ランダム回答法における繰返しの影響と有限母集団修正について

逆瀬川　浩　孝
高　橋　宏　一

(1974年7月　受付)

Effects of Repetition and Finite Population Corrections in Randomized Response Models

Hirotaka Sakasegawa, Koiti Takahasi

In this paper, we first consider the effects of two or more trials per respondent in the Warner randomized response model with respect to both the precision of estimation and the confidentiality of respondents. Secondly, we give the formulas for the finite population corrections in the Warner randomized response model and the Simmons unrelated question randomized response model.

The Institute of Statistical Mathematics

§1. はじめに

面接調査において回答者が自分の回答を調査員に知られたくないような質問項目の調査法として，Warner (1965) はランダム回答法 (randomized response technique) を提案した．母集団の各人はAかBのいずれかのクラスに属するものとし，母集団におけるAの比率の推定を目的とする．実際にAに属している人は自分がAであることを知られたくないといった場合に（実際にBに属している人は自分がBであることを知られてもかまわない），直接的に「あなたはAですか？　それともBですか？」と質問したのではAに属している人は回答を拒否したり，あるいは偽ってBと答えたりすることが考えられ，通常の標本比率による推定には，かなりの偏りが入ってくるであろう．こうした偏りを避けるために，たとえば「あなたはAですか？」という質問の書かれたカードと，「あなたはBですか？」と書かれたカードを $p:1-p$ の割合でまぜたものを用意する．回答者にその1組のカードを渡し，よくきってもらってランダムに1枚をぬきとってもらう．そのカードにどちらの質問が書いてあるかを，回答者は調査員に知らせずに，たゞその質問に対して“yes”であるか“no”であるかだけを回答する．答が“yes”であっても調査員にはどちらの質問に対して“yes”なのかわからない．“no”という答についても同様である．しかしサンプル全体についての“yes”あるいは“no”の比率から，母集団におけるAの比率，π_A と書くことにする，の一致不偏推定量を得ることができる（たゞし，$p\neq\frac{1}{2}$のとき）．これが Warner の提案したランダム回答法である（以下，このモデルを Warner のモデルとよぶ）．もちろんいずれかの質問を確率的に選んでもらうための道具としてはカードに限らない．

このランダム回答法は，Campbell and Joiner (1973) が簡潔に紹介しているように，いろいろな方向に改善，拡張されている．その一つは Horvitz et al. (1967) によるものである．Warner の方法で，「あなたはBですか？」という質問は「あなたは $\bar{\text{A}}$（Aでない）ですか？」

ということであり，結局，「あなたはAですか？」という質問と「あなたは $\bar{A}$ ですか？」という質問を確率的に選んでもらって回答してもらったことになる．これに対して Horvitz et al. (1967) は「あなたはAですか？」という質問と，この質問とは直接関係のない「あなたはCですか？」という質問を利用する．すなわち，この2つの質問を確率的に選んでもらって，選ばれた質問に "yes" か "no" で答えてもらう．この場合も回答者はどちらの質問が選ばれたかは調査員に知らせない．母集団におけるCの比率が既知のときはその必要はないが，未知のときにはAの母集団比率を推定するためには2組のサンプルを用いて，第1組と第2組の間では2問の選択確率をちがえる必要がある．以下，このモデルを Simmons のモデルとよぶ．

