

大規模入学者選抜における CBT-IRTへの懸念

林 篤裕

(名古屋工業大学 社会工学専攻
& アドミッションオフィス長)

e-mail: hayashi.atsumuro@nitech.ac.jp



資料掲載URL: stat.web.nitech.ac.jp/haifu/#JART2108

1. はじめに

- ◆ 情報化社会 ==> 大学入試でもICTの利活用を模索
- ◆ 「CBT活用検討部会」: 2020年4月～2021年3月
 - ◆ (独)大学入試センター 大学入学共通テスト企画委員会の下
 - ◆ 大規模試験におけるICTの利活用を検討
 - ◆ 報告書: 2021年3月に公表
- ◆ 「CBT活用有識者会議」: 2019年4月?～2020年3月
 - ◆ 「数理的見地から見た選抜試験におけるCBT・IRTの利活用に関する考察 --- 顕在変量と潜在変量とのハザマで ---」
 - 2019年10月25日 ==> 今回の発表に
- ◆ 検討事項
 - ◆ 大規模入学者選抜試験とICT(CBT)の親和性: 2節
 - ◆ 日本における入試成績を潜在変量で表現すること: 3, 4節 2

2. 大学入試とICT環境(CBT)

- ◆ 紙と鉛筆による試験(Paper Based Testing, PBT)
- ◆ コンピュータを用いた試験
(Computer Based Testing, CBT)
- ◆ PBT・CBTの利点・欠点: 詳細は報告書を参照ください
[一例] CBTに関して
 - ◆ 出題にマルチメディアが利用可能
 - ◆ 動的出題(Adaptive, CAT)も実装可能
<=== 短時間で測定できる。逐次検定的(統計学)。
 - ◆ 受験者への設問のデリバリーが容易・安価
<=== 特に大規模試験では効果

2. 大学入試とICT環境(CBT): 懸念

- 装置不具合の危惧 = DNCリスニング機器(少部品点数)でさえ発生・対応してきた過去(2006年～)
 - ◆ 部品点数と故障率は正比例
 - ◆ 共通試験における英語リスニング試験の試験監督を担当したことのある者なら危惧するはず
 - ◆ 紙の頑健性・利便性を凌駕する環境が提供できるのか?
- ◆ 膨大なItem Bankが必要
 - ◆ 管理(統計量、履歴、プレテスト等)も必須
 - ◆ Itemは原則非公開 <===> 日本の試験文化との親和性
- アジアでの失敗例: TOEFLのItem窃盗(2000年前後)
 - ◆ Item自動生成の研究が過去にはあったがその後聞かない

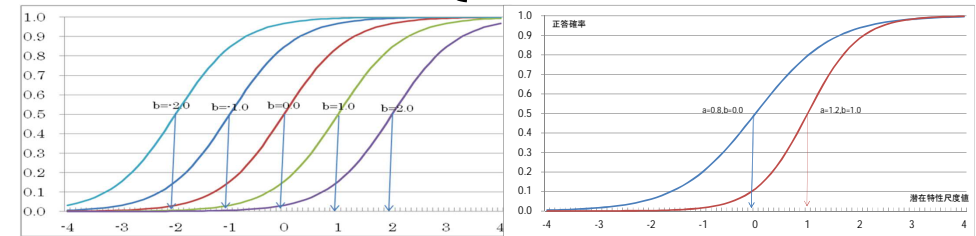
2. 大学入試とICT環境(CBT)

- ◆ 欧米では可能なのになぜ日本では動きが鈍いのか?
- ◆ 試験文化の違い
 - ◆ 入試問題が問おうとしている内容・質に対する要求度
 - ◆ 多肢選択式設問の作題技術
- ◆ 大学入試の競争性・意味合いの違い
 - ◆ 欧米: ローステークス
 - 大学終了段階(出口)の学修度合いを厳密に評価
 - ◆ 日本: ハイステークス
 - 1点(時にはそれ未満)で合否が分離
 - 先の高大接続答申(2014)にも提言はある
 - その具体的な方策に名案はない
 - Stanineは救世主になれるのか?

5

3. IRTを日本の大学入試に用いる際の注意点

- ◆ IRTのモデル: シンプル。理解も容易?
- ◆ 適用条件(前提条件、強い条件)
 - ◆ 単発的な多肢選択式設問をイメージ → 大問形式への提案もある
 - ◆ 単答式の出題を対象としている → 日本の出題文化としての
大問形式は対象外
 - ◆ 個々のItem(設問)の独立性
 - ◆ 能力の1次元性 <==== {
 - 多様性を測ろうとしている時代に?
 - 複数単元が包含された科目には?



