

[研究論文]

項目応答理論に基づく母数推定法とテストの分析

Parameter estimation methods and an analysis of the examination
based upon item response theory

田中 武之・山川 修・菊沢 正裕

T. Tanaka, O. Yamakawa and M. Kikusawa

Keywords:IRT, parameter estimation, logistic model, EM method, maximum likelihood method.

1. はじめに

学生や受験者（以下被験者とよぶ）の学力・能力をテストによって客観的に測定する理論として、項目応答理論（Item Response Theory, 以下IRTとよぶ）がある。IRTは、テストの特性を表す項目特性関数を用いた統計的手法により、被験者の学力（特性値とよぶ）の推定や、テストの特性の分析、さらには被験者に応じたテストの構成法などに広く応用される。IRTに関するこれまでの先行研究では、実データに基づく統計的推定や応用例は多く発表されているが、数値解析的な検討を行った例は少数である。本論文の前半では、IRTにおける項目特性関数中の母数推定法について、計算手法と結果の妥当性を検証する目的で、2つの典型的なアルゴリズム（直接法とEM法）の特徴を比較検討する。後半では、2003年度に実施した情報科学の期末試験結果をデータ化し、実際にIRTを適用してテストの特性の分析を行う。今回、新たに推定のためのプログラムを書きおこして計算した結果、直接法とEM法の数値計算上の利点・欠点、2母数および3母数ロジスティックモデルの特徴、被験者特性の最尤推定値とベイズ推定値の傾向が明らかになった。

2. 母数推定アルゴリズムの比較検討

この節では、IRTにおける項目母数および特性値の一般的な推定方法について述べる。IRTという項目とはテストの中の小問のことである。IRTの理論や応用に関しては、文献1),2)を参照されたい。母数推定に関してはこれまで各種の方法が提案されているが²⁾、ここでは数値解析の観点から直接法とEM法の特徴を比較検討する。

受理日 2004. 5.18

所 属 福井県立大学情報センター

2. 1 項目特性関数

まず、学力テスト結果を $n \times m$ 行列 $U = (u_{ij})$ で表現する。ただし、 i は被験者、 j は項目を表す添字であり、 u_{ij} は次の2値をとる確率変数である。

$$u_{ij} = 0 \quad (\text{被験者 } i \text{ が項目 } j \text{ に誤答したとき})$$

$$u_{ij} = 1 \quad (\text{被験者 } i \text{ が項目 } j \text{ に正答したとき})$$

行列 U は応答パターン行列とよばれる。ここで、 n は全被験者数、 m は全項目数である。($i = 1, 2, \dots, n, j = 1, 2, \dots, m$)

対象とする被験者それぞれについてテストによって測定される量 (たとえば学力) を潜在特性値、あるいは単に特性値とよび θ で表す。特性値 θ の被験者が項目 j に正答する確率として次のような式を用いる。

$$P_j(\theta) = \Pr[u_{ij} = 1 | \theta] = c_j + \frac{1 - c_j}{1 + \exp(-Da_j(\theta - b_j))} \quad (1)$$

(1)式は3母数ロジスティックモデルとよばれる。母数 a_j は識別力、 b_j は困難度、 c_j は擬似偶然水準といい、 D は $D = 1.7$ の定数である。ここで、 c_j は $0 \leq c_j < 1$ の範囲にある。(1)式で $c_j = 0$ としたものは2母数ロジスティックモデル、また、 $a_j = a$ (定数)、 $c_j = 0$ としたものは1母数ロジスティックモデルとよばれる。ロジスティックモデルは、特性値が低い被験者は正答率が低く、逆に特性値が高い被験者は正答率が高いという傾向を表すモデルの一つである。このほかにもいろいろなモデルが提案されているが²⁾、実用上、ロジスティックモデルが使用されることが多い。IRTでは一般に、(1)式のように、特性値 θ の関数になっている正答率 $P_j(\theta)$ を項目特性関数とよんでいる。

項目特性関数の形を仮定すると、被験者 i が項目 j に u_{ij} を応答する確率は次のように表される。

$$f(u_{ij} | \theta_i, \lambda_j) = \begin{cases} P_j(\theta_i) & (u_{ij} = 1) \\ 1 - P_j(\theta_i) & (u_{ij} = 0) \\ 1 & (u_{ij} \text{ が欠測値のとき}) \end{cases} \quad (2)$$

ここで、 $\lambda_j = (a_j, b_j, c_j)$ である。記号 $f(\cdot | \cdot)$ は条件付確率あるいは条件付確率密度関数を表すもので、以降の式で多用する。確率変数名が異なる場合、異なる関数を表すことに注意されたい。なお、(2)式で、 $1 - P$ の形では数値計算の際、丸め誤差の影響が大きいため別の形で表現する必要がある。(付録1)

項目応答理論に基づく母数推定法とテストの分析

項目間の局所独立性を仮定すると、被験者*i*が応答 $u_i = (u_{i1}, u_{i2}, \dots, u_{im})$ を返す確率は

$$f(\mathbf{u}_i | \boldsymbol{\theta}_i, \Lambda) = \prod_{j=1}^m f(u_{ij} | \theta_i, \lambda_j), \quad \text{ただし } \Lambda = (\lambda_j). \quad (3)$$

となる。(2), (3)式で項目母数 λ_j および特性値 θ_i がともに未知であるため、応答パターン U の情報に基づいて、合計 $3m+n$ 個の未知母数を推定することになる。一般に統計解析では、性質が不明な母数群の推定には、次節のように最尤推定法が用いられることが多い。IRTの分野でよく用いられるLOGIST³⁾(同時最尤推定法)や、BILOG⁴⁾(周辺最尤推定法)というソフトウェアも最尤推定法に基づくものである。

2. 2 被験者特性値の推定

この節ではIRTによる被験者特性値 θ の推定について述べる。実際の手順としては、項目母数の値を求めた後の処理になるが、項目母数の推定については次節以降で述べることとして、この節では項目母数が定まっているという条件下で特性値 θ の推定法を示す。

推定には最尤推定法またはベイズ推定法が用いられる。最尤推定値を求める場合はそれぞれの被験者*i*について、

$$\log L(\theta_i | \mathbf{u}_i, \Lambda) = \sum_{j=1}^m \log f(u_{ij} | \theta_i, \lambda_j) \quad (4)$$

を最大にする θ_i が最尤推定値になる。ただし、全問正解または全問不正解の被験者については、特性値の最尤推定値を定めることができない。

ベイズ推定値を求める場合には同様に

$$\log f(\theta_i | \mathbf{u}_i, \Lambda) = \log L(\theta_i | \mathbf{u}_i, \Lambda) + \log g(\theta_i) - \log f(\mathbf{u}_i | \Lambda) \quad (5)$$

を最大にする θ_i がベイズ推定値(点推定値)になる。ところで、(5)式の右辺第3項は θ_i に関わらないので最大化の計算に寄与しない。つまり、ベイズ推定値と最尤推定値の違いは右辺第2項の有無だけとなるため、両者の計算は似たような手順で行える。また、ベイズ推定値によれば、全問正解または全問不正解の被験者についても特性値の推定値が定まる。

2.3 項目母数の推定（直接法）

項目間の局所独立性、および、被験者の特性値 θ_i が独立かつ同一の分布に従うと仮定すると、未知母数 $\Theta = (\theta_i)$ 、 $\Lambda = (\lambda_j)$ のもとで応答 U が出現する条件付確率は、

$$f(U|\Theta, \Lambda) = \prod_{i=1}^n f(\mathbf{u}_i | \theta_i, \Lambda) = \prod_{i=1}^n \prod_{j=1}^m f(u_{ij} | \theta_i, \lambda_j) \quad (6)$$

と表される。(4)式の U に実測値を与え、 Θ, Λ の関数とみたときの f を尤度といい $L(\Theta, \Lambda | U)$ と表す。すなわち、

$$L(\Theta, \Lambda | U) = \Pr[U = (\mathbf{u}_{ij}) | \Theta, \Lambda] = f(U | \Theta, \Lambda) \quad (7)$$

である。尤度の対数をとって

$$\log L(\Theta, \Lambda | U) = \sum_{i=1}^n \sum_{j=1}^m \log f(u_{ij} | \theta_i, \lambda_j) \quad (8)$$

を最大にするように θ_i, λ_j を定める方法が古典的な最尤推定法である。

しかし、このようにして θ_i, λ_j を推定することは、統計的に好ましくない性質があることが知られている。通常、統計的推定ではデータ数 n が増加すると推定誤差が小さくなるのが一般的であるが、この場合、 n が増加すると同時に未知母数 θ_i の数も増加するため、推定値の安定性が向上しないためである³⁾。

そこで、 θ_i が未知の場合には、これに適当な事前分布 $g(\theta)$ 、例えば標準正規分布を仮定し、次式のように θ の周辺尤度関数をもって最大化をはかる方法が考えられる。

$$\begin{aligned} L(\Lambda | U) &= f(U | \Lambda) = \prod_{i=1}^n f(\mathbf{u}_i | \Lambda) = \prod_{i=1}^n \int_{-\infty}^{\infty} f(\mathbf{u}_i, \theta | \Lambda) g(\theta) d\theta \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} \prod_{j=1}^m f(u_{ij} | \theta, \lambda_j) \cdot g(\theta) d\theta \end{aligned} \quad (9)$$

この対数をとって、

$$\log L(\Lambda | U) = \sum_{i=1}^n \log \int_{-\infty}^{\infty} \prod_{j=1}^m f(u_{ij} | \theta, \lambda_j) \cdot g(\theta) d\theta \quad (10)$$

