大学教育における高度統計処理の応用による教育情報データマイニング

Educational data mining by using the advanced statistical method in university education

神谷 達夫

要旨

試験得点の分析のために、平均点を基準とする手法は広く使用されている。平均点を基準とする評価法は、成績の評価が容易であるのみでなく、成績を評価される者にも理解が容易である。しかし、平均点を基準とした分析が有効であるのは試験得点の分布が一様分布の場合に限られる。ところが、近年の学生の試験成績には一様分布とならない場合が見られるようになった。特にユニバーサルアクセス化した大学においては、従来では想定できなかったような学力層の学生が存在するようになり、極端に広い学力分布に対応した授業方法の開発が必要となっている。

本研究では、試験得点を EM アルゴリズムによる分布推定によって分析することにより学生をグループ化し、広範な学力分布を持つ学生集団に対しての教育方法を検討するための指標を得る方法を提案する。さらに、本論文では意思決定法として使用される AHP 手法と EM アルゴリズムを共に用いることによる授業評価アンケートの結果の信頼性評価法を提案する。

キーワード: 試験得点、データマイニング、EM アルゴリズム、混合正規分布, AHP

Keywords: examination score, data mining, EM algorithm,
contaminated normal distribution, AHP, consistency index

1. はじめに

近年、学生の基礎学力の低下が指摘されている[1]。特に、ユニバーサルアクセス化した大学においては通常想定できないような学力層の学生が存在し、学生の学力は広範囲に分布するようになった。

受理：2006 年 10 月 1 日

このような状態で通常の試験により得点を評価した場合、試験得点は一様分布とならず、いくつかのグループに分かれることになる。一方、試験得点による成績評価には、偏差値のように試験得点の平均点を基準にする方法が多く用いられている[2]。平均点を基準とする成績評価法は、評価結果の算出が容易であるのみでなく、成績を評価される者にも理解が容易であるため、多用されているものと思われる。しかし、平均点を基準とした成績評価法は、試験得点が一様分布となる場合に対して有効であり、学力が複数のグループに分かれる場合の成績評価に用いることができない。

広範な学力分布を持った学生に対応するためには①学力別クラス編成とする、②特別な試験問題を作成して試験成績の分布を一様にさせる配慮をする、③一様分布にならない試験問題でも学生の学力を分析できる手法を導入するなどの方法が考えられる。本学では基礎学力向上のための特別授業が設けられており、2004 年度は学力別クラス編成を採用した。しかし、学生が自己の学力水準を明確に認識できる学力別クラス編成では、学生の学習意欲に対して負の影響を与える事例が観察された。また、多くの教科を学力別クラス編成とする場合、教員数の確保が問題となる。特別な問題を作成することによる対応法は社会科学系の科目で一定の成果を上げているが、明確に正解の決定される問題が多数を占める教科の場合、この手法の採用は容易でない。これらの問題に対応するために 、単一の授業であっても複数の学力グループに属する学生を正しく評価し、その学力の向上を測定する方法の開発が必要であると考えられる[4,5]。

本研究は最尤推定法の一種である EM アルゴリズム[3]を用いることにより、広範な学力分布を持った学生集団に対応できる学力評価法を提案する。さらにこの EM アルゴリズムと共に意思決定法の一種である AHP[6]を用い授業評価アンケートの信頼性を検証する方法も提案する。

2.EM アルゴリズムによる試験得点分析法

2.1 EM アルゴリズム

EM アルゴリズムは不完全データ問題を解くために用いられる汎用の反復法の一種である[3]。試験得点が一様分布とならない場合、受験者の学力がどの分布に属するかの情報を試験得点から直接知ることはできない。このような場合、試験得点には受験者がどの分布に含まれるかという情報が潜在データとして含まれていると解釈され、学生の試験得点分布の分析は不完全データ問題となる。本研究では、EM アルゴリズムを用いることにより、試験得点という観測データから学生の学力分布を混合正規分布に最尤推定し、得点分布の傾向を分析した。

試験得点が C 個の 1 変量正規分布の混在した混合正規分布であると仮定し、得点 $\{x_i\}\ \ i \in \{1,2,\ldots,N\}$ の分布が形成されているとする。推定すべきパラメータを θ とすると、$\{x_i\}$ の確率密度は次の式(1)となる。

$$p(x \mid \theta) = \sum_{c=1}^{C} \pi_c g_c(x \mid m_c, {\sigma_c}^2) \tag{1}$$

