

等化,あるいはリンクングについて

光永 悠彦

(名古屋大学大学院教育発達科学研究科)

2021/9/17

日本テスト学会第19回大会

実行委員会企画録画講演

「テストイングの現状と将来展望」

- 等化の必要性と尺度の不定性
- 等化係数による等化の方法
- その他の等化の方法
- 等化計画
- 等化の前提と実践上の注意点
- リンキング

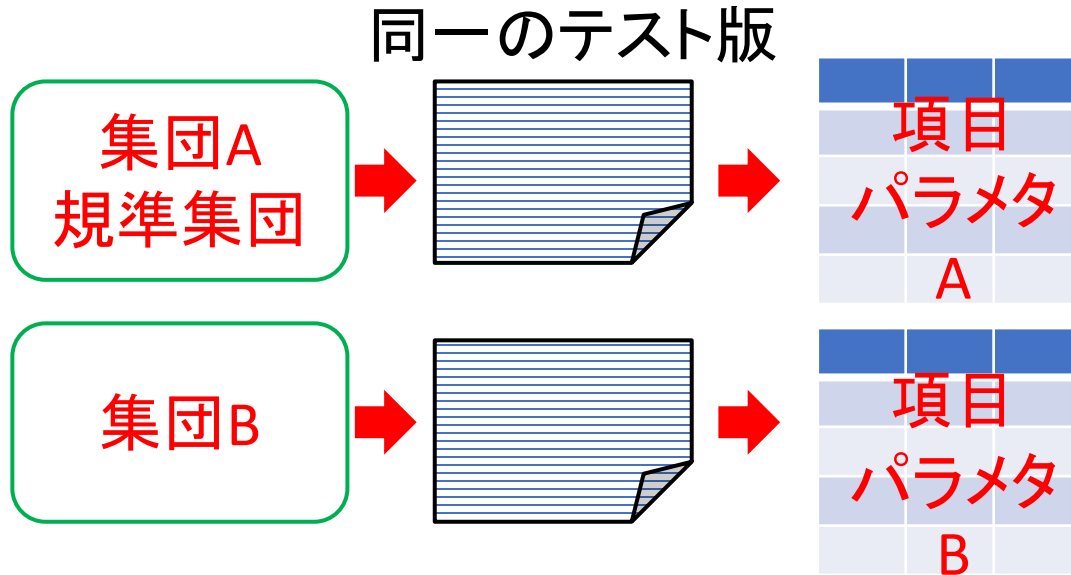
等化の必要性和尺度の不定性

等化,あるいはリンクングについて

- 複数のテスト版の間でスコアの意味が同じになるように、共通の尺度上でスコアを表す変換操作を「等化」と呼ぶ。
- 「等化」は、それぞれの尺度の意味が同一である場合に行われる。→ 尺度の同一性が仮定されない場合の共通尺度化は「リンクング」と呼ばれる。
- 項目反応理論(IRT)を用いた尺度化を行う場面において、複数の集団から得られた正誤データを用いて項目パラメタを推定した結果を、統一された尺度上で表示することが行われる。

等化の必要性

集団における平均的能力の違いによって推定される困難度が変わってくる。



集団Aの尺度で表現された集団Bの困難度は、もとの集団Bの困難度に1を足すことで得られる

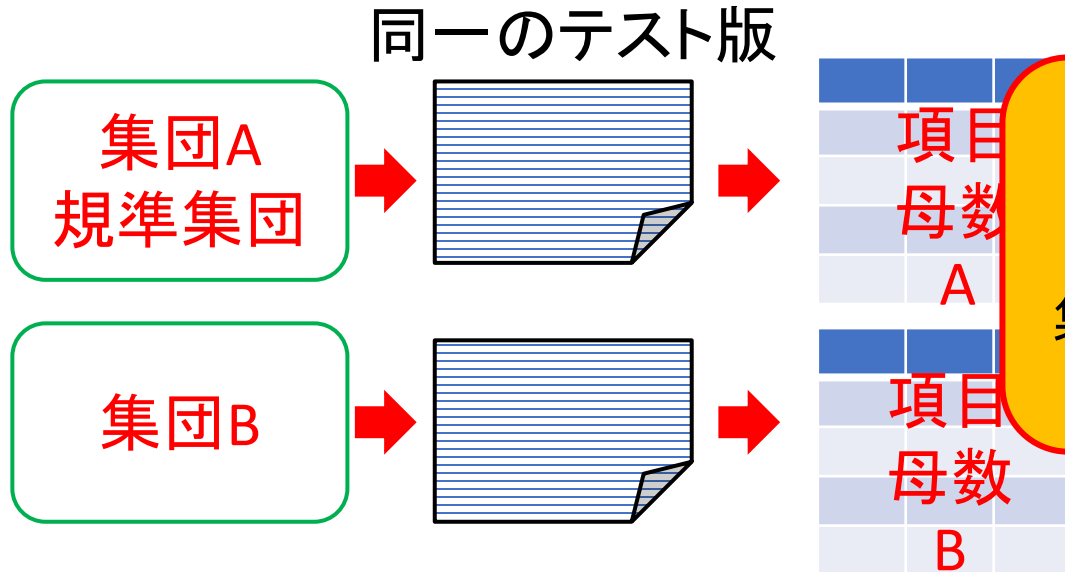
集団A	識別力	困難度
項目1	0.8	-1.2
項目2	1.2	-0.8
項目3	0.8	0.4
項目4	1.0	1.2

集団B	識別力	困難度
項目1	0.8	-2.2
項目2	1.2	-1.8
項目3	0.8	-1.4
項目4	1.0	0.2

集団Bでの困難度の方が集団Aより低い
= 集団Bの方が集団Aより能力が高い

等化の必要性

集団における平均的能力の違いによって推定される困難度が変わってくる。



集団Aの尺度上における
集団Bにおける困難度 =
集団Bにおける困難度 + 1

集団A	識別力	困難度
項目1	0.8	-1.2
項目2	1.2	-0.8
項目3	0.8	0.4
項目4	1.0	1.2
集団B	識別力	困難度
項目1	0.8	-2.2
項目2	1.2	-1.8
項目3	0.8	-1.4
項目4	1.0	0.2

集団Aの尺度で表現された集団Bの困難度は、もとの集団Bの困難度に1を足すことで得られる

集団Bでの困難度の方が集団Aより低い
= 集団Bの方が集団Aより能力が高い

尺度変換による能力値尺度の共通尺度化

- 2つの「平行テスト」A・Bを考える。
- A・B双方を受験した受験者について考える。
- 受験者の能力値 θ について、標準正規分布であると仮定。
- 受験者の能力以外の要因で、 θ がばらつかないと仮定。
- 両方を同時に受験した者の θ_i に関して

$$\frac{\theta_{iB} - \mu_B}{\sigma_B} = \frac{\theta_{iA} - \mu_A}{\sigma_A}$$

という関係が成り立つ(μ_A 及び μ_B はA・Bそれぞれのテスト版における θ の母平均, σ_A 及び σ_B は母標準偏差)。

ここから

$$\theta_{iA} = \frac{\sigma_A}{\sigma_B} \theta_{iB} + \mu_A - \frac{\sigma_A}{\sigma_B} \mu_B$$