(野口 裕之(2017)から引用) 6

3. IRTを日本の大学入試に用いる際の注意点

- ◆ [懸念]
 - ◆ 単答式の出題=知識を問う設問になりがちではないか?
 - ◆ 思考力を測るのに大問形式は有効な手段(日本の文化)
- ◆ 変換点(Scaled Score):
 - ◆ 選抜試験における変換点の受容が進むか?: 疑問
 - ◆ 資格試験(TOEFL、TOEIC等)での普及で理解が進んだのか??
- パラメータの推定
 - ◆ 目的関数(尤度関数)が複雑(多変数多峰性)で最適解が求まる保証がない。擬似的に求まってよように振る舞っているだけ。
 - ====> {
 - ◆ モデルの特性
 - ◆ 計算技術の問題ではない(EM、MCMC、Bayes、...)
 - ◆ 計算手法が改良されたら数値が変わる可能性

7

3. IRTを日本の大学入試に用いる際の注意点

- ◆ 顕在変量(Manifest Variable)
 - ◆ 回帰分析(RA), 主成分分析(PCA),
(Regression Analysis) (Principal Component Analysis)
 - ◆ 線形計算、行列計算で実現: 解は一意に確定
- ◆ 潜在変量(Latent Variable)
 - ◆ 因子分析(FA), IRT, 潜在クラス分析(LCA),
(Factor Analysis) (Latent Class Analysis)
 - ◆ 測定できない。解は複数(無限個)。
確認の手段がない。恣意性の入り込む余地。
 - ◆ 推定したパラメータの真偽は誰が保証できるのか?
- 「計算結果=真値」であるということに対する
考え方・捉え方の違いのように思われる
★ 批判ではない。このような特性を知った上で運用すべきと考えているだけ。 8

3. IRTを日本の大学入試に用いる際の注意点

◆ パラメータの推定

- ◆ 尤度関数を最大にするパラメータ θ を求める。
- ◆ 解析的には解けず、繰り返し計算で求める必要。
- ◆ 初期値依存。Newton-Raphson法等の最適化手法の弱点。
- ◆ 計算機精度にも依存。
- ◆ パラメータ数も指定する必要がある。
- ◆ Local Maxは求まるが、Global Maxが求まる保証はない。

● θ の組み合わせは無限個存在する ==> 一種の違和感

- ◆ IRT: $L(\theta|\mathbf{u}) = f(\mathbf{u}|\theta) = P(u_1, u_2, \dots, u_m|\theta) = \prod_j P_j(\theta)^{u_j} Q_j(\theta)^{(1-u_j)}$

$$\ln L(\theta|\mathbf{u}) = \sum_j (u_j \ln P_j(\theta) + (1-u_j) \ln Q_j(\theta))$$

- ◆ LCA: $L = f(y; \phi) = \prod_{c=1}^C \left\{ \sum_{i=1}^n \pi_c f_c(y_i|\theta_c; x) \right\}$

$$\log L = \log f(y; \phi) = \sum_{c=1}^C \log \left\{ \sum_{i=1}^n \pi_c f_c(y_i|\theta_c; x) \right\}$$

3. IRTを日本の大学入試に用いる際の注意点

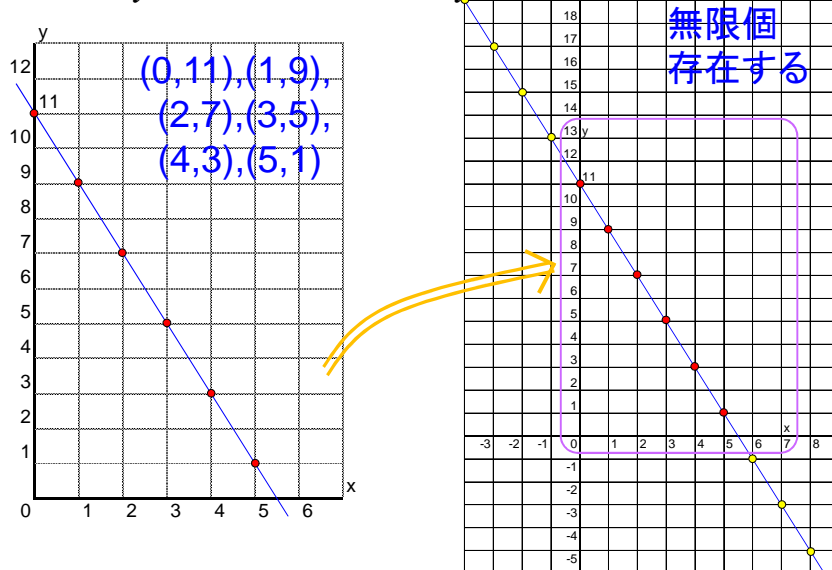
◆ パラメータの推定(続き)

◆ FA

- $Z = AF + E$
(pxn) (pxk)(kxn) (pxn)
 - Z: 測定値。pxn行列。顕在変量。
 - A: 因子負荷量。pxk行列。
 - F: 共通因子。kxn行列。
 - E: 独自因子。pxn行列。
- } 潜在変量
- k: 縮約次元数。陽に利用者が指定する。
 - $I = TT^{-1}$ を用意すると、 $Z = AF + E = AIF + E = ATT^{-1}F + E = A^*F^* + E$
 - 回転の不定性
 - AとFの組み合わせが無限個存在する ==> 一種の違和感

