試験の得点調整における分位点差圧縮手法の改良

福嶋　　裕
高橋　美貴
樽磨　和幸
佐藤　　仁

1．序　論

科目の選択を認める試験や、期日を変えて同じ科目の試験を実施する場合、試験者は、異なる被験者集団に異なる問題を出題し、その成績を対等に比較しなければならない。その比較の公平性を確保するために、試験者は大変苦労する。

この問題に対処する手段として、さまざまな標準得点が工夫され、使われてきた[1)]。その代表例が、通常「偏差値」と呼ばれるZ得点である。これは、素点をXとし、その平均値をμ、標準偏差をσとするとき、一次変換、

$$Z = \frac{10\left(X-\mu\right)}{\sigma} + 50 \qquad (1)$$

で与えられる。この標準得点は極めて簡単に計算できる。また、被験者も試験者も、この得点に不満を感じることは少ない。そういう利点から、この得点は多くの試験現場で広く利用されている。だが、これによる得点調整は、平均値を50、標準偏差を10に統一するため、それらの統計的なゆらぎは無視される。その結果、補正が過剰になることが少なくない。そのような典型例を**図１a**に示す。Kolmogorov-Smirnovの２標本検定[2)]によると、S_1とS_2の標本の母集団が一致する確率は10^{-7}より小さい。だが、Z_1とZ_2で示すこれらのZ得点は、きわめてよく一致する。これとは対照的に、素点の分布の形が著しく異なるときには、Z得点で十分な補正ができるとは限らない。しかも、仮に調整不足が判明しても、それを再補

1）実際の試験で得点調整を行う場合には、試験の得点に関する社会的先入観や披験者の心理的影響などさまざまな要因に配慮する必要があり、それらを含めた研究も行われている（たとえば 石塚 2002、村上 2003）。

2）Kolmogorov-Smirnovの２標本検定ならびに適合度検定の手法は、たとえば、武藤（1995）を参照のこと。また、以下のサイトも参考になる。
Kolmogorov-Smirnov Test, http://www.physics.csbsju.edu/stats/KS-test.html

正する手段はない。そのような典型例を**図1b**に示す。この図で、S_3とS_4の標本の母集団が一致する確率は、いずれも10^{-7}以下である。だが、これらの平均値は50に近く、標準偏差は10に近いため、Z得点による補正で分布の不一致が改善されることはほとんどない。

Z得点ほど一般的ではないが、T得点と称する標準得点を使うこともある。T得点は、素点Xの累積相対度数分布を$F_k(X)$とし、平均値が50で標準偏差が10の正規分布関数を$N(T, 50, 10)$として、

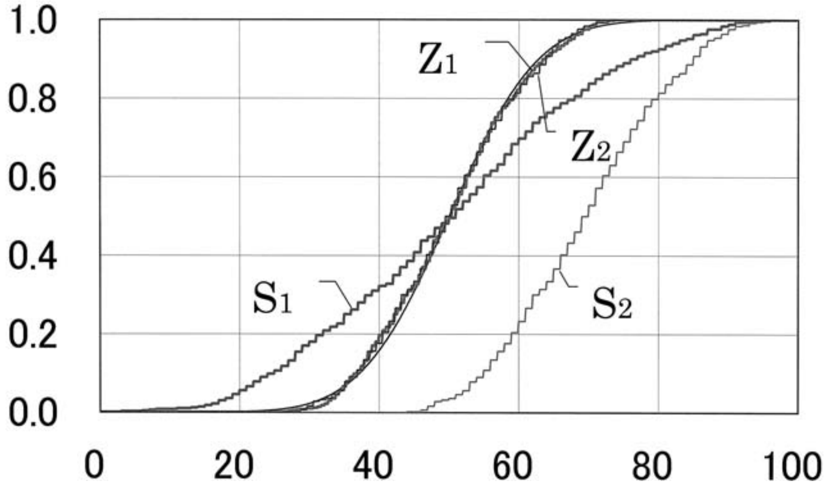

図1a　Z得点が過剰補正となる例

折れ線のS_1とS_2は、1点刻みの100点満点で作成した各500個の模擬素点の累積相対度数分布を、Z_1とZ_2はそれらのZ得点の累積相対度数分布を表わす。また、曲線は正規分布$N(x, 50, 10)$を表わす。Z_1とZ_2は、正規分布とほぼ完全に一致する。

