

合否入れ替わり率のブートストラップ法による区間推定

前川真一*

菊地賢一**

要約

一般に、複数のテストを用いて選抜を行う場合、 N 人の受験者に関して各テストの重み付き合計点を算出し、その上位から N_0 人の合格者を選抜する方式が取られることが多い。この場合のテスト j の寄与を示す合否入れ替わり率とは、 N_0 人の合格者の中で、もしもテスト j の得点を除いて選抜を行った場合には不合格となる受験生の割合である。

本稿では、実際の大学入試のデータからブートストラップ法を用いて合否入れ替わり率の標準誤差を推定した結果、多くの場合、合否入れ替わり率は安定した指標であることを示した。また、その信頼区間を求め、入学者選抜試験の経年的変化との関連について考察すると共に、合否の入れ替わりをもとに在学生の得意とする教科・科目を同定し、学生および入試システムの診断目的に利用することを提案した。

キーワード： 合否入れ替わり率、入れ替わり率、選抜試験、ブートストラップ法

*大学入試センター 研究開発部 進学適性研究部門

〒153 東京都目黒区駒場 2-19-23 e-mail: mayekawa@rd.dnc.ac.jp Telefax: 03-5478-1297

**大学入試センター 研究開発部 情報処理研究部門

1 はじめに

一般に、複数 (m 個) のテストを用いて選抜を行う場合、 N 人の受験者に関して各テストの重み付き合計点を算出し、その上位から N_0 人の合格者を選抜するという方式が取られることが多い。すなわち、

$$t = \sum_{j=1}^m \alpha_j x_j \quad (1)$$

により受験生を順序づけ、その上位 N_0 人を合格者とする。ただし、 x_j は個人のテスト j における得点、 α_j は合計点算出のための各テストの重みである。

この様に、複数の試験を用いて受験生を選抜する場合に、選抜に際する個々のテストの貢献度を知することは、受験生にとっても、大学側にとっても重要である。通常、受験生側からは、各テストの重要度はそのテストの総合点に対する配点という形でしか認識されていない。他方、大学側で選抜試験に携わる人々の間では、各テストの重要度を示す指標はこれまでにいくつか開発されてきている。例えば、各テストの分散を s_j^2 で表せば、

1. 各テストの配点比、 $\alpha_j / \sum \alpha_j$
2. 各テストの標準偏差の比、 $\alpha_j s_j / \sum \alpha_j s_j$
3. 各テストの分散の比、 $\alpha_j^2 s_j^2 / \sum \alpha_j^2 s_j^2$
4. 相関係数、 $r_{tj} = \text{corr}(t, \alpha_j x_j)$
5. 共分散比、 $\text{cov}(t, \alpha_j x_j) / \text{var}(t)$
6. R^2 の増分、 t をすべての変数へ回帰したときの決定係数 (この場合は常に 1) と x_j を除く他の全てのテストに回帰したときの決定係数との差

等があるが、竹内 (1986) は、各テストの共分散比の和が 1 となることとその加算性 (一つのテストがいくつかの下位テストから構成されている場合、下位テストの共分散比の和がそのテストの共分散比になること) から共分散比を推奨している。(これらの指標の特徴に関しては Kikuchi & Mayekawa, 1995 を参照。)

上記の指標は全受験者 (合格者 + 不合格者) を対象とした指標であるが、各テストの配点を微妙

に変えることにより合否判定が逆転するような、所謂ボーダーライン層 (平・池田, 1994) に着目した指標として、本稿で取り扱う合否入れ替わり率がある。

選抜に対するテスト j の効果を示す合否入れ替わり率 (θ_j) とは、合格者の中で、もしもテスト j の得点を除いて選抜を行った場合には不合格となる受験生の割合である。すなわち、

$$t_j = \sum_{k=1}^{j-1} \alpha_k x_k + \sum_{k=j+1}^m \alpha_k x_k \quad (2)$$

により受験生を順序づけた場合、 t による順位では上位 N_0 人に入っているが、 t_j による順位では上位 N_0 人に入らなかった者の N_0 に対する割合である。

この合否入れ替わり率に関しては、古くは、吉澤・藤本 (1975) にその提案があるが、直感的に分かりやすいこと、また、国立大学入学選抜研究連絡協議会におけるセミナーでの紹介 (池田, 1991、清水, 1993、1994、田栗, 1995) の影響もあり、近年様々な応用研究がなされてきている。(たとえば、鈴木・高野, 1992、高崎, 1992、清水, 1995、高野, 1995、豊田, 1995、等がある。) また、2 段階選抜試験において所謂足切りを行う場合にその適正足切り倍率を決定するという観点から行われた熊本他 (1988) や矢野他 (1990) の研究、および、テストの妥当性の指標として提案された Taylor & Russell (1939) や Thomas & Owen (1977) の指標とも関連している。

この様に広く用いられるようになってきた合否入れ替わり率の統計的性質を検討することは重要であろう。合否入れ替わり率の統計的性質に関しては、Kikuchi & Mayekawa (1995) で詳しく論じられているが、本稿の目的は、実際のデータから計算した合否入れ替わり率の安定性とその経年的な変化の傾向をブートストラップ法により推定した信頼区間を用いることによって検討することにある。本稿では、解析的に求めることの難しい合否入れ替わり率の信頼区間を、Kikuchi & Mayekawa (1995) で提案されている方法を利用して実際のいくつかの募集単位のデータから推定する。そして、その信頼区間を用いることによって、その募集単位の受験制度の変遷と合否入れ替わり率の経年的変化との関係について議論する。また、合否入れ替わ