を最大化することで、項目母数 λ_j を推定する方法をここでは直接法とよぶことにする。3母数モデルの場合、 $3m$ 変数の最大化問題を解くことになる。数値積分の分点数を K とすると、項目特性関数 $P(\theta_i | \lambda_j)$ の評価回数は $n \times m^2 \times K$ のオーダーとなる。このように、周辺尤度

項目応答理論に基づく母数推定法とテストの分析

関数を用いる場合、積分を含む最大化問題を解かねばならないため、項目数 m が増加するにつれ莫大な計算量が要求される。

なお、ロジスティックモデルは実質的に $a(\theta - b)$ の関数の形になっており、尺度と位置に関して θ と a, b は相対的關係にあるので、一意的に決定できない。したがって、推定の際にはあらかじめ、 a, b の尺度と位置を定義しておくか、または θ の平均と分散を設定しておくなどの必要がある。今回は後者の方法を取り、 θ に標準正規分布（平均0，分散1）を仮定した。

2.3 項目母数の推定（EM法）

一方、IRTの母数推定にはEM法⁶⁾とよばれる計算アルゴリズムがよく用いられる。EM法は確率論的な観点から構成されたアルゴリズムであり、以下に示すようにEステップとMステップを繰り返して最適解を求める方法である。

Eステップでは各被験者の特性値 θ_i の確率分布の近似値を評価する。まず、 $f(\theta_i | u_i, \Lambda)$ の未知母数 Λ に対して近似値 $\Lambda^{(0)}$ を与えたものを $h(\theta_i | u_i, \Lambda^{(0)})$ とする。

$$h(\theta_i | \mathbf{u}_i, \Lambda^{(0)}) = f(\theta_i | \mathbf{u}_i, \Lambda) \Big|_{\Lambda = \Lambda^{(0)}} \quad (11)$$

これをベイズの定理を用いて変形し、

$$h(\theta_i | \mathbf{u}_i, \Lambda^{(0)}) = \frac{1}{s_i} f(\mathbf{u}_i | \theta_i, \Lambda^{(0)}) g(\theta_i). \quad (12)$$

を計算する。ここで、

$$s_i = \int_{-\infty}^{\infty} f(\mathbf{u}_i | \theta, \Lambda^{(0)}) g(\theta) d\theta. \quad (13)$$

である。もし、 $\Lambda^{(0)}$ が真の項目母数ならば、 h は θ の事後分布を与えることに注意されたい。この確率分布 h は後の積分のために、各点での値をあらかじめ数値的に評価しなければならない。一般に、無限遠において指数的に減少する正值関数の無限区間における積分については台形則が最良近似を与えることが知られており、分点のとり方が適当であれば台形則で十分良い積分近似値が得られる（付録3参照）。たとえば、等間隔台形則による場合、

$$h_{ik} = h(\theta | \mathbf{u}_i, \Lambda^{(0)}) \Big|_{\theta = d_k} = \frac{1}{s_i} \prod_{j=1}^m f(u_{ij} | \theta, \Lambda^{(0)}) \Big|_{\theta = d_k} g(\theta) \Big|_{\theta = d_k} \quad (14)$$

$$s_i = \sum_{k=-K}^K \prod_{j=1}^m f(u_{ij} | \theta, \Lambda^{(0)}) \Big|_{\theta = d_k} g(\theta) \Big|_{\theta = d_k} \quad (15)$$

となる。ここで、 d_k は原点をはさんで等間隔に離散評価したときの分点を表し、 $2K+1$ 個の分

点はあらかじめ固定しておく。すなわち、 $d_{k+1} - d_k = e$ (定数), $d_0 = 0$ とする。(14)式の h_{ik} は $\Lambda \sim \Lambda^{(0)}$ の近似のもとで、特性値 θ_i の条件付確率分布の離散近似値を与えるものである。

Mステップでは、Eステップで計算した確率分布 h を使って推定値 $\Lambda^{(0)}$ を改良する。実測値 U , 欠測値 θ の条件下でのパラメータ Λ に関する完全データ尤度関数を $M(\Lambda | U, \theta)$ と表し、 $\log M$ の θ に関する条件付期待値をとると、

$$\begin{aligned}
 E[\log M | U] &= E[\log f(U, \theta | \Lambda) | U] = \sum_{i=1}^n E[\log f(u_i, \theta | \Lambda) | U] \\
 &= \sum_{i=1}^n \int_{-\infty}^{\infty} \log f(u_i, \theta | \Lambda) \cdot h(\theta | u_i, \Lambda^{(0)}) d\theta = \sum_{i=1}^n \int_{-\infty}^{\infty} \log \{f(u_i | \theta, \Lambda) g(\theta)\} \cdot h(\theta | u_i, \Lambda^{(0)}) d\theta \\
 &= \sum_{i=1}^n \int_{-\infty}^{\infty} \log f(u_i | \theta, \Lambda) \cdot h(\theta | u_i, \Lambda^{(0)}) d\theta + (\Lambda \text{ not implied}) \\
 &\approx \sum_{i=1}^n \int_{-\infty}^{\infty} \log f(u_i | \theta, \Lambda) \cdot h(\theta | u_i, \Lambda^{(0)}) d\theta = \sum_{j=1}^m \sum_{i=1}^n \int_{-\infty}^{\infty} \log f(u_{ij} | \theta, \lambda_j) \cdot h(\theta | u_i, \Lambda^{(0)}) d\theta \\
 &\approx \sum_{j=1}^m \sum_{i=1}^n \sum_{k=-K}^K \log f(u_{ij} | \theta, \lambda_j) \Big|_{\theta=d_k} \cdot h_k
 \end{aligned} \tag{16}$$

となる。ここで、 d_k は(14)式と同じく数値積分の分点を表す。(16)式を用いて、期待完全データ尤度関数 $E[\log M | U]$ を最大にするような λ_j を求めるのがMステップである。式の形から、各 j について個別に最適解 $\lambda_j^{(1)}$ を求めればよい。つまり、3母数ロジスティックモデルでは、3変数関数の最大化を m 回行うことになる。一般に、任意の $\Lambda_j^{(0)}$ に対して上の方法で導かれた $\Lambda_j^{(1)}$ は、ある仮定の下で

$$L(\Lambda^{(1)} | U) \geq L(\Lambda^{(0)} | U) \tag{17}$$

を満たすことが示される。よって、 $\Lambda^{(0)}$ を $\Lambda^{(1)}$ で置き換えてEステップに戻り、再びMステップを計算すると L の値が増加する。以下同様にEステップとMステップを繰り返して、 $\Lambda^{(0)} \rightarrow \Lambda^{(1)} \rightarrow \Lambda^{(2)} \rightarrow \dots$ を順に計算して L を最大にする最適解を求める方法がEM法である。

2.4 直接法とEM法の比較

直接法とEM法の数値解析上の特徴を表1に示す。まず、計算量について考察すると、2.2節と2.3節で述べたように、項目特性関数 $P(\theta_i | \lambda_j)$ の評価回数に関して、直接法の計算量は繰り返し数1回につき $n \times m^2 \times K$ のオーダーであるのに対し、EM法では1回につき $n \times m \times K$ のオーダーである。したがって、EM法の方が1回当たりの計算量では有利である。しかし、EM法は直接法に比べ収束が遅く多数の繰り返しを必要とするので、一概にどちらが有利とはいえない。(表1) 繰り返し数の多寡は、目的関数が期待対数周辺尤度か真の周辺尤度かの違いに起因している。そこで、パソコンを使って計算時間を測定した。(パソコンの性能は表2脚注を参照)

項目応答理論に基づく母数推定法とテストの分析

計算の条件は, $n = 400$, $m = 30$, 相対誤差 10^{-6} とした. 数値積分に関しては, EM法では分点数101点, 分点間隔0.2の等間隔台形則(付録3)を用いた. 一方, 直接法では高速に計算させるため自動積分ルチン(SSL2)⁷⁾を用いた関係上, 分点数は131~645の範囲で変動する.

以上のような条件下での計算時間を表2に示す. 2母数モデルでは直接法のほうが有利となったが, 3母数モデルではEM法のほうが有利な結果となった. 2母数モデルでEM法が遅い原因は収束速度が遅い(繰り返し数が多い)ためである. 3母数モデルで直接法が遅いのは, ヘッセ行列のランクが高いことが影響しているためで, ランクが60階から90階に上がったことにより, 計算時間がEM法より10倍以上かかっている. また, 直接法をEM法と同条件の等間隔台形則で計算した場合, 自動積分の場合よりさらに2~3倍の時間がかかる結果となった.

推定値の収束速度は, 多変数関数の最大化問題を解くアルゴリズムにも依存するが, 今回は汎用の数値計算ライブラリ(SSL2)⁸⁾を用いた. 多変数関数の最大化の概要については付録2を参照されたい. EM法の収束を加速するアルゴリズムも提案されているが⁹⁾, 期待対数尤度に基づく限り大幅な収束速度向上は困難と考えられる. しかし, 次の2点において, EM法から得られる情報は大きい.

第1点は, 特性値 θ_i の確率分布(事後分布)が得られることである. EM法のステップが進行するとともに確率変数ベクトル Θ が法則収束することから, (11)式の h の極限分布 $h(\theta_i | u_i, \Lambda^{(\infty)})$ と特性値の推定値の事後分布 $f(\theta_i | u_i, \Lambda)$ は一致する. したがって, 推定値の平均や分散, 区間推定など, 推定値の確率分布が必要な場合には, EM法収束後の(14)式の h_{ik} を使って各種統計量の近似評価が行える. この方法は推定値の確率分布を解析的に表現することが困難な場合に有効である.