ただし、$c \in \{1,2,\ldots,C\}$で、$g_c(x \mid m_c, {\sigma_c}^2)$は平均$m_c$、標準偏差$\sigma_c$の正規分布を示す。ただし、$c \in \{1,2,\ldots,C\}$で、$g_c(x \mid m_c, {\sigma_c}^2)$は平均$m_c$、標準偏差$\sigma_c$の正規分布を示す。また、$\pi_c$は各正規分布が混合正規分布に含まれる割合を示す混合率で、式(2)のような条件を持っている。

$$g_c(x \mid m_c, {\sigma_c}^2) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp(-\frac{(x-m_c)^2}{2{\sigma_c}^2}) \tag{2}$$

また、π_cは各正規分布が混合正規分布に含まれる割合を示す混合率で、式(3)のような条件を持っている。

$$\sum_{c=1}^{C} \pi_c = 1 \tag{3}$$

パラメータθを含んだ$p(x \mid \theta)$の対数尤度は式(4)のようになる。

$$L(\theta) = \sum_{c=1}^{N} \log p(x_i \mid \theta) \tag{4}$$

この対数尤度を最大にするパラメータθを求めることにより、混合正規分布への分布推定が完成する。しかし、θを解析的に求めることは難しく、反復法を用いたEMアルゴリズムを適用する必要がある。

反復tサイクル目のパラメータθの近似値を$\theta^{(t)}$とすると、$L(\theta)$は式(5)のようにQ関数とH関数の差により表現できる。

$$L(\theta) = Q(\theta^{(t)}, \theta) - H(\theta^{(t)}, \theta) \tag{5}$$

一方、H関数には式(6)のような性質があるため、$L(\theta)$の極大値はQ関数によって支配され、$L(\theta)$の極大値はQ関数の極大値と一致する。

$$H(\theta^{(t)}, \theta) \le H(\theta^{(t)}, \theta^{(t)}) \tag{6}$$

Q関数は式(7)のようになる。

$$Q(\theta^{(t)}, \theta) = \sum_{i=1}^{N} \sum_{c=1}^{C} P(c \mid x_i, \theta^{(t)}) \log p(x_i, c \mid \theta) \tag{7}$$

ただし、$P(c \mid x_i, \theta^{(t)})$は事後確率で、確率密度からベイズ則より式(8)のように求められる。

$$P(c \mid x_i, \theta^{(t)}) = \frac{p(x_i, c \mid \theta^{(t)})}{p(x_i \mid \theta^{(t)})} = \frac{\pi_c{}^{(t)} g_c(x_i \mid m_c{}^{(t)}, \sigma_c{}^{2(t)})}{\sum_{c=1}^{C} \pi_c{}^{(t)} g_c(x_i \mid m_c{}^{(t)}, \sigma_c{}^{2(t)})} \tag{8}$$

また、対数確率密度 $\log p(x_i, c \mid \theta)$ は式(9)のようになる。

$$\log p(x_i, c \mid \theta) = \log \pi_c - \frac{1}{2}\log(2\pi) - \log \sigma_c - \frac{(x_i - m_c)^2}{2\sigma_c{}^2} \tag{9}$$

Q 関数に含まれる π_c には式(3)のような制約条件があるため、Q 関数の極大値を求めるために Lagrange の乗数法を用いる。Lagrange 関数を式(10)とする。

$$\tilde{Q}(\theta^{(t)}, \theta) = Q(\theta^{(t)}, \theta) - \lambda\left(\sum_{c=1}^{C} \pi_c - 1\right) \tag{10}$$

式(10)の極値条件と式(3)を連立させて解くことにより、更新パラメータ $\theta^{(t+1)}$ に関する更新式(11)(12)(13)を得ることができる[7]。

$$\pi_c{}^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N} P(c \mid x_i, \theta^{(t)}) \tag{11}$$

$$m_c{}^{(t+1)} = \frac{\sum_{i=1}^{N} P(c \mid x_i, \theta^{(t)}) \cdot x_i}{\sum_{i=1}^{N} P(c \mid x_i, \theta^{(t)})} \tag{12}$$

$$\sigma^2{}_c{}^{(t+1)} = \frac{\sum_{i=1}^{N} P(c \mid x_i, \theta^{(t)})(x_i - m_c{}^{(t+1)})^2}{\sum_{i=1}^{N} P(c \mid x_i, \theta^{(t)})} \tag{13}$$