り率を求める際の操作によって在学生に得意科目のラベル付けをおこない、学生個人および入試システムの診断目的に利用する方法についても検討する。

なお参考のため、母集団の得点分布として多変量正規分布を仮定した場合の合否入れ替わり率の母集団特性に関する議論を付録 A とした。

2 ブートストラップ法

母集団のパラメータを推定するための統計量の性質を調べるためには、仮定された確率分布から統計量の標本分布を解析的に求めることが必要である。たとえば、母集団確率分布として F を考え、その未知パラメータ θ を $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \dots, X_n)$ で定義される統計量で推定する場合を考える。ただし、 $\{X_1, X_2, \dots, X_n\}$ は F から無作為に抽出した n 個の標本である。通常、この推定量の性質を知るためには、統計量 $\hat{\theta}$ の分布を解析的に導き、推定の誤差を捉える指標となる推定量のバイアス、および標準誤差

$$\text{bias}(\hat{\theta}_n) = E_F(\hat{\theta}_n) - \theta \quad (3)$$

$$SE(\hat{\theta}_n) = \sqrt{E_F(\hat{\theta}_n - E_F(\hat{\theta}_n))^2} \quad (4)$$

を評価することになる。ただし、 E_F は確率分布 F に関する期待値である。

しかし、母集団の分布やパラメータの構成方法が複雑である場合には、解析的にその性質を求めることが困難である場合が多い。そのような解析的アプローチの困難な問題を解決するために、Efron(1979) によってブートストラップ法が提唱された。ブートストラップ法は、計算機による数値計算を利用しているが、近年の計算機の発達により、急速に普及し多くの論文が発表されている。(ブートストラップ法の現状に関しては、小西、1990、汪他、1992、久保川他、1993、にまとめられている。) このブートストラップ法の基本的アイデアは以下のようなものである。

まず、未知の母集団確率分布 F を経験分布 F_n によって推定する。ここで、経験分布 F_n は、 n 個の観測値(データ) $\{x_1, x_2, \dots, x_n\}$ の各点で、それぞれ確率 $1/n$ となる離散分布である。 F_n は、既知であるため、 F_n から無作為に標本を作成す

ることが可能である。この大きさ n の標本(ブートストラップ標本)を $\{x_1^*, x_2^*, \dots, x_n^*\}$ とし、 $\hat{\theta}_n^* = \hat{\theta}(x_1^*, x_2^*, \dots, x_n^*)$ を考えれば、この統計量の分布が $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \dots, X_n)$ の分布を近似していると考えられる。従って、モンテカルロ法を用いて $\hat{\theta}_n^*$ の分布を求めれば、それが θ の分布を推定していることになる。すなわち、(3)、(4) 式の F を F_n で置き換えたものが、ブートストラップ推定量である。

F_n はデータを基にして計算された経験分布であるため、上記の無作為標本を作成する過程は、データから重複を許して n 個の標本を無作為に抽出することに等しい。従って、この手続きはリサンプリングと呼ばれることがある。モンテカルロ法はこのデータからのリサンプリングで得られるブートストラップ標本ごとに統計量を計算することになる。すなわち、モンテカルロ法の反復数を B とすれば、各ブートストラップ標本(それぞれ大きさ n) から $\{\hat{\theta}_{(1)}^*, \hat{\theta}_{(2)}^*, \dots, \hat{\theta}_{(B)}^*\}$ を計算し、その分布(ブートストラップ分布)で統計量の分布を推定することになる。従って、バイアス、および標準誤差のブートストラップ推定量は、十分大きな B に対して

$$\text{bias}^*(\hat{\theta}_n) = \sum_{i=1}^B \hat{\theta}_{(i)}^* / B - \hat{\theta}$$

$$SE^*(\hat{\theta}_n) = \sqrt{\sum_{i=1}^B (\hat{\theta}_{(i)}^* - \sum_{i=1}^B \hat{\theta}_{(i)}^* / B)^2 / (B - 1)}$$

で近似できる。ただし、 $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ とする。

本稿で取り扱う合否入れ替わり率は、母集団に正規分布を仮定した場合でさえ、その性質を解析的に求めることは困難である。そこで、本稿では実際のデータを利用して、ブートストラップ法により、合否入れ替わり率の標本分布の推定ならびに区間推定を行うことにする。ただし、ブートストラップ法を用いるためには、データが独立同一分布に従う標本であることが仮定されている必要がある。本稿では、各募集単位、各日程の試験ごとにそれぞれの母集団があると仮定して、受験者はその中から無作為に抽出されたものとして、以下の議論を進める。

