等価選択肢数を用いた教養共通テスト分析の一例

上　原　宏　行

Abstract

情報理論で扱う情報エントロピーの応用例である等価選択肢数を用いて、教養共通テストの分析を行った。その結果、等価選択肢数－正答率図に等価選択肢数最大曲線と等価選択肢数最小曲線を加えることで、誤答の選択割合などが一目で確認できることが分かった。また、テスト全体の問題の難易度分布も容易に確認できる。これらの考え方を示すとともに、この方法の有用性について議論する。問題数が集まれば、問題再利用時にグループ毎に問題を選ぶことで、出題を最適化することができる。

キーワード：等価選択肢数、多肢選択問題、情報エントロピー

1．はじめに

愛知学院大学では、2023年度に文部科学大臣が認定する「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）」に申請[1]するため、2022年度から当プログラムが動き出した。教養部が担当する学部に対しては、教養教育科目である「情報科学Ⅰ・Ⅱ」と「教養セミナーⅢ・Ⅳ」1)の4科目でプログラム内容をカバーするように担当範囲が振り分けられ、筆者も情報科学Ⅰ・Ⅱを担当している。

プログラムでカバーする内容やレベルなどの確認を行う手段としてBS放送の放送大学を活用することがある。そこには、当プログラムに該当する「数理・データサイエンス・AIリテラシー講座」や「データサイエンスの技術」など[2]の関連講義が放送されており、その中に「情報理論とデジタル表現」という講義があった。その第6回の講義で情報理論におけるエン

トロピーの応用例として等価選択肢数[3] が紹介されており、興味深い指標であると感じた。

通常、データのばらつき具合を表現する指標として、分散や標準偏差を使用することがある。これらは気温（℃）のような間隔尺度や質量のような比例尺度に対して用いることは可能であるが、高校生、大学生、社会人のような名義尺度や成績評価である AA、A、B、C のような順序尺度に対して用いることはできない。しかし、等価選択肢数は名義尺度や順序尺度において、データのばらつき具合を出現頻度に関する指標として求めることができる[3]。

そこで、愛知学院大学キャリアセンターと教養部が共同で実施している教養共通テスト2) を題材として、その適用から得られる情報について調べた。このような多肢選択問題のテストがどの程度適切なレベル・問題難易度なのか、正答率だけ見ても分かりにくいので、この判断材料としての有用性についても検討した。

先行研究として、等価選択肢数を U-L 指数（上位下位項目弁別指数）など他の指数と比較した論文[4] はあるが、ここで行おうとする多肢選択問題の問題分析に関する論文は筆者が調べた限りでは見いだせていない。この論文では、多肢選択問題が抱えている問題点などについて議論するのではなく、等価選択肢数を用いることで、多肢選択問題の問題難易度や選択肢の特徴など、どのような情報が得られるのかを議論する。

２．等価選択肢数

まず、この論文で中心的な役割を担う等価選択肢数について説明する。エントロピーは統計力学において定義されているが、情報理論では式（1）により情報エントロピーHを定義している。ここでは、n 個の選択肢から 1 つ解答を選ぶ場合、i 番目の選択肢が選ばれる確率を p_i としている。

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

この情報エントロピーHを用いて、等価選択肢数 O を 2^H と定義している。なお、確率については

$$\sum_{i=1}^{n} p_i = 1$$

を満たす。

この関数 H は 2 択の場合、

$$H = -p_1 \log_2 p_1 - (1 - p_1) \log_2 (1 - p_1)$$

と表され、p_1=1/2のとき、H=1となり、p_1=0と1のとき、H=0となる[3)]。このように不確実性が最大のp_1=1/2で最大となっており、統計力学におけるエントロピーの特徴と同じである。

