流通科学大学論集－経済・経営情報編－第 17 巻第 2 号，87-104(2009)

多値 Rasch モデルのカテゴリ閾値に関する一考察

Category Thresholds of the Polytomous Rasch Model

平 越 裕 之*
Hiroyuki Hirakoshi

評定尺度モデルと部分採点モデルは，共に多値順序カテゴリカルデータを扱う多値 Rasch モデルである。これらのモデルを適用する際には，モデル適用の前提や結果の解釈において，カテゴリ閾値の解釈とモデル間でのカテゴリ閾値解釈の相違を把握し，適用の妥当性を検討する必要がある。本稿ではこれらのモデルの多値データの扱いに焦点を絞り，カテゴリ閾値とカテゴリ特性の解釈を中心にしてモデルの特徴の一部を明かにする。

キーワード： 評定尺度モデル，部分採点モデル，多値 Rasch モデル，カテゴリ閾値.

Ⅰ. はじめに

評定尺度モデル(rating scale model, 以降 RSM)[1)]と部分採点モデル(partial credit model, 以降 PCM)[2)]は，共に多値順序カテゴリカルデータを扱う確率モデルであり，どちらのモデル構成も Rasch モデル[3)]に準じていることから，一般に Rasch モデル族(Rasch model family)と呼ばれている。非 Rasch モデルである項目反応理論(item response theory, 以降 IRT)体系では，段階反応モデル(graded response model, 以降 GRM)[4)]が多値カテゴリカルデータを扱う代表的なモデルとして位置づけられている。

多値順序カテゴリカルデータ Rasch モデル(polytomous Rasch model, 以降 PRM)と非 Rasch モデルの最大の相違は，モデルに適合した条件で行われるテストにおいて，項目や被験者によらず，測定に客観性(objectivity)を持っているか否かである。ここでいう客観性とは，モデルパラメータの分離によって獲得されている，どのような項目や被験者を測定しても，測定値に影響を及ぼさず常に客観的な値を測定できるという特徴である。Rasch モデル族では未知パラメータの十分性(parameter sufficiency)の概念が崩れないようにモデルを構成し，客観性を保つようにモデル構築されているが，IRT 体系では，このような規定を持っていない。そのため，モデルの良否と言う意味では，Rasch モデルのほうが優れているという意見もあるが，得られたデータを説明しやすいモデルは，より自由度の高い IRT 体系のモデルであるとも言える。なお，本稿では Rasch モデルと非 Rasch モデル間でのモデル良否についての議論は行わない。

さて，PRM の RSM と PCM においては，Rasch モデル族としての特徴を受け継いだカテゴリ生起

* 流通科学大学情報学部、〒 651-2188　神戸市西区学園西町 3-1

(2008 年 9 月 2 日受理)

確率が規定されているが，カテゴリ生起確率はカテゴリ閾値(category thresholds)によってのみ規定される。また，RSMとPCMは，本質的には同様のモデル構成がなされているが，複数項目間のカテゴリ閾値の扱いに違いがある。どちらも同じ順序カテゴリカルデータを扱うRaschモデル族であるが，カテゴリ閾値の扱いの違いやカテゴリ閾値で規定されるカテゴリ特性の解釈を明確にした上でモデルを適用することが必要である。しかし，モデル適用者へのこれらの示唆は，これまで乏しい状況である。本稿ではRaschモデル族としてのRSMとPCMのカテゴリ閾値の構成，解釈に焦点を絞り考察を行う。

Ⅱ．評定尺度モデル（RSM）と部分採点モデル（PCM）

本章ではRSM及びPCMが一次元能力測定テストに適用されることを想定する。つまり，受験者が項目を受験し，結果として項目のカテゴリを得るということを想定する。モデル上，受験者は能力パラメータ，項目は項目困難度パラメータ，カテゴリはカテゴリ閾値パラメータを持っているが，これは潜在特性であるので得られたデータから推定することとなる。能力パラメータは値が大きいほど能力が高く，困難度パラメータは値が大きいほど項目は困難であり，能力と困難度は同次元の指標であると考える。また，カテゴリについては項目得点と対応させ，数値の低いほうを得点が低いと考えることとする。カテゴリ閾値パラメータは，次に示すモデル式等によって解釈を与える。

RSMでは，受験者iが項目jに対してカテゴリkの得点となる確率を次のように表している[1)]。

$$P_i(x=k\mid\delta_j)=\frac{\exp\{\kappa_k+k(\beta_i-\delta_j)\}}{\sum_{l=0}^{m}\exp\{\kappa_l+l(\beta_i-\delta_j)\}} \tag{1}$$

β_i：受験者iの能力値

δ_j：項目jの困難度

τ_m：カテゴリmの閾値（全項目共通），$\sum\tau_m=0$

$\kappa_m=-\sum_{k=1}^{m}\tau_k,\ \kappa_0\equiv0$

一方，PCMでは，受験者iが項目jに対してカテゴリkと得点する確率を次のように表している[2)]。

$$P_i(x=k\mid\delta_j)=\frac{\exp\{\kappa_{j,k}+k(\beta_i-\delta_j)\}}{\sum_{l=0}^{m}\exp\{\kappa_{j,l}+l(\beta_i-\delta_j)\}} \tag{2}$$

β_i：受験者iの能力値

δ_j：項目jの困難度

$\tau_{j,m}$：項目jのカテゴリmの閾値，$\sum_m\tau_{j,m}=0$

$\kappa_{j,m}=-\sum_{k=1}^{m}\tau_{j,k},\ \kappa_{j,0}\equiv0$

PCM におけるカテゴリ k 生起確率式(2)は RSM における式(1)とほとんど同じ構成となるが，RSM ではカテゴリ閾値が全項目に渡り共通であるのに対し，PCM では各項目でカテゴリ閾値が違っている。なお，ここで利用した表現以外にもモデル定義は可能であるが，本質的には差異がないので，本稿では記号も含め，以上のような表現を用いる。

モデル導出時 [1),2)]においては，RSM と PCM では，それぞれ別のアプローチがとられていると考えることもできるが，どちらも Rasch が示唆したモデル式 [3)]を元にしており，結果的に導出されたモデルは，形式的にはほとんど同じである。全ての項目のカテゴリ閾値 τ が同じであるか，個別項目ごとに違うかが RSM と PCM の相違と表現される。このため，RSM においては全ての項目においてカテゴリ数が同一である必要があるが，PCM ではその制限はない。モデル導出時の概念がどうあれ，最終的にはモデル上の相違のみが二つのモデル相違であるので，このモデル上の相違のみを念頭に RSM と PCM についての議論を進めることとする。

二つのモデルで，項目の閾値がどのような役割を果たしているのかについて概説する。多値順序カテゴリカルデータを扱うため PRM においては，次のような概念を導入している。一つの m カテゴリ多値項目（生起するカテゴリは，$0,1,2,\cdots,m-1$）を，仮想的に独立した $m-1$ 個の 2 値項目（$1,2,\cdots,m-1$）と考え，それぞれの 2 値項目の困難度をカテゴリ閾値とする。カテゴリ閾値の平均を項目困難度とする。ただし，カテゴリ k が生起するという場合を，1 番目の 2 値項目から k 番目の 2 値項目の全てに正答していて，$k+1$ 番目の 2 値項目から $m-1$ 番目の 2 値項目までの全てに誤答している，と考えることとする。このように $m-1$ 個の 2 値項目を左から右に順に並べて，どこかで 2 つに分けたとき，左の 2 値項目群には全て正答していて右の 2 値項目群には全て誤答しているという場合のみ，つまり完全ガットマンスケールパターン [5)](以降ガットマンパターン)で反応が生起することだけを考える。ガットマンパターンで反応が生起する $m-1$ 個の 2 値項目は独立であるので，本来ガットマンパターンばかりが生起するわけではないが，多値項目を 2 値項目に仮想的に対応させる際に，ガットマンパターンが出現するという条件付確率としているのである [1),2)]。