という関係がわかる。

$$K = \frac{\sigma_A}{\sigma_B}$$
$$L = \mu_A - K\mu_B$$

とおくことによって、以下の式を得る。

$$\theta_{iA} = K\theta_{iB} + L$$

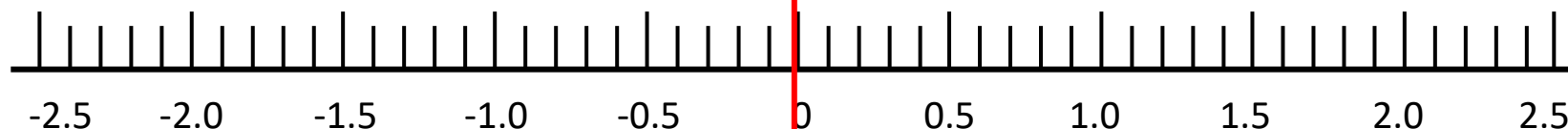
等化係数

*K*倍して*L*を足す、という操作で、
テスト版A受験者の*θ*の尺度上で、
テスト版B受験者の*θ*の尺度が
表現できる。

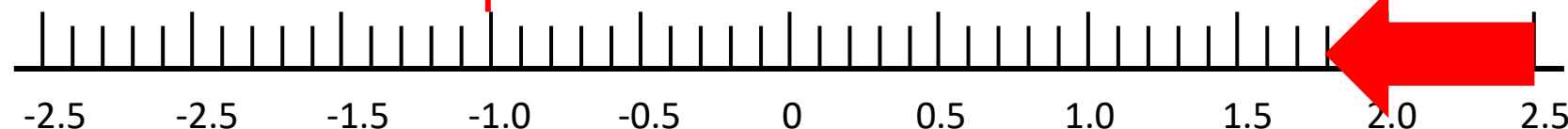
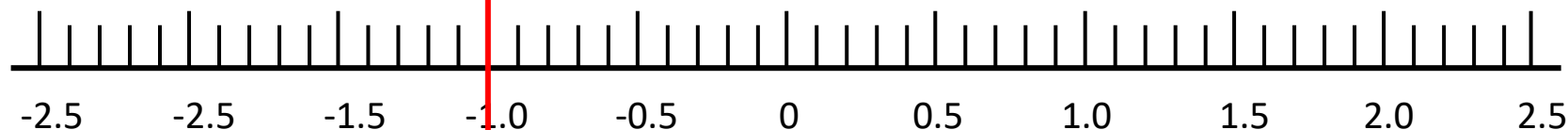
尺度変換による能力値尺度の共通尺度化

集団A

(=標準集団)での θ 尺度



集団Bでの θ 尺度

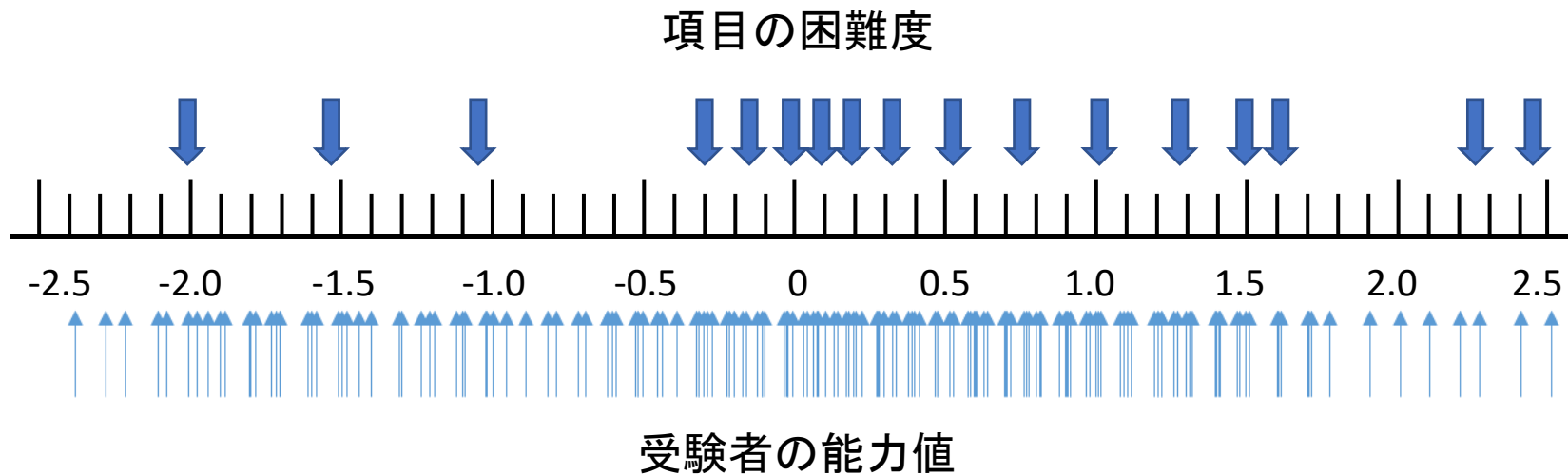


集団Bでの θ 尺度を集団Aの尺度で表したものの

集団Bでの困難度の方が集団Aより低い
= 集団Bの方が集団Aより能力が高い

IRTにおける困難度と能力値の関係

- IRTでは、項目困難度と受験者の能力値 θ を、それぞれ独立に、同一の尺度上で推定する。
- IRTに基づくテストであれば、 θ の尺度変換は、**困難度 (項目パラメタ)の変換を行う**ことで達成される。



尺度の不定性

$$\theta^* = K\theta + L \longrightarrow \theta = \frac{1}{K}\theta^* - \frac{L}{K}$$

項目反応関数

$$\begin{aligned} P_j(\theta) &= \frac{1}{1 + \exp\left(-Da_j(\theta - b_j)\right)} \\ &= \frac{1}{1 + \exp\left(-Da_j\left(\frac{1}{K}\theta^* - \frac{L}{K} - b_j\right)\right)} \\ &= \frac{1}{1 + \exp\left(-D\frac{a_j}{K}\left(\theta^* - (L + Kb_j)\right)\right)} \end{aligned}$$

$$a_j^* = \frac{a_j}{K}$$

$$b_j^* = Kb_j + L$$

と変換すれば

$$P_j^*(\theta^*) = \frac{1}{1 + \exp\left(-Da_j^*(\theta^* - b_j^*)\right)}$$

$P_j^*(\theta^*)$ と $P_j(\theta)$ は値が一致する。

IRTによる項目パラメタの尺度変換

等化係数 K, L を用いて、項目パラメタの値を変換することができる。

$$\begin{aligned}a_{jB}^* &= a_{jB}/K \\ b_{jB}^* &= Kb_{jB} + L\end{aligned}$$

テスト版Aの尺度上で表現された、テスト版Bの項目パラメタの変換後の値：
 a_{jB}^* (識別力パラメタ) b_{jB}^* (困難度パラメタ)

テスト版Bのサンプルから得られた項目パラメタ(変換前の値)：
 a_{jB} (識別力パラメタ) b_{jB} (困難度パラメタ)

等化係数による等化の方法

等化係数 K と L の求め方

- Mean/Sigma法 (Marco, 1977)
 - 困難度パラメタのみを用いた方法.
- Mean/Mean法 (Loyd & Hoover, 1980)
 - 識別力パラメタを加味した方法.

【特性曲線変換法】(characteristic curve transformation method)

- Haebara法 (Haebara, 1980)
- Stocking-Lord法 (Stocking & Lord, 1983)
 - 規準集団から得られた項目特性曲線に, 他のテスト版から得られた項目特性曲線を近づける方法.