この対数尤度を最大にするパラメータ θ を求めることにより、混合正規分布への分布推定が完成する。しかし、θ を解析的に求めることは難しいため、EM アルゴリズムを用いた数値計算によって θ を求める。

2.2 試験得点の分析

本学 1 年対象の授業である「一般数学」の小テストの得点は 100 点満点で 25 点に最大の分布を持つ結果となった(図 1)。試験問題は高等学校科目の数学 I に含まれる基本的な事項を問う問題である。高校 3 年生に同じ問題を解かせた場合、本学学生らより短時間で満点となる場合があった。

この得点に対し EM アルゴリズムを用い得点分布を混合正規分布に再尤推定した。式(1)における C の値を 4 とし、推定される混合正規分布は 4 つの正規分布を含む。その結果、平均点 19 点に最も大きな混合率 44%を持つ分布、平均点 47 点に次に大きな混合率 43%を持つ分布の存在することが分かった(表 1)。

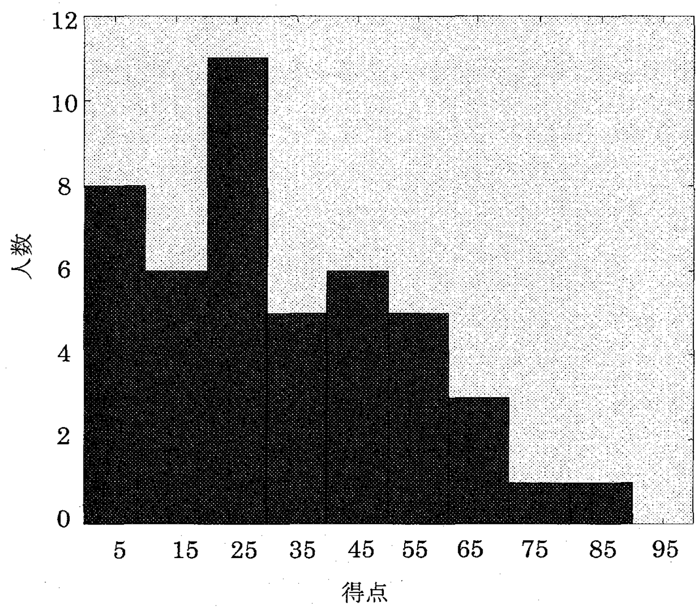


図 1　本学の得点分布

表 1　本学得点分布の推定結果

平均点 m_c	標準偏差 σ_c	混合率 π_c
1.5	1.6	8.5%
19	7.8	44%
47	12	43%
82	2.0	4.1%

EM アルゴリズムによる推定結果から、ヒストグラムから定性的に読み取れるような内容を定量的に示すことができるようになり、試験結果の分布についての定量的判断が可能になった。例えば、なんらかの理由により全く授業を理解することのできない学生を含むと考えられる最低点の分布に 8.5%程度の学生が属するということが定量的に判断できる。

全く同一の試験ではないが、他大学(理系)での同程度かやや高い難度の試験における得点のヒストグラムは正規分布に近い形状となった(図 2)。この得点を本学の試験得点の分析と同様の条件で 4 つの正規分布から成る混合正規分布に最尤推定した(表 2)。同大学は理系であり、入試のために数学を学んだ学生が多いため、本学より得点が高いものと思われる。

推定の結果、平均点 60 点に混合率 70%の分布が存在し、60 点を合格点とするならば、この試験は受験した学生に対する難易度として適当であったと思われる。この得点分布の尖度は 0.036、歪度は-0.25 であり、主となる 60 点平均の分布はほぼ正規分布に従っていると考えられる。しかし、混合率と歪度が示すように低い得点の学生のグループが存在し、この学生に対しては補習などの特別な措置が必要であると考えられる。