この論文では，こうしたランダム回答法について2つの注意を述べる．一つは各回答者に同じ質問について2回（以上）繰返して試行してもらう場合についての注意である．Horvitz et al. (1967) は上記の無関係な質問を利用するモデルについて，各回答者にランダムに質問を選ぶことを独立に2回してもらい，それぞれについて選ばれた質問に答えてもらう場合も考察している．2回繰返しをする理由については，はっきりした形の叙述はなされていないが，たとえばAに属する人がAに関する質問にあたったときに，どちらに答えているかが調査員にはわからないという事情があるにもかゝわらず，さらにBであると偽る可能性を考慮したようなモデルを用いると，モデルに含まれているパラメターを推定するためには2回の繰返しが必要になるといったことが主な理由のようである (Gould et al. (1969) 参照)．一方，回答者が正しい方法でカードをぬき出し，そのカードに書かれている質問に正しく回答してくれる場合には，2回繰返しということは，1回の場合にくらべて，π_A の推定の精度をかなり向上させることになる．この点で問題が生じる．ランダム回答法を用いることによって，直接的に「あなたはAですか？」と質問したときに生じる回答拒否や偽りの回答による偏りを防ぐことができると期待される理由は，p：$1-p$ の割合で2つの質問のいずれかをランダムに選び，どちらの質問かを調査員に知られることなしに回答するのであるから，回答者はランダム回答法においては正直に答えてくれるであろうというところにある．したがって p が0や1に近い場合には，どちらか一方の質問だけが選ばれる機会が多くなりランダム回答法の本来の意義が失われることになる．逆に p が $1/2$ に近いほど回答者の協力は得られやすいことになる．この観点から各回答者に2回（以上）繰返しをしてもらうことの影響を考えてみると，後述するように1回の試行では p：$1-p$ の割合で質問が選ばれているときでも，2回（以上）繰返したことにより，実質的には p が $1/2$ からはなれることになることがわかる．したがって，2回（以上）繰返しをするときには，この事情を回答者に理解してもらうことが調査をする側の正しい態度であろう．単に方法を説明しながら1度回答してもらい，説明がすんだところで，「では，方法がわかっていたゞけたでしょうから，もう一度お願いします」という形式で2回繰返しをしてもらうのでは，上で述べた事情に回答者は気がつかない可能性がある．この論文で述べる第1の注意というのは，2回（以上）の繰返しの推定の精度にもたらす影響と，回答者の協力の得やすさにもたらす影響の双方を調べ，その結果を考察することである．

第2の注意はランダム回答法における有限母集団修正のことである．ランダム回答法は本来，すでに述べたように，直接質問したのでは回答拒否や偽りの回答による偏りが避けられないような質問項目の調査法として提案されたものである．しかし，Campbell and Joiner (1973) が述べているように，ランダム回答法は統計や確率の教材として適当なものと考えられる．2項分布，最尤推定法，信頼区間といった基本的な事項を具体的な問題の中でとり上げることができること，あるいは適当な道具を使ってランダムな変動を体得できることなどが主な理由である．しかし同時に，どちらかといえば，まだ実際に種々の統計的手法の利用にせまられていない段階で学ぶことからくるかもしれない統計学に対する関心の低さを，ランダム回答法を自

分達の身近な問題に応用することによってひき上げるという利点も十分考えられる．上記の文献中にもあるように，学級や学校の寮や，あるいは大学といった単位でランダム回答法を実施してみるのもよいかもしれない．こうした場合に，教育的見地からも実際的見地からも母集団の規定をはっきりさせておくことが重要である．たとえば学級をもっと大きな母集団からサンプルと考えるのか，それとも学級そのものを母集団と考えるのかということである．目的が明きらかに自分達の学級でどうなっているかを調べることである場合には，学級を有限母集団として推論すべきである．これまでのランダム回答法に関する論文では無限母集団の場合を扱っている．母集団からのサンプリングによる変動と，サンプルがランダムに質問を選択することによる変動にわけて，有限母集団修正がどのような形で入ってくるかを基本的なモデルについて考察する．