Mean-Sigma法

θ の 変換

$$\frac{\theta_{iB} - \mu_B}{\sigma_B} = \frac{\theta_{iA} - \mu_A}{\sigma_A} \longrightarrow \theta_{iA} = K\theta_{iB} + L$$

$$K = \frac{\sigma_A}{\sigma_B}$$
$$L = \mu_A - K\mu_B$$

μ_A 及び μ_B はA・Bそれぞれのテスト版における θ の母平均、
 σ_A 及び σ_B は母標準偏差。

b_j の 変換

$$\frac{b_{jB} - \mu_{bB}}{\sigma_{bB}} = \frac{b_{jA} - \mu_{bA}}{\sigma_{bA}} \longrightarrow b_{jA} = Kb_{jB} + L$$

$$K = \frac{\sigma_{bA}}{\sigma_{bB}}$$
$$L = \mu_{bA} - K\mu_{bB}$$

μ_{bA} 及び μ_{bB} はA・Bそれぞれのテスト版における b_j の母平均、
 σ_{bA} 及び σ_{bB} は母標準偏差。

Mean-Sigma法

$$K = \frac{\sigma_{bA}}{\sigma_{bB}} \quad \text{の代わりに} \quad \hat{K} = \frac{s_{bA}}{s_{bB}}$$
$$L = \mu_{bA} - K\mu_{bB} \quad \text{を用いる。} \quad \hat{L} = \bar{b}_A - \hat{K}\bar{b}_B$$

\bar{b}_A 及び \bar{b}_B はA・Bそれぞれのテスト版における b_j の平均,
 s_{bA} 及び s_{bB} は標準偏差。

同一のテスト版による
項目パラメタ推定値

集団A	識別力	困難度
項目1	0.8	-1.2
項目2	1.2	-0.8
項目3	0.8	0.4
項目4	1.0	1.2

(\bar{b}_A, s_{bA})

$$a_{jB}^* = a_{jB} / \hat{K}$$

$$b_{jB}^* = \hat{K} b_{jB} + \hat{L}$$

集団B	識別力	困難度
項目1	0.8	-2.2
項目2	1.2	-1.8
項目3	0.8	-1.4
項目4	1.0	0.2

(\bar{b}_B, s_{bB})

$$\hat{K} = \frac{s_{bA}}{s_{bB}}$$

$$\hat{L} = \bar{b}_A - \hat{K} \bar{b}_B$$

テスト版Aの尺度上で表現された、テスト版Bの項目パラメタの変換後の値:

a_{jB}^* (識別力パラメタ) b_{jB}^* (困難度パラメタ)



等化係数を用いて変換

テスト版Bのサンプルから得られた項目パラメタ(変換前の値):

a_{jB} (識別力パラメタ) b_{jB} (困難度パラメタ)

a_{jB}^* (識別力) b_{jB}^* (困難度)

集団A	識別力	困難度
項目1	0.8	-1.2
項目2	1.2	-0.8
項目3	0.8	0.4
項目4	1.0	1.2

	識別力	困難度
項目1	0.8	-1.2
項目2	1.2	-0.8
項目3	0.8	0.4
項目4	1.0	1.2

テスト版Aの尺度上で表現された、テスト版Bの項目パラメタの変換後の値

a_{jB} (識別力) b_{jB} (困難度)

集団B	識別力	困難度
項目1	0.8	-2.2
項目2	1.2	-1.8
項目3	0.8	-1.4
項目4	1.0	0.2

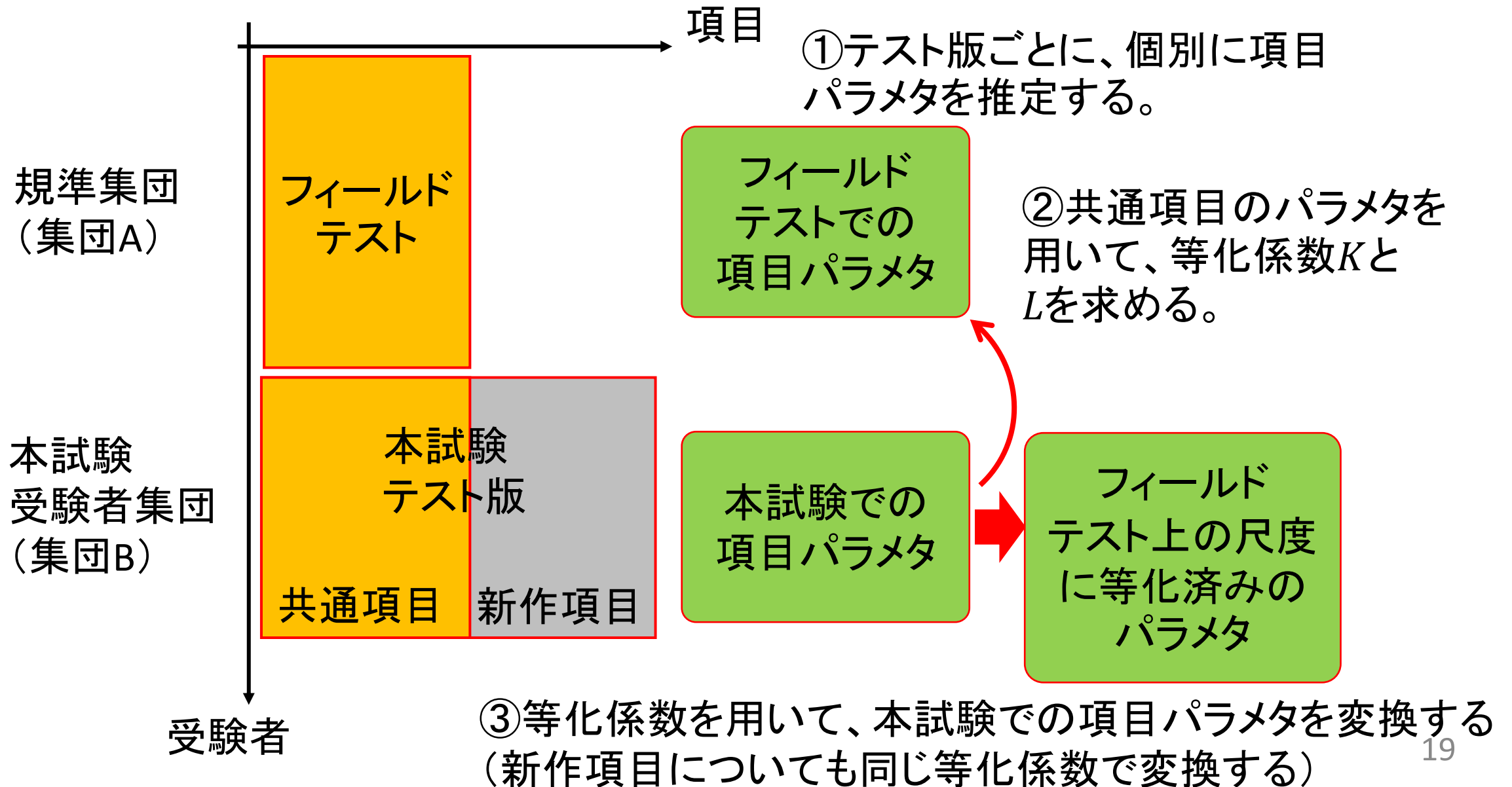
$$a_{jB}^* = a_{jB} / \hat{K}$$

$$b_{jB}^* = \hat{K} b_{jB} + \hat{L}$$

$$\hat{K} = \frac{S_{bA}}{S_{bB}}$$

$$\hat{L} = \bar{b}_A - \hat{K} \bar{b}_B$$

個別推定による等化



Mean-Mean法

$$a_{jB}^* = a_{jB}/K \quad \longrightarrow \quad K = a_{jB}/a_{jB}^* \quad \longrightarrow \quad K = \mu_{aB}/\mu_{aA}$$
$$L = \mu_{bA} - K\mu_{bB}$$

$$\hat{K} = \frac{\bar{a}_B}{\bar{a}_A}$$
$$\hat{L} = \bar{b}_A - \hat{K}\bar{b}_B$$

\bar{b}_A 及び \bar{b}_B は、A・Bそれぞれのテスト版における b_j の平均、
 \bar{a}_A 及び \bar{a}_B は、A・Bそれぞれのテスト版における a_j の平均。