RSM では，1 テスト中に多値項目が複数あっても，これらのカテゴリ閾値は同じパラメータであり，どの項目もカテゴリ閾値の平均値としての困難度が違うだけで，カテゴリの特性は基本的には同じものとなる。これに比して PCM では，1 テスト中の複数の多値項目では，違うカテゴリ閾値が想定されており，同時に項目によってカテゴリ数も異なっており，カテゴリ特性は項目間で異なっている。RSM では必然的に各項目のカテゴリ数は同じでなくてはならない。RSM と PCM におけるカテゴリ閾値の違いは，まさに以上の点のみである。

なお，多値項目を仮想的に複数の 2 値項目に対応付け，2 値項目の困難度をカテゴリ閾値としながらも，ガットマンパターンが生起する条件付確率で多値項目を表現するので，独立した 2 値項目の困難度を多値項目のカテゴリ閾値と単純に照らし合わせてカテゴリ閾値の数値的な特性を考えることはできない。言い換えればガットマンパターンの場合という条件付確率を求めていることで，多値項目

のカテゴリ閾値は，仮想的に対応させた独立した2値項目の困難度と式上は対応しているが，付帯条件を意識せずに解釈することは危険である。そのため，多値項目のカテゴリ閾値は，対応する独立した2値項目の困難度の特性から一旦切り離して考察する必要がある。

Ⅲ．多値 Rasch モデル(PRM)におけるカテゴリ閾値の解釈

本章では，カテゴリ閾値を中心に PRM の項目の解釈を行う。最初に一項目だけを考えてカテゴリ閾値の考察を行い，その後に多項目に拡げて RSM と PCM についての違いを含めて議論する。

1．一項目における閾値と隣接カテゴリ間の関係

RSM においても PCM においても，一つの項目に注目すれば，どちらも閾値の解釈は同じである。ここではまず，一項目だけを考えた場合の PRM カテゴリ閾値と隣接カテゴリの関係について考察する。式(1)を被験者についてもある能力 β を想定して書き直し，

$$P(x=k \mid \delta)=\frac{\exp\{\kappa_k+k(\beta-\delta)\}}{\sum_{l=0}^{m}\exp\{\kappa_l+l(\beta-\delta)\}} \tag{3}$$

β:受験者の能力値

δ:項目の困難度

τ_m:カテゴリmの閾値,$\sum \tau_m=0$

$$\kappa_m=-\sum_{k=1}^{m}\tau_k,\kappa_0\equiv 0$$

とする。式(3)は RSM においても PCM においても共通である。Ⅱ章でも述べたように，式(3)は，一つの多値項目を仮想的に複数の2値項目に対応させ，ガットマンパターン反応を条件としたときの確率式である。このことから，対応付けた複数の2値項目困難度の物理的意味を解釈し，その単純投影として多値項目のカテゴリ閾値を解釈するといった報告もなされている[6),7)]が，2値項目は本来独立であり，ガットマンパターンだけが出現するという条件は無い。そのため2値項目の困難度解釈を単純に多値項目のカテゴリ閾値解釈へ拡張せず，その違いを明確にして議論する必要がある。また，多値項目のカテゴリ特性を，カテゴリ数分の独立した2値項目の総点獲得確率と等価する方法[8),9)]も報告されている。しかし，多値項目と複数の独立した2値項目では，項目としての内容，性質が変わるため，確率的に等価であっても項目どうしの特性は異なる。このようなことから多値項目カテゴリ閾値を，仮想的に対応する2値項目困難度と対応付けて解釈する際には，その間の相違を十分意識した議論が必要である。

次に多値項目が複数の2値項目と仮想的に対応する場合を具体的に考える。ここでは例として，一つの3カテゴリ項目を2つの2値項目で仮想的に構成する場合を考えてみる。この場合，表3-1のように，2値項目1に誤答(0と表現)で2値項目2に正答(1と表現)しているという場合は3カテゴリ項目では考えない，つまり確率事象としては考えず，標本空間に含めない。これは，反応が（2値項目

1，2 値項目 2）＝（0,1）は本来起こりえるが，多値項目の特性構成上は含めないということである。この時点で独立した 2 値項目の困難度と多値項目カテゴリ閾値は意味が異なることとなる。二つの 2 値項目では表 3-1 の 4 つのパターンが全て起こりえるが，3 カテゴリ項目では，そのうちの 3 つのパ

表 3-1　3 カテゴリ項目と対応する二つの 2 値項目

2値項目 1	2値項目 2	3カテゴリ項目
0	0	0
0	1	N/A
1	0	1
1	1	2

ターンしか考えないということになるのである。このことをもって，『ガットマンパターンが起こりやすいほうが，つまり 2 値項目 1 が 2 値項目 2 よりも易しいほうが，多値項目モデル構成上，適っていると言える』という指摘[6),7)]もあるが，仮に 2 値項目 1 の困難度が 2 値項目 2 の困難度よりも大きいとしても，全てのガットマンパターンは起こり得て，結果的にカテゴリ生起確率が定義可能である。つまり，ガットマンパターンが起こりやすいほうが良いという指摘は，モデルに適合しているかどうかという観点からは的を射ない。よりガットマンパターンが起こりやすい状況，つまり 2 値項目 1 のほうが 2 値項目 2 よりも困難度が低い場合でも，受験者反応が（2 値項目 1，2 値項目 2）＝（1,0）となる確率を多値項目構成に利用しているだけなのであり，（2 値項目 1，2 値項目 2）＝（1,0）が起こりやすいか起こりにくいかはモデルの規定ではない。もし，2 値項目の困難度についての条件がモデル上で必要なのであれば，モデル上で明確にすべきことであり，PRM モデル構築上必須要件として組み入れられる事項となるが，PRM の構築過程において，対応する 2 値項目についてそのような困難度順序の条件は全く必要とされない。つまり，Rasch モデル族としての優良な性質を持つか持たないかに，困難度順序はなんら影響を及ぼさないということであり，このような条件を必要とせず PRM 構築が可能なのである。このことを言い換えれば，『PRM の多値項目構成において，カテゴリ閾値の定義域は$(-\infty,\infty)$であり，閾値の大小順序にモデル上の規定はない』ということである。

2 値項目の場合には，項目特性曲線の形状は一意であり，どのような困難度であっても平行移動によって完全に一致することがモデル上の特徴である。この特徴によって客観性がもたらされている。一方，PRM の多値項目の場合，2 値項目の項目特性曲線に相当するカテゴリ特性曲線の形状は，カテゴリ閾値の値によってさまざまに変化する。このことが客観性に代表される Rasch モデル族の優良な特徴とかけ離れているように感じられることが，よりいっそう閾値の解釈や順序性の議論を深くしているとも言える。しかし，PRM のカテゴリ特性曲線は，モデル上では，その形状がさまざまに変化する柔軟性が許容されているのである。PRM は，カテゴリ閾値範囲が規定されなくとも，受験者群及び項目群の分布によらず客観的な推定値が得られるモデル特徴を持っている。

