

項目プールの拡充法と その留意点

明治学院大学心理学部
川端一光

第8回『日本テスト学会賞』記念講演会およびワークショップ
「項目反応理論を用いたテスト運用の実際」

於：法政大学 市ヶ谷キャンパス

本発表の話題

- 項目プールを拡充する際の留意点について
—いろいろなデザイン例を紹介をしながら
- IRTによって運営される大規模試験での適用を想定。
- 「理論的にできる」とことと「安定的に実践する」ということを区別して考え、特に、後者について解説。

参考書との関連性

- 第11章: 等化
- 第14章: R による等化

をより実践的に読むための
話題提供

「IRTによるテスト運営への意欲度」 を測定する6項目 ※勿論、**架空**です

質問1:基準尺度を作成するために十分なモニタや、試験会場を確保できますか？（その予算を割けますか？） ---- **YES or NO**

質問2:難易度、信頼性の観点から、等質なテストを作成し続ける予算、人材・資源は準備できますか？ ---- **YES or NO**

質問3:項目プールを拡充するために必要なモニタや、試験会場を持続的に確保できますか？（その予算を割けますか？） ---- **YES or NO**

質問4:組織内の意思決定者(あなた)は、IRTについてよく理解していますか。あるいは理解を補助する専門家を配置できますか。 ---- **YES or NO**

質問5:CTTのテストを運営するよりも、コストが格段に高くなることを把握していますか。 ---- **YES or NO**

質問6:質問1から質問5の回答が全てYESであったあなたに質問です。IRTでいったんテストを開始すると、得点が等化されているだけに、**後戻りができません**。それでもIRTで運営しますか？ ---- **それでも YES or NO**

「**それでもYES**」のあなたこそが、
事故のないテスト運営が可能です！

※少し大げさかもしれませんが、そのぐらいの覚悟でいた方が安全ではないでしょうか。

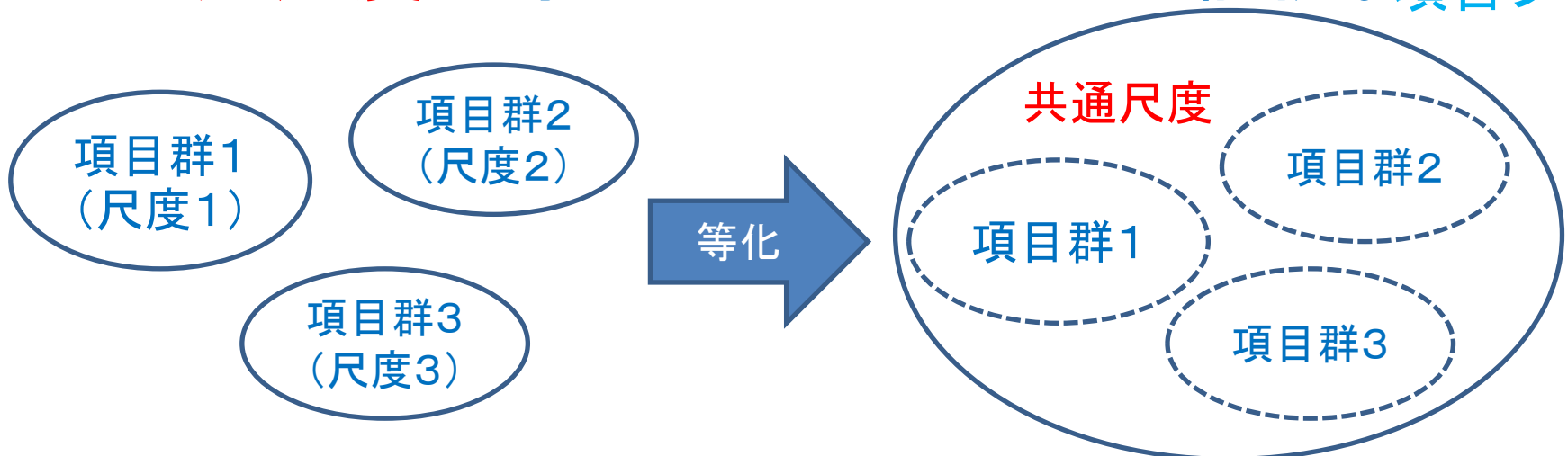
• 「IRT＋等化」は魔法の杖ではない

— 入念な事前準備、厳しい前提条件が満たされてこそ実現可能。

— High Cost and High Return

項目プールとは・・・

- 識別力パラメタ、困難度パラメタが**既知**である項目の集まり(プール)
- 単に既知であるだけでは不十分で、**共通尺度**に等化されていることが前提。項目プール



項目データベース(プール)に含める 項目情報

- ユニークな項目番号
- CTT, IRTの項目統計量
- 項目内容そのもの、選択枝, 正答枝
- 項目の測定領域、内容的分類
- 項目の露呈回数(過去の使用回数)
- この先、当該項目を含める予定のフォームの情報
- 必要回答時間

など...