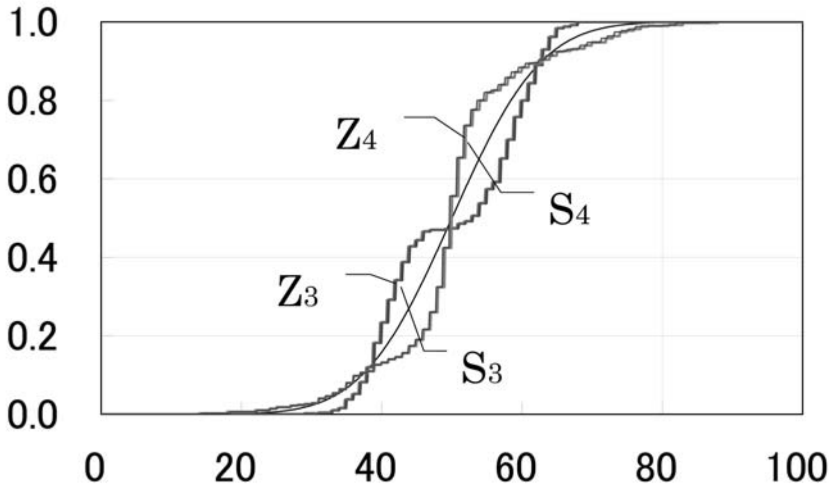

図1b　Z得点では補正できない例

折れ線のS_3とS_4は、1点刻みの100点満点で作成した各500個の模擬素点の累積相対度数分布を、Z_3とZ_4はそれらのZ得点の累積相対度数分布を表わす。また、曲線は正規分布$N(x, 50, 10)$を表わす。Z得点による補正の効果は認められない。

$$F_k(X) = N(T, 50, 10) \qquad (2)$$

が成り立つ T、すなわち、X に対応する等分位点を指す。これを使って調整すれば、得点分布を高い精度で基準とする正規確率分布に一致させることができる。しかし、その調整は、試験で得たデータのうちの順位情報以外をすべて無視するもので、常に過剰なものとならざるを得ない。T 得点が過剰補正になることを示す典型例を**図 2** に示す。この図の S_3 と S_4 は、**図 1 b** に示した標本の分布で、T_3 と T_4 はそれらの T 得点分布である。いずれの場合も T 得点は基準の確率分布 $N(T, 50, 10)$ とほとんど完全に一致する。

大学入試センター試験では、「分位点差縮小法」と称する、Z 得点や T 得点とは異なる方法を、選択科目間の得点調整に採用している（前川　2001）。この方法は、平均点が最も高い科目の累積相対度数分布を基準とし、他の科目の累積相対度数分布の分位点差を一定の比率で縮小するものである。大学入試センターは、平均点が最も高い科目の平均点を μ_{max}、最も低い科目の平均点を μ_{min} とすると、$\mu_{max} - \mu_{min} \geq 20$ であればこの調整を実施するとしている。

そのとき、分位点差の縮小比率を $\dfrac{15}{\mu_{max} - \mu_{min}}$ とする。この得点調整法は、不公平感の払拭と過剰補正の回避のために、次のような点で配慮がなされている。

① 「試験問題の難易差に基づく平均点の差が20点以上の場合に調整を実施する」ということで、調整実施の要件が確定している。

② 「平均点の差を15点以下に抑える」ということで、調整の目標が確定している。

③ 素点に含まれる平均点や標準偏差の揺らぎの情報が消失していない。

その点で、よく工夫された手法といえるだろう。しかし、調整実施の要件にしても、調整の目標にしても、統計的有意水準を明示したものではない。また、平均点の差に注目するだけで、分布形の広がりや歪みの違いは無視されている。したがって、その点でなお重大な欠

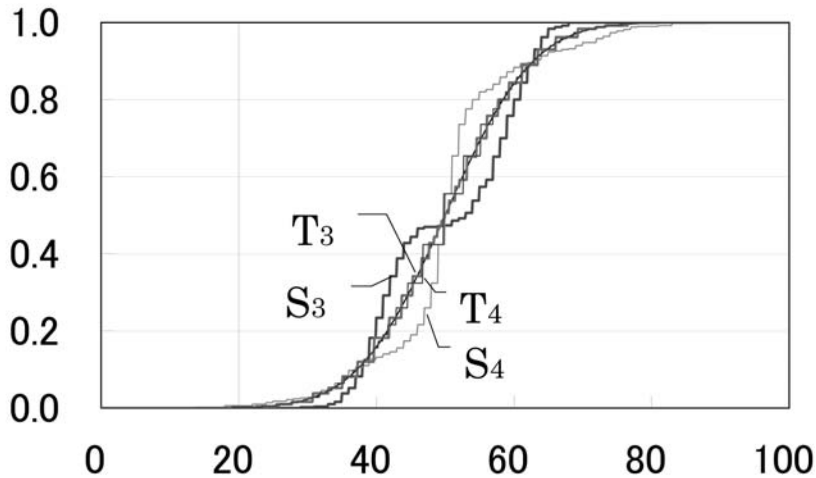

図 2　T 得点による過剰補正の例

折れ線の S_3 と S_4 は図 1 b に示した模擬素点の累積相対度数分布を、T_3 と T_4 はそれらの T 得点の累積相対度数分布を表わす。また、曲線は正規分布 $N(x, 50, 10)$ を表わす。T_3 と T_4 は正規分布とほぼ完全に一致する。

