

## 偽りの回答の確率を考慮したランダム回答法

木更津工業高専                      大内 俊二 ( Shunji Ohuchi )

千葉大学自然科学研究科            A. W. Alhassan

千葉大学理学部                      田栗 正章 ( Masaaki Taguri )

### 1. 序

ある母集団において、他人には知られたくないような性質を持ったグループの比率  $\pi$  を、質問による調査によって推定しようとする場合に、直接質問したのでは、回答者は答えを偽ったり、わからないと答えたりしてしまうものである、このようなことを防ぐために考えられたのがランダム回答法 ( Randomized Response Technique ) である。その一つに Warner 法 ( Warner [1965] ) があるが、そこではカード等を用いて、質問を確率的に選択するようになっているため、質問者には回答者がどのような質問に答えたのか判らないようになっており、そのため比較的正直な答えが得られる。しかし、現実には、ランダム回答法を適用してもなお、回答者が答えを偽る可能性は高い ( Chi, Chow and Rider [1972] 参照 )。そ

こでわれわれは、ランダム回答法が適用されてもなお偽回答が存在するような状況を考えて（具体的には、回答者が偽りの回答をする確率を導入して）、比率  $\pi$  の不偏な推定量  $\hat{\pi}_{pr}$  を提案し、オリジナルな Warner の推定量  $\hat{\pi}_w$  との比較・検討を行った。その結果、 $\hat{\pi}_{pr}$  が平均二乗誤差の意味で  $\hat{\pi}_w$  より良くなるためには、比較的大きな標本の大きさが必要であることがわかった。しかし、 $\hat{\pi}_w$  の場合、デザイン・パラメータの決定や  $\hat{\pi}_w$  の分散の推定が行えないため、 $\hat{\pi}_w$  による  $\pi$  の推定は現実には出来ない。そこでここでは、 $\hat{\pi}_{pr}$  を用いて実際に  $\pi$  を推定する手順を与えた。さらに、繰り返し質問による方法を考え、 $\hat{\pi}_{pr}$  を改良する推定量  $\hat{\pi}_{tt}$  を提案し、同様の解析を行った。

## 2. Warner のモデル

### 2. 1 Randomised Response Technique とは

直接質問したのでは、正直に答えにくい種類の質問に対して、偽りの回答や D.K.（わからない）への逃避を防ぐために考えられた調査法で、現実には、特殊な病気にかかっている人、麻薬常用者、政治家に賄賂を贈っている企業などが、ある母集団に占める割合を調査するために適用されるものである。

## 2. 2 Warner法

母集団の各人はグループ A (他人には知られたくない性質を持っているグループ) かグループ  $A^c$  のいずれか一方に属するものとする。調査者は次の 2 種類の質問を用意し、それらが確率的に選ばれるように、Q. 1 の選ばれる確率  $p$  ( $0 < p < 1$ ) を決めておく。

Q. 1 : 「あなたは A ですか?」 (確率  $p$  で選ばれる)

Q. 2 : 「あなたは  $A^c$  ですか?」 (確率  $1 - p$  で選ばれる)

回答者は自分の選んだ質問に対して、Yes か No のみを調査者に答える。起こりうる回答パターンを模式的に表すと、次のようになる。

	A ( $\pi$ )	$A^c$ ( $1 - \pi$ )
Q. 1 ( $p$ )	Yes	No
Q. 2 ( $1 - p$ )	No	Yes

確率変数  $X_i$  を次のように定義する。

$$X_i = \begin{cases} 1 & (i \text{ 番目の回答者の答えが Yes のとき}), \\ 0 & (i \text{ 番目の回答者の答えが No のとき}). \end{cases}$$

このとき、次式が成り立つ。

$$P(X_i = 1) = p\pi + (1-p)(1-\pi),$$

$$P(X_i = 0) = p(1-\pi) + (1-p)\pi.$$

### 2. 3 Warnerの推定量 $\hat{\pi}_w$

Warnerのモデルにおいては、次の2つの仮定がおかれている。

《仮定1》 すべての回答者は正直に答える。

《仮定2》  $p \neq 1/2$  .

$n$ 人の回答者の中でYesと答えた人数を  $m (= \sum_{i=1}^n X_i)$  とし、 $\hat{\lambda} = m/n$  とおくと、 $\pi$ の最尤推定量  $\hat{\pi}_w$  は次で与えられる。

$$\hat{\pi}_w = \frac{\hat{\lambda} - (1-p)}{2p-1}.$$

推定量  $\hat{\pi}_w$  は、次の性質をもつ。

a)  $E(\hat{\pi}_w) = \pi$  (不偏推定量),

b)  $V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}.$

### 2. 4 偽りの回答を考慮する必要性

下の表1 (Chi, Chow and Rider[1972]による) は、回答者

者が、ランダム回答法を信頼して、どの程度本当の答えをするかということについて、台湾とNorth Carolinaで行った調査（標本数は各々692人，4571人）の結果である。これによれば、それぞれ36.7%，15.4%が、信頼できないと答えしており、さらに、D.K.まで含めると、それぞれ52%，23.7%が、ランダム回答法を適用しても安心して本当の答えを言えないということになる。

表 1 (RRT Trick Response Table)

Question: Did you think we could figure out which question was answered ?