架空のデータベース例

	A	B	C	D	E	F	G	H	I	J
1	項目番号	項目内容	出題領域	正答	提示回数	正答率	I-T相関	a	b	c
2	ITEM0001	XXXXXXXXXXXXXXXXXXXXXX	A	1	3	0.82	0.43	1.33	-1.52	0.02
3	ITEM0002	XXXXXXXXXXXXXXXXXXXXXX	B	5	1	0.42	0.59	1.53	0.83	0.01
4	ITEM0003	XXXXXXXXXXXXXXXXXXXXXX	A	4	1	0.62	0.26	0.98	-0.34	0.03
5	ITEM0004	XXXXXXXXXXXXXXXXXXXXXX	C	2	1	0.23	0.53	1.12	2.38	0.21
6	ITEM0005	XXXXXXXXXXXXXXXXXXXXXX	E	2	3	0.43	0.36	0.86	1.43	0.15
7	ITEM0006	XXXXXXXXXXXXXXXXXXXXXX	B	1	2	0.64	0.32	0.79	-2.53	0.2

・これはごく簡単な例。実際のテスト運営の状況に合わせて、より多くの情報(変数)を含めていく必要がある。

・どのような情報を保存しておくことが好ましいかについて、**等化計画の相当先を見越した上で、データベースのフォーマットを作成することが必要**。→後になってのデータベースの改修は、非常に煩雑で、かつ、**危険**。

※更新作業時以外には上書き保存ができないような工夫も必要

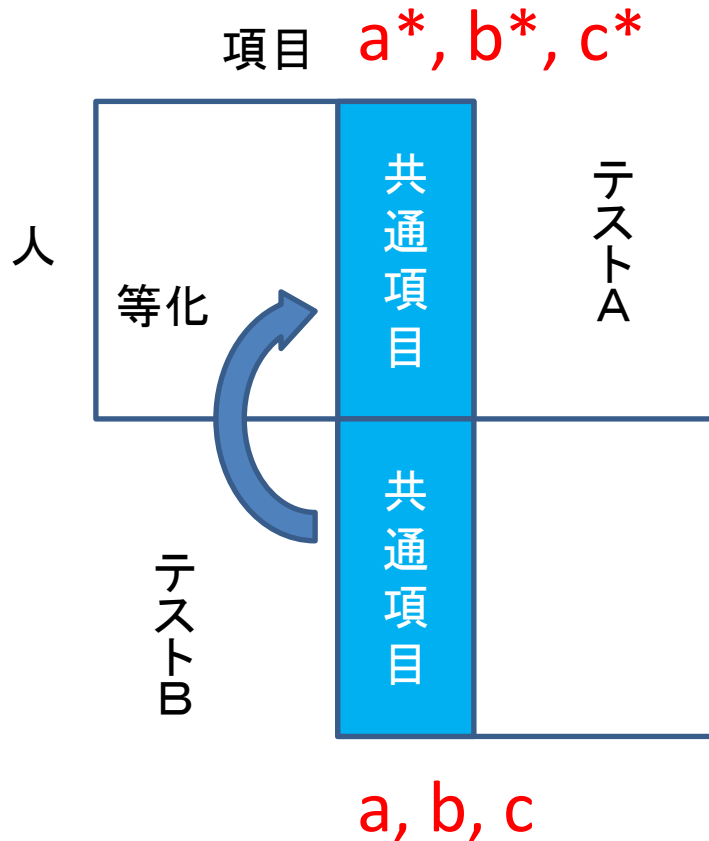
・項目プールが**大規模**で、さらに項目情報が十分用意されているのならば、緻密な複数版自動構成を行うことが可能かもしれない。