3 ブートストラップ法による 合否入れ替わり率の 標本分布の推定

3.1 実際に用いたブートストラップ法

本節では、実際のいくつかの募集単位データをを用いて、ブートストラップ法により、合否入れ替わり率の標本分布を推定する。多変量正規分布など、得点分布に何らかの理論分布を仮定している場合には、それに従う乱数に基づいてモンテカルロ法を行うことなどの方法によって、標本分布を推定することになる。(Kikuchi & Mayekawa, 1995 を参照。)しかし、実際のデータには正規分布を仮定する事に無理がある場合が多く、そのような場合には単純なモンテカルロ法は利用できない。従って、まずデータから母集団分布をノンパラメトリックに(多項分布として)推定し、その推定された分布から乱数を発生させることになる。これは、手持ちのデータからリサンプリングを繰り返すことに他ならない。ここでは、1次試験と2次試験という二つのテスト($m=2$)を用いる選抜に際して、2次試験の効果をみるための合否入れ替わり率の標本分布を推定するための手続きを示す。

1. $l=1$
2. N 人から N 人のサンプル(ブートストラップ標本 $(x^*_i, y^*_i), i=1, 2, \dots, N$)を復元抽出により取る。
3. サンプルを y^* の値、および、 $\alpha x^* + y^*$ の値により並べ替え、標本入れ替わり率(r_l^*)を、 $\alpha x^* + y^*$ の上位 N_0 に入ったサンプルのうち、 y^* の上位 N_0 に入らなかったものの割合として計算する。
4. $l=l+1$
5. $l \leq B$ ならば 2-4 を繰り返す。
6. $l > B$ ならば B 個の r_l^* の分布(ブートストラップ分布)を合否入れ替わり率の標本分布とみなす。

本稿では、上の繰り返し(レプリケーション)の数を $B=10000$ としてリサンプリングを行った。

また、各パラメーターは以下のようになる。

N = 募集単位の受験者数

N_0 = 合格者数

p = 受験倍率 = N/N_0

α = 1次試験の重みを1としたときの

$$= \left(\frac{\text{2次試験の配点}}{\text{1次試験の配点}} \right)^{\text{2次試験の相対重み}} \times \left(\frac{\text{2次試験の標準偏差}}{\text{1次試験の標準偏差}} \right)$$

x = 標準化された2次試験成績

y = 標準化された1次試験成績

3.2 結果

実験に使用したデータは、A大学の募集単位1～3についての1990年から1992年までの前期日程と後期日程のデータ、同じくA大学の募集単位4についての1986年から1994年までのデータ(1990年以降は前期日程と後期日程)、そしてB大学の一つの募集単位についての1981年から1994年のデータ(1990年以降は前期日程のみ)である。図1に、B大学、1989年の場合の標本入れ替わり率の度数分布を示す。この場合の標本入れ替わり率の分布はほぼ対称である。他の場合についても同様であった。

標準誤差 SE を B 個の r_l^* の標準偏差、バイアスを、 B 個の r_l^* の平均一すべてのデータから計算した入れ替わり率 r 、と定義し、平均、標準誤差、バイアスを求めた結果を、表1、表2に示す。

表1、表2において、それぞれのパラメーターの値が系統だっていないため、合否入れ替わり率の標本平均と標準誤差をパラメーターの値に関して比較することは難しい。しかし、受験者数 N に関しては、受験倍率が同程度であれば、受験者数 N が大きくなると、標準誤差は減少する傾向がある。このことから、受験者数 N が増えるほど標本入れ替わり率は安定するといえる。バイアスに関しては、特に偏った傾向はみられない。標準誤差は、ごく大きく括ると、受験者数が少ない場合(500人以下)には4%から8%、1000人を越えると2%以下である。

本節で求めた合否入れ替わり率の標本分布の利用の仕方の一例として、標準誤差を利用した信頼区間の例を示す。標本分布は、図1に見られるよ

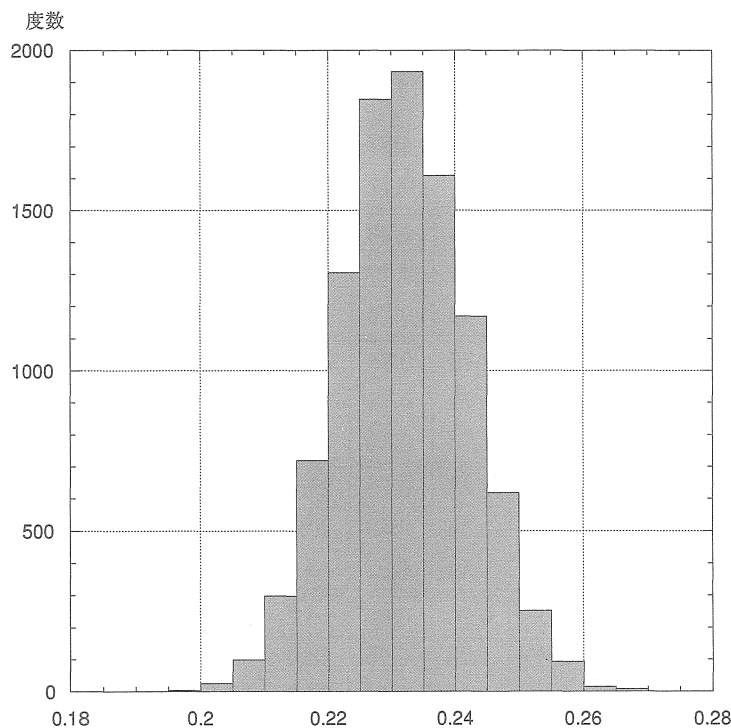


図1 ブートストラップ法による合否入れ替わり率の標本分布の例（B大学、1989年）

うに左右ほぼ対称で、正規分布で近似することができる。正規分布で近似した場合、たとえば95%信頼区間は、

標本値 $\pm 1.96 \times$ 標準誤差

となるが、その区間を表1、表2に示した。実際、Kikuchi & Mayekawa(1995)では、正規分布近似ではなく、パーセンタイルによって信頼区間を求めているが、それと比較してもほとんど同じ信頼区間となっている。(同文献で使用しているデータは、本稿のA大学の募集単位1～3に対応している。)