次に，このように柔軟に変化する多値項目を持つ PRM が持ち得る客観性について，隣接カテゴリ間の生起確率から議論する。

式(3)における隣接するカテゴリ k 及び $k+1$ の生起確率間には，2 値 Rasch モデルの正誤確率と同じ特性がある。2 値 Rasch モデルでは，ロジットコレクト（logit correct: 正答確率を誤答確率で除して対数をとったもの）は，受験者能力 β，項目困難度 δ とし，$\ln(\cdot)$ を自然対数として，

$$\ln\left(\frac{\exp(\beta-\delta)}{1+\exp(\beta-\delta)}\frac{1+\exp(\beta-\delta)}{1}\right)=\beta-\delta \tag{4}$$

と表されるが，PRM においても隣接カテゴリ間の生起確率比は，

$$\frac{P(x=k+1\mid\delta)}{P(x=k\mid\delta)}=\frac{\exp\{\kappa_{k+1}+(k+1)(\beta-\delta)\}}{\sum_{l=0}^{m}\exp\{\kappa_l+l(\beta-\delta)\}}\frac{\sum_{l=0}^{m}\exp\{\kappa_l+l(\beta-\delta)\}}{\exp\{\kappa_k+k(\beta-\delta)\}}$$

$$=\frac{\exp\{\kappa_{k+1}+(k+1)(\beta-\delta)\}}{\exp\{\kappa_k+k(\beta-\delta)\}}$$

$$=\exp\{\beta-\delta-\tau_{k+1}\} \tag{5}$$

であり，

$$\ln\left[\frac{P(x=k+1\mid\delta)}{P(x=k\mid\delta)}\right]=\beta-\delta-\tau_{k+1} \tag{6}$$

と表される。$\sigma_k=\delta+\tau_k$ と表現すれば，

$$\ln\left[\frac{P(x=k+1\mid\delta)}{P(x=k\mid\delta)}\right]=\beta-\sigma_{k+1} \tag{7}$$

となり，2 値 Rasch モデルと同様な関係が，隣接カテゴリの生起確率に存在していることがわかる。これは RSM と PCM で共通の特徴である。また，この関係は次のような直観的方法からも示すことができる。例として 3 カテゴリの場合の隣接するカテゴリ間の関係を示す。

表 3-2　カテゴリ閾値を困難度として持つ二つの 2 値項目

困難度($\delta+\tau_1$)の2値項目	困難度($\delta+\tau_2$)の2値項目	3カテゴリ項目
0	0	0
0	1	N/A
1	0	1
1	1	2

3 カテゴリ項目の場合，カテゴリ閾値は τ_1,τ_2（$\tau_1=-\tau_2$）の二つであり，項目困難度にカテゴリ閾値を足した $\delta+\tau_1,\delta+\tau_2$ は，前述のように対応する 2 値項目二つの困難度とそれぞれ等しい。つまり，表 3-2 のような関係にある。しかし，3 カテゴリ項目は表 3-2 に示す 3 カテゴリ項目の 0,1,2 の部分だけが標本空間であるので，単純に二つの独立した 2 値項目の生起確率と対応させることができないこ

とから，対応する二つの 2 値項目の(0,0),(1,0),(1,1)の生起確率を全事象生起確率として基準化して，カテゴリ生起確率を定式化している。この結果，隣接するカテゴリ間での生起確率比は，次のような方法でも記述できる。

カテゴリ 0 は，対応する二つの 2 値項目の正誤が(0,0)であり，同様にカテゴリ 1 は(1,0)，カテゴリ 2 は(1,1)である。ある能力 β を持った受験者が 3 カテゴリ項目を受験したとき，

$$\frac{\text{カテゴリ 1 生起確率}}{\text{カテゴリ 0 生起確率}} = \frac{\Pr(\tau_1\text{項目に正答}\mid\tau_2\text{項目に誤答})}{\Pr(\tau_1\text{項目に誤答}\mid\tau_2\text{項目に誤答})}, \tag{8}$$

$$\frac{\text{カテゴリ 2 生起確率}}{\text{カテゴリ 1 生起確率}} = \frac{\Pr(\tau_2\text{項目に正答}\mid\tau_1\text{項目に正答})}{\Pr(\tau_2\text{項目に誤答}\mid\tau_1\text{項目に正答})} \tag{9}$$

とも表せるので，

$$\frac{P(x=1\mid\tau_1,\tau_2)}{P(x=0\mid\tau_1,\tau_2)} = \frac{\exp\{\beta-(\delta+\tau_1)\}}{1+\exp\{\beta-(\delta+\tau_1)\}}\frac{1+\exp\{\beta-(\delta+\tau_1)\}}{1} = \exp\{\beta-(\delta+\tau_1)\}, \tag{10}$$

$$\frac{P(x=2\mid\tau_1,\tau_2)}{P(x=1\mid\tau_1,\tau_2)} = \frac{\exp\{\beta-(\delta+\tau_2)\}}{1+\exp\{\beta-(\delta+\tau_2)\}}\frac{1+\exp\{\beta-(\delta+\tau_2)\}}{1} = \exp\{\beta-(\delta+\tau_2)\} \tag{11}$$

と表現できる。式(10)，式(11)は，それぞれ $\delta+\tau_1, \delta+\tau_2$ という困難度を持つ二つの 2 値項目正誤確率比と全く同じ式である。PRM においては，隣接するカテゴリ間の生起確率比率は，2 値項目と同様の式で規定できるということである。これが PRM の特徴であり，2 値項目と同様の客観性に関する特性は，PRM の隣接する 2 つのカテゴリ間にのみ存在していることを示している。

２．一項目における閾値とカテゴリ特性

本節では，PRM 項目のカテゴリ閾値の値によって，一つの項目全体のカテゴリ特性がどのようになるのか考察する。これまで見たように，多値項目と対応する 2 値項目は密接な関係がある。しかし，この関係は 2 値項目一つと，多値項目において対応する隣接カテゴリ間つまり段階(step)とだけで成立することである。項目全体としてのカテゴリ特性は，対応する 2 値項目の困難度とは違った特性を持つことになる。

例として 3 カテゴリ項目を想定する。項目困難度 δ と閾値 τ_1(または τ_2)が決まれば項目のカテゴリ特性はすべて決定される。ここでは，$\tau_1=-\tau_2$ の値を変化させて，項目の特性を議論することとする。まず，能力 β の受験者が，項目困難度 δ，カテゴリ閾値 $\tau_1=-\tau_2$ である項目を受験した際の期待得点を図 3-1 に示す。図 3-1 では，閾値 $\tau_1=-\tau_2$ を－10 から 10 まで 1 ずつ変化させている。なお，$\tau_1\geq 3$ のときには，図上でほとんど区別ができないので $\tau_1\geq 3$ の範囲では凡例をまとめている。これより，$\tau_1=-\tau_2$ の値がどのような値であっても期待得点は β の単調増加関数となっていることがわかる。τ_1 と τ_2 の値が $\tau_1>\tau_2$，つまり $\delta+\tau_1>\delta+\tau_2$ である場合には，$\beta-\delta=0$ の近傍で期待得点は急激に増加し，それ以外ではほとんど変化がない。逆に $\tau_1<\tau_2$，つまり $\delta+\tau_1<\delta+\tau_2$ であれば，2 段階に期待得点が増加し，特に $\delta+\tau_2$ が $\delta+\tau_1$ より十分大きい場合には，$\beta-\delta=0$ の近傍では期待得点にほと