等化計画の概略

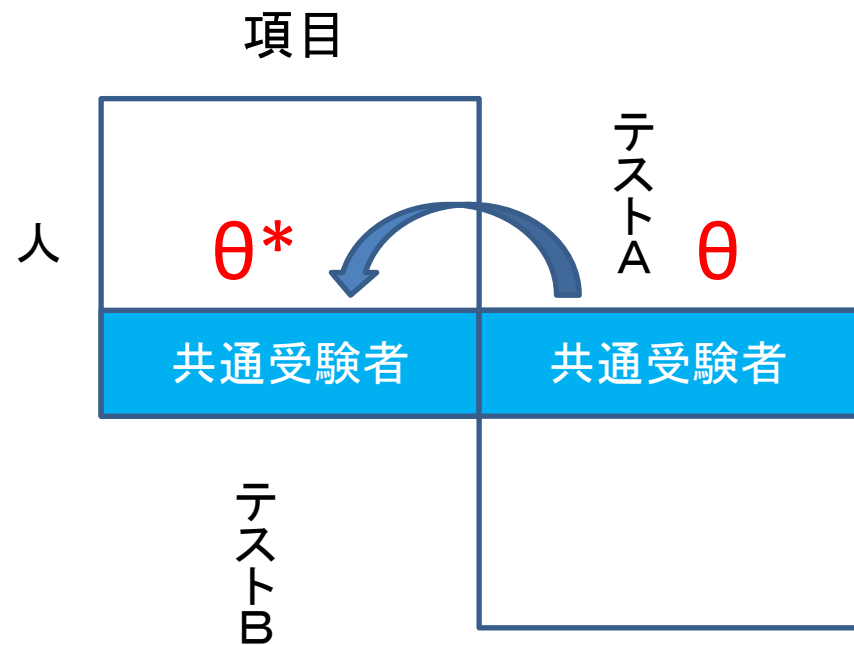
※後の説明のため

- 共通項目法

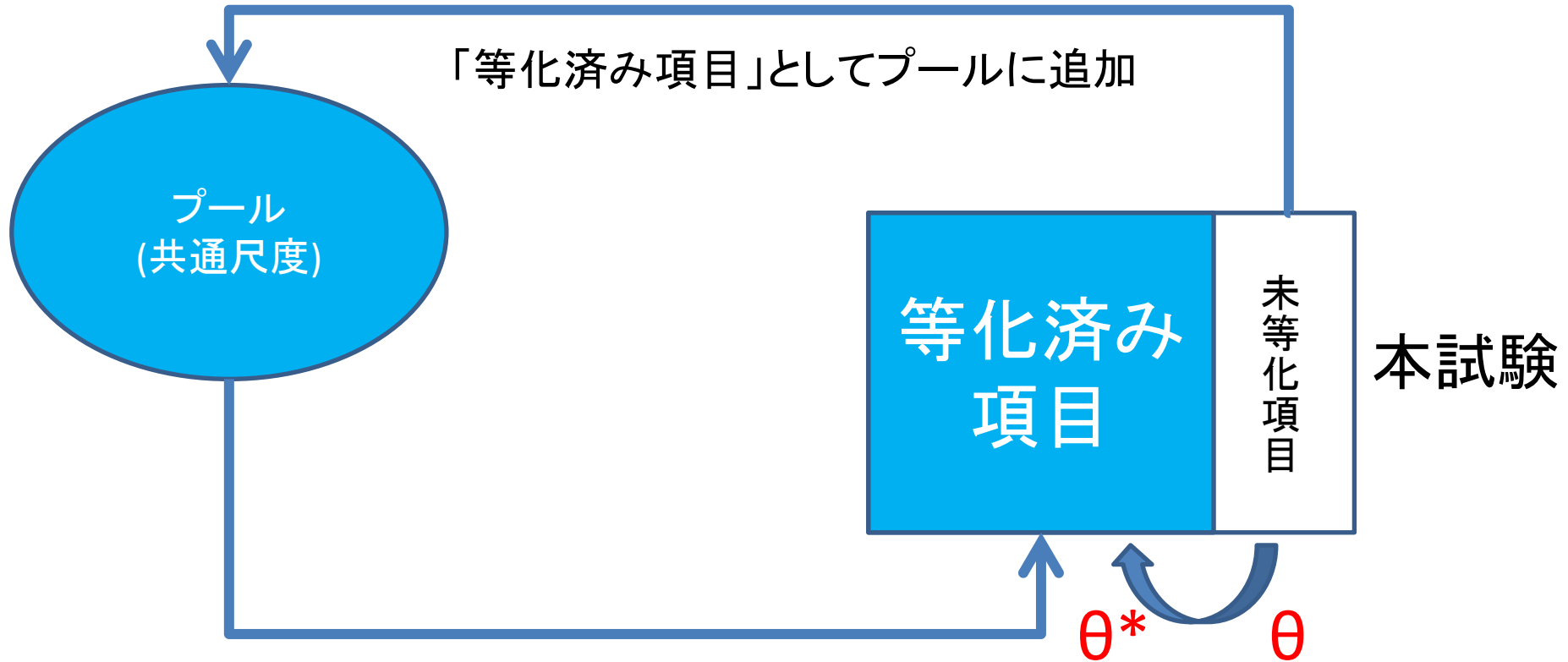
*(アスタリスク)は等化後という意味



- 共通受験者法



拡充法1. 本試験に未等化項目を 忍ばせる

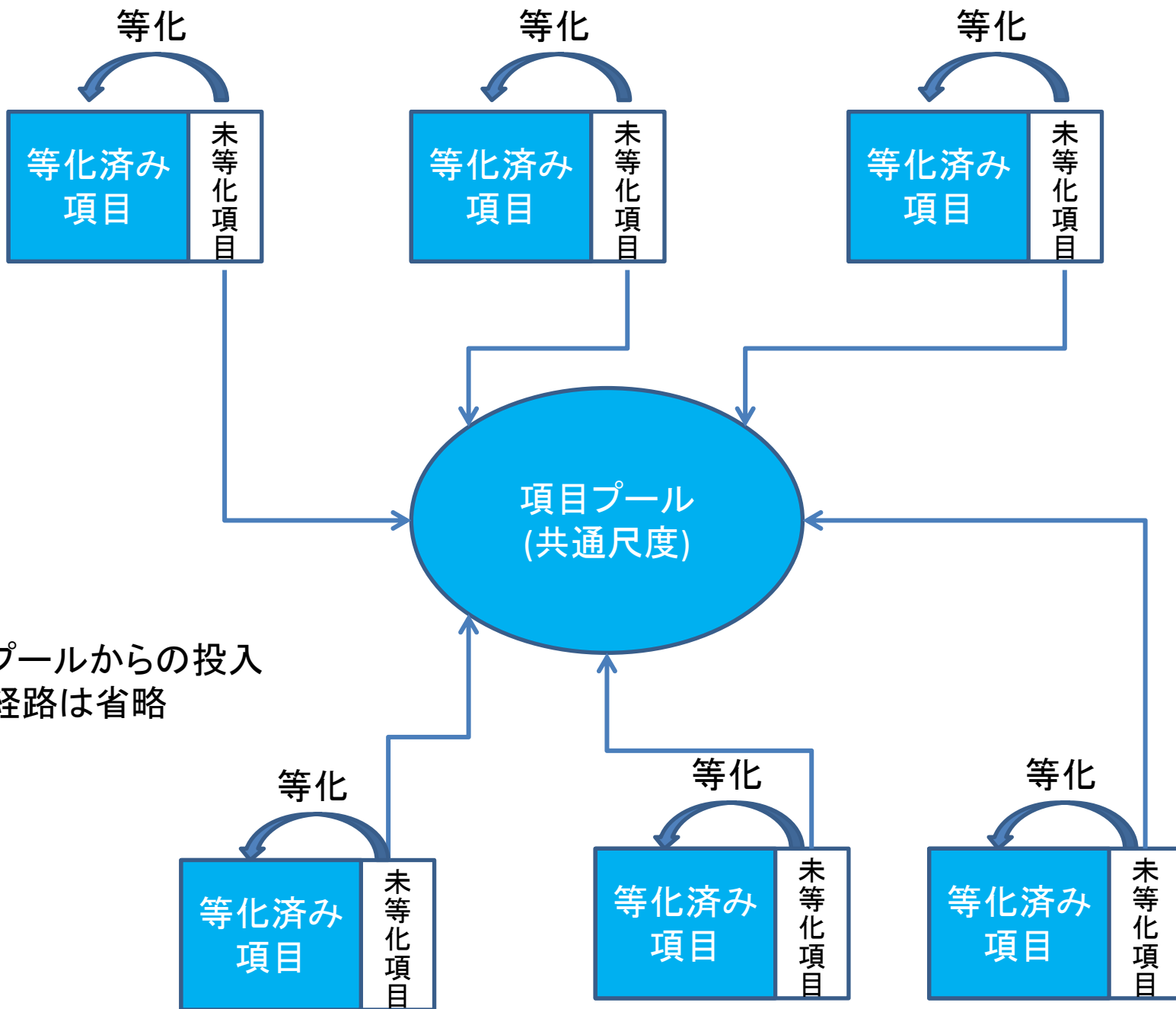


版構成の後投入

共通受験者法で等化

$$\theta^* = k\theta + l$$