Kikuchi & Mayekawa(1995)はその年の標本値が前年度の信頼区間に含まれているか否かにより、選抜に関わる種々の条件（選抜方法、受験倍率、試験問題の質や量等）が本質的に変化したか否かを見ることを提案している。表1、表2の場合、A大学募集単位4の後期日程（受験者数が少ない上に倍率が高いため入れ替わり率が特に不安定と思われる）を除くと、ほぼ半数の募集単位において、何らかの変化が生じたことを示唆している。

4 考察

前節で計算した信頼区間について考察する。まず、倍率 α が約10倍と非常に高く1次試験と2次試験の相関係数 ρ が0.3以下と小さいA大学募集単位4後期日程の場合には、1991年から1994年までのすべての年度において前年の信頼区間に入った。(表1参照)これは、もとの標準誤差が8%前後と大きいと考えられる。このように倍率が高く合格者が少ない場合には、入れ替わり率の誤差が大きいと、信頼区間によって年度間の比較を行うことは適当でない。

入れ替わり率の経年変化を見るために、継続的にデータのあるA大学募集単位4とB大学のデータに着目してみる。ただし、分離・分割方式導入後の前期日程の試験は、従来の試験を継続しているものとみなす。表1より、A大学募集単位4においては、1987～1988、1990～1992年で入れ替わり率が前年の信頼区間に入らず、前節に述べた選抜に関わる条件のうちいずれかに関して何らかの変化があったものと思われる。条件を選抜方法の観

表 1: ブートストラップ法による合否入れ替わり率の標本分布 (A 大学)

募集単位	N	p	α	相関 係数	r	r_ℓ^* の 平均	$bias$	SE	95% 信頼区間	共分 散比
1990 年募集単位 1 前期	1807	2.93	1.88	0.73	31.4	31.6	0.2	1.5	28.6–34.3	67.4
1991 年募集単位 1 前期	1643	2.60	1.69	0.79	22.2	21.8	-0.4	1.3	19.6–24.8	64.3
1992 年募集単位 1 前期	1744	2.70	1.44	0.76	19.8	19.9	0.1	1.2	17.5–22.1	60.2
1990 年募集単位 2 前期	1405	3.66	1.73	0.71	28.9	28.6	-0.3	1.8	25.4–32.4	65.4
1991 年募集単位 2 前期	1063	2.79	1.79	0.74	27.0	26.3	-0.7	1.7	23.7–30.4	66.0
1992 年募集単位 2 前期	1126	2.91	1.73	0.75	25.6	25.7	0.1	1.7	22.2–28.9	65.2
1990 年募集単位 3 前期	574	3.24	1.57	0.78	29.4	29.3	-0.1	2.7	24.1–34.6	62.4
1991 年募集単位 3 前期	513	2.88	1.64	0.80	20.8	21.3	0.5	2.3	16.3–25.3	63.5
1992 年募集単位 3 前期	463	2.37	1.54	0.82	15.4	15.7	0.3	2.0	11.4–19.3	61.7
1986 年募集単位 4	507	2.07	0.76	0.66	14.7	15.0	0.3	1.7	11.3–18.1	42.0
1987 年募集単位 4	772	2.66	1.23	0.75	18.6	18.4	-0.3	1.8	15.0–22.2	55.8
1988 年募集単位 4	637	3.32	1.85	0.64	34.9	34.0	-0.9	2.7	29.5–40.2	67.9
1989 年募集単位 4	1023	3.89	1.31	0.73	29.7	29.6	-0.1	2.2	25.4–33.9	57.6
1990 年募集単位 4 前期	920	4.26	1.82	0.65	35.2	36.4	1.2	2.8	29.6–40.7	67.3
1991 年募集単位 4 前期	791	3.34	1.38	0.71	28.7	28.0	-0.7	2.3	24.2–33.2	59.2
1992 年募集単位 4 前期	813	3.44	1.29	0.76	20.8	21.8	1.1	2.1	16.6–24.9	57.1
1993 年募集単位 4 前期	701	2.84	1.51	0.78	23.9	24.0	0.1	2.1	19.7–28.0	61.4
1994 年募集単位 4 前期	573	2.43	1.46	0.77	22.9	22.8	-0.1	2.0	18.9–26.8	60.4
1990 年募集単位 1 後期	621	5.31	1.51	0.64	48.7	48.2	-0.5	3.8	41.3–56.2	62.3
1991 年募集単位 1 後期	833	6.66	1.89	0.66	40.0	39.5	-0.5	3.6	33.0–47.0	68.2
1992 年募集単位 1 後期	674	4.96	1.51	0.63	38.2	37.8	-0.4	3.3	31.9–44.6	62.3
1990 年募集単位 2 後期	448	5.82	1.31	0.56	44.2	43.7	-0.5	4.5	35.3–53.0	58.6
1991 年募集単位 2 後期	471	6.04	1.33	0.59	38.5	38.8	0.4	5.0	28.6–48.3	58.8
1992 年募集単位 2 後期	438	5.41	1.24	0.59	28.4	30.5	2.1	4.5	19.7–37.1	56.8
1990 年募集単位 3 後期	180	5.14	1.18	0.65	34.3	35.0	0.8	7.3	20.0–48.6	54.9
1991 年募集単位 3 後期	240	6.67	1.12	0.66	41.7	41.9	0.2	6.9	28.2–55.1	53.4
1992 年募集単位 3 後期	182	3.96	1.00	0.64	19.6	19.5	-0.1	5.6	8.6–30.6	50.0
1990 年募集単位 4 後期	247	9.88	0.84	0.29	64.0	64.0	-0.0	8.5	47.3–80.7	43.5
1991 年募集単位 4 後期	335	10.81	0.97	0.23	61.3	65.0	3.7	8.6	44.5–78.1	48.9
1992 年募集単位 4 後期	409	12.78	1.20	0.28	62.5	60.2	-2.3	7.5	47.9–77.1	57.0
1993 年募集単位 4 後期	371	11.59	0.84	0.28	65.6	63.6	-2.0	8.5	48.9–82.3	43.0
1994 年募集単位 4 後期	310	9.69	1.08	0.28	56.2	55.9	-0.3	7.6	41.4–71.1	52.8