次に、等価選択肢数のイメージを明らかにするため、いくつかの計算例を示す。

（ア）　選択肢数は4で、1つの選択肢だけが選ばれる（p_1=1、p_2=p_3=p_4=0）場合[3)]

$$H = -1 \times \log_2 1 = 0$$

$$O = 2^0 = 1$$

等価選択肢数は1となる。

（イ）　選択肢数は4で、2つの選択肢が等確率で選ばれ、他の2つの選択肢は選択されない（p_1=p_2=1/2、p_3=p_4=0）場合[3)]

$$H = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2} = 1$$

$$O = 2^1 = 2$$

等価選択肢数は2となる。

（ウ）　選択肢数は4で、3つの選択肢が等確率で選ばれ、残り1つの選択肢は選択されない（p_1=p_2=p_3=1/3、p_4=0）場合[3)]

$$H = -\frac{1}{3}\log_2 \frac{1}{3} - \frac{1}{3}\log_2 \frac{1}{3} - \frac{1}{3}\log_2 \frac{1}{3} = \log_2 3$$

$$O = 2^{\log_2 3} = 3$$

等価選択肢数は3となる。

（エ）　選択肢数は4で、全ての選択肢が等確率で選択される（p_1=p_2=p_3=p_4=1/4）場合

$$H = -\frac{1}{4}\log_2 \frac{1}{4} - \frac{1}{4}\log_2 \frac{1}{4} - \frac{1}{4}\log_2 \frac{1}{4} - \frac{1}{4}\log_2 \frac{1}{4} = 2$$

$$O = 2^2 = 4$$

等価選択肢数は4となる。

（オ）　選択肢数は4で、各選択肢の選ばれる確率がp_1=1/2、p_2=p_3=1/4、p_4=0の場合[3)]

$$H = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{4}\log_2 \frac{1}{4} - \frac{1}{4}\log_2 \frac{1}{4} = \frac{3}{2}$$

$$O = 2^{\frac{3}{2}} = 2\sqrt{2} \approx 2.83$$

等価選択肢数は約2.83となる。

このように、(ア)～(エ）では、等価選択肢数は実際に選択された選択肢数と一致している。

また、実際に選択された選択肢数が等しい場合でも、選択確率に偏りのある（オ）は選択確率の等しい（ウ）より等価選択肢数は小さくなる。このように、「選択肢の等価な数」を表していることが分かる。ただし、この指標は、正答率とは直接の関係がないことに注意しなければならない。

３．正答率に対して取り得る等価選択肢数の範囲

容易に分かるのは等価選択肢数が１の場合と４の場合である。等価選択肢数が１の場合、単一の選択肢が選ばれたことを意味している。そして、選択された選択肢が正答の場合と誤答の場合の２通りがあり、それぞれの正答率は１と０である。一方、等価選択肢数が４の場合、（エ）の場合に対応しており、どの選択肢が正答でも正答率は0.25である。しかし、これ以外の場合、複雑になるので次の例を考える。

（カ）　選択肢数は４で、そのうち３つの選択肢は等確率で選ばれ、残り１つの選択肢もわずかな確率で選ばれる（p_1=p_2=p_3=0.33、p_4=0.01）場合、

$$H = -0.33\log_2 0.33 - 0.33\log_2 0.33 - 0.33\log_2 0.33 - 0.01\log_2 0.01 \approx 1.65$$

となり、等価選択肢数は$O \approx 2^{1.65} \approx 3.14$となる。この場合、$p_4$が正答であるという稀なケースもありうる。

放送大学「情報理論とデジタル表現」では、選択肢の数はn個の場合で説明されており、等価選択肢数が１の場合と、nの場合の正答率を直線で結んだ線の間が、データが取り得る範囲の目安としている。したがって、それをもとにデータが存在する目安となる領域は等価選択肢数１と４での正答率を直線で結んだ２直線の間の領域（図１）とすると、大体のイメージができる。ただし、（カ）の場合で、p_4が正答である場合には、この領域から外れることに注意が必要である。

等価選択肢数は「何個の選択肢の中から解答が選ばれたのか」を示す指標であるので、問題難易度の目安となる指標と考えることもできる。図１の（i）付近にデータがある場合、問題は容易であることを意味している。等価選択肢数が１に近く、正答率が１に近いため、ほとんどの解答が正答であることを意味している。一方、図１の（ii）付近にデータがある場合、問題が難しすぎることを意味している。問題が分からず、４つの選択肢から、「サイコロを振るよう」にして解答した場合に例えることができる。（i）と（ii）の間の領域では、等価選択肢数は２～３であり、選択肢がある程度機能しており、正答率も適度に高い領域となるので、問題の難易度は中程度とみなせる。解答に迷うと正答率が下がり、等価選択肢数は増えることが考えられる。多数の問題からなるテストでは、等価選択肢数－正答率図にデータをプロットす