んど変化がなく，$\beta-\delta=\tau_1$，$\beta-\delta=\tau_2$ の近傍で，ちょうど二つの 2 値項目の正答確率と同様の変化が認められる。これは，二つの閾値 τ_1,τ_2 が昇順で大きく離れているため，Ⅲ－1 で述べた対応する二つの 2 値項目においては，ガットマンパターン出現の確率合計が 1 に近く，τ_1,τ_2 を困難度とする二つの独立した 2 値項目の合計得点を求めているのに近づいてくるからである。

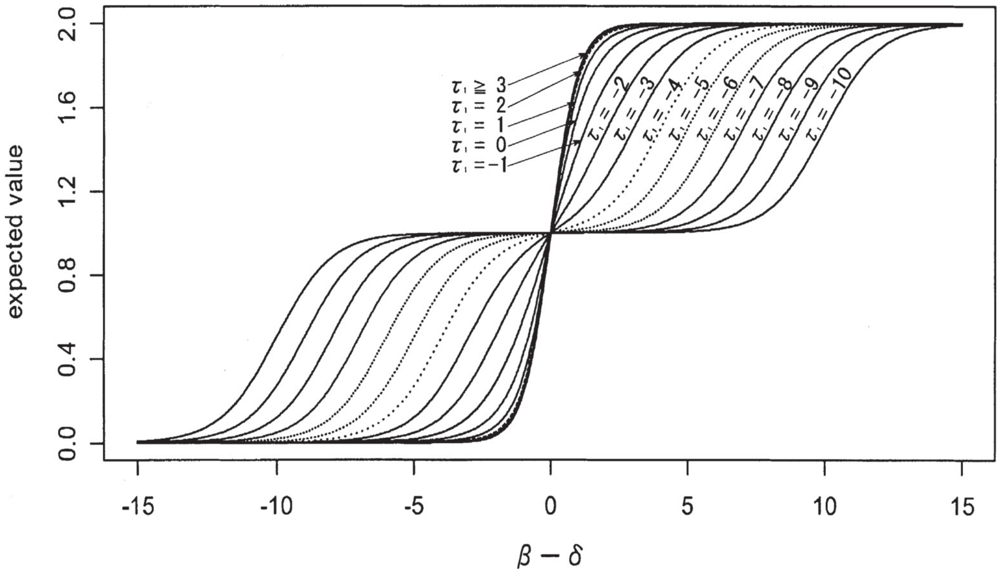

図 3-1　3 カテゴリ項目（満点 2）の期待得点

なお，前述の通り，二つの閾値 τ_1,τ_2 が降順に離れていっても，多値項目の期待得点は，図 3-1 のように，$\beta-\delta=0$ 近傍で急激に増加するという特徴を持つだけである。これは，対応する二つの 2 値項目で(1,0)との反応が非常に起こりにくくなるため，(0,0)から(1,1)へ一気に推移してしまうということを表している。

以上のように図 3-1 からは，項目閾値の値によって，項目の期待得点の変化が様々に決定されることがわかる。このことは，RSM では 1 テスト内に限って閾値がどの項目も同じであるので，どの項目の期待得点関数も形状は同じであり，困難度によって左右に平行移動したものとなるが，PCM においては同一テスト内においても項目ごとに閾値が違うため，項目間で期待得点関数の形状そのものが様々に変化しているということを示している。

図 3-2 には，図 3-1 に対応する情報関数値を示す。この図においても $\tau_1\geq4$ のときは図上でほとんど区別ができないので $\tau_1\geq4$ の凡例をまとめている。これより，τ_1 の値が大きくなればなるほど $\beta-\delta=0$ 近傍で大きな弁別力(discrimination power)を示し，逆に τ_1 の値が小さくなれば，$\beta-\delta=0$ 近傍ではほとんど弁別力が無く，$\beta-\delta=\tau_1,\beta-\delta=\tau_2$ 近傍で，2 値項目と同様の弁別力を示していることがわかる。これらも図 3-1 と同様の理由によっている。このように図 3-2 からも，項目閾値の値

によって，項目の情報関数の変化が様々に決定されることがわかる。

このカテゴリ特性の柔軟性は，PRM が持ち合わせている特徴であり，Rasch モデル族としての制約，例えば式(7)のような関係と対抗するものではない。モデル上は柔軟性が認められている。

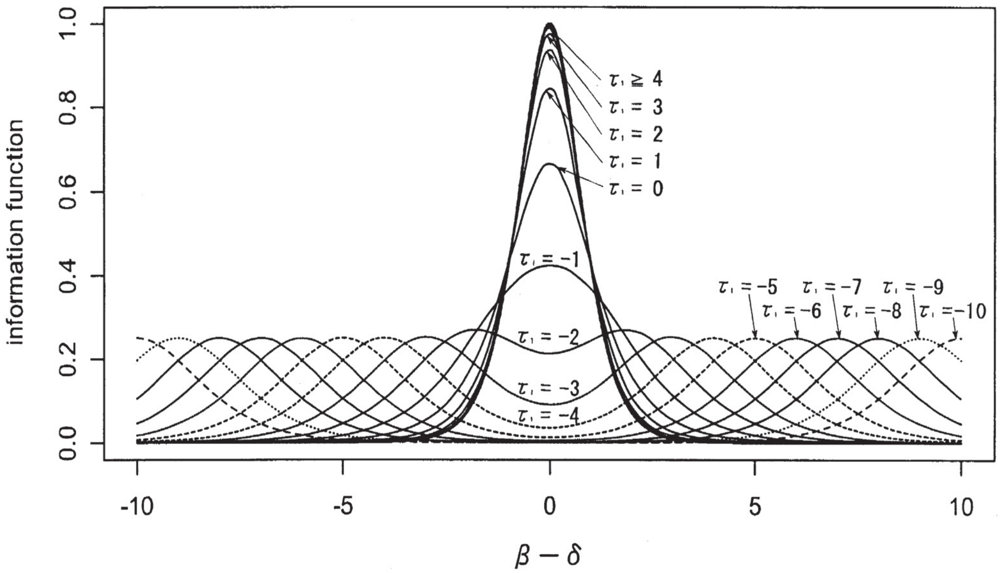

図 3-2　3 カテゴリ項目（満点 2）の情報関数値

3．複数項目における閾値とカテゴリ特性

通常テストには複数項目が設定される。PRM において，一項目は柔軟性に富んだカテゴリ閾値設定が可能であるが，項目が複数個ある場合に，このことがどのように影響を与えるかについて述べる。一項目に限ってみれば，PRM ではカテゴリ閾値が無限に考えられ，それによって項目の特性も無限に規定される。最初に RSM について，その後 PCM について項目単位でのカテゴリ特性について議論し，最後に RSM と PCM を比較することとする。

RSM においては，カテゴリ閾値は全項目で同じと規定されているので，全ての項目のカテゴリ特性は同様と考えられている。仮に項目ごとに潜在カテゴリ閾値が違うとしても，RSM 上では同一として推定することになるので，図 3-1 の期待得点曲線は全ての項目で同一であると解釈する。このことは，RSM においては各項目が同一の潜在カテゴリ閾値を持つようにテストを設計することが求められている，とも言い換えられる。全ての項目のカテゴリ閾値が同じであるということは，どのような能力を持った受験者のペアにとっても，任意の項目一対の期待得点の大小は変わらないという大きな特徴を持つことになる。つまり，ある受験者 a にとって項目 i よりも項目 j の期待得点のほうが高ければ，他の全ての受験者にとっても項目 i よりも項目 j の期待得点のほうが高いということであり，項目単位で見た場合には単純にやさしい項目，難しい項目と位置付けることが可能となる。これは 2 値 Rasch モデルにおける受験者能力と項目困難度を一次元尺度上に並べたマップと同様のマップ構築が，RSM

