多肢選択問題に当て推量で解答した場合の正答確率分布についての理論的基礎

大竹　洋平*

概　要

本稿は，基礎知識の確認等の教育評価で利用される「多肢選択式問題」についての理論的考察であり，確率論的な解析を中心とする。授業を担当し様々な評価方法を併用する中で，多肢選択式試験に対して抱いた問題意識について述べる。多肢選択式試験については，定量的な取り扱いが可能で，テスト理論研究の蓄積がある。多肢選択式試験の理論と実際の適用，さらに当て推量による解答についての先行研究をまとめる。確率論的な解析は，多肢選択式試験問題に対して，受検者がランダムに回答を選択する場合の，正答数の確率分布を求める。選択肢数が全ての問題項目で同じ場合には「二項分布」に，選択肢数が問題項目ごとに異なっている場合には「複合二項分布」になることを，具体例も用いながら示す。それら確率論的な考察を通して理論研究の困難性を示すことで，今後の研究の可能性についても議論する。実践的な示唆として，教育評価における多肢選択式問題の可能性について議論する。特に，多肢選択式試験をカスタマイズすることによる試験への解答を通した学習と，論述形式なども含め様々な評価方法を併用することの重要性について論じる。

キーワード：多肢選択式，当て推量，ランダム解答，二項分布，複合二項分布

1.　はじめに

1.1　本稿の構成

本稿は，基礎知識の確認等の教育評価で利用される「多肢選択式問題」についての考察である。最初に1章では，江戸川大学で講師を担当している授業の評価方法の概要と，多肢選択式試験に対して抱いた問題意識について述べる。次に2章では，多肢選択式試験に関する先行研究をまとめる。テスト理論における用語の整理を行ったうえで，多肢選択式試験の適用と当て推量についての研究をまとめ，さらに，項目反応理論の当て推量パラメータについて論ずる。続く3章，4章，5章が，多肢選択式試験問題に対して，受検者がランダムに回答を選択する場合に，正答数がどのような確率分布になるかという理論解析である。3章では，選択肢数が全ての問題項目で同じ場合（二項分布）を，4章では，選択肢数が問題項目ごとに異なっている場合（複合二項分布）を扱う。5章では，この確率論的な考察の困難性と，確率分布の近似についての残された問題点について議論する。最後に6章で，教育評価における多肢選択式問題の可能性について議論する。

1.2　アカスキについてのイントロダクション

本学における「アカデミック・スキル演習（通称「アカスキ」）」は，2017年度より全学で共通して実施されており，大学入学後の二年間を同じ教員が通しで担当する必修科目である。本学の実践は，様々な観点で新規性の高いものである。詳しい理念や実践の詳細は荒谷ほか（2018a, 2018b）などを参照してもらうとして，多くの大学で初年次教育科目が導入されてきている中で

2018年11月30日受付
* 江戸川大学 基礎・教養教育センター非常勤講師　総合教育科学・情報学

も，二年間を通じて行うという点だけをみても，「初年次教育」という枠を超えているし，日本語分野と情報分野を融合させているという意味でも，また，多様な背景をもった教員の連携という意味でも，他に類をみない。

具体的な授業の評価（西岡ほか，2015; Suskie, 2009 齋藤訳 2015）においても，共通の尺度が求められることにより，ルーブリックによる評価（Stevens and Levi, 2013 井上ほか訳 2014）を行っている。ルーブリックの各項目については，学期の当初に受講学生にも示している。ルーブリックの元となる授業内資料としても，幾つかの形式があるので，以下に列挙する。

♣授業内小テスト
標準的にイメージされる授業内テストである。解答の正誤がほぼ一意に定まる，多肢選択式，穴埋め式，短文解答式などと，ある程度の幅をもって正答を認める記述式などがある。

♣スキルテスト
タイピングテスト，情報系ソフトの操作などのパフォーマンスを測るテストを行う。

♣レポート等
論文や書籍の内容を要約するレポートや，意見文の論述など成果物による評価である。授業内のその場で書いて提出するものや，宿題として課して後日提出するものなどがある。

♣平常点
単に出席点ではなく，授業への参加度・貢献度，他者との協同学習活動の様子などを総合的にみている。

各課題について，全教員で共通の教材を用いるとともに，各教員が独自の裁量で用意する教材等もある。特に，全学で共通して行う部分は，共通問題への解答による到達度評価に頼る部分もあり，実施方法も紙であるものとCBT（Computer Based Test）で行うものとがある。そこでは，多肢選択式試験も併用されている。

本稿では，多肢選択式テストによる評価の方法を理論的に解析することを通して，学習の成果や到達度が測定できるかどうかを考察する。特に，多肢選択式テストにランダムに解答した場合に，どのような正答数が得られるかという確率計算できるような理論的基礎を与える。また，多肢選択式テストを補完するうえでの，（選択式ではない）記述式の穴埋め式テストの役割についても検討する。

1.3　アカスキの中での問題意識

授業中に行う小テストでは，多肢選択式問題への解答を学生に求める場面も多い。学生の授業に向かう態度には，色々なタイプがおり，

◇授業に向けて準備（予習）をしてきたり，授業で学んだことを復習してきたりする学生
◇授業時間中はきちんと参加するが，授業外では，ほとんど勉学を行わない学生
◇授業を聞いていない学生（欠席・遅刻者，居眠り等も含む）