$$a^* = a/k, b^* = kb + l$$



拡充法1の留意点

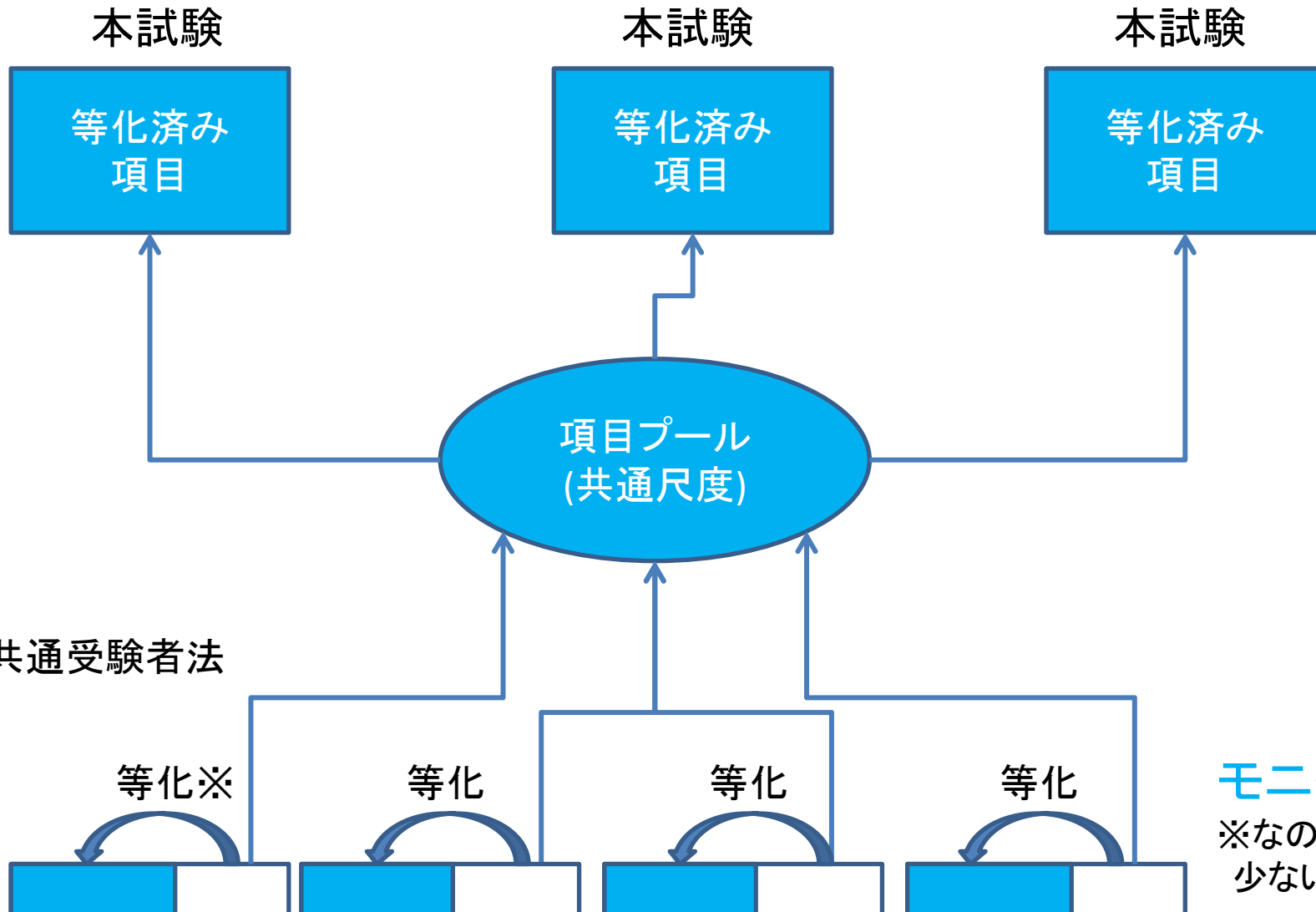
【利点】

- 本試験受験者の情報全てを利用して(共通受験者法で)、等化が可能。等化が安定的に行える。
- 本試験に忍ばせるので、プール拡充のための受験者確保を行わなくて良い。コスト減。

【留意点】

- 多数の受験者に未等化項目が露呈する。項目の賞味期限が早くに来てしまう。
- 本試験での新規項目数を増やすと、個人の能力推定に利用できる等化済み項目数が減ってしまう(試験時間は一定なので)。