§2.　ランダム回答法における繰返しの影響

この節では簡単のために母集団は無限母集団とする．母集団はクラスAとクラスBに分割されており，クラスAの比率は π_A，クラスBの比率は $\pi_B=1-\pi_A$ とする．π_A の推定が目的である．母集団から大きさ n のサンプルをとる．「あなたはAですか？」という質問と「あなたはBですか？」という質問を用意する．サンプル中の各人は，カード，サイコロ，ルーレット，乱数表などを使用して，2つの質問の中から1つをランダムに選び出す．第1の質問を選ぶ確率および第2の質問を選ぶ確率は全員に共通で，それぞれ p, $q=1-p$ とする．各人は選び出された質問に対して，“yes” か “no” かを答える．どちらの質問が選び出されているかは調査員には知らせないのであるから，推定は n 人中の “yes” と答えた人数 y のみにもとずいて行われる．この場合の π_A の最尤推定量 $\hat{\pi}_A$ は

$$\hat{\pi}_A=\frac{\dfrac{y}{n}-q}{p-q},\ \left(\text{ただし}\ p\neq\frac{1}{2}\right) \tag{1}$$

であり

$$E\ (\hat{\pi}_A)=\pi_A, \tag{2}$$

$$\mathrm{Var}\ (\hat{\pi}_A)=\frac{1}{n}\left\{\pi_A\ (1-\pi_A)+\frac{pq}{(p-q)^2}\right\} \tag{3}$$

である（Warner (1965)）．$\hat{\pi}_A$ は最小分散不偏推定量である．

さて，各人は質問をランダムに選び，それに対し回答することを独立に t 回繰返すものとする．一般的に各回の第1の質問を選ぶ確率は異なってもよいものとし，第 i 回目の第1の質問を選ぶ確率を p_i，第2の質問を選ぶ確率を $q_i=1-p_i$ とする．“yes” を 1，“no” を 0 で表わすことにする．$K=(\delta_1\ (K),\ \delta_2\ (K),\ \cdots,\ \delta_t\ (K))$，$\delta_i\ (K)=0$，あるいは 1，とする．第 i 回の回答を表わす確率変数を X_i として，$X\equiv(X_1,\ X_2,\ \cdots,\ X_t)=K$ となる確率 $\lambda\ (K)$ は

$$\lambda\ (K)=\pi_A\prod_{i=1}^{t}p_i^{\delta_i(K)}\,q_i^{1-\delta_i(K)}+(1-\pi_A)\prod_{i=1}^{t}p_i^{1-\delta_i(K)}\,q_i^{\delta_i(K)} \tag{4}$$

となる．これを簡単に

$$\lambda\ (K)=C\ (K)\ \pi_A+B\ (K) \tag{5}$$

と表わす．すなわち

$$C\ (K)=\prod_{i=1}^{t}p_i^{\delta_i(K)}\,q_i^{1-\delta_i(K)}-\prod_{i=1}^{t}p_i^{1-\delta_i(K)}\,q_i^{\delta_i(K)}, \tag{6}$$

$$B\ (K)=\prod_{i=1}^{t}p_i^{1-\delta_i(K)}\,q_i^{\delta_i(K)} \tag{7}$$

とおいた．大きさ n のサンプル中の回答結果 K の度数を $n\ (K)$ とする．$D\ (K)=B\ (K)/C\ (K)$ とおくと π_A に対する最尤方程式は

$$\sum_K \frac{n\ (K)}{\pi_A+D\ (K)}=0 \tag{8}$$

となることが容易にわかる．また，これから求められる最尤推定量 $\hat{\pi}_A$ の漸近分散は

$$\frac{1}{n}\left(\frac{1}{\sum_K C^2\ (K)\ /\lambda\ (K)}\right) \tag{9}$$

となることも容易にわかる．

この一般式を，$t=2$，$p_1=p_2=p$，$q_1=q_2=q$ の場合に適用すると (8) より

$$\hat{\pi}_A=\frac{\dfrac{n_{11}}{n_{11}+n_{00}}-\dfrac{q^2}{p^2+q^2}}{\dfrac{p^2}{p^2+q^2}-\dfrac{q^2}{p^2+q^2}} \tag{10}$$