など，それぞれである。授業に向かう態度と，一対一に対応するわけではないが，ほぼ同様の分類として，小テストへのぞむ態度としても，

♡授業で学んだことを復習したうえで，小テストにのぞむ学生
♡復習等はせず，授業で学んだことだけを頼りに，小テストにのぞむ学生
♡授業を聞いていないままで，小テストにのぞむ学生

などを見て取ることができる。我々教員は，小テストの受験態度を観察することはできるが，評価として考慮するのは小テストの解答用紙に書いてある内容である。そこには，答案用紙に表現された，正答，誤答，無回答しかみることができない。（誤答のパターンは，幾つかありうるので，そこから学生の学習状況についてのさらなる情報を得ることができるよう，今後検討していきたい。）我々教員が本当にみたいのは，授業で学んだことを活かすことができているか否かである。小テストへの取り組み状況と，テスト項目への反応パターンを組みあわせて，この論点を整理すると，表1のようになる。

本稿では，アカデミック・スキル演習の授業で多用している「多肢選択式試験」で，学生が授業

表 1　小テストへの取り組み態度と、解答の正誤との対応関係

学生	小テストへの取り組み態度		
無職		考えて解答している	当て推量による解答
テスト項目に	正答	達成	偶然
	誤答	未達成	偶然
	無解答	混乱・未達成	無気力

で学んできたかどうかを，教員が測れているかを検討する。特に，当て推量で解答し偶然に正答している学生と，学んできたうえで解答し正答に達している学生とを区別できることを目指し，まずは，当て推量で回答した場合にはどのくらいの正答率になるか，そして，（小テストは複数の問題で構成されるため,）小テスト冊子全体の正答数は，どのくらいになるかという確率分布を計算することにする。

2.　多肢選択式試験に関する先行研究

大学入試センター試験に代表されるように，マークシートなどによる多肢選択式のテストを考えると，わからない問題があったときに，ランダムに回答を選択する（鉛筆を転がして適当にマークするなど）ことも考えられる。多肢選択式であれば，ランダムに回答を選択したとしても，一定の確率で正解する可能性があるからである。本稿のテーマは，このランダムに回答を選択する場合についての確率論的な考察である。

2.1　用語の定義およびテストの形式について

まずは，用語の定義を与えておく。（1 章「はじめに」では，イントロダクションであるので，日常用語も用いていたが，ここからは明確に学術用語を使い分けていく。）基本的に，日本テスト学会が発行している『テストスタンダード』（日本テスト学会 , 2007）に準拠した表記とする。例えば,「試験の問題に解答する受験者」という日常用語としては自然な表現も，テスト理論における正式な表現としては,「テストの（質問）項目に回答する受検者」のような表記になる。特に，問題は項目（item）と表現されることに注意を要する。

前章で述べた授業内テストについて，試験の解答形式を分類・列挙してみよう。

池田（1992）『テストの科学』は,客観テストを,以下の五つの形式に分類している（池田 , 1992：76)。

1. 真偽形式（○×形式）
2. 多枝選択式（択一式）
3. 組み合わせ式（はめこみ式）
4. 並べ換え式
5. 完成式（穴埋め式）

同書では,「純粋な意味で代表的な客観テストといえば，組み合わせ式か択一式テストがその双璧（池田 , 1992：75)」と述べられている。ちなみに,「完成式問題を，空白に埋めるべき語句をあらかじめ提示しておき，その中から選ばせるようにすると，それは組み合わせ式（はめこみ式）と同じことになります（池田 , 1992：77)」と述べているが，この点は，学習効果の点で違ってくると考えるので，のちに 6 章の考察で議論する。

また，肥田野（1972）『心理学研究法 7：テスト 1』の 2 章には，項目の形式として，以下のような分類がある（肥田野 , 1972：35-44)。

1. 論文形式（essay form）
2. 短答形式（short-answer form）
3. 真偽形式（true-false form）
4. 組合せ形式（matching form）
5. 配列形式（arragement form）
6. 多肢選択形式（multiple-choice form）

客観テストとはいえない論文形式（1.）についての記載が加わっているだけで，本質的に同じ分類をしている。（また，並べ替え式を配列形式と，完成式を短答形式と言い換えており，順序が変っているだけで，同じものである。）

「多枝選択式」とは,『テスト・スタンダード』の記述では,「あらかじめ選択枝として複数用意

された質問に対する答のリストの中から適切なものを選ぶという形の質問形式。選択式のひとつ。(日本テスト学会 , 2007： 218)」とある。同じシリーズの，池田（1972）『心理学研究法 8：テスト 2』の 2 章で，著者である渡辺は，多肢選択形式の欠点として，以下のように述べている。

> 多肢選択形式の項目を作成するとき，もっともむずかしいのが，このまよわしの作成である。まよわしの作成の出来いかんが，その項目の良否・有効性を決めるといっても過言ではない（池田 , 1972：43）

さらに，多肢選択形式の項目作成上の注意点も幾つかあげている。同書では，一般的注意として，

- 項目はできるだけ単純明快に表現する
- 正答が一義にきまるのに必要な条件をおとさない
- 不必要な語句の導入は避ける
- 末梢的で特殊な内容を項目としない
- 項目に解答するのに必要な能力水準を，テスト目的やテストしようとする集団の能力水準にあうようにする
- 正答をするのに不適切な手がかりがあってはならない