陥があるといわなければならない。この調整方法では調整が不可能である例を**図3**に示す。図から明らかなように、S_5、S_6、S_7の標本分布は著しく異なる。Kolmogorov–Smirnov の2標本検定によると、これらの標本の母集団が一致する確率は1％にも達しない。しかし、平均点の差は1点未満であるため、大学入試センター方式では調整の対象から除外される。

我々は、上に述べたような従来の得点調整法の欠陥を考慮し、新しい調整法を提案した（福嶋ほか　2005)。それは、基準として設定した確率分布関数と、素点の累積相対度数分布を比較し、一定の比率で分位点差を圧縮するものなのだが、圧縮比率を Kolmogorov–Smirnov の適合度検定によって決定するところに特徴がある。この方法によれば、得点調整の要件も目標も統計的有意水準の形で明示される。そして、従来の得点調整法では避けられない問題がおおむね解決できる。しかし、この方法にも不備な点や修正を要する点が残っている。本稿では、それらの点について議論する。まず第2章で、前に提案した得点調整法の概要を述べ、残された問題点を洗い出す。そして第3章で、その問題点の解決策を新たに提案する。

2．旧提案の概要と残された問題点

我々が、福嶋ほか（2005）で提案した得点の調整法は、試験の得点のデータを統計標本とみなし、その母集団の確率分布が基準となる確率分布関数に一定の有意水準で一致するように分位点差を圧縮するものである。以下に、その調整手順の概略を示す。

調整は、まず、基準とする確率分布関数 F の設定から始める。この関数は、技術的に

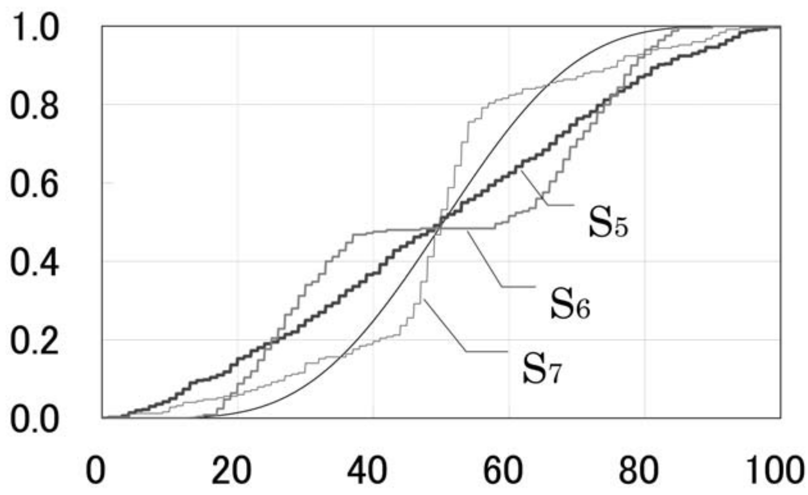

図3　「分位点差縮小法」で調整の対象外となる例

折れ線のS_5、S_6、S_7は、平均値がほぼ一致する模擬素点の累積相対度数分布を表わす。また、曲線は、素点の標本S_5、S_6、S_7をまとめたものに、ベータ分布型の確率分布関数を当てはめた結果を表わす。これらの模擬素点は、大学入試センター方式の得点調整では、調整の対象外とされる。なお、模擬素点の標本は、1点刻みの100点満点で、S_5のデータセットは500個、S_6、S_7については各250個作成した。

は、値域が［0, 1］で、逆関数 F^{-1}が容易に数値計算できる連続非減少関数であれば、何であっても差し支えない。しかし、試験の目的が被験者の資質の測定であることを考えると、正規分布か、それに近い形の常識的な分布を採用することになる。次に、調整実施の要件とする統計的有意水準 α を設定する。これは、自動的に調整の目標ともなる。これだけの準備をした後、注目する被験者集団 H_k について、素点の標本 H_k の累積相対度数分布 F_k を、基準の関数 F と比較する。そして、 Kolmogorov–Smirnov 統計量 D_k と、有意水準 α に対応するその臨界値 D_α の関係をみる。その結果、もし $D_k \leq D_\alpha$ なら調整の必要はないと判断する。それに対して、 $D_k > D_\alpha$ なら調整の必要があると考え、 F_k と F の分位点差を一定の比率で圧縮するための係数 c_k を探す。この値は、素点 X を

$$U = F^{-1}\left(F_k\left(X\right)\right) + c_k\left(X - F^{-1}\left(F_k\left(X\right)\right)\right) \tag{3}$$