が得られる．たゞし，$n_{11}=n\ (1,\ 1)$，$n_{00}=n\ (0,\ 0)$ である．$n_{11}+n_{00}=0$ のときは推定値は求まらない．また，(9) より $\hat{\pi}_A$ の漸近分散は

$$\frac{1}{n\ (p^2+q^2)}\left\{\pi_A\ (1-\pi_A)+\frac{p^2q^2}{(p-q)^2}\right\} \tag{11}$$

となる．この漸近分散は次の意味で (10) で与えられている最尤推定量の分散の近似式にもなっている．すなわち，厳密には，$n_{00}+n_{11}$ は試行数 n，生起確率 p^2+q^2 の2項分布に従うが，この分布の0を打ち切った正の2項分布に従うものとし，かつ $n_{00}+n_{11}$ の逆数の期待値を $\dfrac{1}{n\ (p^2+q^2)}$ で近似したことに対応している．

推定量 (10) は，$1\leqq l\leqq n$ なる任意の l に対して，$n_{11}+n_{00}=l$ という条件のもとで π_A の不偏推定量であることがわかる．また (11) はつねに (3) より小さいことがわかる．さらに (11) は，$\pi_A<\dfrac{1}{2}-\dfrac{\sqrt{3}}{6}\fallingdotseq 0.2113$（あるいは $\pi_A>\dfrac{1}{2}+\dfrac{\sqrt{3}}{6}\fallingdotseq 0.7887$）のとき，

$$\frac{1}{2}\left(1-\sqrt{1-4\frac{\pi_A\ (1-\pi_A)}{1-2\pi_A\ (1-\pi_A)}}\right)<p<\frac{1}{2}\left(1+\sqrt{1-4\frac{\pi_A\ (1-\pi_A)}{1-2\pi_A\ (1-\pi_A)}}\right) \tag{12}$$

の範囲の p が使用されている場合には (3) の n を $2n$ でおきかえたものより小さいことがわかる．

このように Warner のモデルで各回答者に2回の試行（1回目と2回目の質問の選択確率が同じ）をしてもらう場合，1回だけのときにくらべ推定量の分散が小さくなるのは当然として，π_A が0あるいは1に近く，p が0.5に近いときには，繰返しなしでサンプルの大きさを2倍にしたときにくらべても推定量の分散は小さい．さらに調査の費用のことまで考えに入れると，サンプルの大きさをふやすよりも，繰返しをする方が得策であるようにみえるかもしれない．しかし §1 で述べたように回答者の協力の得られやすさという点から繰返しの影響を調べる必要がある．

各人の回答は2つの質問のいずれに対するものか確率的な意味でしかわからないというところにランダム回答法の特徴がある．回答者にとって，自分が "yes" と答えようが "no" と答えようが，その答をしたということによって自分がAであるかBであるかを識別されないというところに調査に協力する理由を見出すことができるのである．そこで，ある回答をしたとい

う条件のもとで，その人がAまたはBに属するという条件付確率を調べる必要が生じる．繰返しのない通常の Warner のモデルでは，

$$Pr\{A|\text{"yes"}\}=\frac{\pi_A p}{\pi_A p+(1-\pi_A)q} \tag{13}$$

$$Pr\{A|\text{"no"}\}=\frac{\pi_A q}{\pi_A q+(1-\pi_A)p} \tag{14}$$

である．Aであることを知られたくないという場合には，この2つの確率を小さく保つことが必要である．ところでAであることを知られたくないという場合，π_A は1/2より小さいのが普通であろう．$0\leqq\pi_A\leqq\frac{1}{2}$ の範囲で，(13)，(14) は，いずれも $\pi_A=\frac{1}{2}$ のとき最大で，そのときの値はそれぞれ p, q である．このことからも p が1/2に近いほど回答者の協力が得られやすいことがわかる．良心的な調査企画者は (13)，(14) をある水準 α 以下におさえるべきであると考えるであろう．α の値は質問の内容に依存するであろう．あまり深刻に考える必要がないならば α は0.7とか0.8でもよかろう．かなり深刻なら α は0.5以下であることがこのましいであろう．(13)，(14) の両者を α 以下におさえるということは