表1 直接法とEM法の特徴

	利点	欠点
直接法	1.種々の数値積分法が適用可能で, 計算精度が高い. 2.ヘッセ行列の精度が増せば収束は早い.	1.項目数が多い場合, ヘッセ行列の評価が困難. 2.アンダーフローに注意が必要.
EM法	1.ヘッセ行列のランクが小さくてすむため, 多母数・多項目でも適用可能. 2.項目母数の標準誤差と特性値の確率分布が同時に求められる.	1.収束が遅い. 計算精度を上げるためには多数の分点が必要. 2.最適化に入る前に確率分布の離散近似を行うため, 計算誤差に注意が必要.

表2 計算時間の比較* (単位: 秒)

	直接法	EM法
2母数ロジスティックモデル	1518.844	1743.375
3母数ロジスティックモデル	18569.797	5683.312

*使用したパソコンのスペック: Pentium4 (3GHz), i865G chip set, FSB 800MHz, 2GB DDR-SDRAM (PC3200 dual channel)
プログラムは Fujitsu Fortran95 V4.0 + SSL2 でコンパイルし, 最適化オプションを適用した.

第2点はEM法の収束とともに、項目母数の推定値の標準誤差が得られる点である。Mステップでの関数最大化の際に現れるヘッセ行列(3×3)は漸近的にフィッシャー情報量と一致する。さらに、EM法が収束した時点でのヘッセ行列の逆行列が、漸近共分散行列になることが示される¹⁰⁾。よって、ヘッセ行列の逆行列の対角要素の平方根をとれば、推定値の標準誤差が得られる。これに対し、直接法において関数最大化に用いるヘッセ行列(90×90)は、フィッシャー情報量と対応していない。

また、直接法には数値計算上の問題がある。(10)式が m 回の乗算後積分という形になっているため、項目数 m が大きい場合にはオーバーフローやアンダーフローを起こしやすい。すなわち、 m が大きいと乗算の繰り返しによって指数部の桁数(11ビット)を越えてしまう。アンダーフローによって関数値が0と評価されると、自動積分アルゴリズムがうまく機能しなくなる。マシンやコンパイラによっては、アンダーフローの例外処理を呼び出さない場合があるので注意が必要である。倍精度実数を用いて直接法で自動積分を使う場合、何らかの工夫がない限り、 $m=100$ 程度が計算精度の限界と思われる。一方、EM法では(14)式の計算精度が保証される限り、 m が大きくても問題が起りにくい。本論文では、(14)式を付録1の形で計算することによって桁落ちによる計算精度の劣化を回避している。

計算精度を比較すると、EM法は直接法より劣る。これは、EM法のEステップ中の h の全分点(101点)のうち、個々の i についてそれぞれ10分の1程度の分点(10~15点程度)しかMステップの積分評価に寄与していないためである。しかし、今回の計算ではEM法と直接法の推定値の差は 10^{-3} 以下であり、実用上問題になるほどの差異は認められなかった。

以上のことをまとめると、項目数や被験者数が多い場合はEM法が有利であり、直接法はEM法の精度を検証する程度に用いるのが現実的であると結論づけられる。

3. 情報科学テストへの適用結果

前節で述べた推定方法を、2003年度「情報科学」期末試験に適用し、項目母数と特性値の推定を行った。情報科学は400名余りの1年生が受講しているが、そのうち期末試験を受験した400名の解答を整理した。試験問題(付録4)は全30問で、解答にあたり教科書¹¹⁾の持ち込みを許可しており、一部を除き各問とも「4つの選択肢から正しいものをすべて選べ」という形式である。採点は、正しい選択肢をすべて選んだ場合にのみ正答とし、1つでも間違えると誤答とした。(この採点は現実に行った成績評価とは無関係である。)よって、被験者が偶然に正答する可能性は低いと考えられる。

表3に2母数および3母数ロジスティックモデルによる識別力 a 、困難度 b 、擬似偶然水準 c の推定値を示す。計算問題やCPUの動作を問う問題(問17, 20, 21, 22)の識別力は高い。一方、用語の意味や単なる知識を問う問題の識別力は低い。また、やさしい問題(特性値が低く

項目応答理論に基づく母数推定法とテストの分析

表3 項目母数の推定値

項目	正答率	2母数ロジスティックモデル		3母数ロジスティックモデル		
		a	b	a	b	c
問1	0.243	0.300	2.357	0.262	2.666	0.000
問2	0.245	0.253	2.719	0.542	2.359	0.131
問3	0.258	0.396	1.721	13.527	1.549	0.207
問4	0.798	0.226	-3.681	0.224	-2.674	0.254
問5	0.143	0.338	3.318	0.301	3.695	0.000
問6	0.185	0.026	33.502	2.238	2.534	0.176
問7	0.468	0.173	0.453	8.954	1.677	0.440
問8	0.395	0.305	0.873	0.794	1.592	0.278
問9	0.903	0.495	-2.961	0.533	-1.954	0.451
問10	0.740	0.210	-3.008	0.550	0.982	0.620
問11	0.820	0.206	-4.430	11.227	1.385	0.803
問12	0.820	0.228	-4.036	0.383	0.042	0.644
問13	0.445	0.151	0.872	9.192	1.593	0.410
問14	0.680	0.585	-0.911	0.581	-0.911	0.000
問15	0.780	0.496	-1.713	0.465	-1.800	0.001
問16	0.895	0.871	-1.929	0.881	-1.927	0.000
問17	0.550	0.670	-0.226	0.689	-0.188	0.013
問18	0.500	0.199	-0.001	0.245	1.176	0.187
問19	0.413	0.388	0.586	0.593	1.107	0.180
問20	0.730	1.214	-0.792	1.257	-0.779	0.000
問21	0.678	1.268	-0.587	1.276	-0.580	0.000
問22	0.610	0.891	-0.421	1.004	-0.275	0.074
問23	0.318	0.515	1.010	0.516	1.015	0.000
問24	0.520	0.434	-0.124	0.397	-0.129	0.000
問25	0.388	0.385	0.765	0.788	1.309	0.225
問26	0.395	0.240	1.084	0.219	1.187	0.001
問27	0.210	0.115	6.858	0.873	2.880	0.182
問28	0.543	0.297	-0.359	0.495	0.876	0.301
問29	0.103	0.257	5.142	0.250	5.268	0.000
問30	0.773	0.196	-3.766	0.160	-4.128	0.086

表4 2母数モデルと3母数モデルの比較

	利点	欠点
2母数モデル	1. 3母数モデルよりも統計的に頑強. 2. 計算が容易で、短時間で推定可能.	1. 特性値が高い、もしくは低い被験者に対して、得られる情報量が少ない.
3母数モデル	1. 2母数モデルよりも細かい項目特性が表現できる. 2. 母数の推定誤差が小さければ、被験者特性について多くの情報量が得られる.	1. 母数の推定に多数の被験者が必要. 2. 統計的に頑強でなく、小数の高特性値被験者の影響を受けることがある.

表5 最尤推定値とベイズ推定値の傾向

最尤推定値	特性値の絶対値を真の値よりも過大に推定する傾向がある。特に高特性値、低特性値の被験者について、真の値からのずれが大きい。
ベイズ推定値	特性値の絶対値を真の値よりも過小に推定する傾向がある。しかし、真の値からのずれは最尤推定値より小さい。