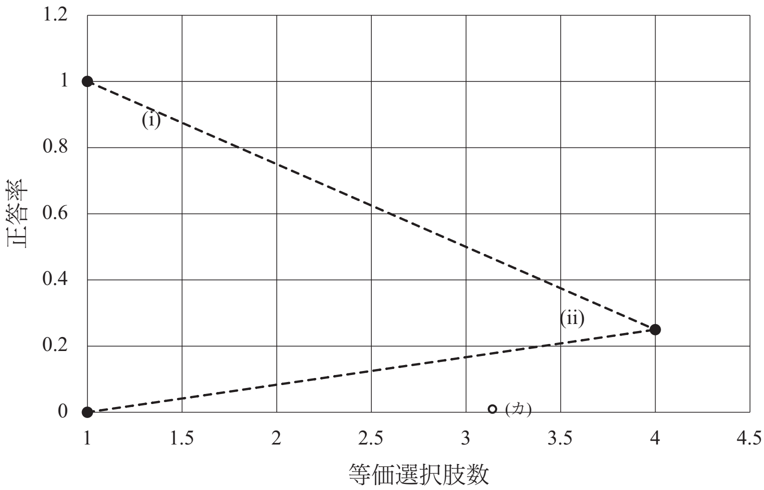

図 1　等価選択肢数に対して取り得る正答率の範囲の目安

ることで、テストの難易度を概観することができる。

図 1 では目安を示したが、ここではより詳細な情報を読み取るため、選択肢数が 4 の場合について取り得る等価選択肢数の上限について検討する。数学的な証明は行わないが、この条件を満たすのは（ア）～（オ）からも推測できるように、正答以外の選択肢が等確率で選ばれる場合である。正答率 p_1 に対する等価選択肢数の最大値を $O_{max}(p_1)$ とすると、

$$O_{max}(p_1) = 2^{-\left[p_1 \log_2 p_1 + (1-p_1)\log_2\left(\frac{1-p_1}{3}\right)\right]} \tag{2}$$

と表せる。この式で表現される曲線を等価選択肢数最大曲線と呼ぶことにする。また、等価選択肢数が最小になるのは、 2 つの選択肢のみが選ばれる場合であり、正答率 p_1 に対するその値を $O_{min}(p_1)$ とすると、

$$O_{min}(p_1) = 2^{-[p_1 \log_2 p_1 + (1-p_1)\log_2(1-p_1)]} \tag{3}$$

と表せる。この式で表現される曲線を等価選択肢数最小曲線と呼ぶことにする。これら式(2)、(3) を図 2 に示す。図 2 には上で求めた（ア）～（カ）の結果も合わせて示している。これらの結果は曲線上または 2 つの曲線間の領域に存在していることが確認できる。

実際のデータが等価選択肢数最大曲線に近い場合、誤答はほぼ等確率に選ばれており、データが等価選択肢数最小曲線に近い場合、誤答として確率の高い選択肢が存在することを意味している。