で変換したときの U に関する累積相対度数分布が $G_k(U)$ になるとして、

$$\sup\left\{\left|G_k - F\right|\right\} = D_\alpha \tag{4}$$

が成り立つようにする。この値を求める具体的な計算法は少し煩雑であり、しかも後に指摘するように修正が必要なので、ここでは示さない 。

この調整法によれば、得点調整の必要性を判断する基準が、統計的有意水準の形で明示される。しかも、その基準は、単に平均点や標準偏差だけに注目するものでなはく、得点分布の形状全般を考慮するものになっている。また、得点調整の結果が所期の水準に達しているかどうかも自動的に検証される。

しかし、この得点調整法にも、不備な点や修正を要する点が残っている。その第１の点は、基準とする確率分布関数 F の設定の仕方である。確かに、体格や運動能力など普遍的な尺度が存在する資質に注目すれば、その測定値は正規分布をなすと認められる。その事実から類推すると、知的資質も適切な尺度で測定すれば、結果が正規分布をなすと期待できる。したがって、得点調整の基準としては、正規分布がふさわしいといえるだろう。しかし、正規分布であればどれでもよいわけではない。たとえば、平均値や標準偏差が素点のデータと著しく異なるような正規分布は、補正が大幅になってしまうので、基準としてふさわしいとはいえない。調整ができるだけ小幅で済むように、素点のデータに即した基準の設定が望まれる。

第２の問題点は、得点調整の要件とする有意水準 α の決め方である。我々の旧提案では、この値は客観的根拠に基づいて一律に決められるものではないとして議論を避けた。もちろん、この値の選定に当たっては、被験者の不公平感が払拭できるかどうか、あるいは、出題者の同意が得られるかどうかといった心理的条件が重要な意味を持つ。しかし、試験者の立場からすれば、そういう条件を勘案するためにも統計的な根拠が欲しいことはいうまでもない。

第３の問題点は、同点者が多数出現する場合に適切な調整ができないことである。この問題は、試験問題に欠陥がある場合だけでなく、得点の階級数に比べて被験者数が多過ぎる場合にも生じる。たとえば、人数が n_k の被験者集団 H_k を対象とする試験において、得点の階級を $X_1 < X_2 < \cdots < X_N$ の N 段階とした場合を考えてみる。このとき、我々の旧提案では、

検定統計量を

$$D_k = \sup\left\{\left|F_k(X_i) - F(X_i)\right|, \left|F_k(X_i) - \frac{J_i}{n_k} - F(X_i)\right|\right\} \qquad (5)$$

とした。ここで、J_i は得点 X_i の同点者数である。この式で、右辺の $\frac{J_i}{n_k}$ は $\frac{1}{N}$ に比例する。それに対して、有意水準 α に対応する Kolmogorov–Smirnov 統計量の臨界値 D_α は、$n_k >$ 40であれば $\frac{1}{\sqrt{n_k}}$ に比例する。したがって、n_k が N^2 よりはるかに大きいなら、$\frac{J_i}{n_k} > 2D_\alpha$ であるような得点の階級が必ず生じ、(5)式の D_k は D_α を超える。その結果、得点の標本 H_k の母集団分布は、基準の確率分布と一致しないという結論に至る。この事態は、大集団を対象とする試験なら、たとえ試験問題が適切であっても避けられない。それは、(5)式が検定統計量として不適当なことを意味するもので、(5)式が修正されなければならないことを示している。

３．得点調整法の新提案

我々が前に提案した得点の調整法には、上で述べたとおり３つの問題点が残っている。それらを解決するための新しい提案を以下に示す。

［基準分布の設定］

m 組の被験者集団、$\mathrm{H}_1, \mathrm{H}_2, \cdots, \mathrm{H}_m$ を対象に問題の異なる試験を実施し、それぞれ得点の範囲を０からAとして採点したとする。そして、それによって得た素点の標本をそれぞれ、$H_1, H_2, \cdots, H_m$ とする。

これらの素点をここで提案する方法によって調整するには、まず基準の確率分布 F を設定する必要がある。第２章で述べたとおり、F は逆関数 F^{-1} の数値計算が容易なら、技術的には何を採用してもよい。しかし、得点の調整幅は小さいほど望ましいから、採点で得た累積相対度数分布にできるだけ近い確率分布を選びたい。そのような要求に応えるものとして、すべての素点を一括した標本、$H = H_1 \cup H_2 \cup \cdots \cup H_m$ の累積相対度数分布に当てはめた確率分布を採用することが考えられる。