$$\max\{p,\ 1-p\}\leqq\frac{\alpha(1-\pi_A)}{\alpha(1-2\pi_A)+\pi_A} \tag{15}$$

と同値である．とくに $\alpha=0.5$ とすると

$$\max\ \{p,\ 1-p\}\ \leqq1-\pi_A \tag{16}$$

となる．π_A が小さければ，この意味では p はかなり1に近くとることも可能であるが，これは1つの必要条件と考えるべきである．すなわち，これまでに発表されている論文でもいわれているように，p が0あるいは1に近いということは一方の質問だけがほとんど選ばれることになり，回答者にとっては，たとえ前述の条件付確率が小さくとも調査に協力することに抵抗を感ずるであろう．したがって (13)，(14) の条件付確率をおさえるということと同時に，どちらの質問を選んでいるかが知られないという2つの面から p をきめる必要がある．

さて，繰返しのある場合の考察に入る．1回目と2回目の質問の選択確率が同じ場合について述べる．(13)，(14) に対応する条件付確率は

$$Pr\{A|\text{"yes, yes"}\}=\frac{\pi_A p^2}{\pi_A p^2+(1-\pi_A)q^2} \tag{17}$$

$$Pr\{A|\text{"yes, no"}\}=Pr\{A|\text{"no, yes"}\}=\frac{\pi_A pq}{\pi_A pq+(1-\pi_A)pq}=\pi_A \tag{18}$$

$$Pr\{A|\text{"no, no"}\}=\frac{\pi_A q^2}{\pi_A q^2+(1-\pi_A)p^2} \tag{19}$$

となる．ここで，たとえば "yes, no" は第1回の回答が "yes"，第2回の回答が "no" であることを表わす．(18)，すなわち第1回と第2回の回答がかわった場合の条件付確率は p に依存しないので，(17) と (19) について考察する．(17)，(19) の右辺の分母，分子をそれぞれ p^2+q^2 で割ってみると，(17)，(19) は，それぞれ (13)，(14) の p, q に $\frac{p^2}{p^2+q^2}$, $\frac{q^2}{p^2+q^2}$ を代入したものと一致している．この条件付確率に関する限りでは，2回繰返しは質問の選択確率を p, q から $\frac{p^2}{p^2+q^2}$, $\frac{q^2}{p^2+q^2}$ に変更して1回だけ行うことに対応している．そして明きらかに (17) と (19) の最大値は (13)，(14) の最大値より大きくなっている．たとえば

$\pi_A=\dfrac{1}{2}$ のとき，(13) は p，(17) は $\dfrac{p^2}{p^2+q^2}$ となるが，$p=0.6,\ 0.7,\ 0.8$ に対応して

$\dfrac{p^2}{p^2+q^2}\fallingdotseq 0.69,\ 0.84,\ 0.94$ となる．

このような繰返しの影響を回答者に説明し，その上で協力してもらうのが調査をする側の正しい態度と考えられるが，実際には困難であろう．しかし少くとも，もし繰返しのあるモデルを利用しようとするならば調査企画者はこの事情を認識して良心的に調査を企画すべきであろう．とにかく回答者がこの事情に気づかないかもしれないことに留意すべきである．

3回以上の繰返しについては実用的でないのでとくに考察しない．

次に §1 で述べた Horvitz et al. (1967) のモデルで2回繰返しのある場合については，ほかに Gould et al. (1969) などで扱われているので省略する．たゞ §1 でもふれたように2回繰返しをとり上げる理由，したがって考察の観点は異なっている．

§3. ランダム回答法における有限母集団修正ついて

§1 で述べたように，かなり小さな集団を対象として悉皆調査を行ないランダム回答法を利用することや，あるいは有限母集団から無視できない抽出率でサンプルを抽出しランダム回答法を利用することが考えられる．ここではランダム回答法から得られる推定量の分散に入ってくる有限母集団修正の形を基本的な2，3のモデルについて求めておく．