をあげており（池田, 1972： 44-45）,また，多肢選択形式の注意として，

- 幹には選択肢に共通する語句を含める
- 幹をできるだけ否定文にしない
- 正答は，その分野の専門家も正答と判断できる選択肢とする
- まよわしは幹に対するもっともらしい答となるように作る
- 「以上のうちいずれでもない」という選択肢は，correct answer を含む項目のみに用いる
- 選択肢をランダムに並べること

などをあげている （池田, 1972： 46-50）。

2.2　多肢選択式テストについて

次に，多肢選択式テストおよびランダム解答についての先行研究をまとめる。

まず，多肢選択式テストの作成について，作成の専門家に対する調査研究をまとめたものに，荒井（2015）がある。

多肢選択式問題への批判は，高校の現場教員からのコメントとしてあげられることが多い。能力をマークシートで判別することに対しての，批判的な意見ということである。例えば，国語の入試における歴史研究をまとめた, 石川巧 (2008)『「国語」入試の近現代史』には，現代文の試験問題について幾つかの指摘がある（石川 , 2008： 200-201)。「正誤判断に偶然性が入り込むからよくない」,「解答を導くまでの過程を評価することができない以上，受験生の総合的能力を正しく評価することができない」,「パズル的思考方法をするようになる」,「そもそも想像力や文章表現力を問うことができない」などというものである。同書では,「教育を語る言説では，マークシート方式が，いかに『論理的思考力や理解力，豊かな情感を養い，人間形成に寄与』しないかという批判だけが常套化していく（石川 , 2008：201)」と，マークシート批判が的を射ていないことを指摘している。

池田（1985）も,「その問題を正しく理解していなくても，当て推量で正解を得る受験生が多いのではないかという懸念が絶えず指摘されている(池田 , 1985: 3)」ことを指摘している。しかし，同論文（池田 , 1985）では,「実際問題として，すべての問題を当て推量で正解する可能性は小さい」としつつも,「無視できるほど当て推量の要素が小さいわけでもない」としており，当て推量の影響の範囲を見積もっている。「与えられた選択肢数および問題数のもとで，当て推量の範囲がどの程度に散らばるか」および「見掛け上の正解数から予想できる真の正解数の範囲がどの程度であるか」を計算し，表にしている（池田 , 1985)。

ここで，用語・概念についても整理しておく必要がある。当て推量による解答の定義を，寺井ら（1961a, 1961b）は,「問題が要求する正当な思考その他の内的過程を経ないで解答を選択する一切の過程」としている。また，当て推量（ゲス , guess）による解答を分類し，

1. 全くでたらめ（ランダム解答）
2. 部分的なでたらめ

(a) 正当な理由＋でたらめ
(b) 誤った理由＋でたらめ
(c) 上記二つの混ざったもの
3. 誤った推論による解答

のように整理している。

すなわち，ランダム解答とは，「全くでたらめに選択肢を選ぶこと」を意味する。先に触れた，池田（1985）で与えられている計算表は，「ランダム・ゲッシング」すなわち「ランダム解答」による確率を計算したものである。

ランダム解答以外にも，ゲス（当て推量）による解答過程は存在する。例えば，櫻井ら（2003）が検証したように，選択肢すべてからランダムに解答を選択するのではなく，正解の可能性の高い選択肢を絞り込むという操作をしたうえで，そこから確率的に選択肢を選び出して解答していることなどである。櫻井ら（2003）では，「正答率を上げたランダム・ゲッシング解答」と名づけている。それは，上記の寺井ら（1961a, 1961b）の分類でいえば，2.（a）にあたるということになる。他にも，大学入試センターによる出題形式についての研究（鈴木ら , 1991）等も参照。

本稿の理論解析では，1．狭義の「ランダム解答」を取り扱う。当て推量による解答の一部である，狭義の「ランダム解答」，すなわち，全ての選択肢の中から全くでたらめに選択肢を選ぶこと，である。

また，古くから，みかけの正解率と，真の正解率との修正公式も与えられている（池田（1972）『心理学研究法 8：テスト 2』，Linn（1989 池田ら訳 1992）『教育測定学（上・下）』などを参照）。

実際の受検者の解答傾向を分析したものも幾つかある。荒井（2005b）では，全問が 5 択の多肢選択式試験である，ある種の公務員試験において，「選択肢 1 と 5 は選ばれにくい」すなわち，最も端にある選択肢は選ばれにくいことを指摘している。また，荒井（2005a）では，実際のデータを用いた分析が行われ，「ゲスによる正当の影響を小さくするためには正答率の高い問題を出せばよい（荒井 , 2005a： 25)」としている。また，「ゲスによる正答を減らすためには，誤答の選択肢がすべて有効であるようにすると良い（荒井 , 2005a：26)」としている。

2.3　IRT での当て推量について

テスト理論で当て推量といえば，項目反応理論（Item Response Theory; IRT と略記）の，当て推量パラメータが連想されるかもしれない。

IRT の詳細については，芝（1991）『項目反応理論：基礎と応用』，池田（1994）『現代テスト理論』，村木（2011）『項目反応理論』，Linn（1989 池田ら訳 1992）『教育測定学（上・下）』，豊田（2012）『項目反応理論［入門編］（第 2 版）』，光永（2017）『テストは何を測るのか ： 項目反応理論の考え方』等を参照してもらうとして，概略を示すと，