Haebara 法

テスト版Aを受験した群において、

能力値 θ_{iA} をもつ受験者が項目 j に正答する確率: $P_j(\theta_{iA})$

テスト版Bを受験した群において、

能力値 θ_{iB} をもつ受験者が項目 j に正答する確率: $P_j(\theta_{iB})$

テスト版Bの項目パラメタをテスト版Aのパラメタに
等化した結果を用いて表現した、能力値 θ_{iA} をもつ
受験者が項目 j に正答する確率:

$$P_j^*(\theta_{iA})$$

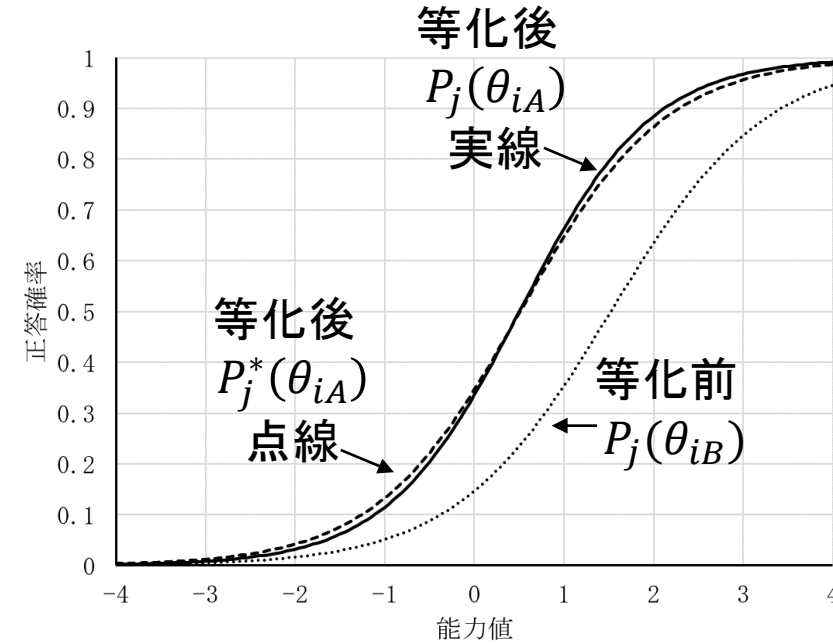
もし、等化が理論通りに行われたとしたら、 θ の全域において

$$P_j(\theta_{iA}) - P_j^*(\theta_{iA}) = 0$$

となるはずであるが、実際にはそうならない。

そこで、両者の差の指標 $\Delta 1_j(\theta_{iA})$ を考える。

$$\Delta 1_j(\theta_{iA}) = \{P_j(\theta_{iA}) - P_j^*(\theta_{iA})\}^2$$

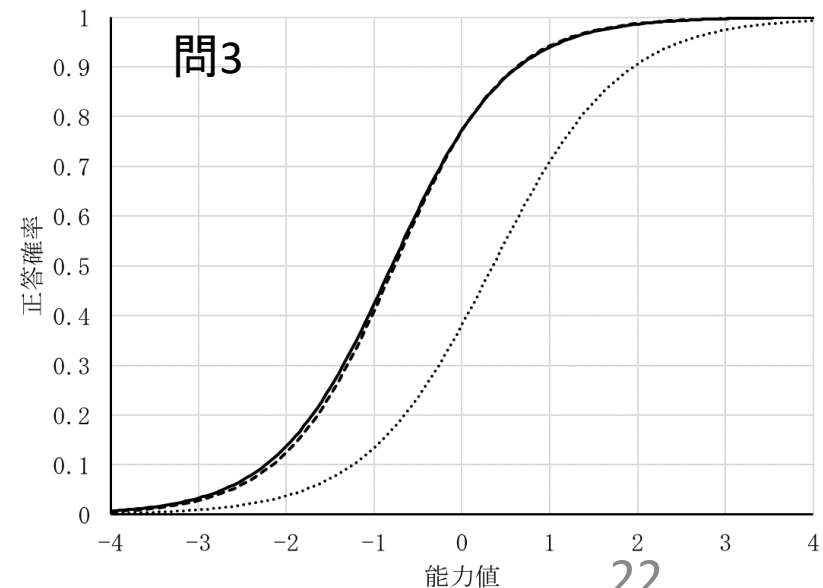
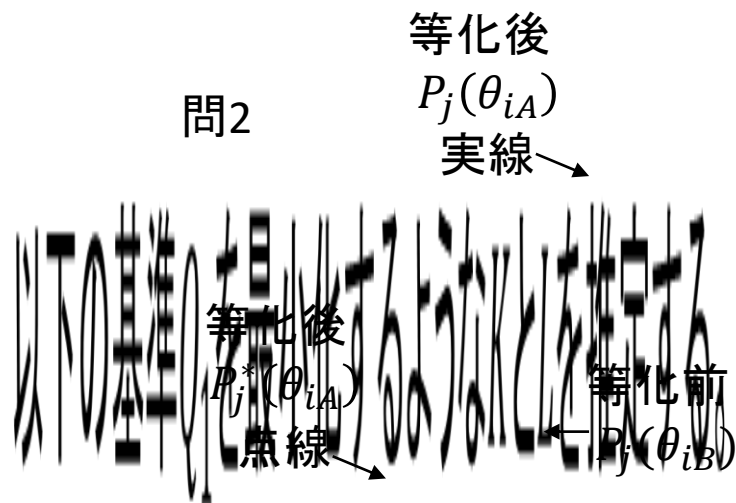
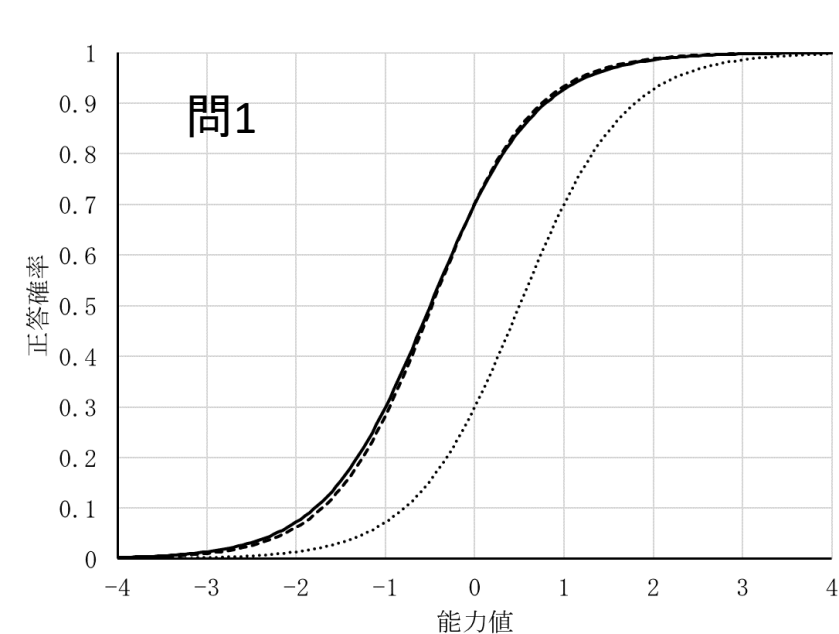


Haebara 法

以下の基準 Q_1 を最小化するような K と L を推定する。

$$Q_1 = \int_{-\infty}^{\infty} \sum_{j=1}^J \Delta 1_j(\theta_A) g(\theta_A) d\theta_A$$

ただし $g(\theta_A)$ は能力値 θ_A の母集団分布を表す。



Haebara 法

テスト版Aで算出されたスコアをテスト版Bの尺度に変換した結果と、テスト版Bで算出されたスコアをテスト版Aの尺度に変換した結果は、同じでなければならない(等化の対称性)。

テスト版Bを受験した群において、
能力値 θ_{iB} をもつ受験者が項目 j に正答する確率: $P_j(\theta_{iB})$

テスト版Aの項目パラメタをテスト版Bのパラメタに等化した結果
を用いて表現した、能力値 θ_{iB} をもつ受験者が項目 j に正答する
確率: $P_j^{**}(\theta_{iB})$

$$\Delta 2_j(\theta_{iB}) = \{P_j(\theta_{iB}) - P_j^{**}(\theta_{iB})\}^2$$

Haebara 法

$Q = Q_1 + Q_2$ を最小化するような K と L を推定する。

$$Q_2 = \int_{-\infty}^{\infty} \sum_{j=1}^J \Delta z_j(\theta_B) g(\theta_B) d\theta_B$$