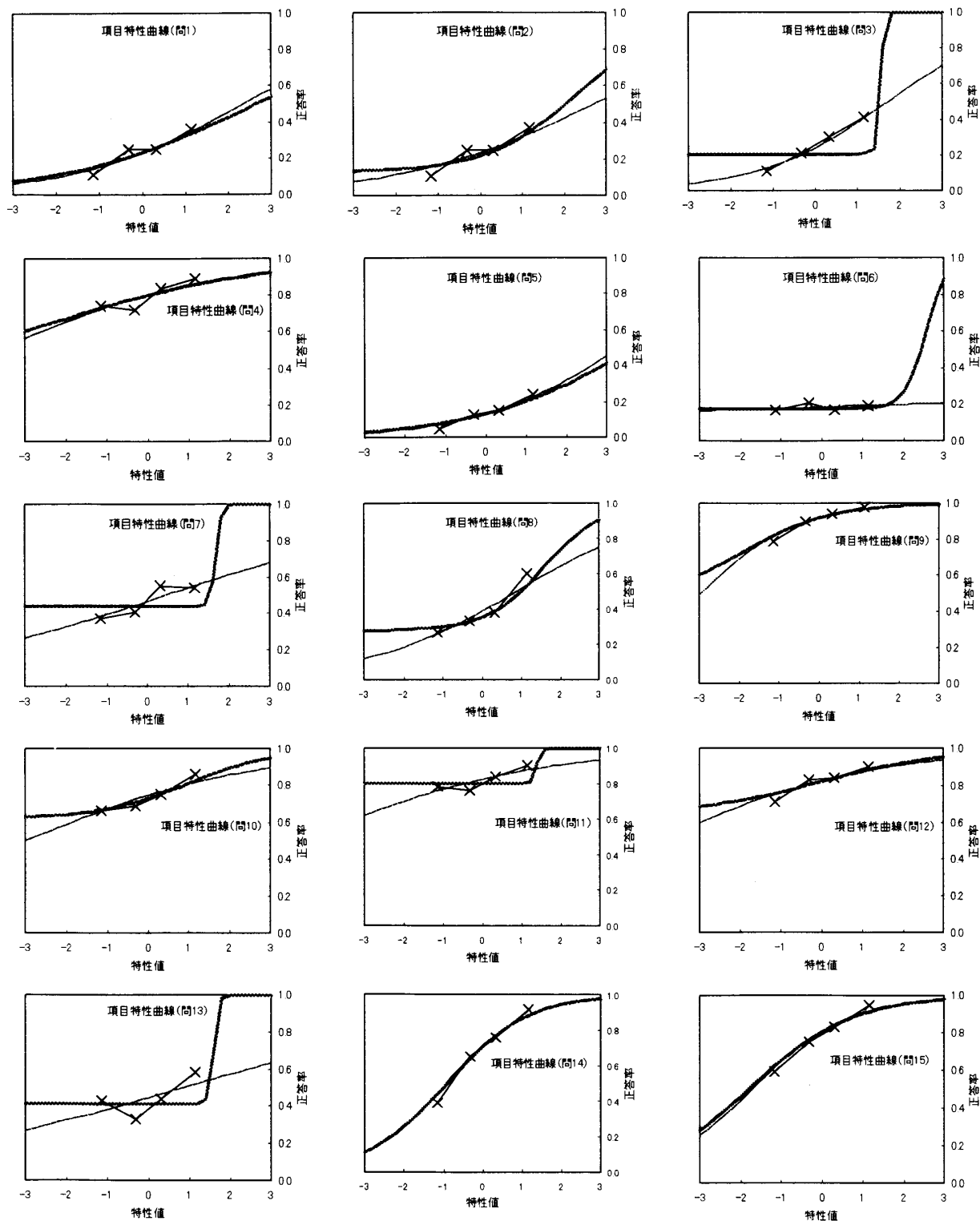


図1 項目特性関数 (問1～問15) (太線は3母数, 細線は2母数, ×印は正答率)

項目応答理論に基づく母数推定法とテストの分析

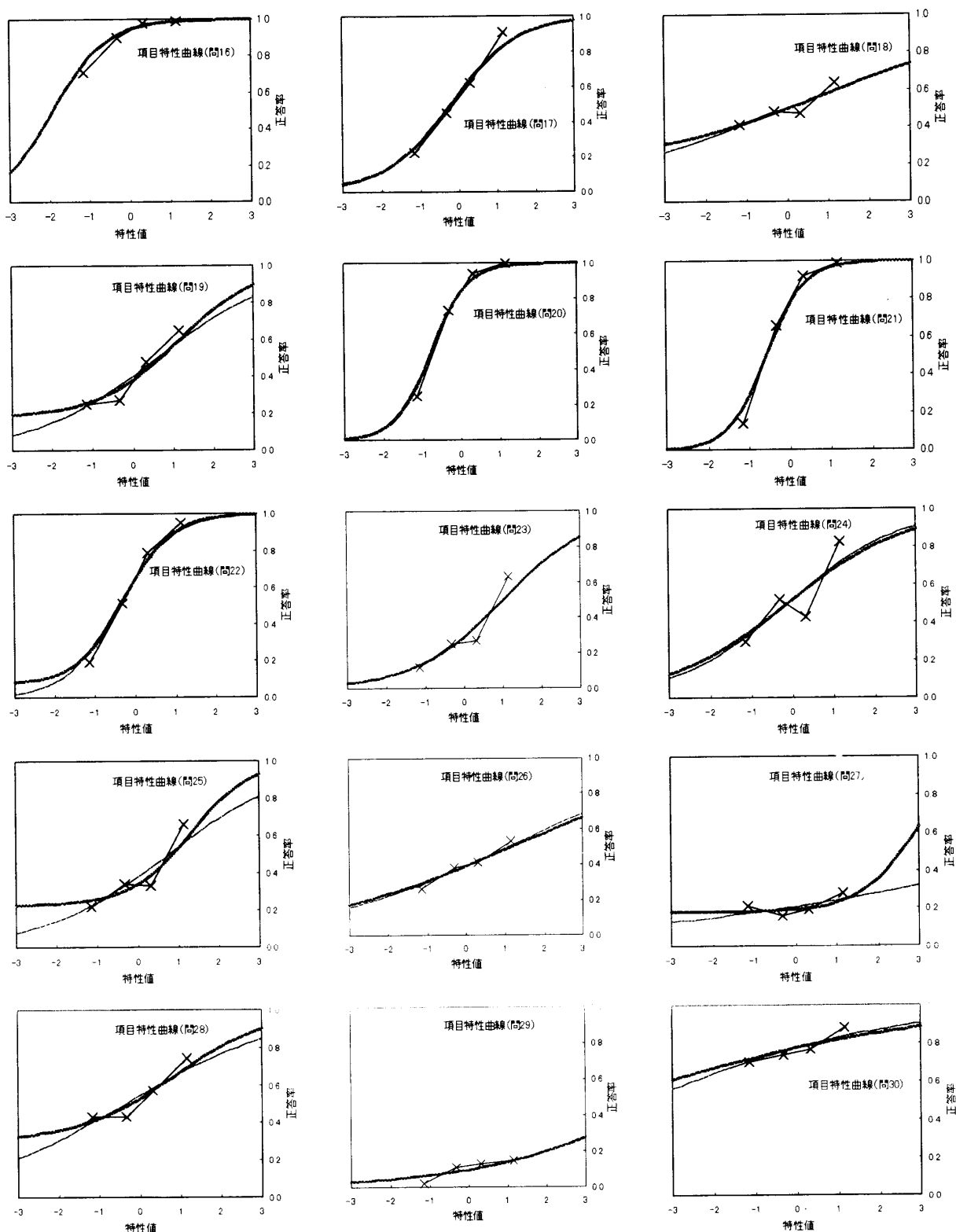


図2 項目特性関数 (問16~問30) (太線は3母数, 細線は2母数, ×印は正答率)

ても正答率が高い項目)や、むずかしい問題(特性値が高くても正答率が低い項目)も識別力が低い。これらのことは図1, 2のグラフを見ると、その傾向がはっきりと読み取れる。

図1, 2は全30項目の項目特性関数をグラフ化したもので、太線が3母数モデル、細線が2母数モデルである。また、4つの×印は、特性値の相対順位において、それぞれ0%~25%, 25%~50%, 50%~75%, 75%~100%の集団に分類したときの正答率である。2母数、3母数モデルとも一致している項目が多いが、一致していない項目(問3, 7, 11, 13)の中には共通の特徴が見られる。特性値の高い被験者の応答パタンの影響を強く受けている点である。そのような被験者は全体の1%程度であるが、3母数モデルは過敏に反応している。すなわち、推定値が少数の高特性値被験者に強く影響されていて、他の大多数の被験者の応答をあまり反映していない。むしろそのような場合には2母数モデルのほうがよくフィットしている。その他の点を含め、2母数モデルと3母数モデルの特徴を表4に示す。

次に、特性値の検討を行う。図3に示すように、2母数モデル(△印)、3母数モデル(□印)とも素点との相関は同程度に高い。素点は単なる点数の合計であるから、この場合0から30までの整数値であるが、特性値は応答パターンから算出されるため、特性値のとりうる値(応答パタンの組み合わせの数)は 2^{30} 通りある。素点が同一であっても応答パターンが違えば特性値は異なるので、特性値は素点よりも詳細に被験者の特性を評価できる。さらに、素点の順位と特性値の順位が逆転するケースも珍しくない。これは、各問の項目特性に照らして識別力の高い問題に正答すると、高く評価されるためである。

また、特性値の推定値は、2母数モデルか3母数モデルのどちらを選ぶかによっても異なり、それによって順位の逆転も起こりうる。(図4)

しかし、特性値の推定に最尤推定法を用いた場合とベイズ推定法を用いた場合で比較すると、両者の順位は一致し、順位の逆転は起こらない。ただし、図5.1(2母数モデル)および図5.2(3母数モデル)からわかるように、全体として最尤推定値は推定値の絶対値を過大に推定し、ベイズ推定値は推定値の絶対値を過小に推定する傾向が見られる。特性値の最尤推定値とベイズ推定値の正規確率プロット(図5.1, 図5.2)によると、2母数・3母数モデルともベイズ推定値の方が妥当な値を出していることがわかる。一般に、最尤推定法は推定量の絶対値が大きいかほどオーバーに推定する傾向があるのはよく知られた事実である。以上に述べたように、特性値の推定方法の違いによる推定値の傾向を表5にまとめた。

最後に、今回の情報科学試験のテスト情報量について検討する。図6は、被験者5名をピックアップして特性値の推定値の確率密度関数を描いたものである。特性値 $\theta = -0.8$ 付近の推定値の分散が最も小さく、それより遠ざかると推定値の分散が大きくなっている。このことはテスト情報量の概念を用いて明快に説明される。IRTにおけるテスト情報量とは、統計学でいうフィッシャー情報量と同一のものであり、次のような式で表される。