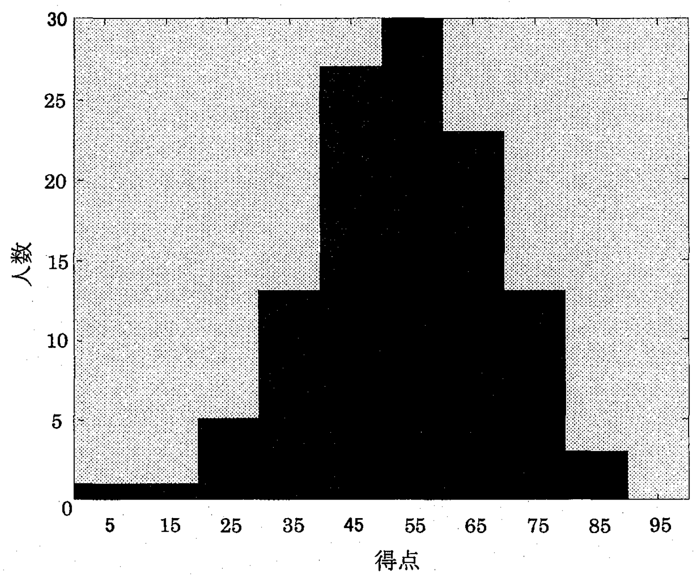


図 2 他大学の得点分布

表 2　他大学の得点分布の推定結果

平均点 m_c	標準偏差 σ_c	混合率 π_c
11	3.0	1.7%
40	6.8	26%
60	11	70%
75	1.3	2.2%

試験の難易度を過去の試験結果により経験的に調整し、目標平均点が 60 点となるように調整した試験の結果は、2 つのピークを持つ分布となった(図 3)。EM アルゴリズムによる推定の結果、63 点に平均値を持つ分布の混合率が 84%となる分布となった(表 3)。したがって、目標平均点を 60 点とする試験問題設計は概ね成功していると考えられる。この試験は①選択式の問題を含め 0 点を防ぐ、②学生に理解が容易であると思わせるような問題により学生の集中力を維持させる、③論述式の問題に組み込むべき語を指定しその語が含まれていることにより加点することなどの手法により、難易度が調整されている。難易度調整の結果、この試験問題は大学の試験として適当でない難度であると大学関係者の一部から指摘を受ける難易度となっている。しかし、このように巧みに調整された問題であっても、混合率から判断して 16%程度の学生が全く解答できていないということが分かり、この分布に属する学生の特別な指導が必要であることが定量的に判断できる。

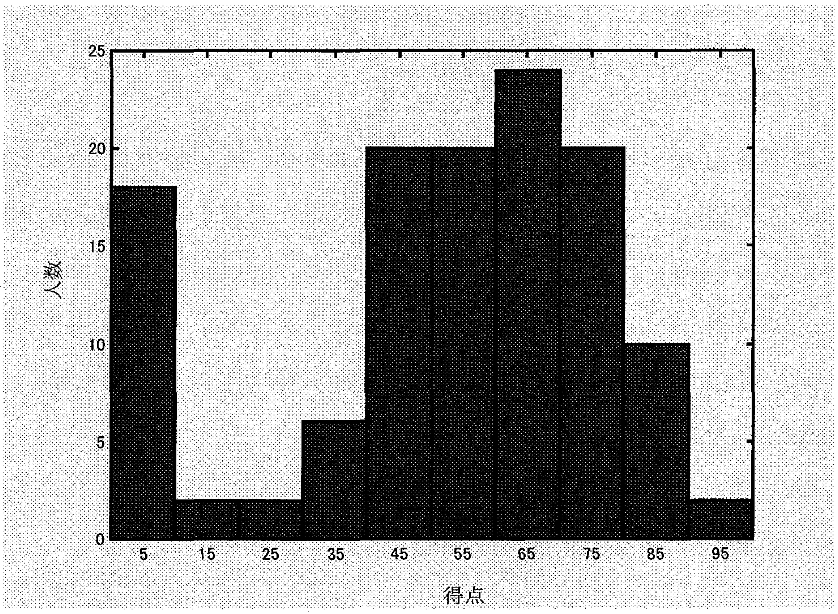


図 3 目標平均点調整した試験の推定結果

表 3　目標平均点調整した試験の推定結果

平均点 m_c	標準偏差 σ_c	混合率 π_c
3.5	8.7	16%
63	16	84%

3.授業評価アンケートの理解度評価

3.1 授業評価アンケートの理解度

学生による授業評価アンケートを実施している大学は 97%程度であるが、そのアンケート結果が十分に数値的に分析されて活用されているとは言いがたい状況である(7)。特に、アンケート結果は記入者の言語能力や精神状態に左右されるため、低学力の学生に対するアンケートの結果は、学生の理解度不足により信頼性に欠く可能性がある。本論文ではこのアンケートの理解度を評価する方法と、学生の理解度を向上させ、信頼性の高いアンケートを作成する方法を提案する。