そういう分布関数の第１の候補として、分布の形を正規型と仮定したものが挙げられる。それは、H の標本平均を μ、標本分散を σ^2 として、

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{(\xi-\mu)^2}{2\sigma^2}\right] d\xi \qquad (6)$$

で与えられる。

体型や運動能力など普遍的尺度で測定できるヒトの資質は、たいてい正規分布をなす。その事実から考えると、これは得点調整の基準としてふさわしいものといえる。しかし、試験の目的によっては、出題者が、正規分布と歪度や尖度の異なる分布を意図的に期待して出題

することもある。そのようなとき、得られた得点分布が期待したものに近いにもかかわらず、正規分布型の基準を採用すると、調整は出題者の意図に沿わないものとなる。そのような問題の解決策として、分布の形をベータ型と仮定する方法が挙げられる。得点の定義域を［0，A］とするベータ分布型の基準分布関数は、

$$p = \frac{\mu}{A}\frac{\mu\left(A-\mu\right)-\sigma^2}{\sigma^2} \tag{7 a}$$

$$q = \frac{A-\mu}{A}\frac{\mu\left(A-\mu\right)-\sigma^2}{\sigma^2} \tag{7 b}$$

$$B = \int_0^A \xi^{p-1}\left(A-\xi\right)^{q-1} d\xi \tag{7 c}$$

として、

$$F\left(x\right) = \frac{1}{B}\int_0^x \xi^{p-1}\left(A-\xi\right)^{q-1} d\xi \tag{8}$$

で与えられる。

［有意水準 α の設定］

人数が n_k の被験者集団 H_k について、得点の標本 H_k の累積相対度数分布を F_k とする。ここでは、得点が素点であっても調整点であってもよい。ただし、 n_k >40であるとする。このとき、基準の確率分布関数 F が H_k の母集団の分布関数と一致するなら、 F から導かれる基準の累積相対度数分布を F_0 として、 Kolmogorov–Smirnov 統計量

$$D_k = \sup\left\{\left|F_k - F_0\right|\right\} \tag{9}$$

が臨界値

$$D_\alpha = \frac{K}{\sqrt{n_k}} \tag{10}$$

を超える確率 α は、

$$\alpha = \Pr ob\left\{D_k > D_\alpha\right\} = 2\sum_{j=1}^{\infty}\left(-1\right)^{j-1}\exp\left(-2j^2K^2\right) \tag{11}$$

で近似できる。

ここで

$$1-\alpha = \Pr ob\left\{D_k \le D_\alpha\right\} = 1-2\sum_{j=1}^{\infty}\left(-1\right)^{j-1}\exp\left(-2j^2K^2\right) \tag{12}$$

を、 K を変量とする確率分布関数と考えてみる。すると、この分布の密度関数は、

$$-\frac{d\alpha}{dK} = 4j^2K\alpha \tag{13}$$

で与えられる。この関数は、**図 4** に示すように、 $K > K_\mu$ の側に尾を引いた非対称形になっている。これをもとに、変量 K の平均値 K_μ と標準偏差 K_σ を求めると、それぞれ、 $K_\mu =$ 0.8687…、および、 $K_\sigma = 0.2603$…を得る。

ところで、一般に、偏差が標準偏差以下であるのは、ごく普通のこととみなせる。一方、

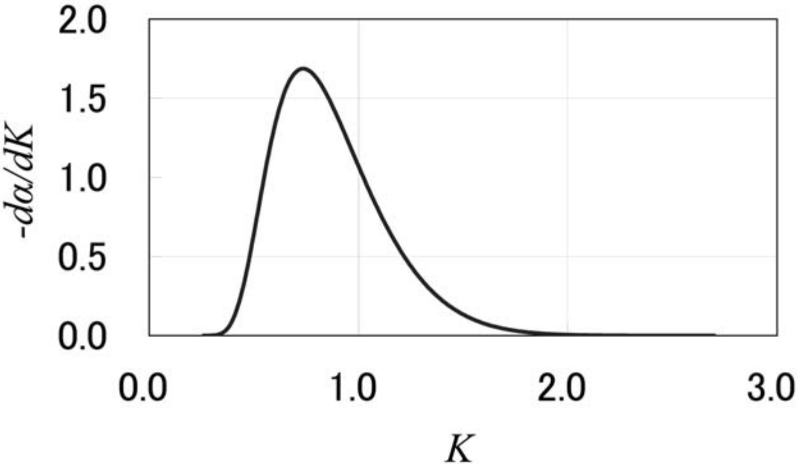

図4　$-\dfrac{da}{dK}$ 分布

(13)式をもとにグラフ化した。K の平均値 K_μ と標準偏差 K_σ は、それぞれ、$K_\mu = 0.8687\cdots$、および、$K_\sigma = 0.2603\cdots$ である。