♠ 1 パラメータ（1PL）モデルでは，項目の困難度パラメータのみを導入
♠ 2 パラメータ（2PL）モデルでは，1PL モデルに加えて，項目の識別力パラメータを導入
♠ 3 パラメータ（3PL）モデルでは，さらに 2PL モデルに加えて，項目の当て推量パラメータを導入

するものである。(ちなみに，当て推量パラメータ（guessing parameter）は，疑似チャンスレベル（pseudo-chance level）とも言われる。そうした方が，意味が分かりやすいかもしれない。)

しかし，この 3PL モデルで考えている当て推量パラメータは，理論的にも実用上もあまり優れたものとは言えない。なぜなら，項目に当て推量で応答することを，全ての受検者に対して仮定しており，非現実的であるからである。

3PL モデルに対しての否定的な見解として，村木（2011）『項目反応理論』の p.49 は，g_j を当て推量パラメータとして，「安定した g_j パラメータの推定値は多数の被験者を必要とする」とし，「不安定な 3PL モデルのパラメータの最終推定値を用いるより，この母数を含まない 2PL モデルがある程度項目反応データに適合するなら，3PL モデルを使う必要はない」と論じている。さらに，「すべての被験者にとって同じ推量パラメータが

適用されるというのも，3PL モデルはモデル構築の面で改善の余地がある」とも述べている。

他にも，池田（1994）『現代テスト理論』の p.67によると，c を当て推量パラメータとして,「経験的には選択肢数分の１よりやや低い値に近づくことが多い」ことと,「必ずしも単調に漸近線に接近するのではなく，一次 c の値よりもさらに低い値に下がって再び，上がってくる」ことも指摘している。その現象は,「その付近の θ の人はまったく当てずっぽうに答えているわけではなく，誤った選択肢を正解と考えて選択している人が多いためと考えられる」と述べている。しかし，3PL モデルでは，そういった詳細にまでふみこんだモデル化をしていないのである。

3　全問が同一の選択肢数の場合

本稿では，適当に（ランダムに）回答を選択した場合に，受験者の得点（期待得点やその確率分布）がどのようになるかを考察する。特に，選択肢数が問題項目ごとに異なる場合について議論していくが，まずは，全ての問題項目が同一の選択肢数である場合には，非常に単純なモデルとなるので，以下で，簡単に示しておこう。

3.1　全て 4 択の場合の例示

例えば，表 2 のように，各問の選択肢数が全て 4である項目 12 問からなるテスト冊子（小テスト）を考える。（この例示は，具体的なイメージをもってもらうためのもので，わかっている方は，次の節へと進んでいただいて構わない。）

表 2　全て 4 択である問題項目が 12 問ある小テストの具体例 1

問題番号	選択肢数	正解
問 1	4	1
問 2	4	4
問 3	4	4
問 4	4	3
問 5	4	3
問 6	4	3
問 7	4	2
問 8	4	2
問 9	4	1
問 10	4	3
問 11	4	4
問 12	4	2

この例では，選択肢数 M=4 であるから，すなわち，正答確率 p=0.25 である。また，問題項目数 N=12 であるから，期待正答数は，Np=3.0 となる。といっても必ず正答数３となるとは限らず，偶然に多く正答できてしまう場合も，少なく正答してしまう場合もありうる。

わかりやすい例として，全て同じ数字を選んだ場合の正答数を表 3 に示す。

表 3　具体例 1 の小テストに全て同じ解答をした場合の正答数

選択肢	正答数
1	2
2	3
3	4
4	3

これは，全て同じ選択肢を選ぶ 4 人（それぞれ，１～４と固定）という標本を選んだことになる。この場合の標本平均は 3.0，標本分散は 0.5 となる。（ちなみに，不偏分散は 2/3 ≒ 0.67 となる。）

ここまで，具体的なイメージを持ってもらうために，このような例示をしてきた。上記の例では，学生が全て同じ選択肢を選ぶとしたが，一般には，全ての問題でランダムに解答を選ぶとして考える方が自然である。また，正解選択肢を具体的に示したが，正解が何番であろうがそれには関係なくなる。

少しだけ一般化すると，正答数の分布は，平均 Np=3.0，分散 $Np(1-p)$ =2.25 の二項分布に従うことになる。

この具体的な確率分布を，図 1 に示す。

3.2　各項目の正答確率分布

では，選択肢数が固定であるという範囲内で，

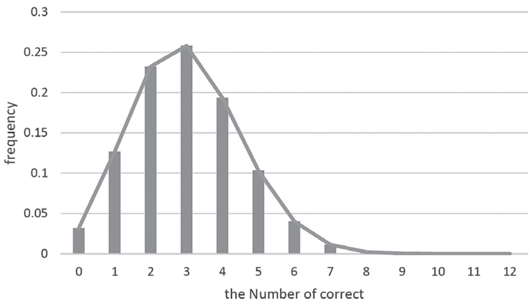

図 1　具体例 1 の正答数確率分布（二項分布 Bin(12,0.25)）

一般の選択肢数の場合に，どうなるかを示そう。選択肢数が M の問題項目にランダムに回答した場合の正答率は，その逆数をとったものとなる。問題項目ごとの正答確率 p は，