拡充法2: モニタ試験を利用する



拡充法2の留意点

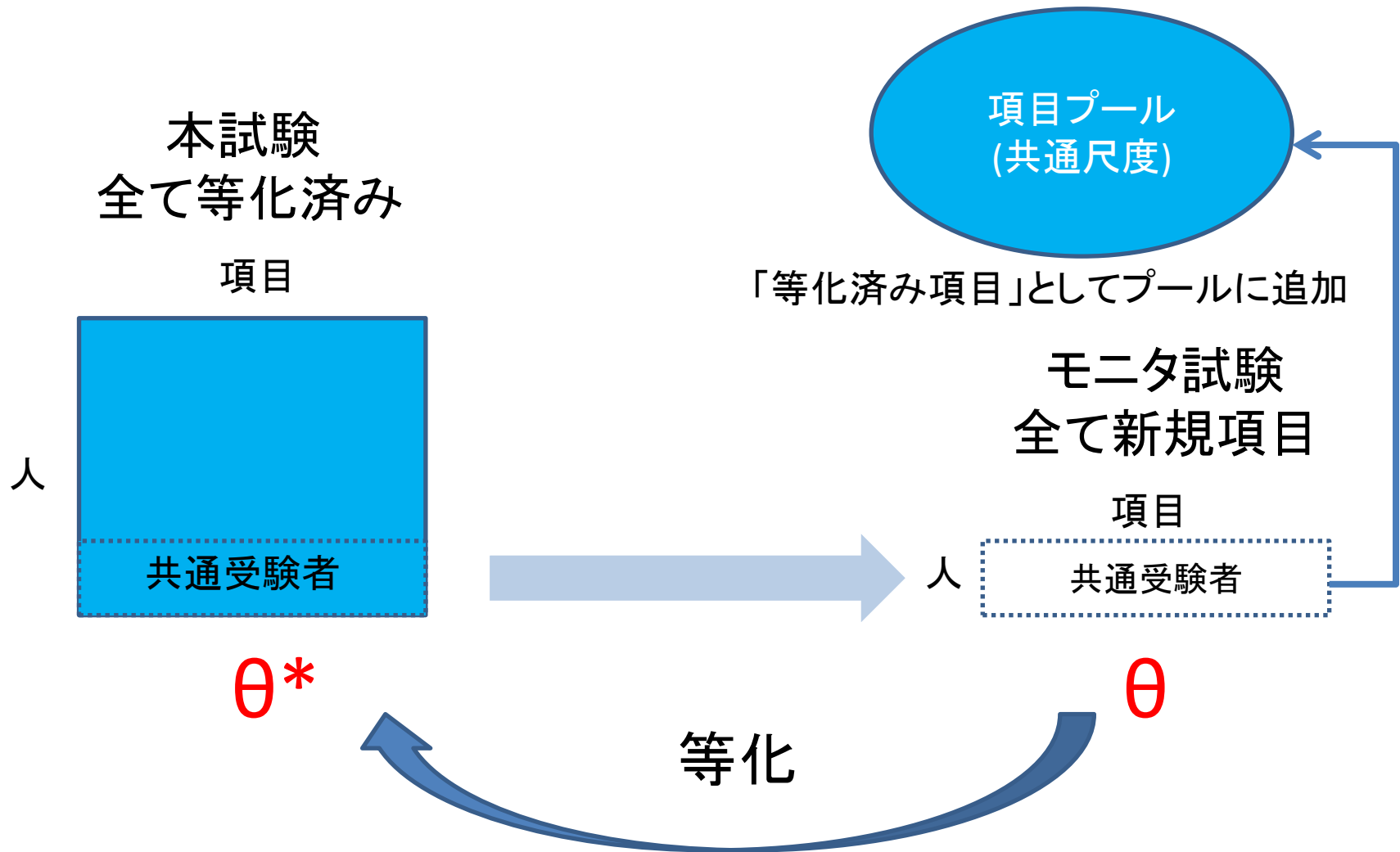
【利点】

- モニタ集団は小規模なので、項目露呈の影響を最小限に抑えられる。

【留意点】

- モニタや試験会場の持続的確保に多大な費用・労力を要する。
- 本試験受験者に対してモニタは少数であるため、等化係数の推定誤差が大きくなる。
- 基準集団に対してモニタ集団の能力が極端に偏っていると、等化係数の推定に影響する可能性。
- 本試験と条件を揃える工夫が必要
 - 受験者の動機づけ, 時間設定