$bias = B$ 個の r_ℓ^* の平均 $- r$ (すべてのデータを用いて計算した値)

第 6 列から第 11 列までは、すべて 100 倍した値

95% 信頼区間 $= r \pm 1.96 \times SE$

表 2: ブートストラップ法による合否入れ替わり率の標本分布 (B 大学)

募集単位	N	p	α	相関係数	r	r_{ℓ}^* の平均	$bias$	SE	95% 信頼区間	共分散比
1981 年	2348	3.36	0.93	0.66	25.2	25.6	0.4	1.3	22.6–27.7	47.7
1982 年	2343	3.38	0.83	0.65	25.4	25.5	0.1	1.3	22.8–28.0	44.2
1983 年	2285	3.35	0.75	0.66	26.0	26.2	0.2	1.3	23.3–28.6	41.5
1984 年	1900	2.76	0.78	0.68	24.7	24.8	0.1	1.3	22.1–27.3	42.8
1985 年	1990	2.95	0.73	0.68	24.9	24.7	-0.2	1.3	22.4–27.4	40.9
1986 年	1783	2.40	0.66	0.65	19.3	19.4	0.1	1.1	17.0–21.5	37.8
1987 年	3471	3.31	1.68	0.51	30.0	30.1	0.1	1.1	27.8–32.1	66.6
1988 年	2730	2.75	1.06	0.68	22.6	22.4	-0.1	1.0	20.6–24.5	51.8
1989 年	2632	2.66	1.23	0.67	23.0	23.2	0.1	1.0	21.0–25.0	56.1
1990 年前期	3074	4.28	1.17	0.63	33.4	33.3	-0.1	1.4	30.7–36.0	54.8
1991 年前期	2133	3.02	1.16	0.67	24.9	24.9	0.0	1.2	22.5–27.3	54.5
1992 年前期	1789	2.40	1.20	0.63	21.4	21.2	-0.2	1.1	19.2–23.6	55.7
1993 年前期	1867	2.55	1.17	0.64	23.7	23.5	-0.1	1.2	21.3–26.0	54.7
1994 年前期	2076	2.79	1.93	0.64	27.3	27.3	0.0	1.3	24.8–29.8	69.0

$bias = B$ 個の r_{ℓ}^* の平均 $- r$ (すべてのデータを用いて計算した値)

第 6 列から第 11 列までは、すべて 100 倍した値

95% 信頼区間 $= r \pm 1.96 \times SE$

点に絞れば、A 大学募集単位 4 では、1987、1990 年に 1 次試験、2 次試験の配点比の変更や分離・分割方式の導入を行っている。つまり、選抜において変化のあった年から数年は、入れ替わり率は経年変化という点からみると安定しないが、数年後には安定すると考えられる。B 大学においては表 2 より、1986～1988、1990～1994 年で入れ替わり率が前年の信頼区間に入らず、選抜に関して何らかの変化があったものと思われる。B 大学では、1987、1990、1994 年に 1 次試験、2 次試験の配点比の変更や分離・分割方式の導入を行っている。1993 年は、入れ替わり率の値が信頼区間に非常に近い点を考慮すると、1986 年を例外として、A 大学募集単位 4 の場合と同様の解釈ができる。特に B 大学のデータをみると、選抜の制度自体が安定している 1981～1986 年の間では、ほとんどの年の入れ替わり率が前年の信頼区間に入っている。このことから、受験者数の多い募集単位では、選抜制度に大きな変化がなければ、入れ替わり率は経年変化をみても安定しているといえよう。