$$p=\frac{1}{M} \tag{1}$$

となる。ちなみに，誤答確率 q は，$q=1-p=1-\frac{1}{M}=\frac{M-1}{M}$ である。各問題項目に回答し，正解する確率を考えるという行為は，ベルヌーイ試行を行うことに対応付けられるから，一問の正答確率の分布は，ベルヌーイ分布 Be(p) に従う。すなわち，テスト得点を x(x=0,1) としたときの確率分布関数（probability distribution function）$f(x)$ は，

$$f(x)=p^x q^{1-x} \tag{2}$$

である。ベルヌーイ分布の確率母関数(probability generating function) $\pi(t)$ は，

$$\pi(t)=pt+q \tag{3}$$

であり，モーメント母関数（moment distribution function）$M(t)$ は，

$$M(t)=pe^t+q \tag{4}$$

となる。ベルヌーイ分布の期待値は p であり，分散は pq である。

3.3　同一選択肢数のテスト冊子全体の正答数確率分布

テスト冊子に含まれる問題項目数が N 問あったときの，正答数を考える。成功確率 p のベルヌーイ試行を N 回繰り返すことに対応するから，正答数の確率分布は，二項分布 Bin(N,p) に従う。すなわち，確率分布関数 $f(x)$ は，

$$f(x)={}_N C_x p^x q^{N-x} \tag{5}$$

である。二項分布の確率母関数 $\pi(t)$ を考えると，

$$\pi(t)=(pt+q)^N \tag{6}$$

であり，モーメント母関数 M(t) は，

$$M(t)=(pe^t+q)^N \tag{7}$$

である。二項分布の期待値は Np であり，分散は Npq である。

4.　選択肢数が問題項目ごとに異なる場合

それでは，本稿の中心テーマである，選択肢数が問題項目ごとに異なる場合について考察しよう。

4.1　異なる選択肢数のテスト冊子の例示

異なる選択肢数で構成される一般の場合の，具体的な例を示そう。(のちに示すように，正答数の確率分布が複合二項分布になる。) 表 4 に示すように，選択肢数がそれぞれ異なる一般の場合で，19 問の問題項目から構成されるテスト冊子である。

表 4　選択肢数が一般の場合 19 問あるテスト冊子の具体例 2

問題番号	選択肢数
問 1	6
問 2	6
問 3	8
問 4	8
問 5	6
問 6	6
問 7	5
問 8	6
問 9	5
問 10	6
問 11	4
問 12	4
問 13	7
問 14	6

問題番号	選択肢数
問 15	7
問 16	4
問 17	4
問 18	6
問 19	4

4.2　各項目の正答確率分布：一般の場合

各項目の正答確率分布については，3.2 節と同様であるが，それぞれの問題項目を区別することが必要となる。

N 問中 i 番目の問題項目について考えると，$i=1,\cdots,N$ において，選択肢数は M_i，ランダムに回答した場合の正答確率 p_i は式（1）に対応して，

$$p_i = \frac{1}{M_i} \tag{8}$$

となる。誤答確率は$q_i=1-p_i=1-1/M_i=(M_i-1)/M_i$である。各項目の正答確率は，ベルヌーイ分布Be($p$) に従うから，確率分布関数$f(x)$ は式（2）に対応して，

$$f(x) = p_i^x q_i^{1-x} \tag{9}$$

であり，確率母関数 $\pi(t)$ は式（3）に対応して，

$$\pi(t) = p_i t + q_i \tag{10}$$

であり，モーメント母関数$M(t)$ は式（4）に対応して，

$$M(t) = p_i e^t + q_i \tag{11}$$

となる。

4.3　異なる選択肢数のテスト冊子全体の正答数確率分布

それでは，異なる選択肢数の問題項目が，N 問あったときの正答数を考える。ただし，同じ選択肢数の項目があってもよい。全ての問題項目の選択肢数が一致するのではない場合，すなわち，選択肢数の異なる項目が含まれる一般の場合を考える，ということである。この場合は，成功確率 p_i が異なベルヌーイ試行を N 回行うことに対応するから，正答数の確率分布は，二項分布 Bin（N,p）にはならず，これは，複合二項分布（compound binomial distribution）と呼ばれる（複合）確率分布となる。

ここで，二項分布の拡張として，多項分布や負の二項分布なども思い浮かぶかもしれないが，それらと複合二項分布とは異なるものであるので，簡単に注記しておこう。

• 多項分布との相違：

二項分布の拡張としては，多項分布（multinomial distribution）が思い浮かぶかもしれないが，それとは異なる。二項分布が，例えば成功・失敗のように，カテゴリーが二個の場合にその事象が得られる回数の確率を表すのに対し，多項分布は，カテゴリーが一般に複数ある場合に，あるカテゴリーの事象が得られる回数の確率を表す。

• 負の二項分布との相違：

さらに言えば，負の二項分布（negative binomial distribution）も，また違った概念である。二項分布が，確率 p のベルヌーイ試行において，n 回繰り返して成功する回数の確率を表すのに対し，負の二項分布は，M 回成功するまでに要した回数の確率を表す。