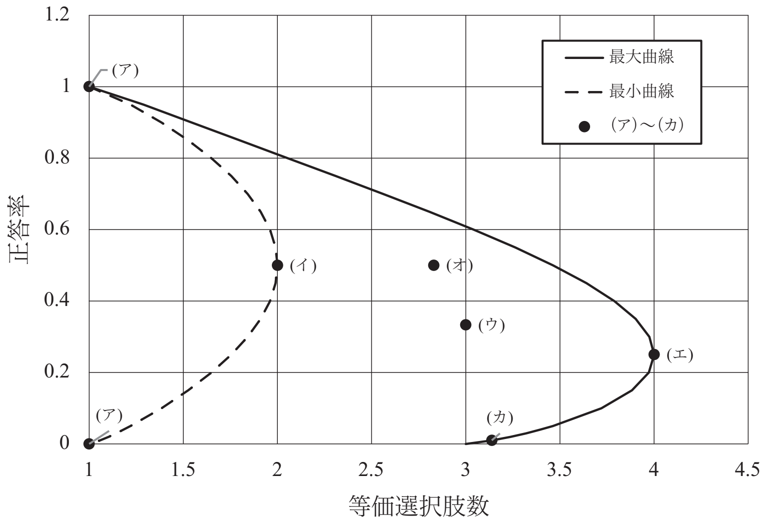

図2　選択肢数が4の場合の等価選択肢数最大曲線と等価選択肢数最小曲線。
図中の●は上で求めた（ア）～（カ）の結果を示す。

4．第1回教養共通テスト

愛知学院大学キャリアセンターと教養部が共同で実施している教養共通テストは、このテストをきっかけに教養教育科目の履修、就職活動に何が必要かを考える機会としての活用［資料1］が期待されており、教養教育科目と数的推理の分野から出題される4択問題である。この論文では令和3年に実施された第1回の結果を使用した。表1の問題番号はオリジナルの番号を踏襲しており、選択確率に直し、情報エントロピーHと等価選択肢数Oを求めた。これらをもとに等価選択肢数－正答率図に等価選択肢数最大曲線と等価選択肢数最小曲線を合わせて図3に示す。各問題ともに最も高い選択確率の選択肢が正答であり、多くは等価選択肢数最大曲線の内側近傍に分布していることが分かる。また、等価選択肢数は1.37から3.75と広範囲に広がっていることから、様々なレベルの問題が混在していることが分かる。ただし、ヒストグラム（図4）で等価選択肢数の出現頻度を調べると、等価選択肢数が3.5以上の問題が多く存在していることも分かる。この点に問題があるかどうかは出題意図により解釈が異なるので、この点について言及はしないことにする。等価選択肢数最大曲線に内接している問題は誤答がほぼ等確率になっているケースであり、この最大曲線から最小曲線側に大きく左側に移動している問題は、比較的高い選択確率の誤答が存在していることを意味している。例えば、問題10では正答率0.492に対して、誤答の中に選択確率0.447が存在している。残り2つの選択確率

表1　教養共通テストの問題ごとの正答率、選択確率と等価選択肢数

		選択確率					
問題番号	正答率	a	b	c	d	H	等価選択肢数
3	0.842	0.136	0.007	0.015	0.842	0.741	1.67
4	0.452	0.198	0.452	0.221	0.129	1.843	3.59
5	0.339	0.337	0.203	0.339	0.122	1.895	3.72
6	0.685	0.018	0.685	0.275	0.022	1.114	2.16
7	0.369	0.334	0.188	0.369	0.108	1.860	3.63
8	0.435	0.218	0.132	0.435	0.215	1.864	3.64
9	0.659	0.659	0.196	0.053	0.092	1.398	2.64
10	0.492	0.042	0.019	0.492	0.447	1.325	2.50
11	0.404	0.178	0.243	0.175	0.404	1.907	3.75
12	0.778	0.055	0.071	0.096	0.778	1.108	2.16
13	0.888	0.888	0.078	0.025	0.009	0.634	1.55
14	0.538	0.193	0.095	0.173	0.538	1.701	3.25
15	0.932	0.012	0.025	0.031	0.932	0.459	1.37
16	0.699	0.089	0.699	0.179	0.034	1.280	2.43
17	0.565	0.040	0.086	0.310	0.565	1.478	2.79
18	0.523	0.079	0.523	0.241	0.158	1.694	3.23
19	0.677	0.677	0.141	0.120	0.063	1.397	2.63

が小さいことから、（イ）に近い状態になっている。また、問題12では正答率0.778に対して、誤答の選択確率は0.055、0.071、0.096と均等に近い分布をしているので、等価選択肢数最大曲線に内接している。このように等価選択肢数最大曲線と等価選択肢数最小曲線を基準にすると、誤答の特徴についても一目で理解することができる。また、図示することで全体のバランスも一目で分かるなど利用価値は高い。当然ではあるが、正答率を全て加えることで、期待値である正答数平均が10.3問であることも分かる。つまり、問題再利用時にこの期待値を基準にすれば、前回の受験者との得点差について議論することも可能である。

資料1　教養部長から学生に対して送信された案内

学生の皆さんへ

教養部長　佐々木真

教養共通テストの案内

　教養部とキャリアセンターが共同して教養共通テストを作成しました。学生の皆さんが愛知学院で何を学んでほしいか、なぜ学ぶのかを記しています。本テストの結果は各科目の成績評価とは関係ありませんが、これが皆さんの学びのきっかけとなってさまざまな教養科目を履修したり、就職活動に際して、何が必要になるかを考える機会としてもらえればと思います。このテストは春と秋に問題を変えて実施します。皆さんの学びの軌跡として活用してください。