ランダム回答法から得られる推定量の分散は，サンプリングによる変動にもとずく部分と，質問をランダムに選択することにもとずく部分に分解され，有限母集団修正はサンプリングによる変動にもとずく部分に対してのみ入ってくるであろうということは当然予想されることである．実際，種々の計算もこの考え方にしたがって行われる．Warner のモデルなどについては，この考え方にしたがって結果もほとんど自明であるが，しかし Simmons のモデルでは結果はそれほど自明でない．その理由は Simmons のモデルでは2組のサンプルを必要とし，それぞれが，有限母集団から非復元で重複のないように抽出されるために2組のサンプルの間の相関を考慮しなければならなからである．

3.1 Warner のモデル

N を母集団の大きさ，n をサンプルの大きさ，N_A を母集団におけるクラスAの大きさ，p を「あなたはAですか？」という質問を選ぶ確率とする．$\pi_A=N_A/N$，すなわち母集団におけるクラスAの比率，また $q=1-p$，すなわち「あなたはB $(=\bar{A})$ ですか？」という質問を選ぶ確率，とする．大きさ n のサンプルの中で"yes"と答えた人数を y とする．目的は π_A（あるいは N_A）の推定である．

さて，N_A に関する尤度関数は次のようになる．

$$L\,(N_A)=\sum_{k=k_1}^{k_2}\frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}\left\{\sum_{j=j_1}^{j_2}\binom{k}{j}\binom{n-k}{y-j}p^{n-k-y+2j}\,q^{k-2j}\right\} \tag{20}$$

たゞし，$k_1=\max\,(0,\ n-N+N_A)$，$k_2=\min\,(N_A,\ n)$，$j_1=\max\,(0,\ y-n+k)$，$j_2=\min\,(y,\ k)$ である．これを最大にする N_A を数値的に求めることは出来る．実際，何組かの N，n，p の組合せ ($N=10$ (10) 50，$n=\dfrac{N}{5}\left(\dfrac{N}{5}\right)N$，$p=0.6$ (0.1) 0.9) に対して求めてみたが，その結果では，§2 の無限母集団の場合の π_A の最尤推定量の式を用いて計算した値の N 倍とほとんど（高々1しか）違わなかった．理論的にはまだその裏付けはしていないが，この

ような理由から，π_A の推定量としては無限母集団の場合の最尤推定量を用いることにした．すなわち

$$\hat{\pi}_A=\frac{\frac{y}{n}-q}{p-q} \tag{21}$$

である．これは無限母集団の場合には π_A の最小分散不偏推定量であった．有限母集団の場合にも π_A の不偏推定量であることは容易に確められる．さて，この不偏推定量の分散は無限母集団の場合は (3) で与えられている．すなわち

$$\frac{1}{n}\pi_A\ (1-\pi_A)+\frac{pq}{n\ (p-q)^2} \tag{22}$$

であった．第1項はサンプリングの変動にもとずく部分で，第2項は質問をランダムに選ぶことにもとずく部分である．

有限母集団の場合には，サンプルに入ってくるクラスAの人数を表わす確率変数を X（これは観測できない）として，

$$\mathrm{Var}\ (\hat{\pi}_A)=\mathrm{Var}\ (E\ (\hat{\pi}_A|X))+E\ (\mathrm{Var}\ (\hat{\pi}_A|X)) \tag{23}$$

を用いて計算すると

$$\frac{N-n}{N-1}\frac{1}{n}\pi_A\ (1-\pi_A)+\frac{pq}{n\ (p-q)^2} \tag{24}$$

であることがわかる．

これを (22) と比較すると，この節のはじめに述べたように，サンプリングの変動にもとずく部分，すなわち右辺の第1項，だけが通常の形の有限母集団修正を受けていることがわかる．もとより悉皆調査の場合には質問をランダムに選ぶことにもとずく第2項だけに帰する．