次に、合否入れ替わり率の標本分布全般に関して、本稿において用いたデータの範囲で検討する。ブートストラップ法により求めた標本分布は、ほぼ左右対称であった。また、バイアスに関しては、特に偏った傾向は見られなかった。ただし、バイアスの値そのものに関しては、比較的小さい値であるため特にバイアスに関して考慮する必要はないものと思われる。標準誤差に関しては、受験者数 N が大きくなると、標準誤差は小さくなっている。このことから、合否入れ替わり率は、受験者数 N が大きくなると安定するといえる。なお、標準誤差の値そのものに関しては全体を通して、比較的小さい値となった。よって、合否入れ替わり率という指標は比較的安定した指標であるといえる。

以上の結果、合否入れ替わり率という指標は、バイアスも小さく、標準誤差の点からも安定した指標であるといえる。ただし、値そのものの利用方法については、今後も検討が必要である。

本論文においては選抜試験に含まれるテストの

選抜への貢献度を示す測度としての合否入れ替わり率に着目し、その信頼区間の利用方法を検討してきたが、合否入れ替わり率の利用方法は上記の問題に限定されているわけではない。第1節に述べたように、あるテストの合否入れ替わり率とは、そのテストが選抜に用いられなかったならば入学しなかった受験生の割合であり、その計算に際しては、 N 人の全受験生にラベルを付与する必要がある。従って、選抜試験に合格し在学中の N_0 人の合格者にも当然のことながらそのラベルは付与されており、在学生の一人一人が、どのテストが無ければ合格しなかった者であるか（あるいはどのテストを除いても合格する者であるか）がわかっている。このことは、合否入れ替わり率の計算に用いられた手続きを利用することにより、在学生の得意科目を同定できることを意味しており、得意科目から見た学生のプロフィールを構成することができる。たとえば、大学入試センター試験5教科と数学と英語の個別学力試験および小論文からなる選抜試験に合格したA君が、小論文が用いられなかった場合には不合格になるとする。このことは、A君が小論文形式の試験を得意としていることを意味するが、また、大学側の立場からは、もしも小論文試験を課さなかった場合にはA君を採れなかったことを意味する。このような情報が全ての在学生に関して利用可能であれば、在学生に対するきめ細かい指導を行うことが可能になると共に、選抜試験科目の変更の試みに際しても、たとえば小論文を廃止すればA君のような学生を採れなくなる、という形できわめて有効な情報をもたらすと考えられる。

このように、選抜試験の結果を単に合格者を選抜することのみならず多目的に活用し、そこに含まれる情報をもとに合格者にラベルを貼ることによりさまざまな目的のために利用していこうという考え方は、岩坪他(1988)が提案しているいわゆる学力型の研究(鈴木、1990、山田、1990、1991、池田他、1992を参照)に始まるものであるが、ここに述べた方法もその流れを汲むものであるということが出来る。本節で述べた方法は技術的にはまだ完成されたものではないが、(たとえば、上記のA君の例の場合、1教科ごとに合否の入れ替わりを判定しているが、同時に2教科以上を選抜から取り除いた場合の合否の入れ替わりを調べるこ

とも可能である)合否入れ替わりの考え方を取り入れることにより、在学生のきめ細かな診断のみならず、選抜試験科目の変更等の入試システムの変更の試みに際しても有用な診断情報を提供する可能性を示している。

付録

A 多変量正規分布を仮定した場合の合否入れ替わり率

本節では、母集団においてテスト得点が正規分布することを仮定した場合に、各テストの配点と受験倍率が与えられた時に、合否入れ替わり率を算出する方法とその標本分布に関して述べる。(詳しくは、Kikuchi & Mayekawa、1995を参照。)

m 個のテスト得点の分布が、平均 $\mu = \{\mu_j\}$ 、分散共分散行列 $\Sigma = \{\sigma_{jk}\}$ の多変量正規分布に従う場合、母集団の特性としての合否入れ替わり率(母入れ替わり率 θ)は以下のように表すことができる。

x を合否入れ替わり率を求めたいテスト(テスト j)の得点を標準化したものとし、 y を選抜に用いられるすべてのテストからテスト j を除いたものの重み付き合計点 t_j を標準化したものとする。すると、 $\{x, y\}$ の分布は、相関係数

$$\rho = \sum_{k \neq j} \alpha_k \sigma_{jk} / \left(\sqrt{\sigma_{jj}} \sqrt{\sum_{k, \ell \neq j} \alpha_k \alpha_\ell \sigma_{k\ell}} \right)$$

の標準2変量正規分布となる。また、 α を x にかかる相対的重み ($\alpha > 0$)

$$\alpha = \alpha_j \sqrt{\sigma_{jj}} / \sqrt{\sum_{k, \ell \neq j} \alpha_k \alpha_\ell \sigma_{k\ell}} \quad (5)$$

とすると、母集団の特性としての合否入れ替わり率(母入れ替わり率 θ)は、

$$\begin{aligned} \theta &= p \times \text{Prob}(y < d \ \& \ \alpha x + y \geq c) \\ &= p \times \int_q^\infty \int_{c-\alpha x}^d f(x, y) dy dx \end{aligned} \quad (6)$$