本論文では、意思決定法として使用される AHP 手法を応用してアンケート結果の理解度を評価した。AHP 手法は一対評価により評価項目の重みを決定するが、重みの決定段階で、回答者の自己矛盾を整合度という値で評価することができる。この整合度により、回答者の状況を評価した。さらに、整合度を EM アルゴリズムにより混合正規分布に最尤推定し、分布状況の把握を可能にした。

3.2 AHP 整合度によるアンケートの理解度評価

意思決定のための手法として AHP(Analytic Hierarchy Process 階層分析法)が知られている(6)。AHP は対立する概念も取り入れることかできることや尺度の違う要素も比較することができることを特徴としており、政策の決定や企業経営上の意思決定などに広く利用されている。

AHP は評価項目の重要度（ウエイト）の「比」を使用している。このような方式は一般に比率尺度(ratio scale)による評価と呼ばれる。AHP は比例尺度による一対比較をもとに、全体としての項目間の比率尺度を決定する方法である。

いま n 個の評価項目 $I1$, …, In があり，その本未のウエイトが $w1$, …, wn であるとする。このとき、項目 Ii と Ij の重要度の一対比較値 aij は、

$$a_{ij} = \frac{w_i}{w_j} \tag{14}$$

という関係を満たす。したがって、一対比較行列 $A=[aij]$は次の式(2)となる。

$$A = \begin{pmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdots & \frac{w_2}{w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & \frac{w_n}{w_n} \end{pmatrix} \tag{15}$$

このAの右側から，ウエイトの列ベグトルを乗じる。式(16)から、ウェイト・ベクトルはAの固有ベクトルであり、λ は固有値であることが分かる。そして、行列Aの最大固有値であるλを求めるとウエイトを算出することができる。

$$\begin{pmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdots & \frac{w_2}{w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & \frac{w_n}{w_n} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \lambda \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \tag{16}$$

現実の一対比較行列Aは(15)の形をしていることは期待できなが、近似的に式(15)が成立していると考えられるならば、Aの最大固有値と固有ベクトルを求めることにより、その固有ベクトルを各評価項目のウェイトとして採用できる。現実の一対比較行列Aの最大固有値を$\lambda_{\max}$、固有ベグトルをvとする。

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \tag{17}$$

このとき固有値と固有ベクトルの関係より式(5)が成立する。

$$Av = \lambda_{\max} v \tag{18}$$

一方、一般に式(6)の成立することが知られている。

$$\lambda_{\max} \geq n \tag{19}$$

行列Aにはn個の固有値が存在し、その和はnとなる。したがって、$\lambda_{\max}$以外の固有値の合計は式(7)となる。

$$\lambda_{\max} - n \tag{20}$$

$\lambda_{\max}$以外の固有値の平均は式(18)の固有値の合計を個数 *n-1* で割った値となり、この値を整合度(consistency index:C.I.)と呼ぶ。

$$\mathrm{C.I.} = \frac{\lambda_{\max} - n}{n-1} \tag{21}$$

式(21)で計算されるC.I.の値は経験的に0.1（場合によっては0.15）以下であればAHPで求めら

れた結果が正当であったと判断される。

3.3 EM アルゴリズムによる整合度分布分析

整合度の分布を前章で述べた試験得点の分析と同様の方法により分析した。整合度が C 個の 1 変量正規分布の混在した混合正規分布であると仮定し、整合度 $\{x_i\}$　$i \in \{1,2,\ldots,N\}$ の分布が形成されているとし、式(1)からの計算法を試験得点の場合と同様に適用する。