拡充法3: 本試験とモニタ試験の共通受験者を利用する



拡充法3の留意点

【利点】

- 試験時間を一定として考えた場合、モニタ試験で提示できる新規項目数が最大となる。

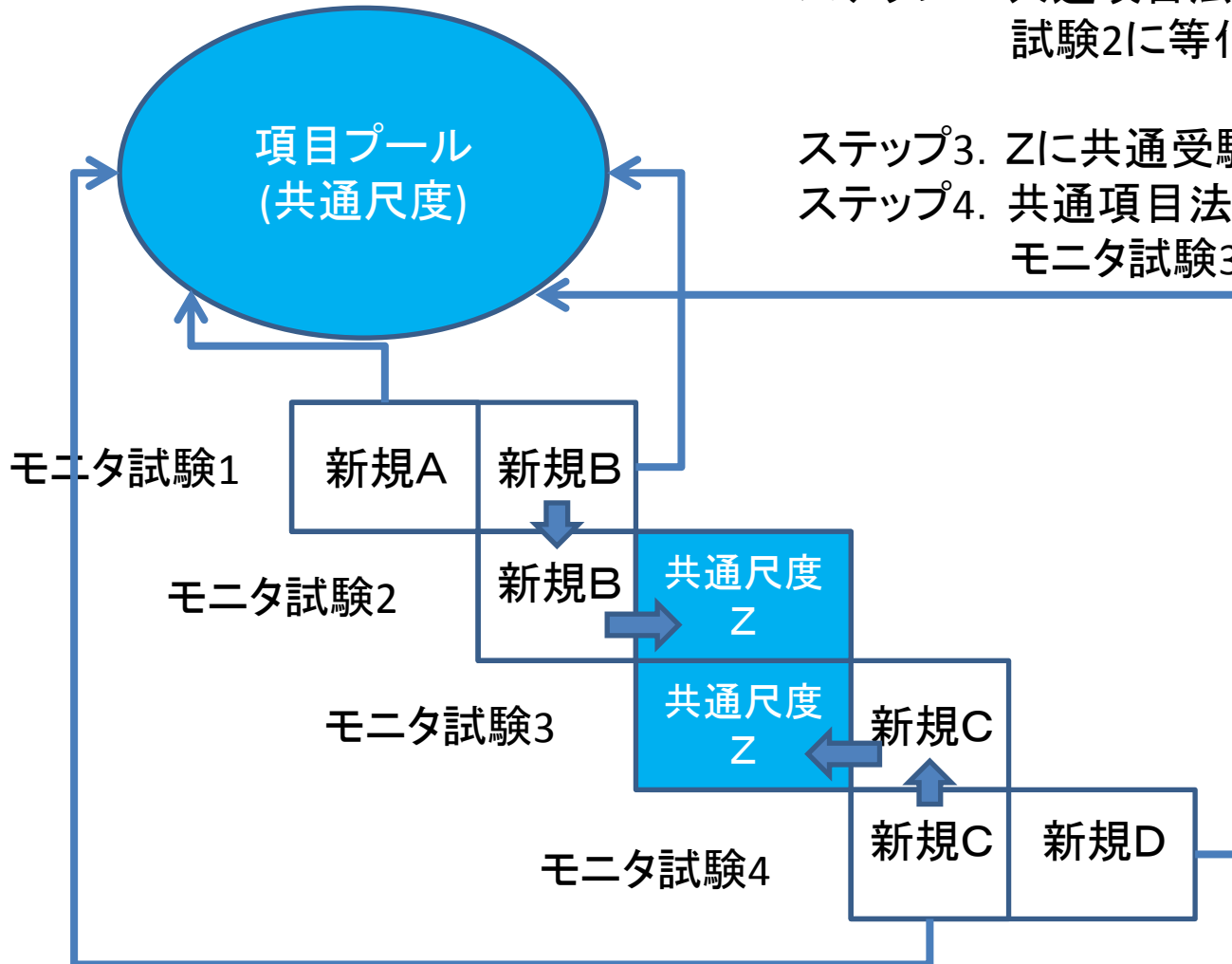
【留意点】

- 本試験とモニタ試験の両方を受験するモニタ(共通受験者)を確保する必要がある。
- 本試験とモニタ試験の間に期間がある場合には、受験者の能力変化について配慮する必要がある。
- モニタ試験を本試験の前に行う場合、等化済み θ を受験者に返せない(受験の動機づけとの関連)。