似たような事例

- ◆ $2x + y = 11$ を満たす x, y の整数解を求めよ。



3. IRTを日本の大学入試に用いる際の注意点

- ◆ IRT: Lord(1952), Lord & Novick(1968)

- ◆ Lord, Frederic M. (1912-2000)
- ◆ Novick, Melvin R. (1932-1986)

◆ FA: Spearman(1904)
 ◆ Spearman Charles E. (1863-1945)

- ◆ LCA: Lazarsfeld(1968)

- ◆ Lazarsfeld, Paul Felix (1901-1976)

- ◆ シンプルなモデル(素朴?)

- ◆ 計算パワーの不十分な時代。モデルの提案として。
- ◆ 今日の学力を表現するにはシンプルなモデルでは?
- 実験室に留めておくべきでは? <=== 実用前段階
- 計算パワーが上がり無理やり計算できる時代だが、真値が求まる保証がなく、また、モデル自身が改善されたわけでもない

3. IRTを日本の大学入試に用いる際の注意点

- ◆ BILOG-MGやRのパッケージ
 - ◆ 「真値」を提示できているのか? できるのか?
 - ◆ ブラックボックス化の怖さ、統計ソフトウェア黎明期(1980年代)にも叫ばれた
- ◆ 十分なサンプルサイズがあれば良いという問題ではない: 無いのはそれ以前だが
- 一種の鞍点を求めているに過ぎないのでは?
 - ◆ 繰り返し計算の宿命

13

4. IRTの活用場面

- ◆ 測定・評価用ツール(教育、学校現場)として
 - ◆ 修正が効く <=== 教員がそばで指導できるから。多人数にも対応化。教員の支援にも使える。
 - ◆ 達成度試験、資格試験: Item Bankも有用だろう
 - ◆ 初等中等教育の現場での利活用はすぐにでも
 - 学習指導要領が完備された国 = 日本
- ◆ 日本の選抜用ツール(入試、合否)としては不安
 - ◆ パラメータの確度が保証できない以上、受験生の進路選択を誤る可能性をはらんでいる
 - ◆ 誰が責任を取れるのか?
 - ◆ IRT(やCBT)は向かない領域と、私は考える

14

5. まとめに代えて: 数理的見地から

- ◆ 西欧諸外国: 出口管理の国 = 入試はローステークス
 - ◆ その国で使えるものが日本でもそのまま使えるとは限らない。文化に依存。
 - ◆ 入試における共通試験の重みは低い。粒度が荒くても利用可能。
- ◆ 日本: 入口管理の国 = 入試はハイステークス ← **独立**
 - ◆ 大規模入学者選抜の実施環境は頑健に構築する必要
 - ◆ 選抜試験: 複数の単元がまとまって一つの科目を構成。大問主義。
 - ◆ 日本の文化に根ざした技術の確立・利活用を。
- ◆ 峻別: 選抜試験と資格試験は個別に検討
- ◆ 潜在変数を影響力の大きい社会制度(入試)に出す恐ろしさ
 - ◆ そこまで危険を犯して導入するメリットはあるのか?
 - ◆ 危ないと判っていて船出をするのか? 子供たちを危険に曝すのか?
- ◆ CBT・Item Bank・IRTとも、将来に向けてさらなる研究を⁵

【参考文献】

- ◆ 丘本 正(1986)、「因子分析の基礎」、日科技連出版社。
- ◆ 芝 祐順編(1991)、「項目反応理論—基礎と応用」、東京大学出版会。
- ◆ 浅野 長一郎 & 江島 伸興(1993)、「潜在構造分析論の現状」、日本統計学会誌 第22巻第3号、PP357-373。
- ◆ 池田 央(1994)、「現代テスト理論」、朝倉書店。
- ◆ 中央教育審議会 (2014)、「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について～すべての若者が夢や目標を芽吹かせ、未来に花開かせるために～」、中央教育審議会答申(中教審第177号)。 https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1354191.htm (2021年7月31日現在)。
- ◆ 宮澤 芳光(2019)、「Computer Based Testing(CBT)を用いたテストの出題」、大学入試センター・アドミッションリーダー研修「入試問題の作成・分析とCBT入門」配布資料、2019年7月20日。
- ◆ 寺尾 尚大(2019)、「CBT・CATとは何か」、日本テスト学会第17回大会、公開シンポジウム2「多面的総合的評価・CBT・アクティブ・ラーニング」配布資料、2019年8月29日。
- ◆ 独立行政法人 大学入試センター(2021)、「大規模入学者選抜におけるCBT 活用の可能性について(報告)」、 https://www.dnc.ac.jp/research/cbt/cbt_houkoku.html (2021年7月31日現在)。

16