項目応答理論に基づく母数推定法とテストの分析

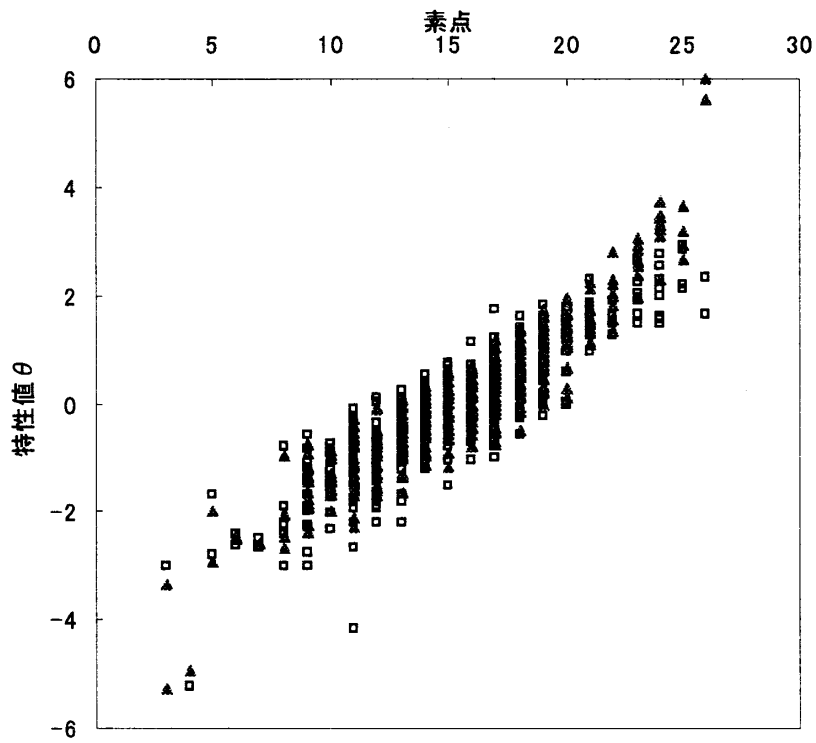


図3 素点と特性値 (□印: 3母数モデル, △印: 2母数モデル)

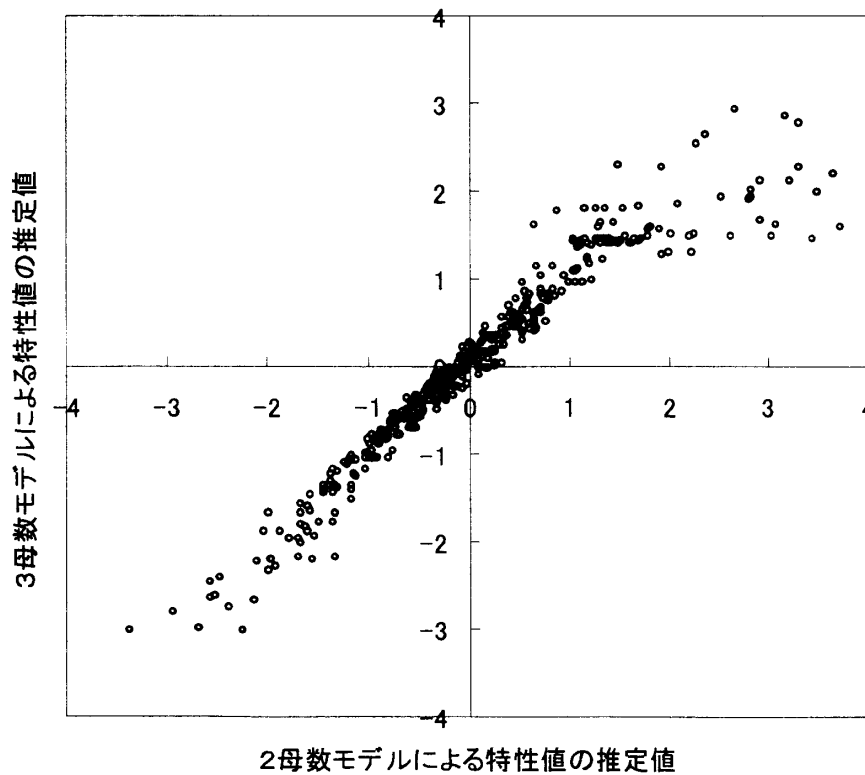


図4 2母数モデルと3母数モデルの特性値の比較

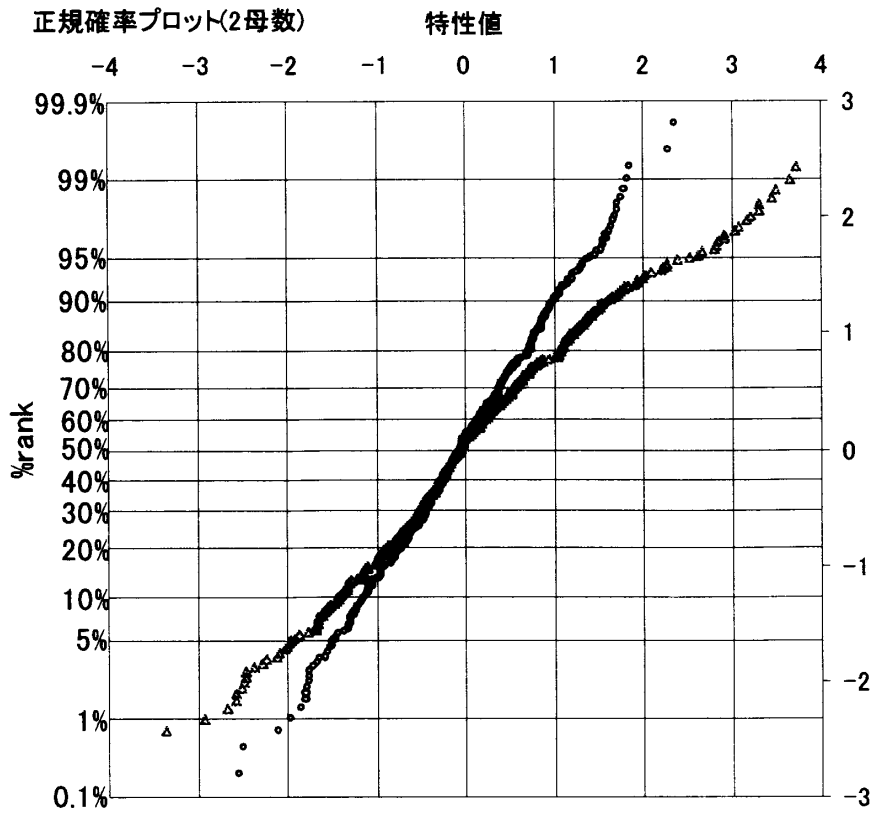


図 5-1 最尤推定値 (△印) とベイズ推定値 (□印) の正規確率プロット(2母数)

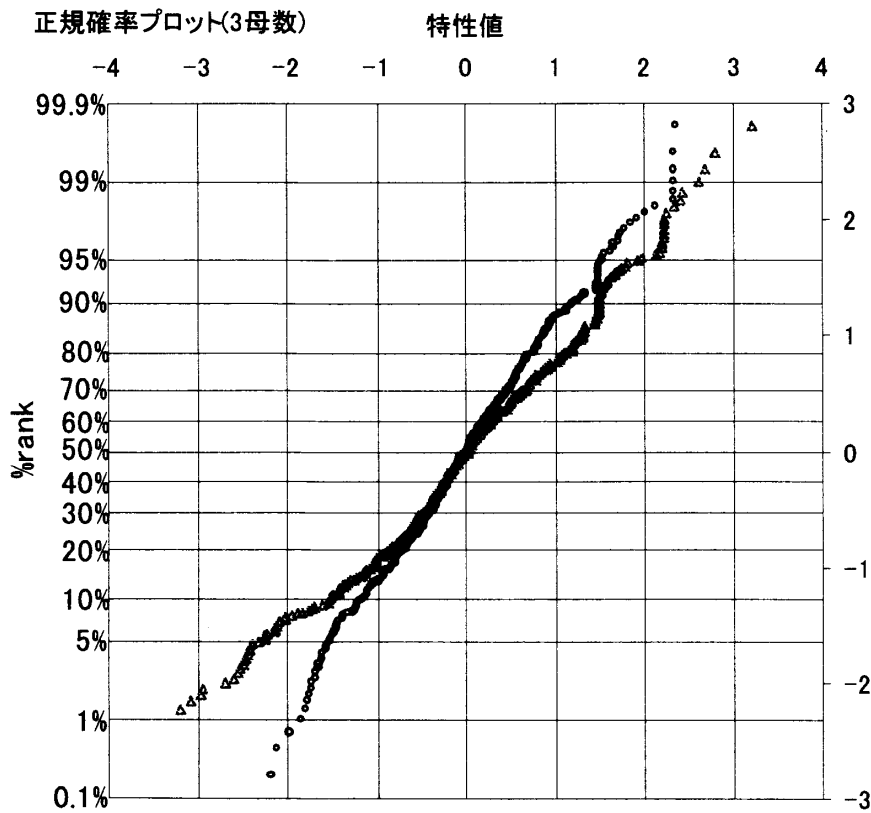


図 5-2 最尤推定値 (△印) とベイズ推定値 (□印) の正規確率プロット(3母数)