ただし $g(\theta_B)$ は能力値 θ_B の母集団分布を表す。

実際の推定にあたっては、求積点を区切って積分を和に置き換え、
区分求積法を用いる。

Stocking-Lord 法

差の指標 $\Delta 1_j(\theta_{iA})$ 及び $\Delta 2_j(\theta_{iA})$ を、以下のように定める。
最適化手法については Haebara法と同様。

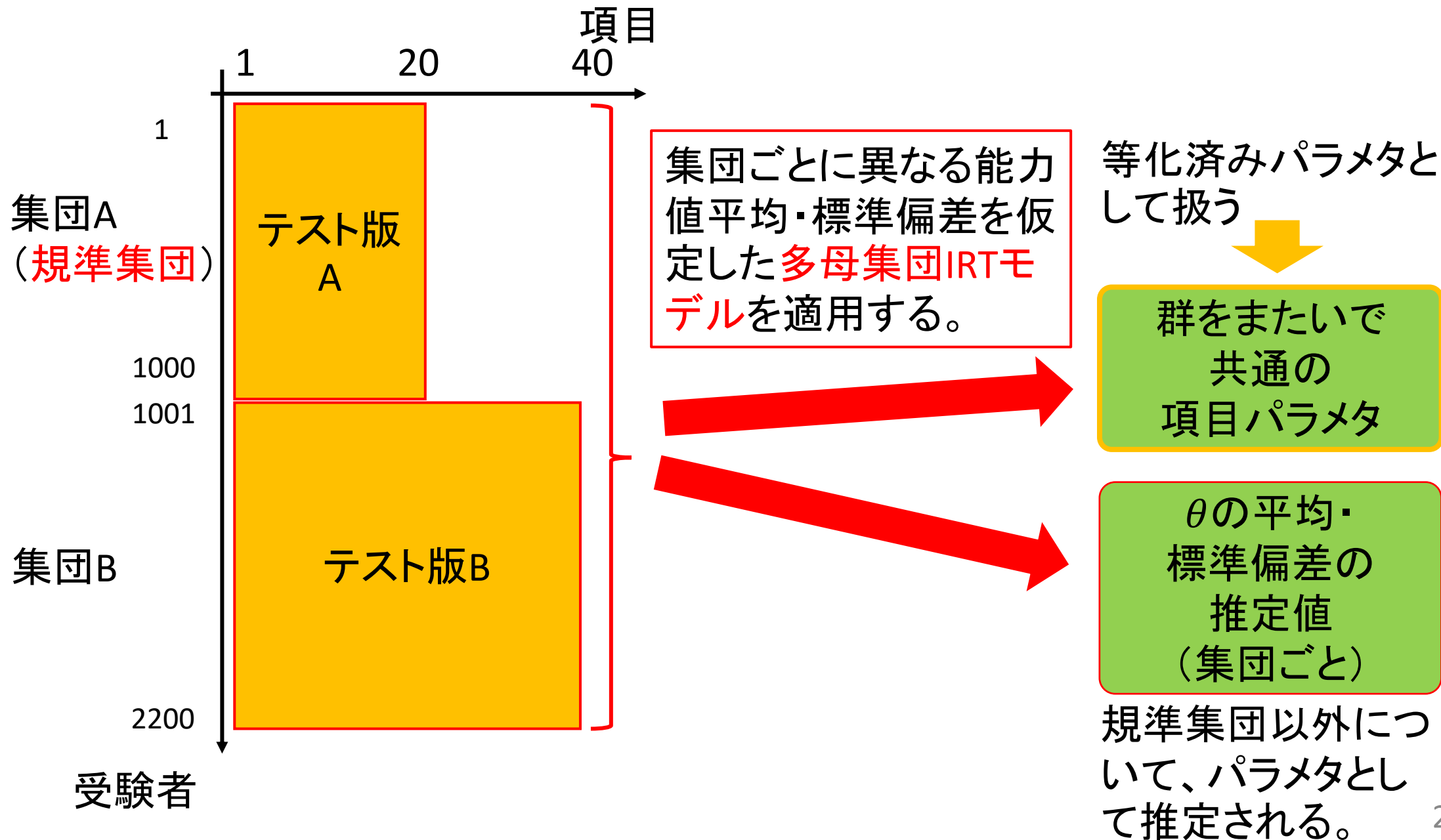
$$\Delta 1_j(\theta_{iA}) = \left\{ \sum_{j=1}^J P_j(\theta_{iA}) - \sum_{j=1}^J P_j^*(\theta_{iA}) \right\}^2$$

$$\Delta 2_j(\theta_{iB}) = \left\{ \sum_{j=1}^J P_j(\theta_{iB}) - \sum_{j=1}^J P_j^{**}(\theta_{iB}) \right\}^2$$

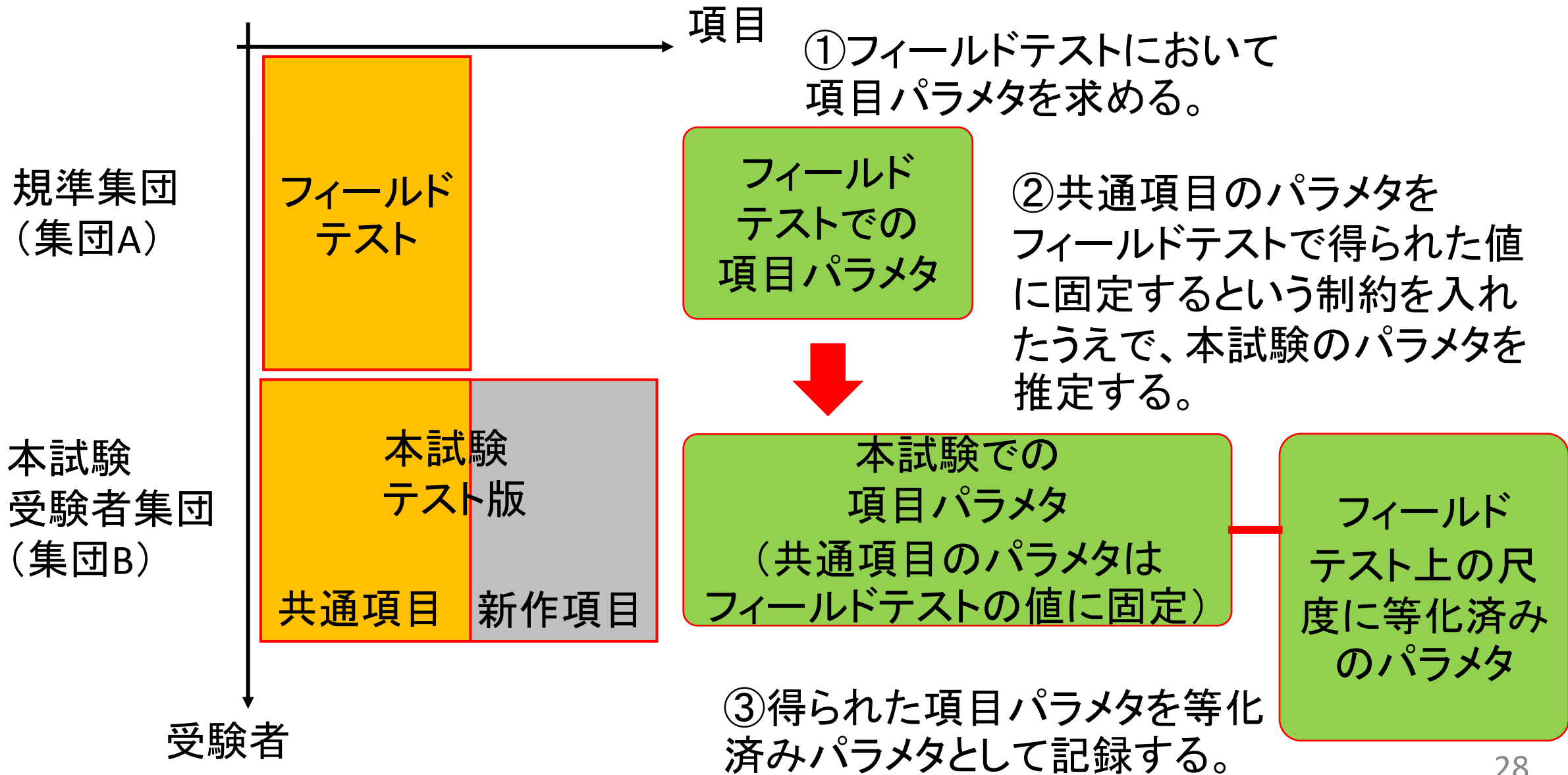
項目特性関数の差ではなく、テスト情報量曲線の差を最小化するような変換を求める。

その他の等化の方法 (同時推定, 項目パラメタ固定法)