の項目でも可能であるということである。ただし，この特徴は一テストにおけるものであることには注意を要する。

一方，PCM では，カテゴリ閾値は各項目で個別であるので，項目ごとのカテゴリ特性は異なっている。また，カテゴリ数そのものの違いも認めている。項目ごとに潜在カテゴリ閾値が違っていれば，PCM では項目ごとに違うものとして推定する。このことから，PCM では，各項目のカテゴリ特性は異なり，例えばある項目は $\tau_1 = -5$ であるが，他の項目は $\tau_1 = 1$ であってもかまわない。これは，図 3-1 でもわかるように，期待得点曲線が項目間で大きく異なっている可能性があるということである。この結果，受験者の能力によっては項目 i と項目 j の期待得点の大小は入れ替わる可能性がある。つまり一テスト内においても 2 値 Rasch モデルや RSM が持ち合わせている，項目単位の単純な困難度規定を持ち合わせていないのである。

この項目特性は，2 値 Rasch モデルは当然持ち合わせており有効な特徴であるが，PRM では複数項目間での『項目困難度』という観点から考えれば，一テスト内に限って RSM はこの特徴を持ち合わせているが，一テスト内に限っても PCM では持ち合わせていない。しかし，RSM が持ち合わせている特徴は一テスト内においてだけであり，運用方法にもよるが，複数のテストにわたる場合には，一組のカテゴリ閾値がそれぞれのテストで別々に推定されることになる。したがって，RSM においても複数テスト間においては項目のカテゴリ特性は異なってしまうと考えられる。その理由を以下に述べる。

RSM においては，一テストにおいて各項目のカテゴリ特性曲線は一致する。しかし，複数のテスト間で同一のカテゴリ閾値を前提としているわけではない。一テストにおいてテスト内全ての項目のカテゴリ特性が一致していることから，PCM と比較してモデルとしては客観性が高いようにも見える。つまり一テスト内においては各項目のカテゴリ閾値が等しいことから，項目の段階のみならず，多値項目全体のカテゴリ特性を含め客観性が成立しているようにも考えられる。客観性について，ここでは 2 値 Rasch モデルと RSM を比較しながらもう一度整理する。

2 値 Rasch モデルでは項目弁別力はパラメータではなく定数であり，テストによらず全ての項目がこの弁別力を持ち合わせているとの前提がある。そのため，複数回のテストにおいて，それぞれが 2 値 Rasch モデルに高い程度に適合しているならば，各項目弁別力はモデル規定の定数に近いと考えられる。実際のテストの全ての項目弁別力が完全に等しいかどうかは定かではないが，モデル規定においては全ての項目の弁別力が等しく客観性が保たれていることから，モデルに高い程度に適合しているデータは客観性が保たれているとみなすことが可能である。一方，RSM においては，カテゴリ閾値に規定がなく，モデル上はパラメータとして扱われる。もし，RSM のカテゴリ閾値がある定数であるとのモデル規定があれば，どのテストにおいても全項目のカテゴリ閾値が同じであるというモデルとなり，多値項目のカテゴリ特性はいずれのテストにおいても完全に一致するので，カテゴリ特性を含めた項目全体として客観性があることにある。しかし，RSM のカテゴリ閾値はパラメータであるので，

データからの推定によってのみ顕在化する。つまり，定数に決められているわけではなく，得られたデータによって推定される値なのであるから行うテスト，得られたデータによって異なることが前提である。これは 2 値 Rasch モデルと大きく異なる点である。RSM において何度もテストを行った際，仮に『どのテストも項目群構成概念が高い程度に同じものであり、同一母集団から抽出された被験者群に対する複数テスト』であったとすれば，いずれのテストにおいてもカテゴリ閾値は共通化して考えることもでき，その場合にはカテゴリ特性は全てのテストの全ての項目において共通である，と考えられるかもしれない。しかし，この考え方には問題がある。テストの項目構成概念を共通化することが困難であることと，それ以上に重要なことは，仮に項目構成概念を各テスト間で共通化できたとしても，モデル上はカテゴリ閾値がパラメータであり，データによって推定される潜在特性であるということである。繰り返しになるが，2 値 Rasch モデルの項目弁別力は定数である。RSM のカテゴリ閾値はパラメータである。つまりモデル上ある値に規定されている値ではなく，データによって推定される値である。幾多のテストにおいてカテゴリ閾値が全く同じ定数であるとの前提がない以上，データによってカテゴリ閾値推定値は異なることが当然である。何回かのテストにおいては，非常に似通った推定値が得られるかもしれないが，例えば回を重ねるごとに，初回のカテゴリ閾値からの乖離が徐々に大きくなったり，あるいは初回だけがずれていてそれ以降は安定していたりといった風に，パラメータは，定数のような参照値・基準値がない潜在特性であるため，何を基準値として議論すればよいのかが不明である。逆に基準値があるのならパラメータではなくモデル上定数化すればよいが RSM ではカテゴリ閾値は定数ではない。

2 値 Rasch モデルではモデル上，いずれのテスト，いずれの項目においても項目弁別力は定数で同じだが，RSM のカテゴリ閾値は一度のテストにおいてのみ各項目共通のパラメータである。パラメータは一度のテストにおいてのみ推定され，テスト回数を重ねるたびに再推定するわけではない。仮に再推定するのであれば，他のパラメータ値にも影響を与え，いつまでも各推定値は再推定を待つこととなり，RSM による測定は破綻する。結論として，RSM は一回ごとのテストにおいて，そのデータからカテゴリ閾値を推定するので，複数テスト間で推定されるカテゴリ閾値が似通っていたとしてもそれは偶然であり，RSM モデル上で一度のテストで推定されるパラメータと規定されている以上，複数テストにおける全カテゴリ特性を含めた項目特性は客観性を持たない。なお，ここでいう客観性は，あくまで多値項目のカテゴリ特性全体での議論であり，複数テストにおいても，次に述べる PCM においても項目のカテゴリ間の“段階”には客観性が存在する。客観性とは偶然の産物ではなく，まずモデル上で厳格に客観性が認められることが必要であり，次にそのモデルにデータが高い程度に当てはまっているとみなせる必要があるということである。

次に PCM の場合である。PCM では各項目が個別にカテゴリ閾値を持つのであるから，複数テストにおいても同様に項目ごとのカテゴリ閾値を想定している。その結果明らかに全カテゴリ特性を含む項目全体には客観性がない。結果的に PRM では，全てのカテゴリ特性を含めてひとまとめにした項

目困難度という観点からは，『どのような能力を持った受験者一対にとっても，任意の項目ペアの期待得点の大小は変わらない』という客観性はないと結論付けることができる。

PRM の多値項目で，全てのカテゴリを含んだ項目困難度という概念を考えれば，このように客観性は保たれない。しかし，多値項目はカテゴリ閾値によって項目全体のカテゴリ特性が変化することを特徴としているのであるから，これは当然であるともいえる。モデル上カテゴリ閾値の値をなんらかの値に決定しておかなければカテゴリ特性は変化することになるが，モデル上ではカテゴリ閾値はパラメータである。PRM においてはⅢ－1 で述べたように，客観性は段階にのみ存在する。項目の各カテゴリのうち隣接するカテゴリに注目すると，式(4), (7), (10), (11)で示したように，隣接カテゴリ生起確率の比は，対応するカテゴリ閾値と多値項目困難度の加算値を困難度に持つ 2 値項目の正誤確率比と同様になる。3 カテゴリ項目を例に挙げれば，次のような式が成立する。多値項目困難度δ，カテゴリ閾値τ_1,τ_2を持つ 3 カテゴリ項目に対して，

$$\frac{\text{カテゴリ1生起確率}}{\text{カテゴリ0生起確率}+\text{カテゴリ1生起確率}}=\text{困難度}\delta+\tau_1\text{の2値項目の正答確率}, \tag{12}$$

$$\frac{\text{カテゴリ2生起確率}}{\text{カテゴリ1生起確率}+\text{カテゴリ2生起確率}}=\text{困難度}\delta+\tau_2\text{の2値項目の正答確率} \tag{13}$$