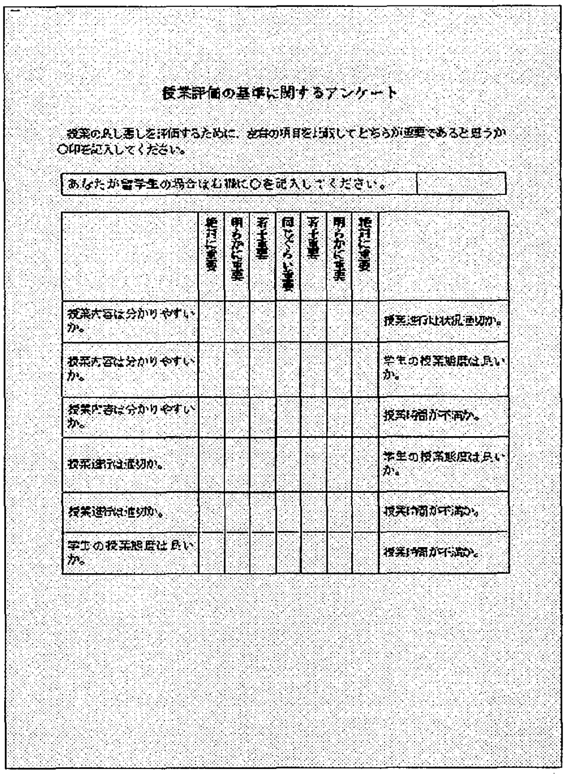
3.4 整合度の分布分析結果

本学の授業において、図 4,5 に示す用紙を用いてアンケート項目の理解度調査を試みた。図 4 に対するアンケート結果を実験 A、図 5 に対する結果を実験 B とする。

回答数は実験 A,B 共に 66 で、有効回答数は実験 A4 が 62 で実験 B にが 62 であった。3 名は一対比較の意味を理解できず、複数に〇を記入していた。実験 A において 1 名は不注意により解答欄が充足していなかった。実験 A では 62 名中 17 名が全て同じぐらい重要に〇を記入し、整合度が 0 となっている(図 6)。有効回答の者で整合度 0 の者と整合度が 04.を超える者を除いた整合度の分布を EM アルゴリズムにより 3 つの正規分布に最尤推定し、混合正規分布モデルを得た(表 4)。全て同一の回答した者 17 名　25.8%と整合度が 0.4 以上の者 5 名 7.6%、無効回答の 4 名 6.06%と推定結果で平均整合度 0.151 に含まれる 28.3%を合計すると、実験 A では 68%程度の学生が十分にアンケート項目を理解していないか適切に回答しておらず、アンケート結果に信頼性を欠く可能性があると判断できる。

一方、実験 B では 63 名中 8 名が全て同じぐらい重要に〇を記入し、整合度が 0 となっている(図 7)。有効回答の者で整合度 0 の者と整合度が 04.を超える者を除いた整合度の分布を EM アルゴリズムにより 3 つの正規分布に最尤推定し、混合正規分布モデルを得た(表 5)。全て同一の回答した者 8 名　12.1%と整合度が 0.4 以上の者 2 名 3.03%、無効回答の 3 名 4.54%と推定結果で平均整合度 0.183 に含まれる 27.3%、平均整合度 0.359 に含まれる 4.79%を合計すると、実験 B では約 52%先の実験とう同様の理由により、学生のアンケート結果に信頼性を欠く可能性がある。しかし、実験 B のアンケートの理解度は実験 A より 16%程度改善されていると考えられる。

ただし、これらの考察では、混合率がその分布に含まれる人数構成比率とほぼ等しいという近似を用いている。厳密には各正規分布に含まれる割合は確率密度関数の計算により求める必要がある。