拡充法4: モニタ試験間で等化

ステップ1. Zに共通受験者法で新規Bを等化
ステップ2. 共通項目法でモニタ試験1をモニタ試験2に等化

ステップ3. Zに共通受験者法で新規Cを等化
ステップ4. 共通項目法でモニタ試験4をモニタ試験3に等化



拡充法4の留意点

【利点】

- 共通尺度の露呈を最小限に、新規項目を大量に確保できる。

【留意点】

- モニタや会場の確保に苦勞する。
- 等化を繰り返すので(chained equating)、等化の誤差が積み重なる。

拡充法5: 固定母数法で等化する

以下の**大行列**を利用し、共通尺度Zの項目母数を固定して、新規項目を自由推定

モニタ試験1	新規A	新規B	欠測		
モニタ試験2	欠測	新規B	共通尺度Z	欠測	
モニタ試験3	欠測		共通尺度Z	新規C	欠測
モニタ試験4	欠測			新規C	新規D

拡充法5の留意点

【利点】

- 共通尺度の露呈を最小限に、新規項目を大量に確保できる。

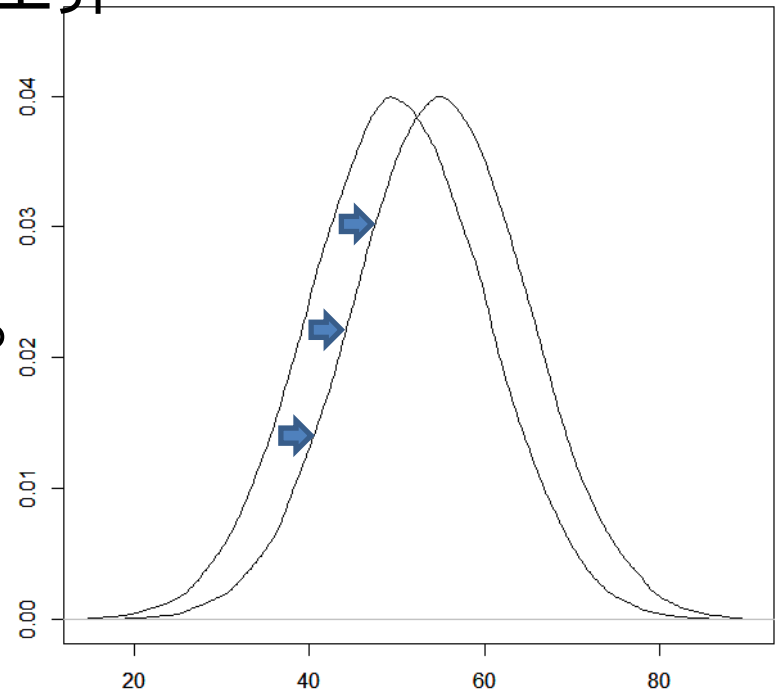
【留意点】

- モニタや会場の確保に苦勞する。
- モニタ集団の能力分布が異なる場合、 θ の分布に対して同じ正規分布を仮定し、項目パラメタの推定を行うことに無理が生じる。
- モニタ試験の数が増えると推定が大変。