となる。ただし、 $f(x, y)$ は相関係数が ρ の標準2変量正規分布の密度関数を表し、 c および d は、受

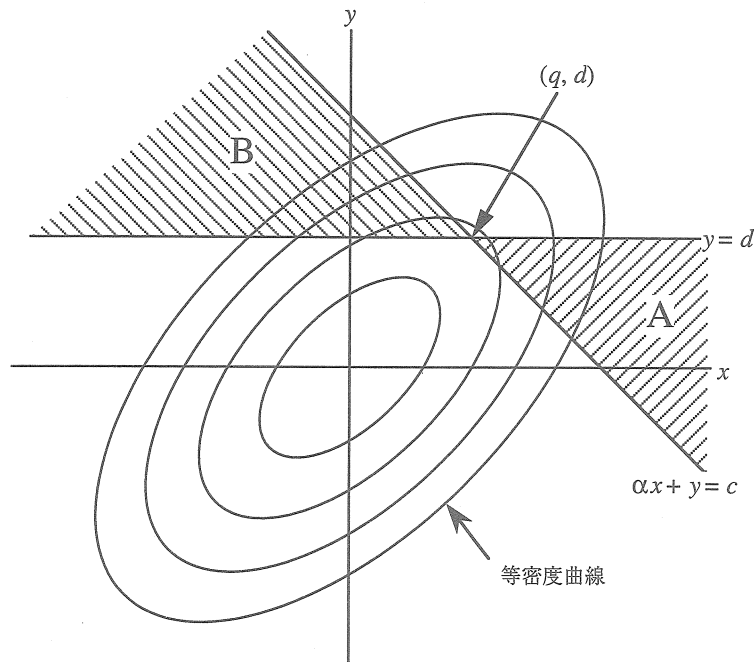


図 2 テスト x の選抜効果を表す合否入れ替わり率

験倍率 p (受験者数 / 合格者数) を与えたときに、

$$\text{Prob}(\alpha x + y \geq c) = \text{Prob}(y \geq d) = 1/p$$

により決まる定数、また、 q は直線 $\alpha x + y = c$ と $y = d$ の交点の x 座標

$$q = (c - d)/\alpha \quad (7)$$

である。すなわち、合否入れ替わり率は図 2 に示された斜線 A の部分上の体積に倍率を掛けたものとして示される。なお、図 2 に示された斜線 B の部分の体積は、斜線 A の部分の体積と等しくなるが、実際のデータでは、同点者がいるため厳密に同じになるとは限らない。この体積 A の計算は 2 変量正規分布の密度関数の 2 変数に関する積分として書き表されるため、何らかの数値計算が必要となるが、基本的には 1 変数の数値積分に帰着させることができる。(矢野他、1990、または、山内、1972、付録 E を参照。) このようにして、Kikuchi & Mayekawa(1995) では、母入れ替わり率およびその標本平均が、相対重み α 、および倍率 p に関して単調増加であり、相関係数 ρ 、に關

して単調減少となることを示している。バイアスを、標本平均-母入れ替わり率 ((6) 式) とすると、全体的な傾向としては、若干正のバイアスを持つことも示している。

この特性を、各テストの選抜に対する重要度を知るための指標として竹内 (1986) や柳井 (1994) が論じている共分散比と比較してみる。重み α に関しては、共分散比、合否入れ替わり率ともに、単調増加する。しかし、相関係数 ρ に関しては、 $\alpha < 1.0$ の時、共分散比は単調増加するのに対して、合否入れ替わり率は単調減少する。また、共分散比は倍率 p に依存しないのに対して、合否入れ替わり率は倍率 p に依存する。このことから、合否入れ替わり率は共分散比に代表される既存の指標と比較して、各テストの選抜に対する重要度という点に関して別の性質を計っている指標であるといえる。この理由は、共分散比が受験者全体の性質を表している指標であるのに対して、合否入れ替わり率が受験者全体の中の特に合否判定ライン付近の受験者に着目した指標であることに起因している。

また、正規分布を仮定して、各パラメーターの値として実際のデータの倍率 p 、重み α および相関係数 ρ を与えて、合否入れ替わり率を計算した値と実際のデータから計算した値を比較してみると、後者の方が大きくなる場合が多い。(Kikuchi & Mayekawa, 1995) 実際のデータの合否入れ替わり率は、得点の分布が合否判定線 $\alpha x + y = c$ 付近が密となるように偏っているため、正規分布を仮定して計算した合否入れ替わり率よりも大きくなるものと考えられる。実際のデータの合否入れ替わり率の推定値として、同じパラメーターの値を持つものとして正規分布を仮定して計算した値は、ある程度の目安とはなるが、そのまま用いることは妥当ではない。

謝辞

電気通信大学の熊本芳朗教授、北海道大学の長谷部清教授には、入試研究の立場から多くのご教示を受けた。記して感謝申し上げる。また、第4節に述べた診断的指標のアイデアは大学入試センターの清水留三郎教授によるものである。