これら確率分布に関しての詳細は，東京大学教養学部統計学教室（1992）『自然科学の統計学』，蓑谷（2010）『統計分布ハンドブック』，芝（1984）『統計用語辞典』などを参照。

複合二項分布についての記載がある文献は，芝（1984）『統計用語辞典』，芝（1991）『項目反応理論：基礎と応用』，池田（1994）『現代テスト理論』など，ごく一部に限られている。それは，確率分布の複合（mixture of probatility distributions）という概念自体が発展的な話題であることにもよる。芝（1984）『統計用語辞典』には，「複合確率分布」として解説がある。また，具体例として考えられる対象が，教育測定関係に多いということにもよる。他には，経営工学関連でも記載がある。

この複合二項分布については，確率分布関数 $f(x)$ を陽に表現するのは困難である。すなわち，

二項分布と対比すると，確率分布関数 $f(x)$ を式(5)のように陽に書くことができないので，確率母関数 $\pi(t)$ による表示を行う。確率分布関数 $f(x)$ を求めるために，確率母関数 $\pi(t)$ から考えていく，ということである。複合二項分布の確率母関数 $\pi(t)$ は，式(6)に対応して，

$$\pi(t)=\prod_{i=1}^{N}(p_i t+q_i) \tag{12}$$

と書ける。よって，この確率母関数を構成したときの，t^x の項の係数が $f(x)$ となる。すなわち，

$$\pi(t)=\sum_{x=0}^{N} f(x)t^x \tag{13}$$

と書けるわけである。複合二項分布のモーメント母関数 $M(t)$ を考えると，式(7)に対応して，

$$M(t)=\prod_{i=1}^{N}(p_i e^t+q_i) \tag{14}$$

である。

複合二項分布の期待値は $\sum_{i=1}^{N} p_i$ であり，分散は $\sum_{i=1}^{N} p_i q_i$ である。

4.4　具体例 2 のグラフ表示

複合二項分布の具体形として，先に例示した具体例 2 の問題冊子に，受検者がランダム解答を行った場合の確率分布を図 2 に示す。

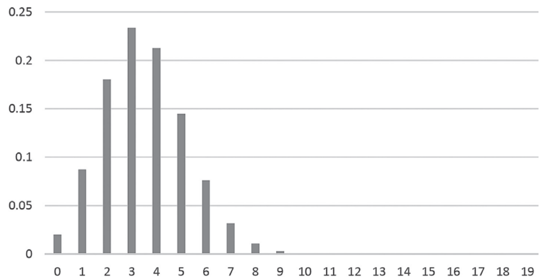

図 2　具体例 2 の正答数確率分布（複合二項分布）

5.　理論研究に向けた視座

本章では，今後の研究に向けて，これまで得られたことを踏まえて，いくつかの論点を整理する。

5.1　複合二項分布を扱うことの困難性

まずは，この複合二項分布を扱うことの困難性について論じる。列挙すると，

- 解析的に表現することの困難性と，
- 具体的にそれぞれの確率を計算することの困難性

が存在する。

複合二項分布は，母関数によって表現されているため，乱数を生成させてその確率分布を求めるなど，確率分布を数値計算によって表示することはできたとしても，解析的に厳密に取り扱う方法を検討した方が望ましいが，それには困難がともなうのである。

さらに，確率分布関数を陽に表現できないために，具体的にデータ解析プログラム（例えば，R 言語）や表計算ソフト（例えば，Excel）で計算する方法も検討していく必要がある。

5.2　二項分布と複合二項分布の差異

実務では，複合二項分布の具体的な計算方法が知られていなかったため，二項分布のパラメータを調整し，複合二項分布に代用することがあった。そこで，二項分布と複合二項分布の差異についても考察する必要がある。また，あまり知られていない複合二項分布を，多用されている二項分布によって近似する方法などについても，検討する必要がある。

6.　実際の授業における多肢選択試験についての考察

ここまで多肢選択式テストに，ランダムに解答した場合に，その正答数の確率分布が，どのようになるかを検討してきた。授業へと応用していく上では，実際の学生による解答の正答数分布と，ランダム解答による確率分布とに差異があるかを検討していく必要がある。ここでは，正答数の分布がどのようになりうるか，という概念図を示しておく。