基準集団の見直し

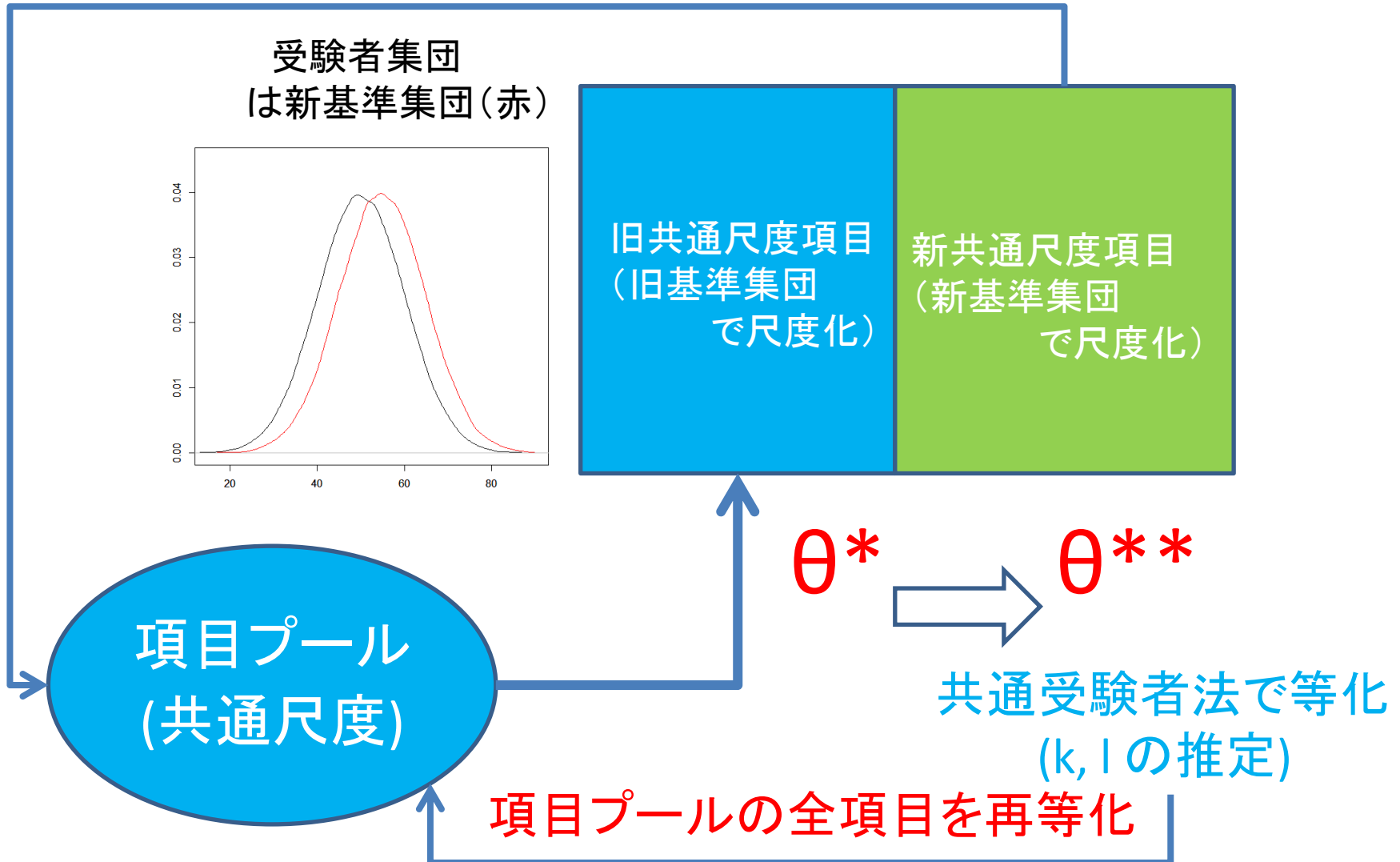
- 試験開始後、試験対策が進み、受験者集団の能力の平均が高くなる可能性
→問題の難易度も経年で少しずつ上昇

- 基準集団の能力分布が相対的に低くなってしまう。
→基準集団の見直し&項目プールの再等化



拡充法6: 項目プールの再等化

新共通尺度に等化後の項目プールに併合



拡充法6の留意点

【利点】

- 旧項目プールを破棄する必要はなく、活用できる。

※簡単に破棄することを考えてはいけない。

【留意点】

- 摂氏→華氏のようにモノサシを変えることになるので、場合によっては、「基準を変えた」ということを受験者に伝える必要がある。例：[得点換算表](#)。
- 旧項目プール中に、新基準集団において簡単すぎる項目が生じてくる。※能力分布が無い部分を識別しても意味がない。
- 経年で試験仕様や作題の傾向が少しずつ変わっている場合もありえる。過去の項目をそのまま使い続けて良いのか熟慮すべき。

安定的に項目プールを拡充するために

- 共通項目数と共通受験者数、モニタ数は**かなりの余裕をもって計画する。**

※ノイズの無いシミュレーション研究の結果を過度に当てにしない。最悪の状況でシミュレーションして、頑健な方法を採用したい。

- 「試験時間の観点、受験者の疲労の観点から、共通項目数を減らしたい……」
- 「共通受験者の確保には費用がかかるので、なんとか減らしたい……」

→妥協することで等化係数の推定精度に悪影響。

→研究の場合、等化が失敗すればお蔵入りで済む。

→実務の場合、等化が失敗すると、その補償のために、場合によっては、莫大な費用がかかる。

長期的な等化計画の重要性

- 一度、試験運営がスタートすると、等化している以上、過去の試験との継続性・整合性を常に考えなくてはならない。
- 5年, 10年先までの中長期的な等化計画を立てる必要がある。
- 露呈回数が増え、プールから破棄される項目もあるので、項目数が確実に増えるような計画を無理なく立てる。
- 等化計画の実行可能性については、予算管理者・作題者・分析家の3者が十分に議論して決めるべき。

参考文献

- 池田央 日本版監訳（2008）テスト作成ハンドブック：
発達した最新技術と考え方による公平妥当なテスト作成・
実施・利用のすべて, Steven M. Downing, Thomas M. Haladyna
編集；教育測定研究所.
- 加藤健太郎・山田剛史・川端一光 (2014) Rによる項目
反応理論, オーム社.

ご静聴ありがとうございました。

※資料は電子媒体でお渡しできます。

※WS終了後にお声掛けください。

Mail to : ikkokawa@meijigakuin.ac.jp