なる関係が成立する。つまり隣接カテゴリ間で上位カテゴリが生起する確率は，(δ＋カテゴリ閾値)を困難度とした 2 値項目の正答確率となる。図 4-1 に，カテゴリ閾値を－10 から 10 に 1 ずつ変化させたときの，隣接するカテゴリ間における上位カテゴリ生起確率を示す。この図からもわかるように，隣接するカテゴリ間で上位カテゴリが生起する確率は，多値項目困難度とカテゴリ閾値の加算値を困難度に持つ 2 値項目の正答確率と全く同じである。このことから，多値項目全体のカテゴリ特性を考えずに，多値項目中で隣接するカテゴリのみに注目すれば，それは 2 値項目と同様である。その結果として，PRM の多値項目は次のような特徴を持つことがわかる。

・PRM の隣接カテゴリにおける上位カテゴリ生起比率に客観性が備わっている。

つまり，

・どの受験者においても，ある項目のカテゴリiと$i+1$のうちカテゴリ$i+1$が生起する確率と，別の（または同一の）項目のカテゴリjと$j+1$のうちカテゴリ$j+1$が生起する確率の大小関係は同じである。

一項目全体としての弁別力は，RSM では試験ごとに，PRM では項目ごとに異なっているようなモデル表現になっているが，隣接カテゴリ間に注目した場合には上位カテゴリ生起確率が，2 値項目と同じロジスティック曲線を描く。このため，一項目全体ではなく，項目中の一つずつの段階を意識した場合には，2 値 Rasch モデルと同様の客観性が隣接カテゴリ間に存在することがわかる。さらに，この客観性はカテゴリ閾値の値によらず備わっているのである。

これらより，PRM における客観性は，一項目全体として捉えるのではなく，一項目一隣接カテゴ

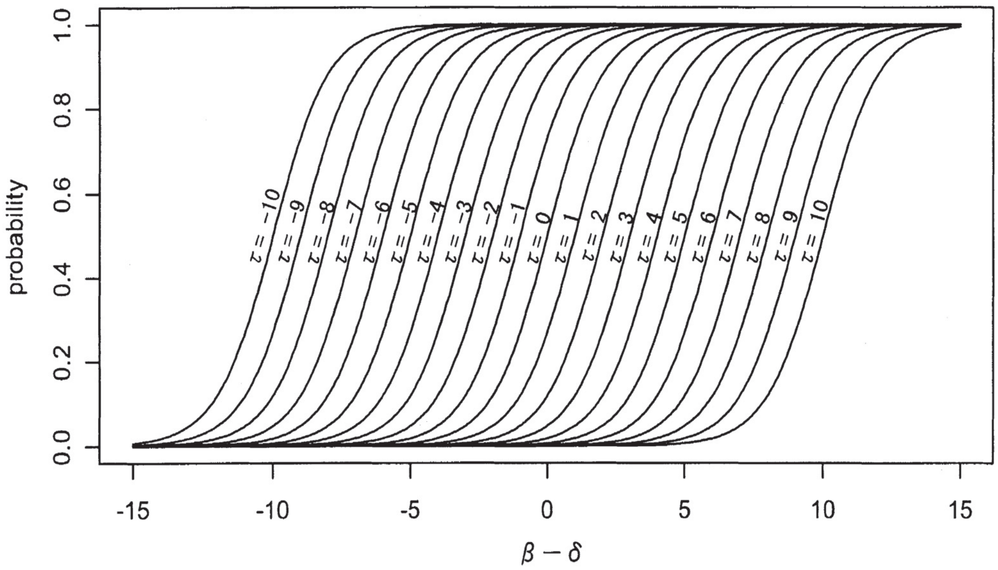

図 **4-1**　隣接カテゴリにおける上位カテゴリ生起確率

リで捉える必要性があることがわかる。基本的には，多値項目全体ではもはや **2** 値項目のような客観性を持ち合わせておらず，多値項目中の隣接カテゴリ，つまり段階こそが **Rasch** モデル族としての客観性を保っているということである。なお，これまでの全ての議論は，**3** カテゴリ項目を例に挙げたが，**4** カテゴリ以上でも議論に相違はない。

Ⅳ．カテゴリ閾値とカテゴリ選択確率

これまで，**PRM** におけるカテゴリ閾値を，モデル規定から考察した。本章では項目のカテゴリ生起確率から，**PRM** を実践運用する際のカテゴリ閾値について考察する。

Ⅲ章で，カテゴリ閾値範囲に規定はないことを示した。しかし，カテゴリ閾値によって項目のカテゴリ特性は柔軟に変化するので，現実的な多値項目として考えると，場合によっては適当でないカテゴリ特性を持つ可能性がある。ここでいう“適当でない”とは，モデル上の制約ではなく，現実テスト，あるいはアンケートなどの多値項目の性質とカテゴリ特性が感覚的にずれているような場合やカテゴリ自身が意味を持たないような場合のことである。これは **PRM** が適用される対象，運用されるテストやアンケート内容，あるいは分野ごとの専門的な考察から判断されうるべきことである。したがって，本章において，適当か適当でないかの判断をするのではなく，判断の材料となる **PRM** 項目のカテゴリ特性について生起確率からの考察を簡潔に行うことにする。

まず，**3** カテゴリ項目においてカテゴリ閾値とカテゴリ生起確率の関係を明確にするため，カテゴリ閾値 τ_1 の値を -2，-1，**0**，**1**，**2** と変化させたときのカテゴリ累積生起確率を図 **4-1**～図 **4-5** に示す。これらの図ではカテゴリ **0**，**1** の順に確率を足し合わせた累積確率を **2** 本の曲線で示している。

そのため$\beta-\delta$が決定すれば，0から全事象確率である1までを3つの部分に分割した形式となり，0から一本目の曲線までの長さがカテゴリ0生起確率を表し，一本目の曲線から二本目の曲線までの長さがカテゴリ1生起確率，二本目の曲線から1までの長さがカテゴリ2生起確率を表している。これらの図より，カテゴリ閾値τ_1が大きくなればなるほどカテゴリ1選択比率は小さくなることがわかる。これはⅢ章での議論と一致する。また，$\beta-\delta=0$，つまり項目困難度と受験者能力が一致する場合には，カテゴリ0とカテゴリ2が同一確率となる。さらに$\beta-\delta=0$で累積確率が0.5となる点を中心に，これらの図は点対称となっている。なお，これはカテゴリ数が3で，二つのカテゴリ閾値τ_1，τ_2が$\tau_1=-\tau_2$であるからであり，4カテゴリ以上の項目ではこの対称性は持たない。

