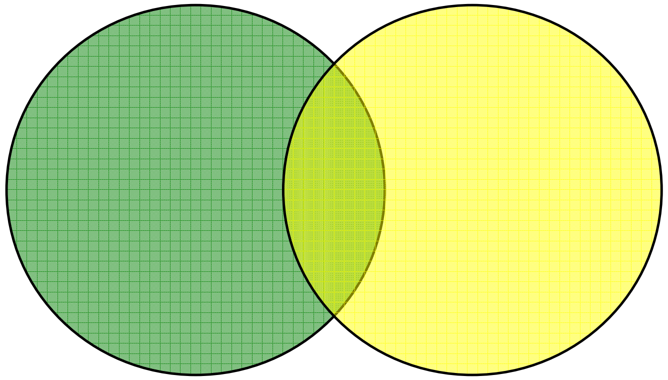
真の値（τ）

CU：Construct Underpresentation

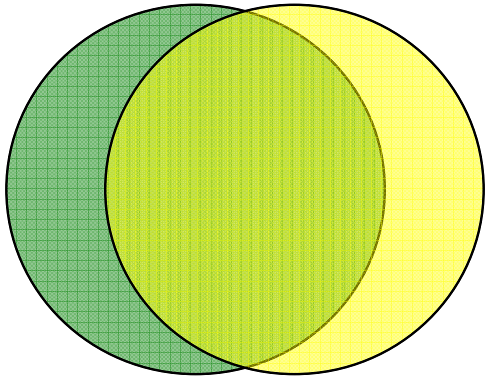
CIV：Construct Irrelevant Variance



- Construct　(知能)
- Test　(頭囲)



- Construct　(知能)
- Test　(脳容量)



- Construct　(知能)
- Test　(知能テスト)
- CU　：創造性、時間をかけた問題解決力
- CIV　：テストワイズネス

理論的関心と実践的関心

- 二つの構成概念はどの程度関連しているか？
  ⇒τとθの相関の推定
- テストは構成概念とどの程度関連しているか？
  ⇒xとθの相関の推定
- テストは現実の基準をどの程度予測するか？

  ⇒xとyの相関の推定

希薄化の修正

$$P(\theta, \tau \mid x, y)$$

$$\propto p(x, y \mid \tau, \theta)\, p(\tau)\, p(\theta)$$

$$= p(y \mid x, \tau, \theta)\, p(x \mid \tau, \theta)\, p(\tau \mid)\, p(\theta)$$

この一般的な式を従来の希薄化の修正式は、
単純化し、点推定値のみ求めている。

$$p(x \mid \tau)\, p(y \mid \xi)\, p(\tau \mid)\, p(\xi)$$

選択母集団の問題

$$p(x, y) = p(x, y \mid w=1)P(w=1)$$

$$+ p(x, y \mid w=0)P(w=0)$$

$$= p(y \mid x, w=1)P(w=1 \mid x)\,p(x)$$

$$+ \boxed{p(y \mid x, w=0)}P(w=0 \mid x)\,p(x)$$

y : gpa　x : 入試得点

$w=1$: 合格　$w=0$: 不合格

$P(w=1 \mid x)$ or $P(w=0 \mid x)$: 大量データにより推定

$p(\mathrm{y} \mid \mathrm{x}, \mathrm{w}=0)$: regression model

テスト使用の究極の妥当性：結果的妥当性

ある意思決定において、テスト情報を用いて決定した場合に予想する効用（期待効用）と、テスト情報を用いないで決定した場合に予想する効用（期待効用）の差をテストの情報価値（Expected Value of Test Information,EVTI)と呼ぶ。

EVTI

$$= \int_X \max_i \left\{ \int u(a_i, \theta) \, p(\theta \mid x) d\theta \right\} p(x) dx$$

$$- \max_i \left\{ \int u(a_i, \theta) \, p(\theta) d\theta \right\}$$

6．まとめ：ベイズ的アプローチの利点

- 単純な場合
  - ほぼ同じ結果(解釈は異なる)
  - 例：単純無作為実験(cf.　Neyman,1923)
- 複雑な場合
  - 異なる結果
  - モデル建立が容易
  - 解釈がしやすい
  - パラメータの分布がわかる
  - 確率的言明が直接手に入る

テスト得点への応用における利点

- テスト応答の発生モデルは単純ではない。
  (正規分布が妥当なのは、潜在変数くらいである。) ⇒複雑なデータモデルを必要とする。(将来の夢:ベイズ的ネットワークによる記述式回答の自動採点)
- テストは常に何かを目的として使用される。すなわち、意思決定の枠組みで考えるべきであり、ベイズ的アプローチはこの枠組みと相性がよい。