項目応答理論に基づく母数推定法とテストの分析

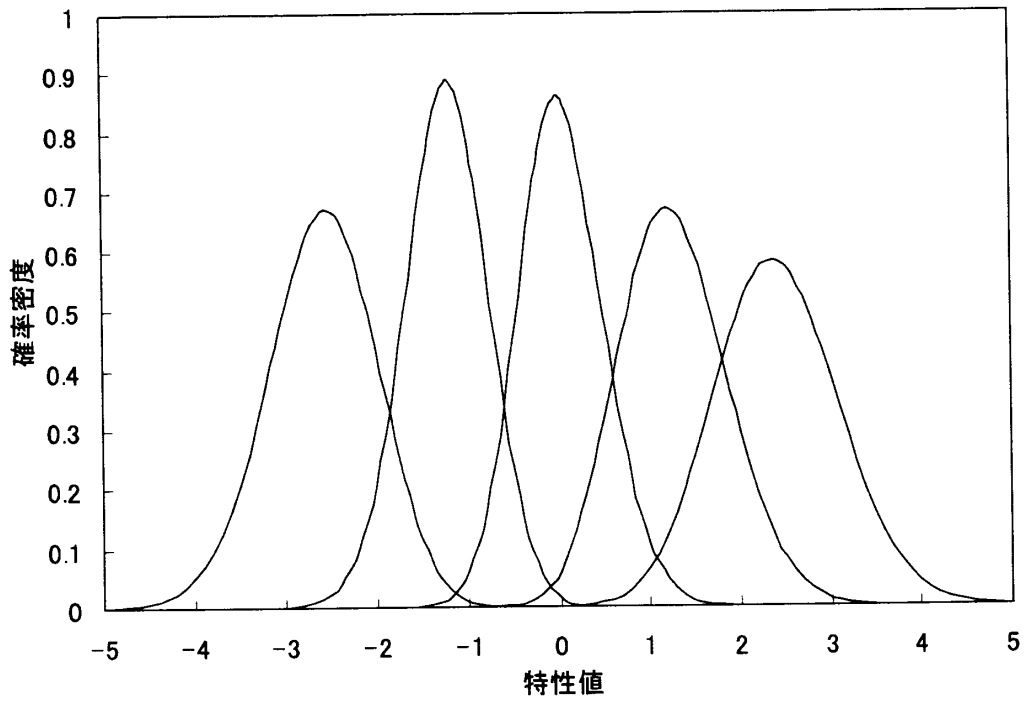


図6 特性値の推定値の確率分布

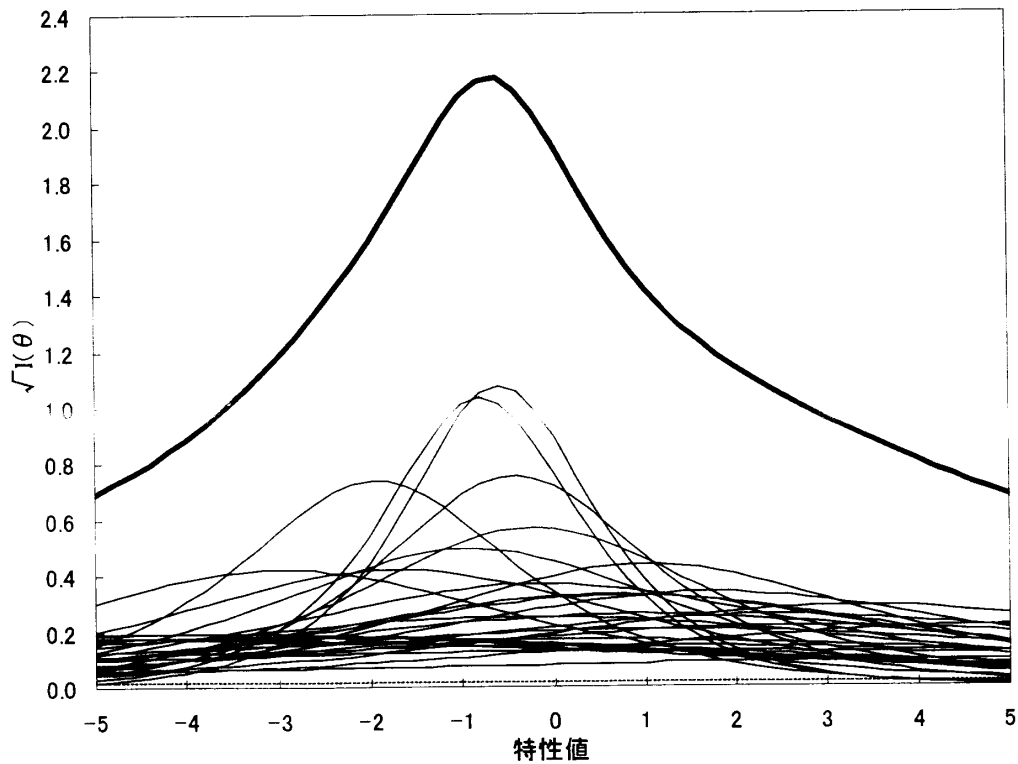


図7 項目情報量 (細線) とテスト情報量 (太線)

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log L(\theta | \mathbf{u}_i) \right)^2 \right] = D^2 \sum_{j=1}^m \frac{a_j^2 (P_j(\theta) - c_j)^2 (1 - P_j(\theta))}{P_j(\theta) (1 - c_j)^2} = \sum_{j=1}^m I_j(\theta) \quad (18)$$

(18)式で、平均は確率変数 u_i に関してとり、 θ はパラメータとみなす。最後の等式で総和記号の中の $I_j(\theta)$ は項目情報量とよばれる。すなわち、テスト情報量は項目情報量の和である。図7に項目情報量（細線）およびテスト情報量（太線）のグラフを示す。図7でテスト情報量のピークが -0.8 付近にあることから、今回の情報科学テストは平均より低め（ $\theta = -0.8$ 付近）の被験者に対して最も測定精度が高かったといえる。

4. むすび

本論文では、IRTにおける項目母数推定法について、数値解析の観点から直接法およびEM法の比較検討を行った。また、母数推定プログラム¹²⁾を開発し、情報科学期末試験のデータを用いて項目母数および特性値の推定を行った。その結果、直接法とEM法の数値計算上の利点・欠点、2母数および3母数ロジスティックモデルの特徴、被験者特性の最尤推定値とベイズ推定値の傾向が明らかになった。今回は推定値の統計的誤差の厳密な評価にまでは至らなかったが、推定値自体は妥当な結果を得たと思われる。

参考文献

- 1) 豊田秀樹, 項目反応理論 [入門編], 朝倉書店, 2002.
- 2) 前川眞一, 4章パラメタの推定, 芝祐順 (編), 項目反応理論—基礎と応用—, 東京大学出版会, 1991.
- 3) Wingersky, M. S., Barton, M. A. & Lord, F. M., LOGIST user's guide, Princeton, NJ: Educational Testing Service, 1990.
- 4) Mislevy, R. J., & Bock, R. D., PC-BILOG 3: Item analysis and test scoring with binary logistic models, Mooresville, ID: Scientific Software, 1989.
- 5) Mislevy, R. J., & Stocking, M.L., A consumer's guide to LOGIST and BILOG, *Applied Psychological Measurement*, 1989, **13**, 57-75.
- 6) Dempster, A. P., Laird, N. M. & Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of Royal Statistical society, Series B*, 1977, **39**, 1-38.
- 7) 秦野甯世, AQEI, SSL II使用手引書, 富士通株式会社, 1987.
- 8) 刀根薫, MINF1, SSL II使用手引書, 富士通株式会社, 1987.
- 9) Ramsay, J. O., Solving implicit equations in psychometric data analysis, *Psychometrika*, 1975, **40**, 337-360.
- 10) 柳井晴夫・繁榊算男・前川眞一・市川雅教, 因子分析—その理論と方法—, 朝倉書店, 1990.
- 11) 菊沢正裕・山川修・田中武之, 情報リテラシー, 森北出版, 2001.
- 12) <http://www.s.fpu.ac.jp/takeyuki/irt>

項目応答理論に基づく母数推定法とテストの分析

付録 1. ロジスティック関数の数値評価

(1)式の $P_j(\theta)$ は次のような形で数値評価する.

$$P_j(\theta) = \frac{1 + c_j \exp(-Da_j(\theta - b_j))}{1 + \exp(-Da_j(\theta - b_j))}$$

$$1 - P_j(\theta) = \frac{1 - c_j}{1 + \exp(Da_j(\theta - b_j))}$$

ここで、指数関数のオーバーフローやアンダーフローと、対数関数の計算誤差に対処するため、次のような関数を導入する.

$$\text{le } x = \log(1 + e^x) \approx \begin{cases} x & (x > 32) \\ \log(1 + e^x) & (-28 \leq x \leq 32) \\ e^x & (x < -28) \end{cases}$$

le 関数の境界値(-28, 32)は倍精度実数の仮数部の桁数および対数関数の計算精度によって決まる. le 関数を用いると、(8)式の被積分関数は次のように表現される.

$$\log P_j(\theta) = \text{le}(-Da_j(\theta - b_j) + \log c_j) - \text{le}(-Da_j(\theta - b_j))$$

$$\log(1 - P_j(\theta)) = \log(1 - c_j) - \text{le}(Da_j(\theta - b_j))$$

$c_j = 0$ のときは上の2式とも右辺第1項を0とする. さらに、次のような変数変換をおこなう.

$$\begin{cases} Da_j = a_j' \\ Da_j b_j = b_j' \\ c_j = \sin^2 c_j' \end{cases}$$

上の変換で、2番目の式は、 a_j が0に近い場合に b_j の発散を防ぐ意味がある. 3番目の式は c_j の値を0から1の範囲に制限するためである. 2母数モデルでは $a_j = 0$ の場合 b_j は不定になるが、 b_j' によって $P_j(\theta)$ が定まる. 3母数モデルでは、 $a_j = 0$ の場合、 b_j' , c_j' は一意的に定まらず、ヘッセ行列の次元が退化する.