次に，$\beta-\delta$の値を−2，−1，0，1，2と変化させたときのカテゴリ累積生起確率を図4-6〜図4-10に示す。これらの図の横軸はカテゴリ閾値$\tau_1(=-\tau_2)$である。いずれの場合においても，τ_1が小さくなればなるほど，カテゴリ1の生起確率が大きくなり，逆にτ_1が大きくなればなるほど，カテゴリ1の生起確率が小さくなっていく様子がわかる。つまり，$\beta-\delta$の値設定した範囲においては，τ_1が小さければ，ほとんどの場合カテゴリ1が生起し，τ_1が大きければ，カテゴリ0の次のカテゴリとしてカテゴリ1を飛ばしてカテゴリ2が生起する様子がうかがえる。これは，図3-1の期待得点の特徴と同様である。

図4-1〜図4-10からわかるように，カテゴリ閾値$\tau_1(=-\tau_2)$が大きくなれば$\beta-\delta$の値によらず，カテゴリ1生起確率はわずかである。項目は3カテゴリ項目ではあるが，実際には2カテゴリ，つまり2値項目と変わらない結果であるともいえるだろう。このような3カテゴリ項目が項目として適当かどうかは，PRM規定からではなく，項目やテストの内容や意図から吟味し，検討すればよい。また逆にカテゴリ閾値$\tau_1(=-\tau_2)$が小さくなればなるほど，カテゴリ1が生起する確率が大きくなり，Ⅲ章で述べた，対応する独立した2値項目二つの総点と意味合いが近づいてくる。この場合もPRM規定上は適っているが，3カテゴリ全体のカテゴリ特性を項目やテストの意味から検討することが必要であろう。

PRMではカテゴリ生起確率やカテゴリ特性は，カテゴリ閾値のみで決定される。PRMのカテゴリ特性は柔軟ではあるが，決して自由に決定できるわけではない。3カテゴリの場合にはカテゴリ特性を決定するカテゴリ閾値は唯一つである。一つのカテゴリ閾値を決めれば，あるいは一つのカテゴリ閾値が決まっていれば，カテゴリ特性は全て決定されるが，逆にカテゴリ特性を所望の特性に自由に決定，あるいは設計してからカテゴリ閾値を導出することは出来ない。4カテゴリ以上においてもカテゴリ閾値数の差異はあるが，カテゴリ閾値がカテゴリ特性の全てを決定することは同様であり，自由なカテゴリ特性の設計は出来ない。カテゴリ閾値によって全てのカテゴリ特性が決定されるのであるから，カテゴリ閾値そのものの値に注目するよりも，カテゴリ閾値から決定されるカテゴリ特性について十分検討しなければならない。カテゴリ特性はカテゴリ数が多くなればなるほど複雑となるので，特に多くのカテゴリ数を持つ項目では注意すべきである。また，本稿では触れないが，データが

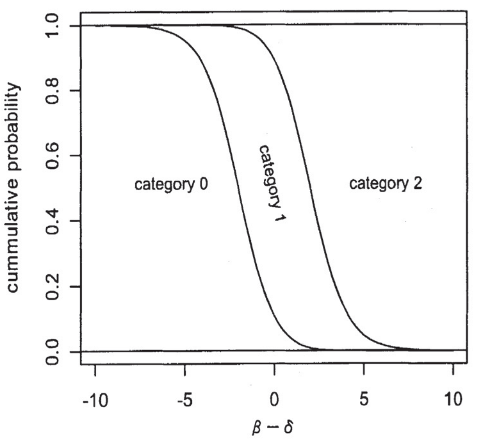

図 4-1　カテゴリ生起確率($\tau_1 = -2$)

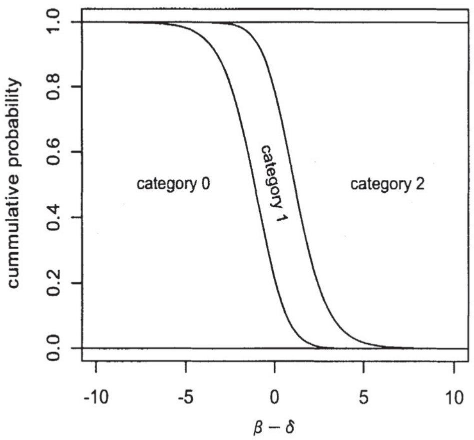

図 4-2　カテゴリ生起確率($\tau_1 = -1$)

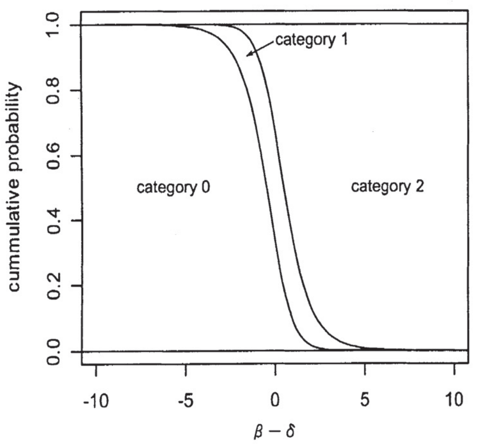

図 4-3　カテゴリ生起確率($\tau_1 = 0$)

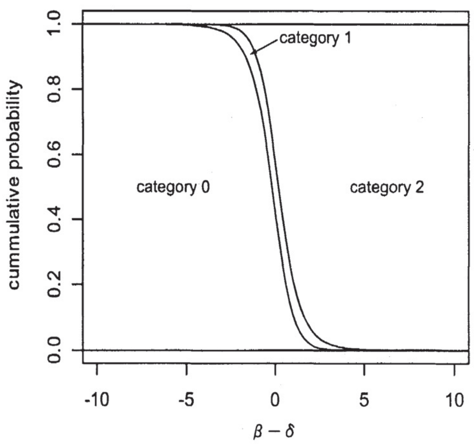

図 4-4　カテゴリ生起確率($\tau_1 = 1$)

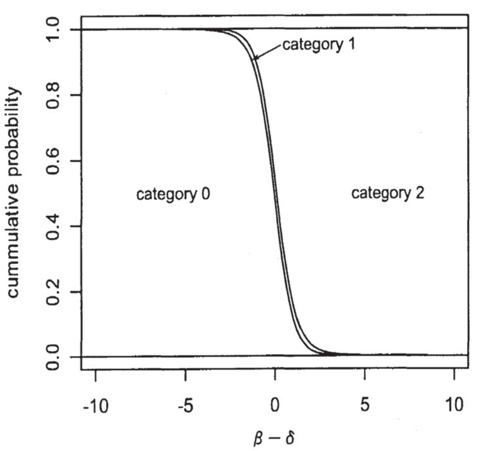

図 4-5　カテゴリ生起確率($\tau_1 = 2$)

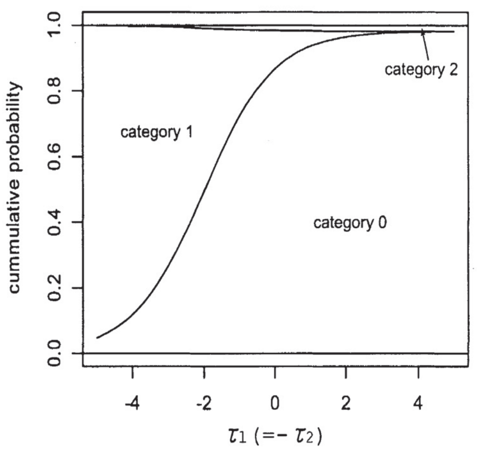

図 4-6 カテゴリ生起確率($\beta - \delta = -2$)