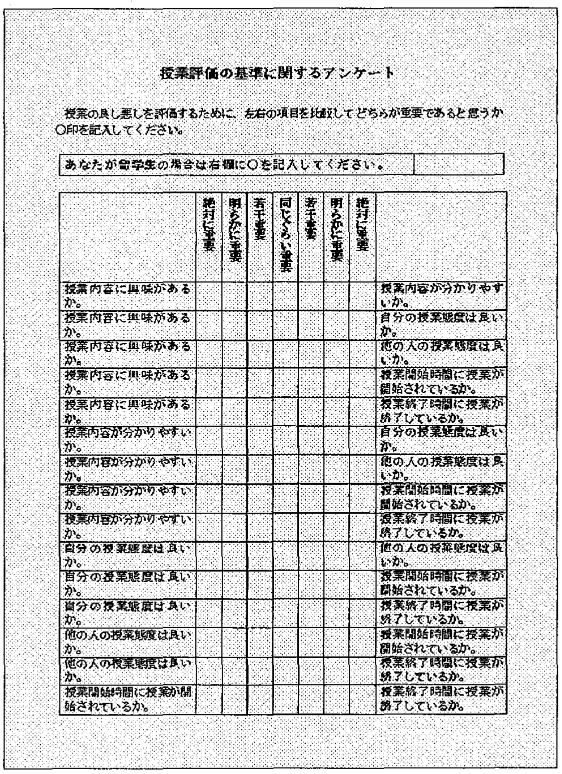
授業評価の基準に関するアンケート

授業の良し悪しを評価するために、左右の項目を比較してどちらが重要であると思うか○印を記入してください。

あなたが留学生の場合は右欄に○を記入してください。	

	絶対に重要	明らかに重要	若干重要	同じくらい重要	若干重要	明らかに重要	絶対に重要	
授業内容は分かりやすいか。								授業進行は適切か。
授業内容は分かりやすいか。								学生の授業態度は良いか。
授業内容は分かりやすいか。								授業時間が不満か。
授業進行は適切か。								学生の授業態度は良いか。
授業進行は適切か。								授業時間が不満か。
学生の授業態度は良いか。								授業時間が不満か。

図 4 実験 A のアンケート用紙



授業評価の基準に関するアンケート

授業の良し悪しを評価するために、左右の項目を比較してどちらが重要であると思うか○印を記入してください。

あなたが留学生の場合は右欄に○を記入してください。	

	絶対に重要	明らかに重要	若干重要	同じくらい重要	若干重要	明らかに重要	絶対に重要	
授業内容に興味があるか。								授業内容が分かりやすいか。
授業内容に興味があるか。								自分の授業態度は良いか。
授業内容に興味があるか。								他の人の授業態度は良いか。
授業内容に興味があるか。								授業開始時間に授業が開始されているか。
授業内容に興味があるか。								授業終了時間に授業が終了しているか。
授業内容が分かりやすいか。								自分の授業態度は良いか。
授業内容が分かりやすいか。								他の人の授業態度は良いか。
授業内容が分かりやすいか。								授業開始時間に授業が開始されているか。
授業内容が分かりやすいか。								授業終了時間に授業が終了しているか。
自分の授業態度は良いか。								他の人の授業態度は良いか。
自分の授業態度は良いか。								授業開始時間に授業が開始されているか。
自分の授業態度は良いか。								授業終了時間に授業が終了しているか。
他の人の授業態度は良いか。								授業開始時間に授業が開始されているか。
他の人の授業態度は良いか。								授業終了時間に授業が終了しているか。
授業開始時間に授業が開始されているか。								授業終了時間に授業が終了しているか。

図 5 実験 B のアンケート用紙

この2つの結果のようにアンケートの項目を変更することにより、アンケート項目の理解を高め、アンケート結果の理解度性を高めることができると思われる。実際のアンケート項目の作成には、アンケート項目を変更して整合度を求める予備実験を繰り返し、十分低い整合度を得ることのできるアンケート項目を決定する手法が有効であると考えられる。