図 3 は，正答数（成績）とその人数（確率分布）を模式的に示したものである。

全員がランダムに解答した場合には，二項分布のような分布となるため，破線（random）のような曲線となる。一方，教育が成功したならば，

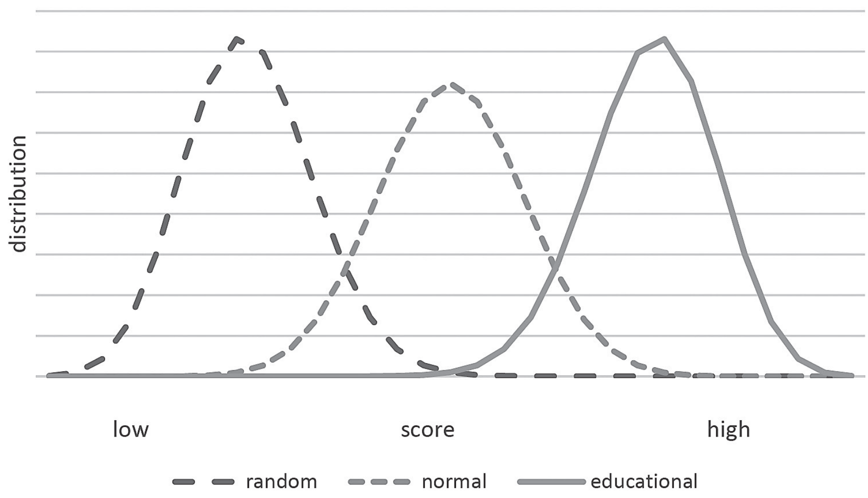

図 3　テストスコア分布の概念図

全ての受検者が一定以上の学力水準に達し，クラス全体の成績が高くなる実線（educational）のような曲線（教育曲線）となることが想定される。（参考までに，わかっているかわかっていないかもランダムに決まるような場合に，正規分布のような曲線となることが見込まれるので，それを点線（normal）で示しておく。）

実際の得点分布の詳細は，今後，検証していくべき課題であるが，最後に本稿のまとめとして，学習効果を考えた上での，テスト利用について検討したい。中島（2018）『学習評価』の第 7 章にもあるように，筆記試験で学習を促す，という視点である。ここでは，特に，多肢選択式テストを補完するうえでの，「選択的」穴埋め式試験の利用について述べたい。

そもそも，多肢選択式テストを作成して出題することの理由を思い出してみる。問題項目に対する正誤が一意に決まることが最大の利点であるが，採点の手間を大幅に減らすことができるということもある。全学で実施することを企図して設計し標準化された小テストであれば，手続きの標準化は（大学入試における共通試験などと同様に）なおさら求められることになる。さらに言えば，実際に幾つかの小テストで試行しているように，CBT（Computer Based Test）や Web テストによって実施することで，いっそう省力化・標準化となり，結果の分析可能性も拡がる（日本教育工学会 , 2012）。

授業内テストの目的は，到達度を確かめることだけでなく，テストを受験することを通しての学習効果をも考えている。すなわち，テストにおいて，問題に答える中で思考を巡らせることであったり，学生自身がテストにおいて解答を書くことによって，その知識が定着するような学習効果などである。テストを受検することを通しての学習効果を目的とすると，客観的・効率的に到達度を測るという目的とは，トレードオフの関係になる点もある。端的に言えば，採点の手間等の労力をいとわなければ，それによって，テストを受けることでの学習効果を高めることができるのである。例えば，多肢選択式試験でも，項目番号を選択するだけでは無く，解答にあてはまりうる用語を語群として列挙したうえで，解答を選択し記入させる,「選択的」穴埋め式も考えられる。これは，多くの授業でも実践されているとは思うが，多肢選択式と穴埋め式の融合形式ともいえる。

さらに，多肢選択式試験を補完するという意味では，論述形式などを併用していることは，冒頭で述べた通りである。文章全体から要点を抜き出すことや，要約させることなども随時行ってきている。ただし，論述形式で，こちらが意図している解答を書かせることは簡単ではなく，得るべき

知識を確実に定着させる意味では，アウトプットを通しての知識獲得が望まれる。すなわち，「選択式ではない」記述式の穴埋め式試験も当然行ってきている。

テストの形式には，様々なものがあるが，それらの利点・欠点を深く理解して，それらを併用することで学生の学習効果を高めるための試みを続けていく必要がある。

補遺：実際の共通試験での具体例

ここでは，実際の大学入試における共通試験の解答方式の具体例を２つあげる。

いずれにおいても，数学は，数字・符号を組み合わせた解答方式である。（便宜的に「数学用」と略記。）その他の科目は，９択を上限とした多肢選択式の解答方式である。

具体例１：日本留学試験（EJU）の解答方式（表 5）

具体例２：大学入試センター試験（DNC）の解答方式（表 6）

表 5　日本留学試験（EJU）の解答方式・選択肢数

試験科目	解答形式および選択肢数の傾向	問題項目数
日本語（記述問題）	記述式	2 問から 1 問選択解答
日本語（読解問題）	4 択固定	25 問
日本語（聴読解問題）	4 択固定	12 問
日本語（聴解問題）	4 択固定	15 問
総合科目	4 択固定	38 問
理科・選択（物理）	選択肢数可変（9 択を上限）	19 問
理科・選択（化学）	選択肢数可変（9 択を上限）	20 問
理科・選択（生物）	選択肢数可変（9 択を上限）	18 問
数学（コース１基本）	選択肢数可変（数学用）	大問構成・小問数可変
数学（コース２上級）	選択肢数可変（数学用）	大問構成・小問数可変

表 6　大学入試センター試験（DNC）の解答方式・選択肢数（2018 年・本試験の主なもの）

試験教科・科目	解答形式および選択肢数の傾向	問題項目数
国語	選択肢数可変 ※複数選択あり	35 問 ※うち複数選択 1 問
英語（筆記）	選択肢数可変 ※組合せ式あり	50 問 ※うち組合せ式 4 問
英語（リスニング）	4 択固定	25 問
公民・地理歴史 （現社・倫理・政経・ 世史・日史・地理 AB）	選択肢数可変 ※世界史 B のみ 4 択固定	32 ～ 36 問
理科基礎 （物・化・生・地）	選択肢数可変 ※物理基礎は部分点あり ※生物基礎は複数選択あり	13 ～ 16 問
理科 （物・化・生・地）	選択肢数可変 ※物理は部分点・生物は複数選択あり	23 ～ 31 問
数学 1・数学 A	選択肢数可変（数学用）	大問構成・小問数可変
数学 2・数学 B	選択肢数可変（数学用）	大問構成・小問数可変

引用・参考文献

荒井清佳（2005a）「多肢選択式問題における当て推量による正答について」『人事試験研究』**194**, 20-26.