多母集団IRTモデルと同時推定

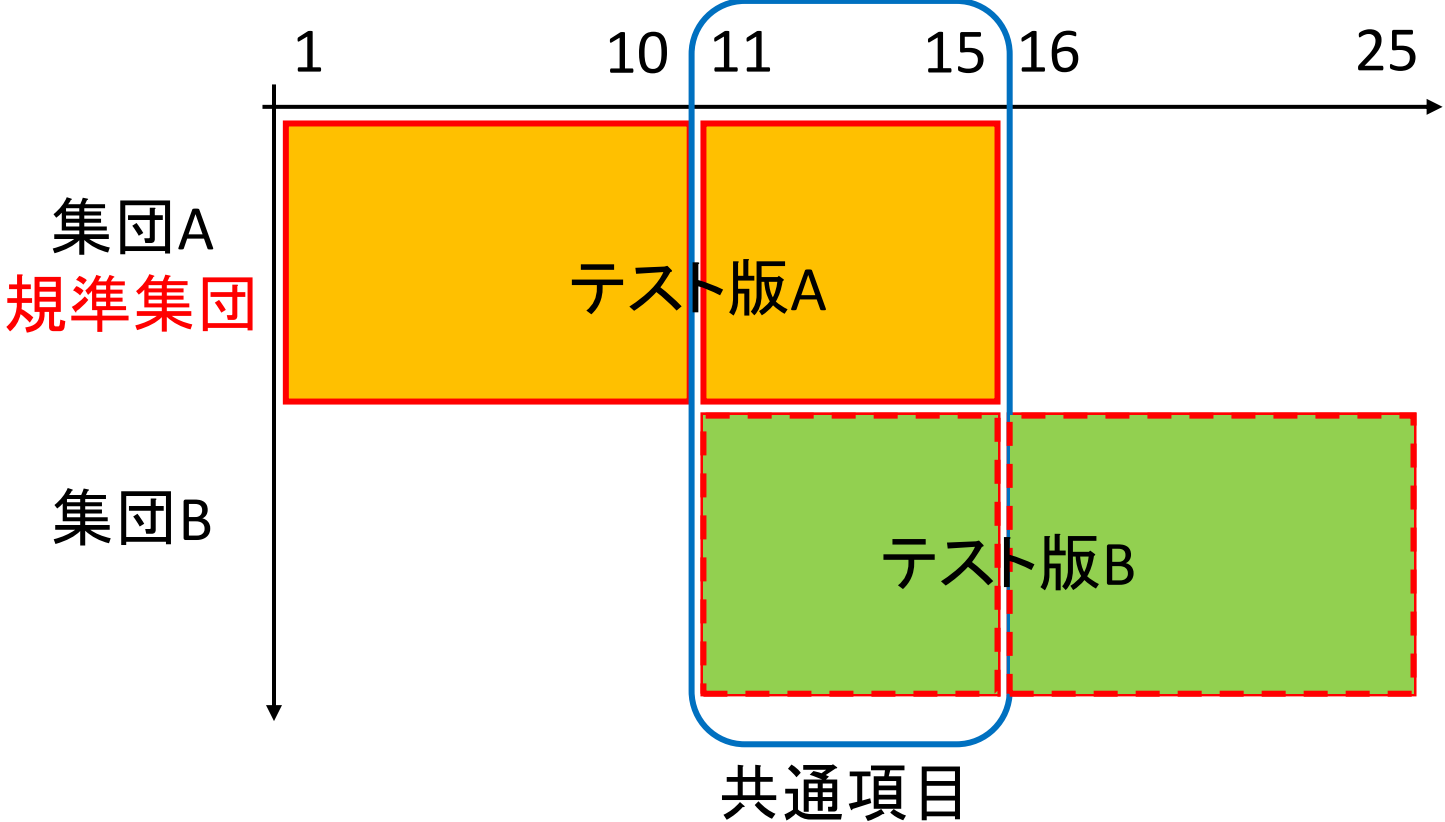


項目パラメタ固定法による等化

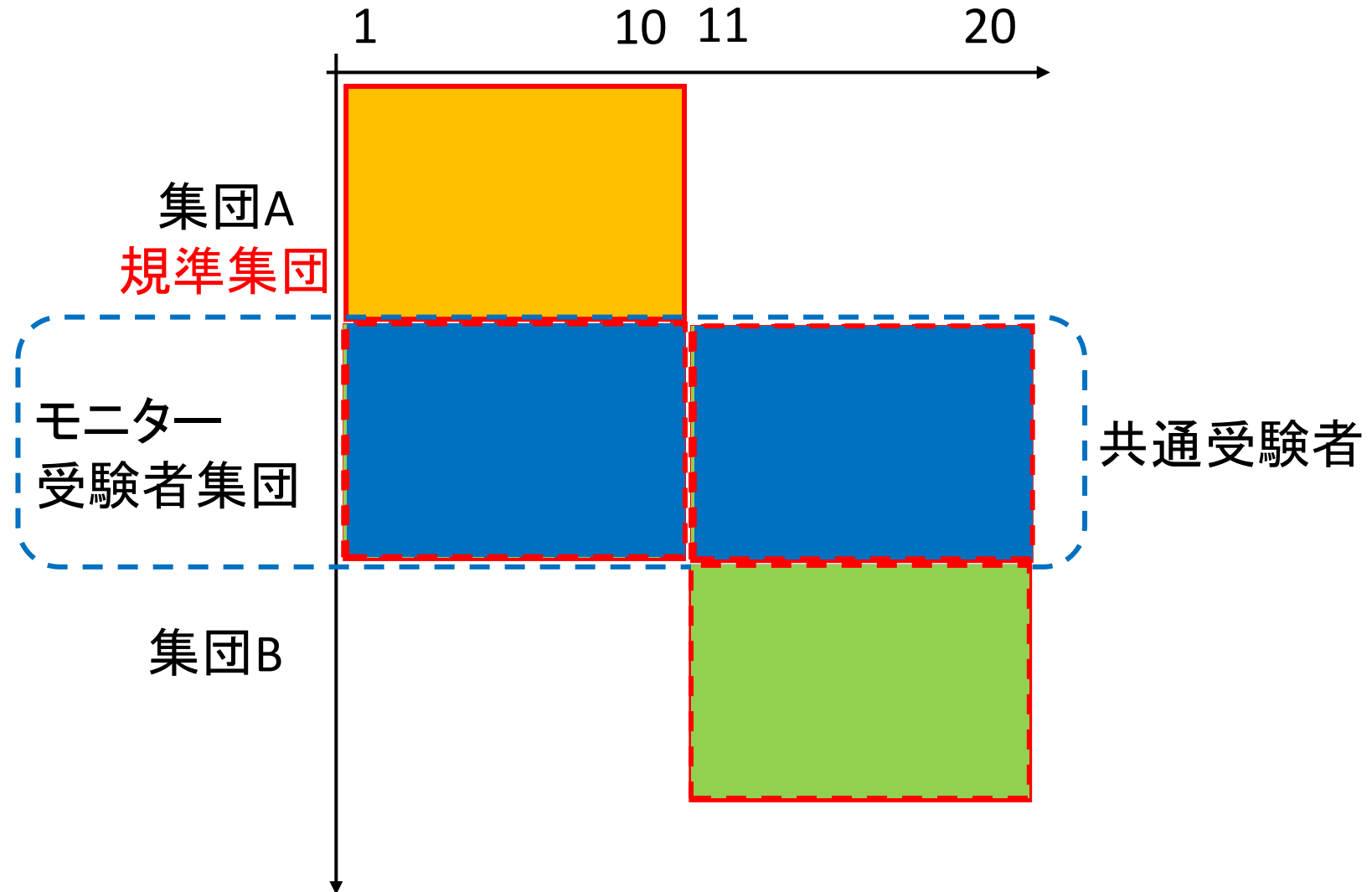


等化計画

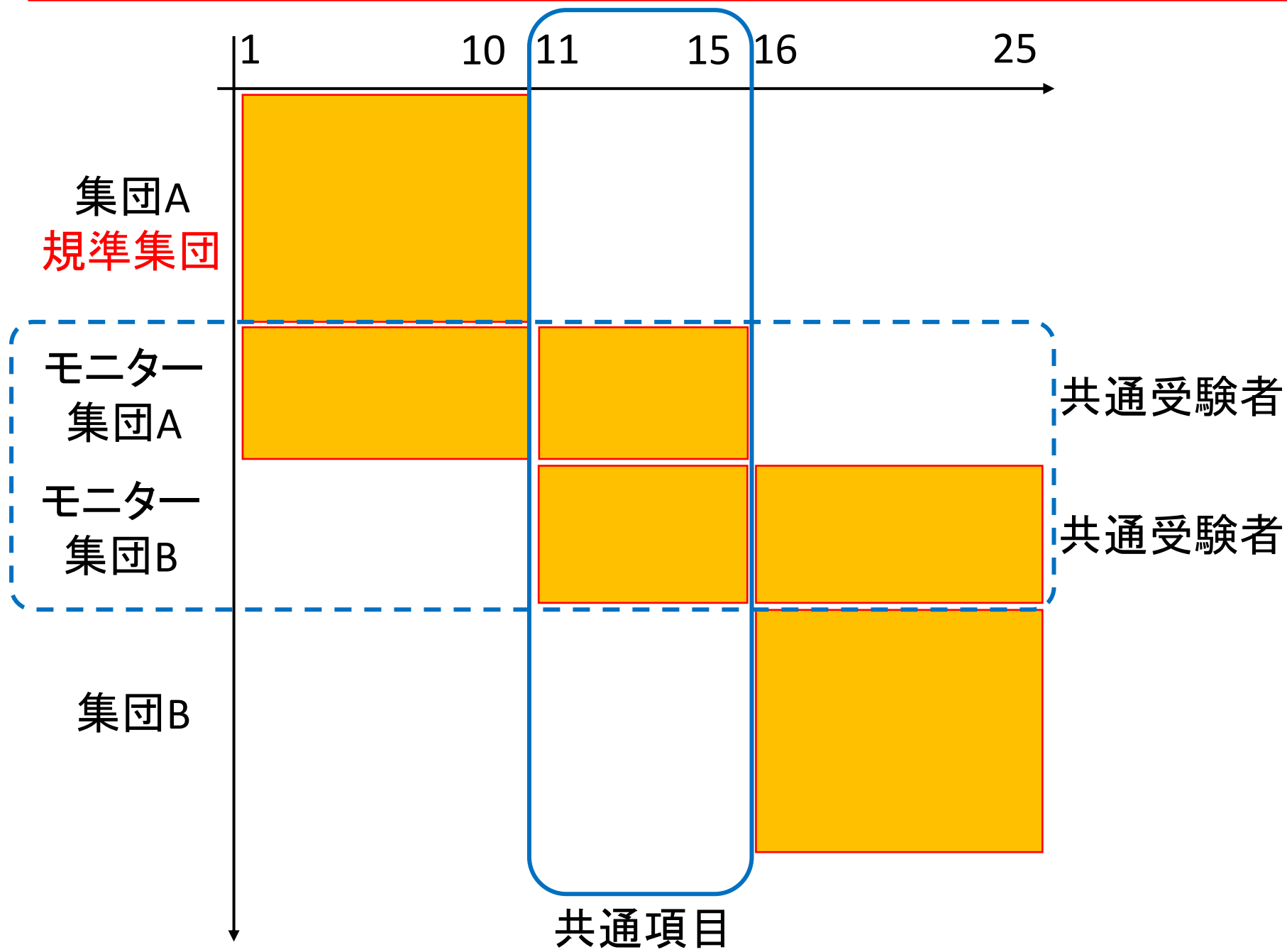
共通項目デザイン



共通受験者デザイン



アンカーテストデザイン



等化の前提と実践上の注意点

「等化の前提」

1. 複数の分冊が、同じ構成概念を測定している。
2. テスト版Aで算出されたスコアをテスト版Bの尺度に変換した結果と、テスト版Bで算出されたスコアをテスト版Aの尺度に変換した結果は、同じでなければならない(等化の対称性)。
3. 複数の分冊は、信頼性が等しい。
4. 能力値の推定値が同じ人をそれぞれの尺度から取り出したときに、測定誤差が等しい。
5. 等化の結果が、集団によらず不変である。テスト版AとBを異なる集団が受験している場合、集団ごとに等化を行った結果(等化係数)が、全体をひとまとめにして等化した結果に一致しなければならない。

光永(2017) テストは何を測るのか, p.164(一部改変)より
川端一光(2014). 等化 加藤健太郎・山田剛史・川端一光[著] Rによる項目反応理論 オーム社, pp.257-258., も参照のこと

テストの実務に係る「等化の前提」