付録 2. 関数値の最大化

$f(x)$ の最大値を求めることは $-f(x)$ の最小値を求めることと同じなので、最小値について述べる. 関数 $f(x)$ が x^* の近傍で無限階連続微分可能なとき、極小点 x^* 近傍でのテーラー級数展開は

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^T B(x - x^*) + O(x^3)$$

で与えられる. ここで、 B はヘッセ行列 (ヘシアン)

$$B = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{x=x^*}$$

である. B が正定値ならば $f(x)$ は一意的な最小値を持つ. x_k を x^* の近傍の点とすると、 x^* は次の式で近似される.

$$x^* \approx x_k - B^{-1} g_k \quad \text{ただし, } g_k = \left(\frac{\partial f}{\partial x_i} \right)_{x=x_k}$$

この式を反復計算して最小値を求める方法は、準ニュートン法として知られている. しかし、 B の逆行列を直接求めることは計算量が大きく効率的でないため、次のような方法が用いられることが多い.

$$B_k p_k = -g_k$$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$B_{k+1} = B_k + E_k$$

ここで、 p_k は勾配 g_k によって定められる探索ベクトル、 α_k はステップサイズ、 E_k は近似ヘッセ行列 B_k を改良する行列である. α_k は1以下の値で B_k の精度に応じて定める. E_k の決定にはいろいろな方法がありうるが⁸⁾、反復計算が収束するためには、 $g_k^T p_k > 0$ を満たすようにすること、すなわち、 B_k の正定値性を保証することが必要である.

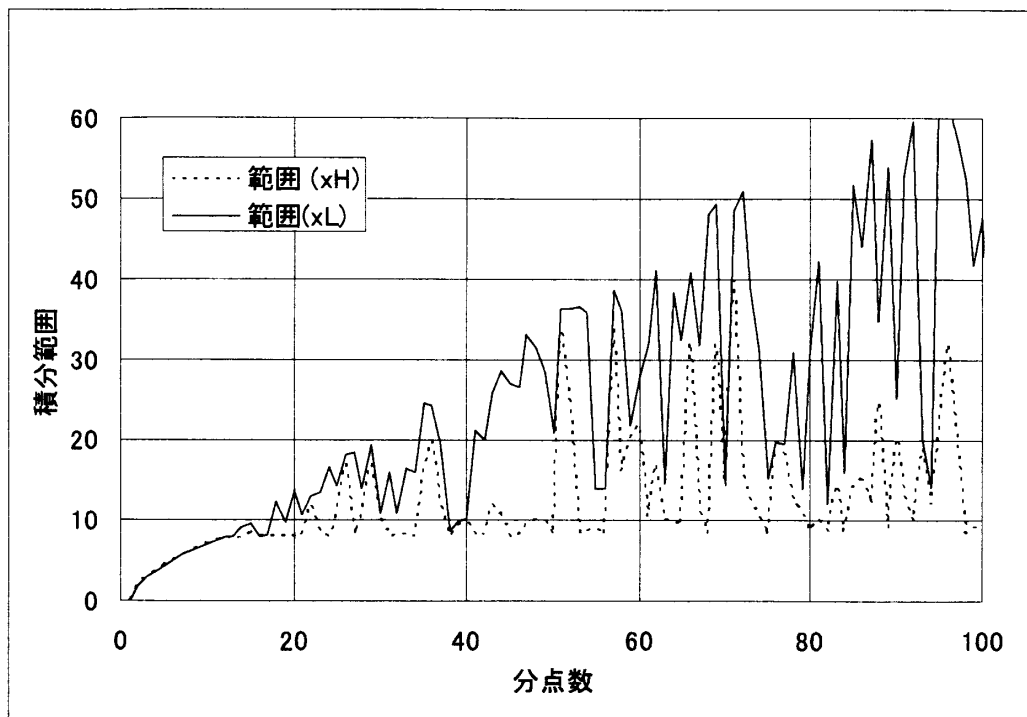
付録3. 数値積分の台形則近似に関する分点数

2.3 節で現れた積分の台形則近似に際して、以下のように積分範囲と分点数を決定した。等間隔台形則による無限区間数値積分は

$$\int_{-\infty}^{\infty} f(x)dx \doteq \sum_{k=-n}^n f\left(\frac{kr}{n}\right) \cdot \frac{r}{n}$$

によって近似される。 r と n は要求精度や計算誤差等によって決まる。今回の数値精度は倍精度実数である。被積分関数に標準正規分布を仮定すると上式の左辺の値は1になるが、これを右辺の台形則によって計算した場合の r と n の関係を図に示す。図の実線(xL)は、分点数 n を固定したとき、誤差を0にするような積分範囲 r の上限である。複数の点で誤差が0になることがあるので、同様に誤差が0になる r の下限を破線(xH)に示した。この図によると、例えば積分範囲が10のとき、必要な分点数は15以上であることがわかる。

EM法に適用する際には、それぞれの被験者特性値の事後分布 h_{ik} の計算に際して上記の値を参考に、全体の分点数と積分範囲を決める。2.3節では $n=50, r=10$ とした。



付録図 標準正規分布曲線の台形則積分における分点数と範囲の関係

項目応答理論に基づく母数推定法とテストの分析

付録 4. 情報科学試験問題

問1 「プロバイダ」の説明として適切なものをすべて選べ。

- a インターネットの通信回線を提供する業者。
- b インターネットへの接続サービスを提供する業者。
- c ビッグロブBiglobe, ニフティNifty, AOLなどの業者。
- d Internet Service Provider, すなわちISPを指す。

問2 行政(G)では、電子調達(GtoB)だけでなく、GtoG(組織内の処理)やGtoC(住民向けサービス)など電子政府にむけた動きが活発化している。この動きに関して述べた次の記述の中で適切なものをすべて選べ。

- a 人手や手間が省けるので経費が削減されるうえ、業務が迅速に処理される。
- b 住民票などを役場に取りに行くかわりに、家のコンピュータから出力できるので便利である。
- c 行政の書類をコンピュータから印鑑なしで取り出せるので個人情報を守られず危険である。
- d 電子政府が完成するまでに情報格差を解消しておくことが大切である。

問3 情報社会に関する次の記述の中で適切なものをすべて選べ。

- a 情報は秘密にしておくことがその価値を高めるとの考えを捨て、情報を開示し透明性を高めることが大切である。
- b シンガポールは多言語、多民族国家であるために英語を公用語とし異文化を受け入れやすい。このことが、同国が情報先進国となった背景にある。
- c 米国は1966年に制定された情報自由法によって開かれた政府を實現しているが、日本では最近になってようやく情報公開法が施行され、2001年から全国で開示請求が始まった。
- d 情報化を推進する行政は、光ファイバーの敷設やパソコンの導入といった基盤整備を優先させてきたが、そのことが日本を情報先進国ならしめた。

問4 WWW上で実行されるアプリケーションとして適切なものをすべて選べ。

- a ICQ b Java c ftp d shockwave

問5 公開鍵暗号に関する次の記述の中で適切なものをすべて選べ。

- a 公開鍵暗号は、暗号化する鍵と復号化する鍵に別なものを使う。
- b 公開鍵暗号は、慣用暗号と違い将来的にも破られる心配はない。
- c 公開鍵暗号を使って暗号化したメッセージを交換する場合、暗号化する鍵を公開する。
- d 公開鍵暗号を使って電子認証を実施する場合、復号化する鍵を公開する。

問6 ドメイン名に関する次の記述の中で適切なものをすべて選べ。

- a ドメイン名とIPアドレスは1対1に対応している。
- b DNSのデータベースは複数のサーバで分散管理されている。
- c TLDは必ず国名に対応している。
- d 現在、本学で使用しているドメイン名はfpu.ac.jpであるが、IPドメインのSLDに自由な文字が使えるようになったので、fpu.jpというドメイン名も登録できる。

問7 電子メールアドレスに関する次の記述の中で適切なものをすべて選べ。

- a インターネットメールアドレスには半角英数字のほか全角英数字が使えるが、半角カナ文字は使えない。
- b 宛先を間違えて送信した場合、有効なメールアドレスでなければ、そのメールは送信者に通知なく消去される。
- c 複数に同じメールを送る場合、複数アドレスを宛先欄に書く以外にCCおよびBCC欄を使うことができる。
- d 宛先を間違えて違う相手に送信してしまった場合、一定時間内なら送信を取り消すことができる。

問8 コンピュータウイルスに関する次の記述の中で適切なものをすべて選べ。

- a ウイルスはプログラムファイルに感染して自己増殖するもので、ワープロやエクセルのデータファイルには感染しない。
- b ウイルスの感染経路はフロッピーディスク等、電子メールの添付ファイル、ネットワークのセキュリティホール等であり、インターネットホー

ムページを表示しただけで感染することはない。

- c ウイルス感染を予防するには、個々のソフトウェア導入前のウイルスチェック、パソコン起動中のウイルス検査、セキュリティ修正の適用、ネットワーク上のファイアウォール設置、メールサーバ上でのウイルス除去等があげられるが、これらの対策を全て施しても今後100%安全とはかぎらない。
- d ウイルスに感染したときには、まず他のコンピュータへの感染拡大を防止するためネットワークケーブルを抜く。次にユーザファイルのバックアップをとってからワクチンソフトによってパソコン本体のウイルスを除去し、そのバックアップファイルを再びパソコンに戻すという手順で復旧する。