3.2 Simmons のモデル

「あなたは A ですか？」という質問と，これとは直接関係のない「あなたは C ですか？」という質問を用意する．第2の質問は回答者が正直に答えることに何ら抵抗を感じないものである．目的は前と同じく母集団におけるクラス A の比率 π_A の推定である．母集団におけるクラス C の比率 π_C は未知とする．

大きさ N の母集団から非復元で大きさ n_1, n_2 の2組のサンプルを2組の間に重複がないように抽出する．第1組のサンプルでは p_1，第2組のサンプルでは p_2 の確率で「あなたは A ですか？」という質問を選択してもらう．したがって，それぞれ $q_1=1-p_1$，$q_2=1-p_2$ の確率で「あなたは C ですか？」という方の質問が選ばれることになる．π_A が推定できるためには $p_1\neq p_2$ が必要である．第1組の大きさ n_1 のサンプル中 “yes” と答えた人数を y_1，第2組の大きさ n_2 のサンプル中 “yes” と答えた人数を y_2 とする．ここでも推定量としては無限母集団の場合の最尤推定量

$$\hat{\pi}_A=\frac{q_2\frac{y_1}{n_1}-q_1\frac{y_2}{n_2}}{p_1-p_2} \tag{25}$$

を用いることにする (Horvitz et al. (1967) 参照)．不偏推定量であることは有限母集団の場合にも無限母集団の場合でも容易にわかる．

この推定量の分散は無限母集団の場合には

$$\frac{1}{(p_1-p_2)^2}\left\{\frac{q_2{}^2\lambda_1\ (1-\lambda_1)}{n_1}+\frac{q_1{}^2\lambda_2\ (1-\lambda_2)}{n_2}\right\} \tag{26}$$

である (Greenberg et al. (1969) 参照). たゞし, $\lambda_i = p_i \pi_A + q_i \pi_C$ $(i = 1, 2)$ である. これをサンプリングの変動にもとずく部分と質問をランダムに選ぶことにもとずく部分に分解すると次のようになる.

$$
\begin{aligned}
&\frac{1}{(p_1 - p_2)^2}\left\{\frac{{q_2}^2}{n_1}\Big((\lambda_1\ (1-\lambda_1) - p_1 q_1\ (\pi_{A\bar{C}} + \pi_{\bar{A}C})\Big)\right.\\
&\qquad \left. + \frac{{q_1}^2}{n_2}\Big((\lambda_2\ (1-\lambda_2) - p_2 q_2\ (\pi_{A\bar{C}} + \pi_{\bar{A}C})\Big)\right\}\\
&\qquad + \frac{\pi_{A\bar{C}} + \pi_{\bar{A}C}}{(p_1 - p_2)^2}\left\{\frac{{q_2}^2}{n_1} p_1 q_1 + \frac{{q_1}^2}{n_2} p_2 q_2\right\}
\end{aligned}
\tag{27}
$$

となる. 第1項がサンプリング, 第2項がランダム回答にもとずく部分である. なお, π_{AC} は母集団におけるクラス A に属しクラス C には属さない部分の比率をあらわしている. π_{AC} も同様の意味である.「A ですか?」という質問と「C ですか?」という質問は直接関係がないといったが, ここでは独立性, すなわち $\pi_{A\bar{C}} = \pi_A \cdot (1 - \pi_C)$, $\pi_{\bar{A}C} = (1 - \pi_A) \cdot \pi_C$ などの成立は仮定していない.