1. 複数の分冊が、同じ構成概念を測定している。
3. 複数の分冊は、信頼性が等しい。
4. 能力値の推定値が同じ人をそれぞれの尺度から取り出したときに、測定誤差が等しい。

これらの前提は、いずれも満たすのが難しい。

- 1. は、立証することが困難である。
- 3. は、実際にテストを実施し、データを取るまではわからない。
→ フィールドテスト等による検討を要する
- 4. は、平行テストに近い複数のテスト版であることを要求しているように見えるが、実務的には平行テストといえない複数のテストを等化したい場合も多い。

テストが仮定する母集団に関する「等化の前提」

5. 等化の結果が、集団によらず不変である。テスト版AとBを異なる集団が受験している場合、集団ごとに等化を行った結果(等化係数)が、全体をひとまとめにして等化した結果に一致しなければならない。

- どのような母集団を仮定して等化を行っているかを、等化の手続きの詳細とともに明確化することは重要である。

リンクング

リンクング

- 等質ではない項目を用いて、共通尺度を得る場合は、厳密な意味で「等化」ができない。
- 厳密ではない等化の方法を「リンクング」とよんで区別している。
- 同じ構成概念を測定しているとは限らない場合に、複数の尺度を一つの尺度上で表現する手法を「リンクング」と呼ぶ(日本テスト学会, 2007, p.224)
- 素点を用いたリンクングの例:「等パーセントイル法」