問9 次の記述に当てはまる言葉を語群から選び記号で答えよ。

- (ア) 不特定の人に再送信を強要するメール
 - (イ) ソフトウェアプログラムのソースコードを利用者に無償公開すること
 - (ウ) 自由に配布できるが、使用には代金が必要なソフトウェア
 - (エ) 自由に配布できて、無料で使用できるソフトウェア
- [語群] a オープンソース b フリーウェア c シェアウェア
d チェーンメール e ミドルウェア f ファームウェア

問10 インターネットの検索エンジン等で情報を得る場合の注意点として適切なものをすべて選べ。

- a その情報をインターネットで調べることが適切かどうかをチェックする。ものによっては、図書館で調べた方がいいような場合も多い。
- b Webページがきれいかどうかをチェックする。きれいに作ってある場合、手間をかけているので情報も信頼できることが多い。
- c 自分に都合が良い情報かどうかをチェックする。自分の考えにあった情報であれば、考慮に値するが、そうでないなら無視したほうが良い。
- d Webページを作った日付が新しいかどうかをチェックする。新しい情報の方が古い情報より考慮に値する場合が多い。

問11 メディアに対する基本的な視点のひとつ「メディアのメッセージは価値観とイデオロギーを伝えている」の以下の解釈のうち適切なものをすべて選べ。

- a TVドラマはストーリーだけでなく、制作者の価値観やイデオロギーを伝えている。
- b 生放送番組は、編集ができないので、価値観やイデオロギーが入り込む余地は少ない。
- c CMは商品の情報だけではなく、それを購入したときにどう生活が変わるかという価値観やイデオロギーを伝えている。
- d ニュース番組は、現実起こったことを題材にしているので、価値観やイデオロギーが入り込む余地は少ない。

問12 インターネットをメディアとしてみた時の特徴として適切なものをすべて選べ。

- a デジタル b クリティカル c ボーダレス d インタラクティブ

問13 次の中で適切なものをすべて選べ。

- a 情報の効果は計ることができ、その単位は[bps]である。
- b 情報システムは情報源、あて先、通報からなる。
- c 情報の意味や形式は、情報源またはあて先によって異なる。
- d 情報の量はシャノンによって定義され、単位は[bit]である。

問14 サイコロを振って得られる目の数(1から6)を情報源とする事象または情報量に関する次の中で正しいものをすべて選べ。ここでサイコロの目の出方に偏りはなく、底を2として $\log_2=1.58$ とする。

- a この場合、情報エントロピーは最大エントロピーより小さい。
- b 1の目が出たときに得られる情報量は2.58[bit]である。
- c 最大エントロピーは3.58[bit]である。
- d 1の目がでる確率は1/6である。

問15 コンピュータの種類と用途に関する次の中で適切なものをすべて選べ。

- a スーパーコンピュータは、他の種類のコンピュータよりあらゆる機能において上位である。
- b ワークステーションは、サーバ/クライアントシステムのサーバとして利用されることが多い。
- c 汎用コンピュータは、銀行などの大組織で汎用的に使用された大型計算機のことである。

d パーソナルコンピュータは、パソコンと呼ばれる前は、マイカーと同じようにマイコンと呼ばれた。

問16 10進数の47について次の中で正しいものを選び。

- a 2進数で表すと00111101である。
- b 2進数で表すと00111011である。
- c 2進数で表すと00110111である。
- d 2進数で表すと00101111である。

問17 ASCIIコードで a は 61(16進数)と表される。これを2進数(下位7ビット)で表したとき、次の中で正しいものを選び。

- a 10100001 b 01010001 c 11000001 d 01100001

問18 デジタル化に関する次の記述の中で適切なものをすべて選べ。

- a アナログ音声をデジタル化するとき、サンプリング間隔が小さすぎると高音を正確に表現できない。
- b MP3は、音声情報圧縮技術規格のLayer3の形式である。
- c 漢字はコンピュータの中では1バイトで表現される。
- d コンピュータの中で漢字は2進数で表現されるので大小を判別できる。

問19 記憶装置に関する次の記述の中で適切なものをすべて選べ。

- a アクセス頻度の高い情報は大容量低速記憶装置に、そうでない情報は小容量高速記憶装置に蓄えるのが、コストとのバランスを考えると効率的である。
- b フロッピーディスクはMOよりも1MB当たりの単価が安い。
- c ハードディスクへの記録はシリンダ、トラック、セクタの3つのパラメータを指定して行う。
- d フラッシュメモリは、内蔵電池が切れるとデータが消滅する。

問20 64メガビットのDRAMを使って256メガバイトのメモリカードを作るにはDRAMが何個必要か。

問21 100メガバイトの画像ファイルをFTTH(光ファイバー)を使って100Mbpsの速度でダウンロードしたとき、転送時間が何秒かかるか計算せよ。ただし、転送は常にここに示した通信速度で行えるものとする。

問22 CPUの動作を説明する以下の文章の空欄(ア)～(オ)に当てはまる語句を[語群]から選び記号で答えよ。

CPUの動作は(ア)が発生するパルスによって1ステップずつ進められる。(イ)が示すアドレスから取り出された命令は(ウ)によって解読され、演算装置などに命令の実行を指示する。演算命令の場合、オペランドで示されるアドレスの内容と、(エ)の内容を使って演算を行い、結果の数値を(エ)に格納するとともに、結果がプラス、マイナス、ゼロのいずれになったかという情報が(オ)に反映される。命令の実行が終わると(イ)の値を一つ増やし、次の命令の取り出しにかかる。分岐命令の場合には現在の(オ)の内容が分岐条件に合致したとき、(イ)にオペランドの内容をセットする。分岐条件に合致しない場合は(イ)を一つ増やす。CPUはこのようなサイクルを繰り返している。

[語群]

- a パソコン b 命令デコーダ c ディスク d メモリ
- e 状態レジスタ f インタープリタ g 命令レジスタ h クロック
- i アキュムレータ j アドレスレジスタ k プログラムカウンタ

問23 OSに関する次の記述の中で適切なものをすべて選べ。

- a Windows, Macintosh, Linuxはどれもハードウェアの名称である。
- b カーネルの主な役割目はプロセスにコンピュータ資源を配分することである。
- c OSは一度インストールを行うと、異なる種類のOSをインストールしなおすことはできない。
- d OSはパソコンの電源を入れたとき最初に起動されるプログラムである。

問24 ウィンドウズで、c:\home\panda\report.xlsと表記されたパス名について、次の中から適切なものをすべて選べ。

- a report.xlsはcドライブの中にある。
- b homeはルートディレクトリである。
- c report.xlsの".xls"の部分は拡張子といい、ファイル名の一部である。
- d homeディレクトリの直下にpanda.docというファイルを作ることはできない。

問25 入出力管理とシステム管理に関する次の記述の中で適切なものをすべて選べ。

- a ログ機能とは、利用者が日々の記録(ログ)を入力するための機能である。
- b 周辺装置を接続する場合、個々の装置に対応するデバイスドライバが必要である。
- c コンピュータの利用を終了するときにはシャットダウンを行い、さらに電源を切る際にはログオフの操作をする。
- d OSの入出力管理機能を利用することで、プログラムは必ずしも個々の装置に特有の制御をしなくてもよい。

問26 プログラミング言語に関する次の記述の中で適切なものをすべて選べ。

- a FORTRANはデータベース作成に向けた言語である。
- b LISPは関数型言語の一つである。
- c C言語は高水準言語であるが、アセンブリ言語なみの細かい処理を記述できる。
- d COBOLのプログラムは記号をあまり用いず、英語の文章によく似た形になっている。

問27 プロトコルに関する次の記述の中で適切なものをすべて選べ。

- a インターネット上で異なったコンピュータがデータをやりとりするための規則をプロトコルという。
- b IPアドレスというのはプロトコルの名前の1つである。
- c インターネットですべて使われている共通のプロトコルはTCPとIPである。
- d 家庭で使えるブロードバンド回線に、電話線(ADSL)、同軸ケーブル(CATV)、光ファイバーなどの様々なものが使えるのは、プロトコルが階層的になっているためである。

問28 コンピュータネットワークが発達してきた主な理由として適切なものをすべて選べ。

- a 電話での通信が高価だったので、電話に代わる通信手段を確保するため。
- b 計算が速いコンピュータ等の資源をネットワーク上で共同で使えるようにするため。
- c メーリングリストやネットニュース等を使ってネットワーク上の人が同じ情報を持ち、議論をしたり共同で仕事をするため。
- d 企業が宣伝広告費を安くするため、ホームページを使って企業情報を提供するため。

問29 家庭からインターネットに接続する場合の次の記述で適切なものをすべて選べ。

- a DHCP サーバを指定すれば、IPアドレス、サブネットマスク、デフォルトゲートウェイの設定は自動的に行なわれるが、DNS サーバの指定は手で設定しなければならない。
- b サブネットマスクが255.255.255.0の場合、IPアドレスのうちネットワーク部として使用する部分が、25ビットであるという意味である。
- c 家庭でインターネットを利用する場合、ホームページを閲覧することは多いが、家庭から情報を提供することは少ないので、ADSLやCATVにおいて下りくらべて上りの伝送速度が遅くてもあまり問題にならない。
- d ブロードバンドを使ってインターネットに自宅のパソコンを接続する場合、インターネット層では設定する項目は多数あるが、トランスポート層とアプリケーション層では設定する項目は無い。

問30 次の説明はTCP/IPプロトコルのどの層の働きについて述べたものか、層の名称を答えよ。

- a 物理的な機器を使って信号を送る。
- b パケットが届かなかったことを検知し、必要があれば再送する。
- c ユーザが使っているソフト同士の通信規約を決める。
- d インターネット上で通信先のコンピュータを特定する。