さて有限母集団の場合の推定量 (25) の分散は次のようになる.

$$
\begin{aligned}
&\frac{1}{(p_1 - p_2)^2}\left\{\left(\frac{N - n_1}{N-1}\right)\frac{{q_2}^2}{n_1}\Big((\lambda_1\ (1-\lambda_1) - p_1 q_1\ (\pi_{A\bar{C}} + \pi_{\bar{A}C})\Big)\right.\\
&\qquad \left. + \left(\frac{N - n_2}{N-1}\right)\frac{{q_1}^2}{n_2}\Big((\lambda_2\ (1-\lambda_2) - p_2 q_2\ (\pi_{A\bar{C}} + \pi_{\bar{A}C})\Big)\right\}\\
&\qquad + \frac{\pi_{A\bar{C}} + \pi_{\bar{A}C}}{(p_1 - p_2)^2}\left\{\frac{{q_2}^2}{n_1} p_1 q_1 + \frac{{q_1}^2}{n_2} p_2 q_2\right\}\\
&\qquad + \frac{1}{(p_1 - p_2)^2}\left\{\frac{2 q_1 q_2}{N-1}\Big(\pi_{AC} + p_1 p_2 \pi_{A\bar{C}} + q_1 q_2 \pi_{\bar{A}C} - \lambda_1 \lambda_2\Big)\right\}
\end{aligned}
\tag{28}
$$

この節のはじめに述べたように, Simmons のモデルでは, 無限母集団の場合のサンプリングの変動にもとずく部分に通常の形の有限母集団修正が各組のサンプルに対応して入り, ランダム回答にもとずく部分は修正を受けないが, そのほかに2組のサンプルの間の相関に起因する項 ((28) の最後の項) が入っていることが特徴である.

3.3 Morton のモデル

Simmons のモデルで, 特に π_C が既知の場合を考えると, π_A を推定するためには, 一組のサンプルで十分である. 更に又,「C ですか」という質問に答えさせる代りに, ある決まった確率で "yes" 又は "no" を答えてもらってもよい. すなわち, 各回答者は, 確率で p_1「A ですか」という質問に対する答 "yes" 又は "no" を, 確率 p_2 で "yes" を, 確率 p_3 で "no" をそれぞれ回答するように決められているものとする $(p_1 + p_2 + p_3 = 1)$. このようなモデルを Morton のモデル (Greenberg et al. (1969) p. 532 脚注) と呼ぶことにしよう.

ここでも n をサンプルの大きさ, y をサンプル中 "yes" と答えた人数とする. 無限母集団の場合の π_A の最尤推定量は, 次式で表されるが, この推定量が無限母集団の場合も有限母集団の場合にも不偏推定になっていることは, 容易にたしかめることができる.

$$
\hat{\pi}_A = \frac{\dfrac{y}{n} - p_2}{p_1}
\tag{29}
$$

この推定量 $\hat{\pi}_A$ の分散は, 前述の2つの推定量の有限母集団に対する修正と同様, サンプリ

ングによる分散に有限修正がかゝった形で表わすことが出来る．

$$\mathrm{Var}\ (\hat{\pi}_A) = \frac{N-n}{N-1}\frac{\pi_A\,(1-\pi_A)}{n} + \frac{(1-\pi_A)\,p_2\,(1-p_2) + \pi_A p_3\,(1-p_3)}{np_1{}^2} \tag{30}$$

無限母集団の場合の分散の形は (30) 式で右辺第１項の有限母集団修正を除いたものである．

統計数理研究所

参　考　文　献

[1] Campbell, C. and Joiner, B.L. (1973): How to get the answer without being sure you've asked the question. The American Statistician, 27, No. 5, 229–231.

[2] Gould, A.L. and Shar, B.V. (1969): Unrelated question randomized response techniques with two trials per respondent. Proceedings of Social Statistical Section, American Statistical Association, 351–359.

[3] Greenberg, B.G., Abul-Ela, A.A., Simmons, W.R. and Horvitz, D.C. (1969) : The unrelated question randomized response model: Theoretical framework. Journal of the American Statistical Association, 64, 520–539.

[4] Horvitz, D.G., Shar, B.V. and Simmons, W.R. (1967): The unrelated question randomized response model. Proceedings of Social Statistical Section, American Statistical Association, 65–72.

[5] Warner, S.L. (1965): Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60, 63–69.