荒井清佳（2005b）「「選択肢 3 番は選ばれやすい」は本当？」『人事試験研究』**196**, 18.

荒井清佳（2015）「多肢選択式問題を作成する上で大切なこと：問題作成の専門家に対する調査結果に基づいて」『日本テスト学会誌』**11**（1）, 21-34.

荒谷大輔 , 鈴木哲平 , 岡田大助 , 福島亜理子 , 田上大輔 , 大竹洋平 , 石野一晴 , 羽村太雅 , 中原真祐子（2018a）「大学教育において求められる「新しい教養」の検討：江戸川大学における「アカデミック・スキル演習」の導入」『江戸川大学紀要』**28**, 1-8.

荒谷大輔 , 鈴木哲平 , 岡田大助 , 福島亜理子 , 田上大輔 , 大竹洋平 , 石野一晴 , 羽村太雅 , 中原真祐子（2018b）「「新しい教養」のための「チーム・ティーチング」：江戸川大学、基礎・教養教育センターの実践例」『江戸川大学紀要』**28**, 9-15.

肥田野直（編）（1972）『テスト I』. 東京大学出版会 . シリーズ「心理学研究法」, No.7.

池田央（編）（1972）『テスト II』. 東京大学出版会 . シリーズ「心理学研究法」, No.8.

池田央（1985）「ランダム・ゲッシングにもとづく正解域の推定表」（大学入試センター）『研究紀要』**11**, 1-40.

池田央（1992）『テストの科学：試験にかかわるすべての人に』. 日本文化科学社 .（2007 年 . 教育測定研究所よりオンデマンド版あり .）

池田央（1994）『現代テスト理論』. 朝倉書店 . 行動計量学シリーズ, No.7

石川巧（2008）『「国語」入試の近現代史』. 講談社 . 講談社選書メチエ, No.405.

Linn, Robert L.（ed.）（1989）*Educational measurement*. 3rd ed. Macmillan Publishing Company（池田央監訳（1992）『教育測定学（上・下）』C.S.L. 学習評価研究所, みくに出版（発売））.

蓑谷千凰彦（編）（2010）『統計分布ハンドブック（増補版）』. 朝倉書店 .

光永悠彦（2017）『テストは何を測るのか：項目反応理論の考え方』. ナカニシヤ出版 .

村木英治（2011）『項目反応理論』. 朝倉書店 . シリーズ「行動計量の科学」, No.8.

中島英博（編著）（2018）『学習評価』. 玉川大学出版部 . シリーズ大学の教授法 , No.4.

日本教育工学会（監修）永岡慶三, 植野真臣, 山内祐平（編）（2012）『教育工学における学習評価』. ミネルヴァ書房 . 教育工学選書 , No.8.

日本テスト学会（編）（2007）『テスト・スタンダード：日本のテストの将来に向けて』. 金子書房 .

西岡加名恵, 石井英真, 田中耕治（編）（2015）『新しい教育評価入門： 人を育てる評価のために』. 有斐閣 .

櫻井捷海 , 藤井光昭, 岩坪秀一, 伊藤圭, 松田稔樹（2003）「筆記解答方式とマークシート解答方式により測定された学力の比較」『大学入試研究ジャーナル』**13**, 63-68.

芝祐順ほか（1984）『統計用語辞典』. 新曜社 .

芝祐順（編）（1991）『項目反応理論：基礎と応用』. 東京大学出版会 .

鈴木規夫 , 山田文康 , 池田輝政 , 赤木愛知（1991）「国語の試験問題の出題形式に関する比較研究」（大学入試センター）『研究紀要』**20**, 1-45.

Stevens, Dannelle D. and Levi, Antonia（2013）*Introduction to Rubrics：An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. 2nd ed. Stylus Publishing（井上敏憲 , 俣野秀典 , 佐藤浩章訳（2014）『大学教員のためのルーブリック評価入門』玉川大学出版部 . 高等教育シリーズ , No.163）.

Suskie, Linda A.（2009）*Assessing Student Learning：A Common Sense Guide*. 2nd ed. John Wiley & Sons（齋藤聖子訳（2015）『学生の学びを測る ： アセスメント・ガイドブック』玉川大学出版部 . 高等教育シリーズ , No.170）.

寺井俊健 , 後藤忠彦 , 柴田昌文（1961a）「択一式試験のゲスの実態（I）」『試験研究』**31**, 43-65.

寺井俊健 , 後藤忠彦 , 柴田昌文（1961b）「択一式試験のゲスの実態（II）」『試験研究』**32**, 34-50.

東京大学教養学部統計学教室（編）（1992）『自然科学の統計学』. 東京大学出版会 . シリーズ「基礎統計学」, No.3.

豊田秀樹（2012）『項目反応理論（入門編）（第 2 版）』. 朝倉書店 . 統計ライブラリー .