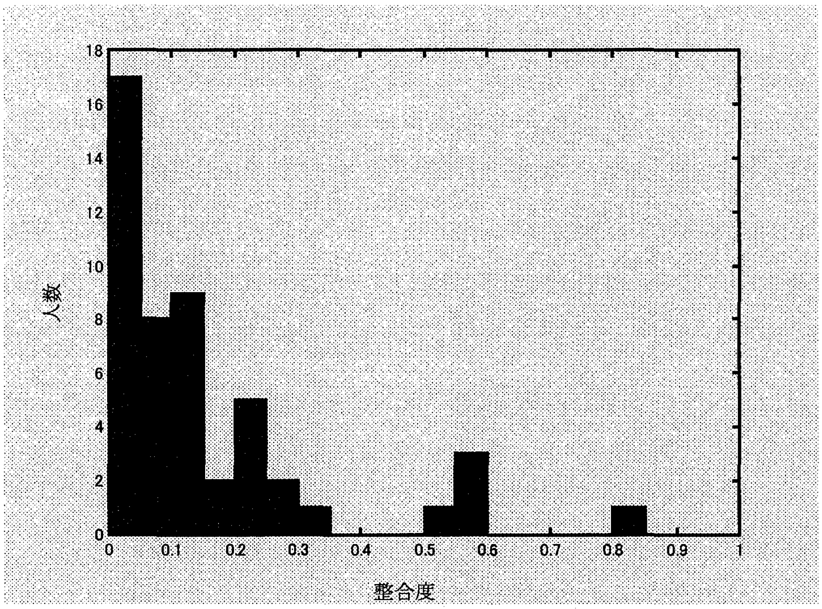


図6 実験Aの整合度分布

表4 実験Aの推定結果

平均値 m_c	標準偏差 σ_c	混合率 π_c
0.0397	0.0163	49.4%
0.151	0.0683	22.3%
0.186	0.283	28.3%

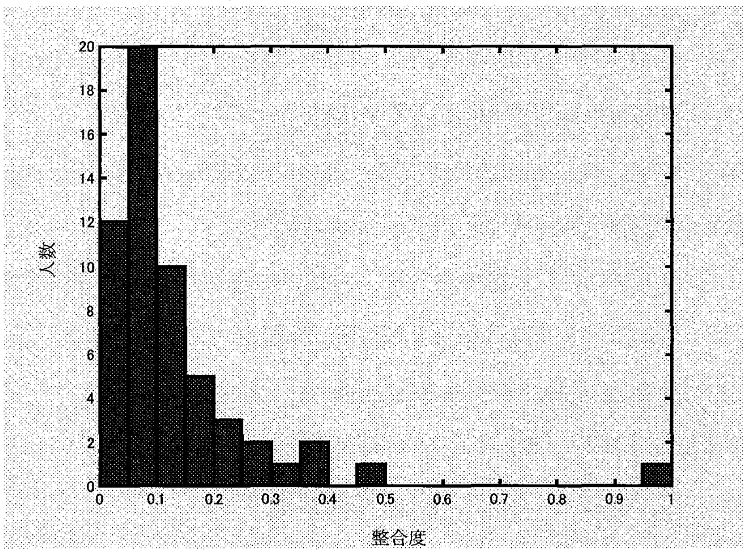


図 7 実験 B の整合度分布

表 5　実験 B の推定結果

平均値 m_c	標準偏差 σ_c	混合率 π_c
0.0615	0.0366	67.9%
0.183	0.0617	27.3%
0.359	0.280	4.79%

4. まとめ

本論文では、試験得点を EM アルゴリズムによる分布推定によって分析すし学生の学力分布をグループ化する手法を提案した。さらに、意思決定法として使用される AHP 手法と EM アルゴリズムを用いることにより授業評価アンケートの結果の信頼性を評価する手法を提案した。

EM アルゴリズムによる試験点数分析の手法を実際の試験点数の分析に用いた結果、EM アルゴリズムを用いて試験得点を混合正規分布に推定する手法は、大きな学力差により試験得点が一様分布にならない場合の分析手段として有効であることが確認された。また、模擬的に作成した授業評価アンケートを用いて実験した結果、AHP と EM アルゴリズムを用いた分析により、アンケート項目を変化させることによる、学生のアンケートに対する理解度の変化を定量的に示すことができた。したがって、本論文で提案したアンケートの理解度評価手法は、学生に理解容易なアンケート項目を作成する手段として有効であると考えられる。

≪参考文献≫

(1) 大谷晃也, 文科系学生の数学の基礎学力と退学率、就職率, 関西外国語大学研究論集, No.82, pp. 191-197 (2005)

(2) 池田央, テストの科学, 日本文化科学社, pp.122-147 (1992)

(3) Dempster A.P., Laird N.M. and Rubin D.B., Maximum likelihood from incomplete data via the EM algorithm, J. Royal Statist Soc, Ser B Vol. 39, pp.1-38 (1977)

(4) 神谷達夫, EM アルゴリズムを用いた試験得点分析, 平成 18 年度大学教育・情報戦略全国大会, pp.- (2006)

(5) 神谷達夫, EM アルゴリズムによる試験得点分布の分析, 教育システム情報学会, pp.- (2006)

(6) Saaty, T.L., The Analytic Hierarchy Process, McGraw-Hill, (1980).

(7) 中野良平, ニューラル情報処理の基礎数理, pp.179-209 (2005)

(8) 学生による授業評価　全大学の 97%が導入, 朝日新聞, 2006 年 6 月 6 日