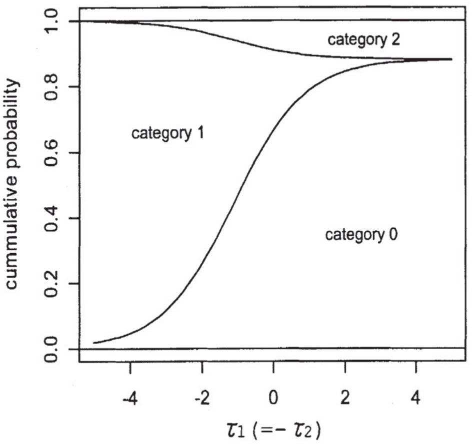

図 4-7　カテゴリ生起確率($\beta-\delta=-1$)

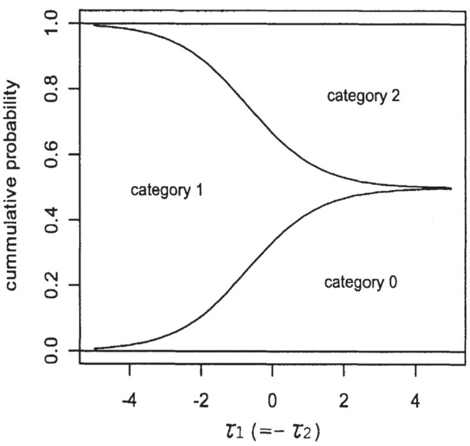

図 4-8 カテゴリ生起確率($\beta-\delta=0$)

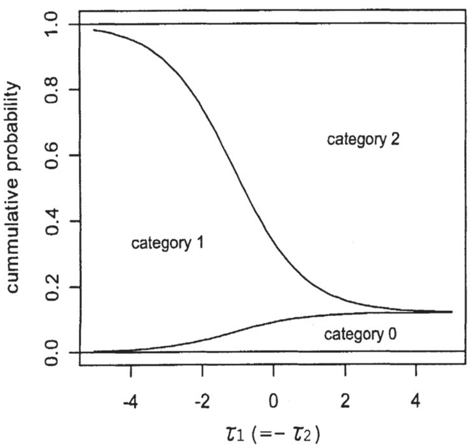

図 4-9　カテゴリ生起確率($\beta-\delta=1$)

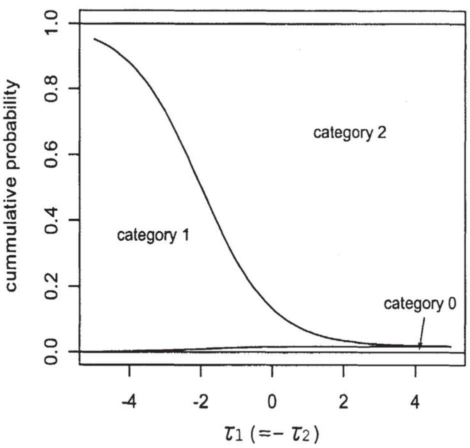

図 4-10 カテゴリ生起確率($\beta-\delta=2$)

PRM に適っているかどうかの検討は必須である。PRM 運用の際にデータが PRM に適っていなければ，Rasch モデルが持つ客観性は発揮できない。

実際に様々な多カテゴリテストやアンケートを設計し，継続的に分析を行う，あるいは能力を測るというような実践的な運用の際には，以上のような PRM の特徴を踏まえ，カテゴリ閾値とカテゴリ特性の関係から項目あるいはテスト全体のカテゴリ閾値の検討を行う必要があり，同時にデータがモデルに適合しているかどうかについても細心の注意を払う必要がある。

V．おわりに

本稿では，PRM におけるカテゴリ閾値から客観性の解釈を行った。PRM においては，多値項目を

規定するカテゴリ閾値の扱いが複雑であるが，多値項目のカテゴリ閾値はモデル上においてはどのような値でもよく，その値によって多値項目のカテゴリ特性が柔軟に決定される。一項目全体として弁別性相当のものを考えたときには，複数の項目でカテゴリ閾値が違えば，弁別力も異なる。RSM では一テストにおいては全ての項目のカテゴリ閾値は同じであるが，複数テストを想定すると，テストごとに異なる可能性がある。PCM においてはそもそも項目ごとにカテゴリ閾値，カテゴリ数が違う。しかし，どちらの PRM のどの項目においても隣接カテゴリ間，つまり段階での生起確率は 2 値項目と同じロジスティック曲線となる。このことが PRM に Rasch モデル族としての客観性を与えている。

本稿において，PRM のモデル上はカテゴリ閾値・カテゴリ特性は柔軟であると結論付けている。これはあくまでモデル上の規定のことである。PRM が適用される対象においては柔軟なカテゴリ特性が現実反応にそぐわないことも想定される。しかし，これはモデル上の規定ではなく，あくまで適用対象における問題である。モデル規定とは別に，個別に議論する姿勢が必要であろう。

また，実際にテストやアンケートに PRM を適用する際，特に RSM においては一テストで共通のカテゴリ閾値を想定しているのであるから，カテゴリ閾値の考察は必須である。まず，テスト全体，あるいは項目ごとに，意図した項目カテゴリ特性かどうかの判断が必要であろうし，複数のテストのことを考えれば，カテゴリ特性はある程度の幅に抑えておいたほうが現実的である。例えば得られたデータを PCM でも検討し，全体の平均としてのカテゴリ閾値と個別項目として考えたときのカテゴリ閾値の違いを検討する必要性もあろう。ただし，適用時にカテゴリ閾値をどのように判断し設計するかは，行うテスト，項目，継続性などにより異なる。これはモデル上で議論することではなく，実際の具体的な運用時に考慮すべき事柄であり，本稿においては様々な実践的適用におけるカテゴリ閾値，あるいはカテゴリ特性の具体的な議論は行っていない。

PRM を実践的に適用する際には，2 値 Rasch モデルの運用以上に十分配慮すべき要素が多い。Rasch モデル族としての特徴を正しく把握し，適用する対象の調査も重ね，慎重に適用することが望まれる。

参考文献

1) Andrich, D. : “A rating formulation for ordered response categories.”, *Psychometrika*, 43, pp.561-573, 1978.

2) Masters, G. N. : “A Rasch model for partial credit scoring.” *Psychometrika*, 47, 2, pp.149-174, 1982.

3) Rasch, G. : *Probabilistic models for some intelligence and attainment tests*, The Danish Institute for Educational Research, 1960. (Reprinted in 1980 by the university of Chicago Press with a Foreword and Afterword by Wright, B. D.:G.ラッシュ，内田良男(完訳): 心理テストの確率モデル，名古屋大学出版会，1985)

4) Samajima, F. : “Estimation of latent ability using a response pattern of graded scores.”, *Psychometrika*,

Monograph Suppliment 17, 1969.

5) Guttman, L. : “A basis for scaling qualitative data.”, *American Sociological Review*, 9, pp.139-150, 1944.

6) Linacre, J. M. : “Optimizing Rating Scale Category Effectiveness.”, *Journal of Applied Measurement*, 3, 1, pp.85-106, 2002.

7) Luo, G. : “The relationship between the rating scale and partial credit models and the implication of disordered thresholds of the Rasch models for polytomous responses.”, *Journal of Applied Measurement*, 6, 4, pp.443-455, 2005.

8) Huynh, H. : “On equivalence between a partial credit item and a set of independent Rasch binary items.”, *Psychometrika*, 59, 1, pp.111-119, 1994.

9) Huynh, H. : “Decomposition of a Rasch partial credit item into independent binary and indecomposable trinary items.”, *Psychometrika*, 61, 1, pp.31-39, 1996.