偏差が標準偏差の2倍を超えるのは希なことと考えなければならない。統計学の常識ともいえるこの観点からすると、得点調整の基準を決める K の値は、

$$K_\mu + K_\sigma < K < K_\mu + 2K_\sigma \tag{14}$$

の範囲内に定めるのが適当といえる。そこで、たとえば、適度な妥協点としてこの範囲の中点をとり、

$$K = K_\mu + 1.5K_\sigma = 1.26 \tag{15}$$

と設定したとする。このとき、得点調整が過剰である確率は、8.4%に抑えられる。

［同点者が多数ある場合の対応策］

被験者集団 H_k を対象とする試験において、得点の階級を $X_1 < X_2 < \cdots < X_N$ の N 段階として採点し、素点の標本 H_k を得たとする。このとき、被験者数 n_k が階級数 N より大きいなら、必ず同点者が存在する。その同点者の資質は、素点で区別ができないからといって完全に等しいとは考えにくい。むしろ、理想的な分解能をもつ試験を想定するなら、すべてが識別でき、異なる実数値で表現できると考えられる。そこで、全被験者について、そのような実数表現が得られたものとし、それを x としてみる。そうすると、ある素点 X_i の同点者の場合、x は X_i の近傍に分布するはずである。もちろん、その分布の形は知り得ない。だが、成績の逆転が起きてはならないことから、異なる素点の同点者に注目したとき、両者の x の分布範囲が互いに重なり合うことはあり得ない。また、常識的観点に立てば、素点は同点者の資質分布の中央値と考えられるから、素点 X_i の同点者の x は、半数が X_i 以下、半数が X_i 以上の範囲に分布すると仮定できる。

このような考え方に基づき、x が X_i 以下の範囲に含まれる確率を $f_k\,(X_i)$ とすると、

$$f_k\left(X_i\right)=\sum_{j=1}^{i}\frac{J_j}{n_k}-\frac{J_i}{2n_k}=F_k\left(X_i\right)-\frac{J_i}{2n_k} \tag{16}$$

となる。ここで、J_i は X_i の同点者数である。

一方、x の分布について Kolmogorov–Smirnov の適合度検定を行うとき、基準とする確率分布 $F(X)$ において、X が X_i 以下の範囲に含まれる確率 $F_0(X_i)$ は、

$$F_0\left(X_i\right)=F\left(X_i\right) \tag{17}$$

で与えられる。したがって、 Kolmogorov–Smirnov 統計量は、第 2 章に示した(5)式ではなく、

$$D_k=\sup\left\{\left|f_k\left(X_i\right)-F\left(X_i\right)\right|\right\} \tag{18}$$

としなければならない。

検定統計量をこのように修正すると、同点者が多数ある場合の問題は解消し、任意の素点の標本に対して、指定した有意水準の目標を達成する適切な得点調整が可能となる。それにはまず、各 X_i について

$$\left|f_k\left(X_i\right)-F\left(F^{-1}\left(f_k\left(X_i\right)\right)+C_{ki}\left(X_i-F^{-1}\left(f_k\left(X_i\right)\right)\right)\right)\right|\le D_\alpha \tag{19}$$

が成り立つ係数 C_{ki} を計算する。この係数は、

$\left|f_k\left(X_i\right)-F\left(X_i\right)\right|\le D_\alpha$ のとき

$$C_{ki}=1 \tag{20 a}$$

$\left|f_k\left(X_i\right)-F\left(X_i\right)\right|> D_\alpha$ のとき

$$C_{ki}=\frac{F^{-1}\left(f_k\left(X_i\right)-sign\left(f_k\left(X_i\right)-F\left(X_i\right)\right)\cdot D_\alpha\right)-F^{-1}\left(f_k\left(X_i\right)\right)}{X_i-F^{-1}\left(f_k\left(X_i\right)\right)} \tag{20 b}$$

で与えられる。次に、これをもとに、分位点差圧縮係数 c_k を $\{C_{ki}\}$ の下限、

$$c_k=\inf\left\{C_{ki}\right\} \tag{21}$$

と定める。そして、素点 X_i を、変換式

$$U_i=F^{-1}\left(f_k\left(X_i\right)\right)+c_k\left(X_i-F^{-1}\left(f_k\left(X_i\right)\right)\right) \tag{22}$$