【等化】

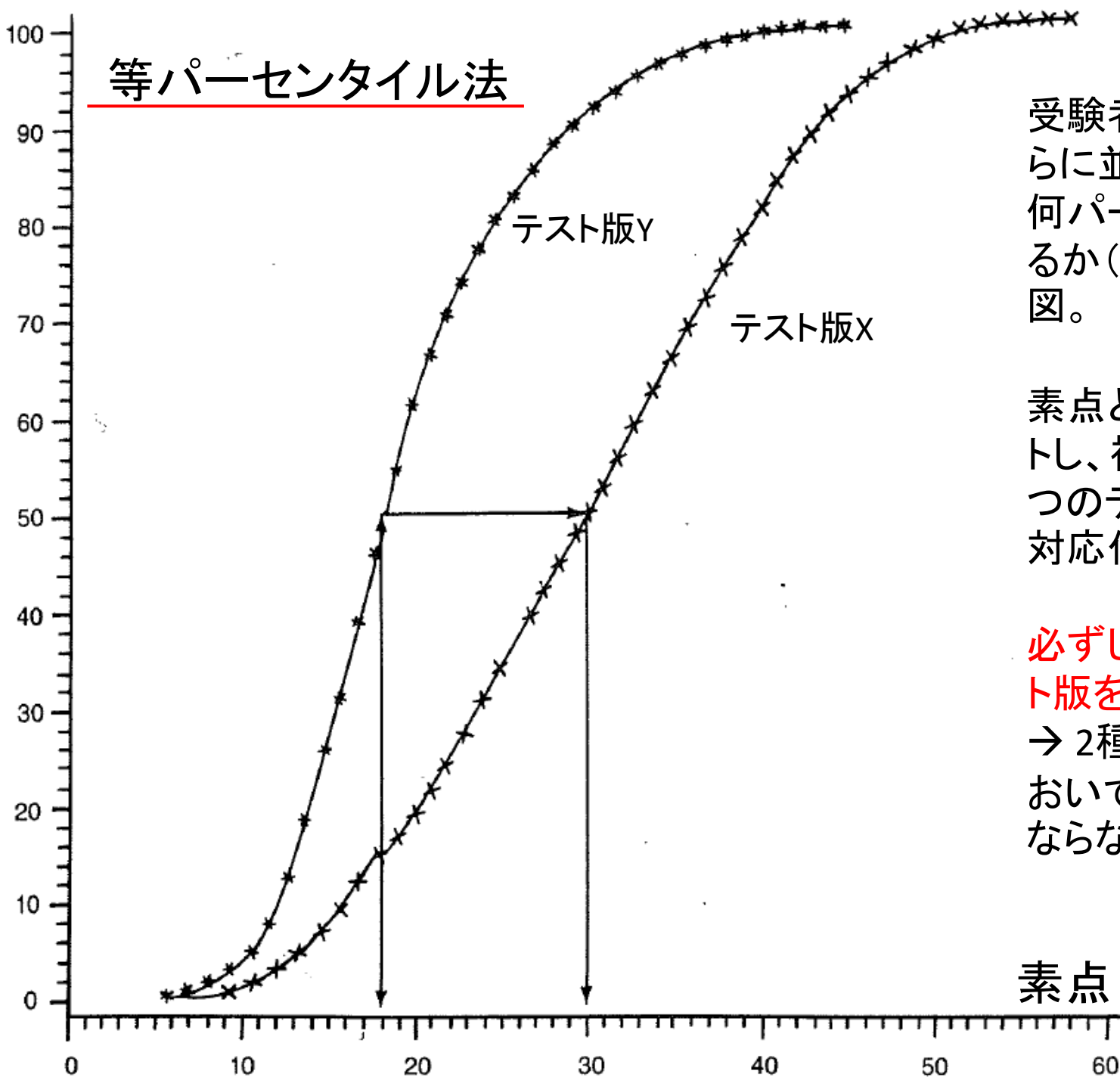
あるテストにおいて同一設計のもとで作成された異なる版の得点を、共通尺度上の得点に変換する統計的な手続きのこと。この手続きはこれらの版がすべて同じ特性を測定していることを前提とする(後略)。

【リンク】(リンキング)

複数のテスト得点を同じひとつの尺度上の得点として表現し、互いに比較可能なようにする手続きの総称。共通尺度である必要は必ずしもない。等化が一番強い仮定のもとで行われるリンキングの一つ。→【等化】

パーセンタイル

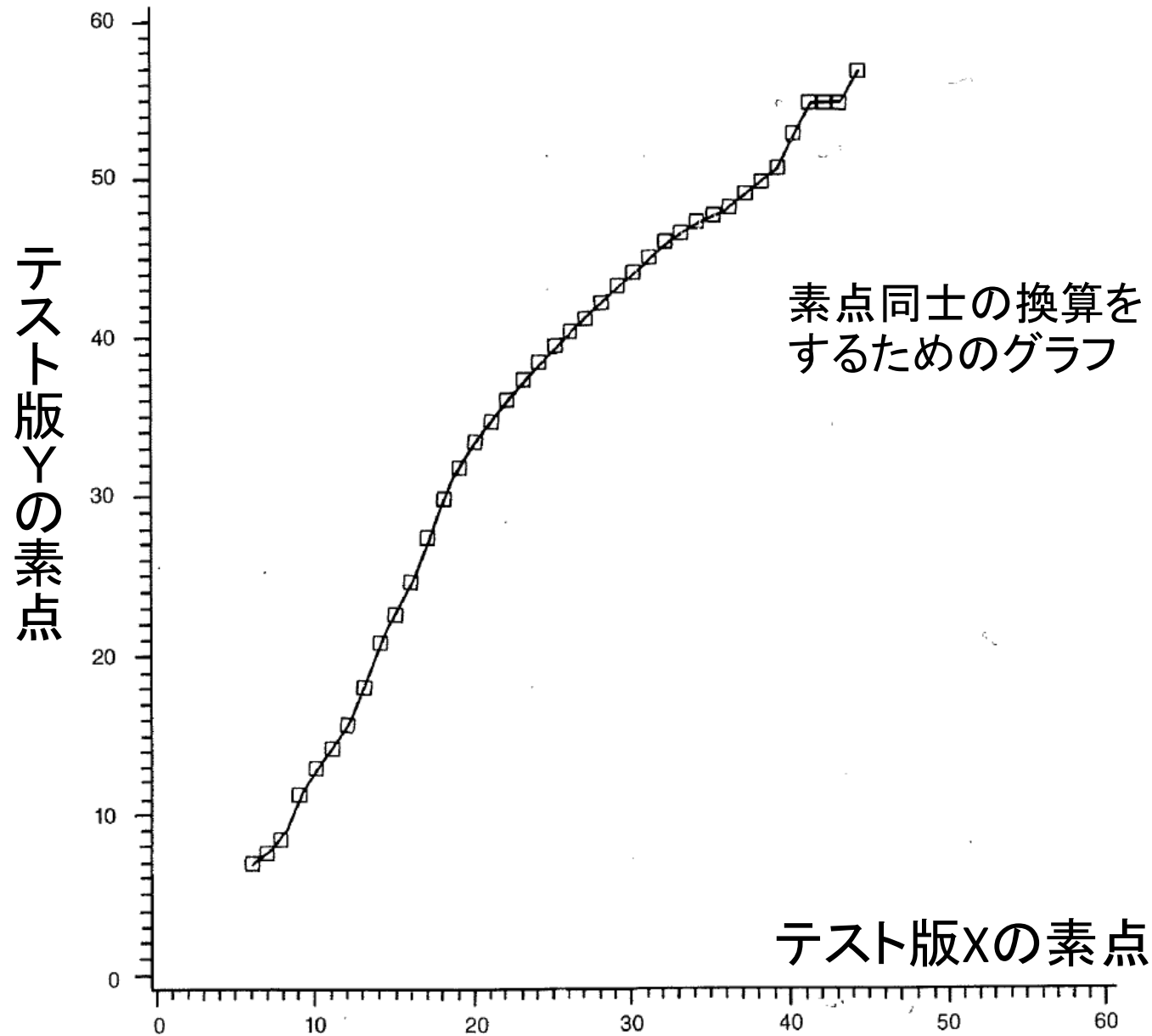
等パーセンタイル法



受験者の素点を小さいほうから並べ替え、素点が下から何パーセントのところに位置するか(パーセンタイル)を表した図。

素点とパーセンタイルをプロットし、補間したグラフにより、二つのテスト版間でのスコアの対応付けができる。

必ずしも同一人物が2種のテスト版を受験していなくてもよい
→ 2種のテスト版の受験者において、著しく能力の分布が異ならなければよい。



Petersen, N.S., Kolen, M.J. & Hoover, H.D. (1989). Scaling, norming and equating. in Linn, R.L. (ed.) Educational Measurement, 3rd ed. (Figure 6.4 and 6.5)

終