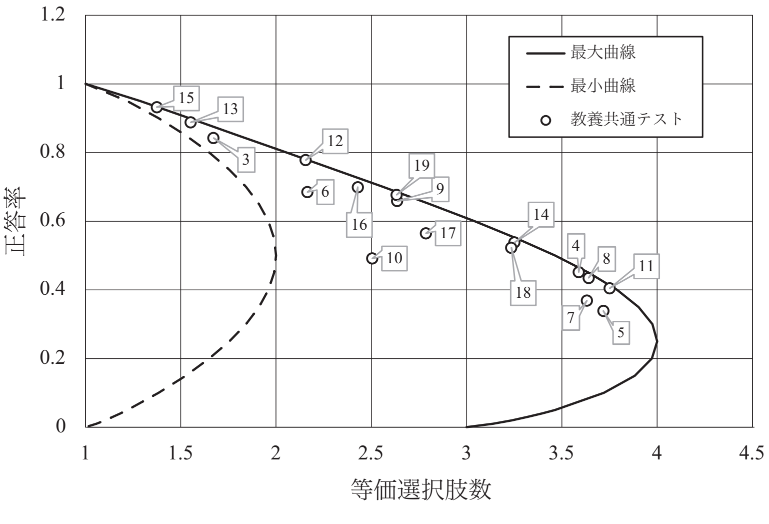

図3　第1回教養共通テストの結果。番号は問題番号である。

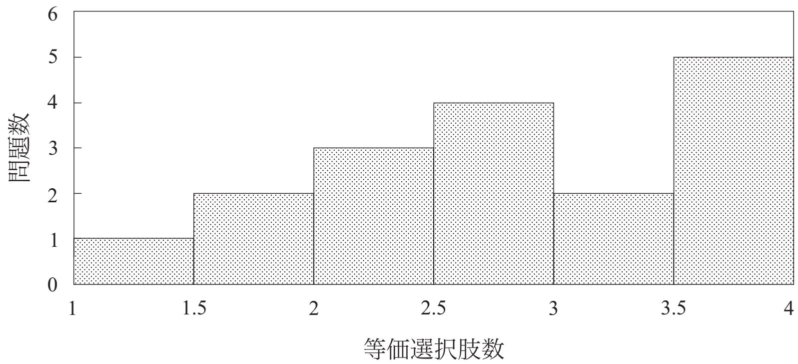

図４　第１回教養共通テストにおける等価選択肢数と問題数の関係

５．まとめ

等価選択肢数－正答率図に等価選択肢数最大曲線と等価選択肢数最小曲線を合わせることで、問題の難易度以外にも誤答の特徴などが容易に理解できることが分かった。最大曲線近傍にあるデータの場合、誤答はいずれもほぼ等確率で選ばれている。それに対して、最小曲線側に大きく左側に移動しているデータは誤答の中に比較的高い確率の選択肢が含まれていることを示している。このように正答率以外に誤答に関する情報も含んでいるので、テスト全体の特徴を理解するのに有用である。また、問題を再利用する場合には、分野・難易度ごとにグループ分けができれば、その中から問題を選ぶことで目的に応じた出題も可能となる。

注

１）2023年度からは「教養セミナーⅢ・Ⅳ」を「情報科学Ⅲ・Ⅳ」に変更して開講する予定である。これにより「情報科学Ⅰ・Ⅱ・Ⅲ・Ⅳ」で当プログラムの内容をカバーする。

２）データの２次利用に関しては実施主体である愛知学院大学キャリアセンター部長と教養部長の許可を得ている。

３）対数関数の定義域には０は含まれていないので正確には不定となるが、極限をとって H=0としている。

参考文献

[1]　愛知学院大学．“データサイエンス教育プログラム”．
https://www.agu.ac.jp/life/data_science/　（参照2023-01-16）.

[2]　放送大学．“放送大学 データサイエンス”．　https://mds.ouj.ac.jp/　（参照2023-01-16）.

[3]　加藤浩，浅井紀久夫．放送大学教材「情報理論とデジタル表現」．放送大学教育振興会，2019.

[4]　齋藤昇．数学教育における創造性に関する態度尺度の開発．全国数学教育学会誌 数学教育学研究．1999, vol. 5, pp. 35–46.