で調整点 U_i へと変換する。

この変換で、被験者の成績が U_i 以下の範囲に含まれる確率 $G_k(U_i)$ は、

$$G_k\left(U_i\right)=f_k\left(X_i\right) \tag{23}$$

になる。また、基準の確率分布 $F(x)$ において、x が U_i 以下の範囲に含まれる確率 $G_0(U_i)$ は、

$$G_0\left(U_i\right)=F\left(U_i\right) \tag{24}$$

である。したがって、U_i の標本 h_k の Kolmogorov-Smirnov 統計量は、

$$d_k=\sup\left\{\left|G_k\left(U_i\right)-G_0\left(U_i\right)\right|\right\}=\sup\left\{\left|f_k\left(X_i\right)-F\left(U_i\right)\right|\right\} \tag{25}$$

ということになる。これと c_k の定義式 (21) から明らかなように、すべての i について、

$$d_k \leq D_\alpha \tag{26}$$

が成り立つ。すなわち、(22)式で与えられる得点 U_i は、所期の目標水準 α に達する調整点になっている。なお、すべての i について $D_k \leq D_\alpha$ であれば得点調整は不必要だが、そのときは、(20)式から明らかなように、$c_k = 1$ となり、$U_i = X_i$ となって調整過剰は起こらない。また、$D_k > D_\alpha$ のときには、$c_k < 1$ となるが、$C_{ki} = c_k$ であるような点が必ず存在し、そこで $d_k = D_\alpha$ が成り立つ。このことは、(22)式に基づく得点調整が必要最小限のものであることを意味する。

この得点調整法の効用を示す模擬実験の結果を**図5**と**図6**に示す。**図5**は、大学入試センター方式では調整の対象とならない例として**図3**に示した標本を、上に提案した方法で調整した結果である。図中の曲線は、素点の標本、S_5、S_6、および、S_7 をまとめたものに、ベータ分布型の関数を当てはめた結果を表わしている。U_5、U_6 および U_7 は、その確率分布関数を基準として計算した調整点である。Kolmogorov-Smirnov の適合度検定によれば、これらの調整点の母集団分布は、いずれも8.4%の有意水準で基準の分布と一致している。それは必要最小限の調整ができていることを意味する。

図6は、Z 得点では補正過剰となる例として**図1a**に示した標本を、上に提案した方法で調整した結果である。図中の曲線は、平均が50、標準偏差が10の正規分布を表わしている。

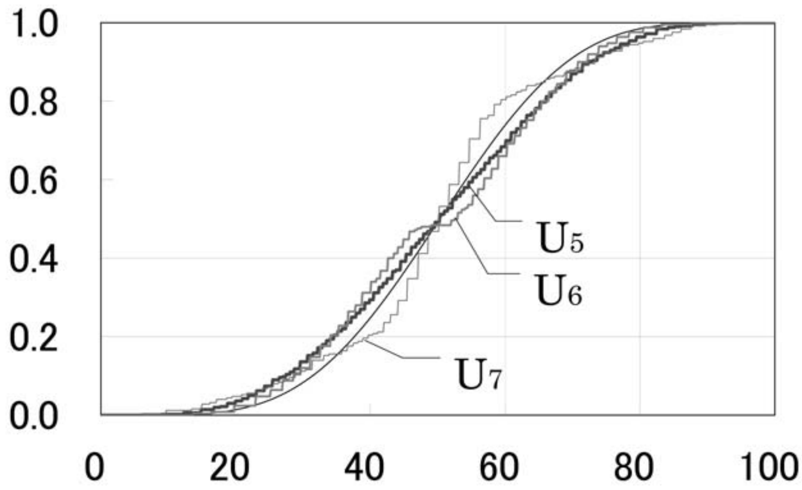

図5　図3の素点を新提案の方法で得点調整した例

折れ線の U_5、U_6 および U_7 は、図3に示した模擬素点の標本 S_5、S_6 および S_7 を(22)式で変換した調整点の累積相対度数分布を表わす。また、曲線は、素点の標本 S_5、S_6 および S_7 をまとめたものに、ベータ分布型の確率関数を当てはめた結果を表わす。図3と比較すれば明らかなとおり、適度な得点調整ができている。

U_1とU_2は、その確率分布関数を基準として計算した調整点で、この場合も、その調整点の母集団分布は、いずれも8.4%の有意水準で基準の分布と一致している。これもまた、必要最小限の調整ができていることを意味する。

4．結　論

我々は、福嶋ほか（2005）でKolmogorov–Smirnovの適合度検定を利用する得点の調整法を提案した。それは、調整の要件と目標が統計的有意水準の形で明示される点、平均点や標準偏差の違いだけでなく分布形全般を考慮した調整ができる点、調整結果が所期の水準に達しているかどうかが自動的に検証される点などで、従来の方法に比べて優れたものになっている。しかし、この提案にも次のような不備な点が残っていた。