参考文献

- 池田輝政 (1991) 選抜方法の変更とその効果 国立大学入学者選抜研究連絡協議会セミナー資料 (第3章)、p33-42.
- 池田輝政・山村滋・岩田弘三 (1992) 受験時点と在学中の5教科成績プロフィールの特徴ーある高等学校についての事例分析ー 大学入試センター研究紀要 No. 21, p59-79.
- 岩坪秀一・池田輝政・岩田弘三 (1988) 大学が重視する入試教科と受験生の学力特性ー共通第1次学力試験の5教科得点を基礎にしてー 大学入試センター研究紀要 No. 17, p101-144.
- 久保川達也、江口真透、竹村彰通、小西貞則 (1993) 統計的推測理論の現状 日本統計学会誌第22巻、p257-312.
- 熊本芳朗・石塚智一・山田文康 (1988) 2変量正規分布の理論による適正足切り倍率のシミュレーション研究 大学入試フォーラム No.10、p182-194.
- 小西貞則 (1990) ブートストラップ法と信頼区間の構成 応用統計学 19巻、p137-162.
- 清水留三郎 (1993) 試験科目・問題の効果の評価 国立大学入学者選抜研究連絡協議会セミナー資料 (第3章)、p24-36.
- 清水留三郎 (1994) 試験科目・問題の効果の評価 国立大学入学者選抜研究連絡協議会セミナー資料 入試データの基礎的分析手法 (第2章)、p11-19.
- 清水留三郎 (1995) 入学者選抜における試験の効果の評価ー合否入れ替り率等を中心にー (第1報) 大学入試研究ジャーナル 第5号、p1-4.
- 鈴木規夫 (1990) 国公立大学入学者選抜試験の効果に関する実証的研究 大学入試センター研究紀要 No. 19, p47-89.
- 鈴木規夫・高野文彦 (1992) 第2次試験の効果の分析 大学入試研究ジャーナル 第2号、p14-20.
- 平直樹・池田輝政 (1994) 入試科目の効果に関する新しい評価法 大学入試研究ジャーナル 第4号、p40-44.
- 高野文彦 (1995) 入れ替り率、全国データの分析 大学入試研究ジャーナル 第5号、p114-118.
- 高崎禎夫 (1992) 小論文よ自信をもてー選抜貢献率と入れ替り率から見た各教科の働きー 大学入試研究ジャーナル 第2号、p48-52.
- 田栗正章 (1995) テスト項目の評価、合否入れ替り率 国立大学入学者選抜研究連絡協議会セミナー資料 (第1章)、p.1-15.
- 竹内啓 (1986) 入試科目の事後の重みの評価について 国立大学入学者選抜研究連絡協議会研究報告書 第7号、p500-501.
- 豊田秀樹 (1995) 入研協共同研究の合否入れ替り率に関する日程別の特徴について 大学入試研究ジャーナル 第5号、p119-123.

- 矢野一幸・大内俊二・田栗正章 (1990) 大学入試における予備選抜倍率についての検討 行動計量学 第 17 巻、第 2 号, p25-33.
- 山内二郎 (編) (1972) 統計数値表 JSA-1972. 日本規格協会, p79-82.
- 山田文康 (1990) 共通第 1 次学力試験の 5 教科得点に基づく学力型の分析 大学入試センター研究紀要 No. 19, p1-45.
- 山田文康 (1991) 共通第 1 次学力試験の 5 教科得点に基づく学力型の分析 大学入試研究ジャーナル 創刊号, p42-47.
- 柳井晴夫 (1994) 多変量データ解析法 ―理論と応用― 朝倉書店
- 吉澤正・藤本洋子 (1975) 入試における選抜基準と合格者の変動 山梨大学工学部研究報告 No. 26, p18-24.
- 汪金芳・大内俊二・景平・田栗正章 (1992) ブートストラップ法 ―最近までの発展と今後の展望― 行動計量学第 19 巻、p50-81.
- Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, p1-26.
- Kikuchi, K. & Mayekawa, S. (1995) On the Sampling Distribution of Swap-Rate. *Behaviormetrika*, 22, No. 2, p185-204.
- Taylor, H. C. & Russell, J. T. (1939) The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 23, p565-578.
- Thomas, J. G. & Owen, D. B. (1977) Improving the use of educational tests as selection tools. *Journal of Educational Statistics*, 2, No. 1, p55-77.

1995年11月14日 受付
1996年 2月 1日 最終修正

Interval estimation of swap-rate using bootstrap method

Shin-ichi MAYEKAWA*

Kenichi KIKUCHI**

Abstract

When an admission is based on a composite score of several subtests, and N_0 students are to be selected out of N applicants, the swap-rate can be used to assess the contribution of each subtest to the selection. The swap-rate of the j -th subtest is defined as the proportion, to N_0 , of the number of students who would fail to pass the admission if the j -th subtest was not used for the selection.

In this paper we estimated the standard error of the swap-rate using bootstrap method and showed that, for most of the datasets, the sample swap-rates are reasonably stable. We also discussed the relationship between the confidence interval of the swap-rate and the annual variations of applicants.

Keywords: swap-rate, admission, entrance examination, bootstrap method

*Scholastic Aptitude Section, Research Division, The National Center for University Entrance Examinations.
2-19-23 Komaba, Meguro-ku, Tokyo 153 Japan. Telefax: 81-3-5478-1297, e-mail: mayekawa@rd.dnc.ac.jp

**Information Processing Section, Research Division, The National Center for University Entrance Examinations.