(i)　試験のデータに即して基準の確率分布を決める合理的手法が示されてない。

(ii)　合理的根拠をもって得点調整の目標を決める方法が示されてない。

(iii)　過度に多数の同点者がある場合、所期の水準に達する調整ができない。

我々は、これらの問題を解消する手立てを第3章で示した。その提案を要約すれば、次のとおりである。

第1の提案は、すべての素点を一括した標本を考え、その標本平均と標本分散から母集団の分布を推定し、その分布関数を検定の基準として採用する考え方である。この方法によれば、素点のデータに即して基準が決まるから、得点の調整幅を最小限に抑えることができる。なお、分布の具体的な形は、正規分布型、または、ベータ分布型が適当と考えられる。

第2の提案は、有意水準αに対応するKolmogorov–Smirnov統計量の臨界値をD_αとするとき、$K = D_\alpha\sqrt{n_k}$を変量と考え、その確率分布の平均値K_μと標準偏差K_σをもとに、$K_\mu + K_\sigma < K < K_\mu + 2K_\sigma$の範囲で$K$を決めようというものである。このようにすれば、普通

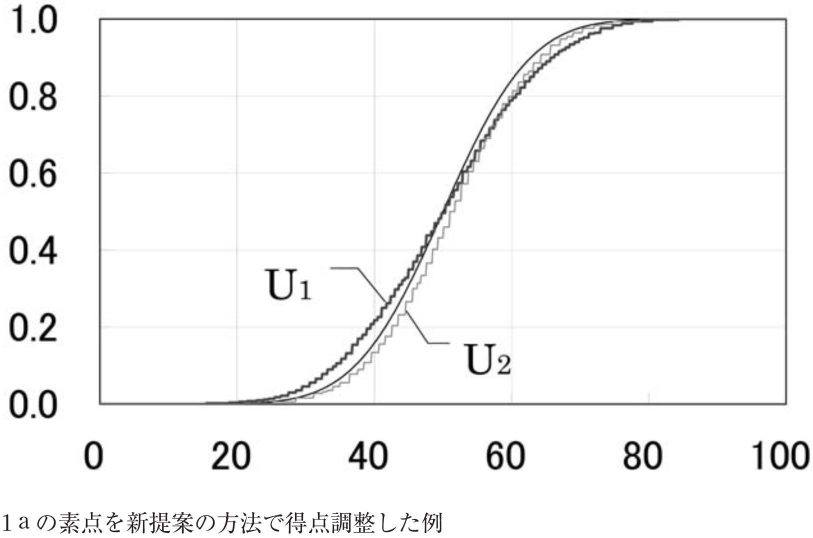

図6　図1aの素点を新提案の方法で得点調整した例

折れ線のU_1とU_2は、図1aに示した模擬素点の標本S_1とS_2を(22)式で変換した調整点の累積相対度数分布を表わす。また、曲線は正規分布$N(x, 50, 10)$を表わす。図1aと比較すれば明らかなとおり、適度な得点調整ができている。

に起こるとはいえない素点の変動は補正の対象となり、普通に起こり得る素点の変動は補正の対象から除かれる。なお、この平均値と標準偏差の具体的な値は、それぞれ、$K_{\mu}=0.87$ と $K_{\sigma}=0.26$ である。

第3の提案は、同点者が多数の場合に対応する方策である。この提案は、被験者の資質が原理的にはすべて区別でき、実数で表現できることを前提とする。それに加え、素点は同点者の資質表現の中央値であると仮定する。そうすると、検定統計量は (5)式で表わされる旧提案の形から(18)式で表わされる正しい形に修正される。そこで、(20)式と(21)式にしたがって分位点差圧縮係数を計算し、それを用いて(22)式に示す分位点差圧縮変換を行う。これが、第3の提案の概要である。この方法によれば、得点調整が不要な場合は素点がそのまま生かされ、調整が必要な場合は必要最小限の調整結果が自動的に得られることになる。

以上の3提案を採用すれば、異なる被験者集団に異なる問題を出題した場合、被験者数が40程度以上なら、常に一定水準の統計的信頼性をもって、試験の成績を公平に比較することができる。

参考文献

Kolmogorov–Smirnov Test, http://www.physics.csbsju.edu/stats/KS–test.html

石塚智一（2002)「大学入学者選抜における評価の標準化に関する研究」大学入試センター研究開発部　平成11年～13年度　共同研究報告書。

福嶋裕、高橋美貴、樽磨和幸、佐藤仁（2005)「試験の得点調整に関する提案」大阪商業大学論集138：１－11。

前川眞一（2001)「大学入試センター試験における選択科目間の得点調整について」計測と制御 40（8）：568－571。

武藤眞介（1995)「統計解析ハンドブック」朝倉書店。

村上隆（2003)「わが国の公的試験における得点等化の導入に向けた心理・教育測定的研究」平成12～14年度科学研究費補助金（12800015）特別推進研究（1）研究成果報告書。