Answer	Percent	
	Taiwan	North Carolina
No trick	48.0	76.3
Yes	36.7	15.4
Don't know	15.3	8.3
Total	100.0	100.0
Number of respondents	692	4571

### 3. われわれの提案するモデル

#### 3. 1 偽りの回答を考慮したモデル

2. 3節の《仮定1》を除いたモデルを考える。すなわち回答者は、正直に答えにくい質問に対しては、確率 $\gamma$ で偽りの回答をすると考える。この場合起こりうる回答パターンは次のようになる。

	A ( $\pi$ )		A <sup>c</sup> ( $1 - \pi$ )
	T ( $1 - \gamma$ )	F ( $\gamma$ )	T
Q . 1 ( $p$ )	Y e s	N o	N o
Q . 2 ( $1 - p$ )	N o	Y e s	Y e s

T : T r u e , F : F a l s e ,  $\gamma$  : 偽りの回答をする確率

《偽りの回答をする確率 $\gamma$ のモデル》

質問 Q . 1 を選ぶ確率  $p$  が 0 または 1 に近くなると、回答者は自分の答えが、どちらの質問に対するものか調査者に判ってしまうというような心理的印象をもつ。このことから、 $\gamma$  に対する自然な仮定として、次の 2 つをおく。

( a )  $\gamma$  は  $p = 1/2$  に関して対称である ,

( b )  $\gamma$  は区間  $( 0 , 1/2 )$  で単調減少である .

この 2 つの仮定を満たす  $\gamma$  のモデルはいろいろ考えられるが、ここでは、次のモデルを想定する。

$$\gamma = a ( 1 - 2 p ) ^ 2 , \quad 0 \leq a \leq 1 .$$

### 3. 2 提案する推定量 $\hat{\pi}_{pr}$

3. 1 節で導入した  $\gamma$  のモデルを想定した場合、推定すべきパラメータは  $\pi$  と  $a$  の 2 つになるので、与えられた標本を 2 つに分けて ( 標本数を  $n_1, n_2$  とする )、それぞれの標本においてランダム回答法を適用することを考える。

{	標本 1 $n_1 ( p_1 )$
	標本 2 $n_2 ( p_2 )$

ここで、 $p_1, p_2$  は、標本 1, 2 で Q . 1 を選ぶ確率とする。

$m_i ( i = 1, 2 )$  を標本  $i$  における Y e s の数とし、 $\hat{\lambda}_i = m_i / n_i$  とおくと、次が成り立つ。

$$E ( \hat{\lambda}_1 ) = \pi ( 1 - 2 p_1 ) ( \gamma_1 - 1 ) + ( 1 - p_1 ) ,$$

$$E ( \hat{\lambda}_2 ) = \pi ( 1 - 2 p_2 ) ( \gamma_2 - 1 ) + ( 1 - p_2 ) ,$$

$$\gamma_i = a (1 - 2 p_i)^2 \quad (i = 1, 2).$$

$p_1 \neq p_2$  のとき、上の 2 式の関係から、モーメント法によって  $\pi$  の推定量を求めると、次を得る。

$$\hat{\pi}_{pr} = \frac{r_2^3 z_1 - r_1^3 z_2}{r_1 r_2 (r_1^2 - r_2^2)},$$

$$z_i = \hat{\lambda}_i - (1 - p_i),$$

$$\gamma_i = 1 - 2 p_i \quad (i = 1, 2).$$

推定量  $\hat{\pi}_{pr}$  は次の性質をもつ。

$$(a) \quad E(\hat{\pi}_{pr}) = \pi \quad (\text{不偏推定量}).$$

$$(b) \quad V(\hat{\pi}_{pr}) = \frac{A_1}{n_1} + \frac{A_2}{n_2},$$

$$A_1 = \frac{q_1 (1 - q_1) r_2^4}{r_1^2 (r_1^2 - r_2^2)^2},$$

$$A_2 = \frac{q_2 (1 - q_2) r_1^4}{r_2^2 (r_1^2 - r_2^2)^2},$$

$$q_i = (1 + r_i) / 2 - \pi r_i (1 - a r_i^2) \quad (i = 1, 2).$$

(c)  $\hat{\pi}_{pr}$  は、 $\pi$  の最尤推定量である。

(d) Warner の推定量との関係

分割された標本のそれぞれから求まる、2 つの Warner の推定量の結合推定量  $\hat{\pi}_c = \alpha_1 \hat{\pi}_{w1} + \alpha_2 \hat{\pi}_{w2}$  を考える。



ここで、

$$\hat{\pi}_{wi} = \frac{\hat{\lambda}_i - (1 - p_i)}{2p_i - 1} \quad (i = 1, 2)$$

とする。このとき  $\hat{\pi}_c$  は、 $\alpha_1 + \alpha_2 = 1$ ,  $\alpha_1 \gamma_1 + \alpha_2 \gamma_2 = 0$  のとき、推定量  $\hat{\pi}_{pr}$  に一致する。

### 3. 3 デザイン・パラメータの最適な値の選択

次の最適化問題を考える。

条件:  $n_1 + n_2 = n$  (一定),

$$p_{i0} \leq p_i \quad (0 < p_{i0} < 1/2; \quad i = 1, 2).$$

の下で、 $V(\hat{\pi}_{pr})$  を最小にする  $n_i$ ,  $p_i$  を求める。ここで、 $p_{i0}$  は現実に使用可能な  $p_i$  の下限値である。また、一般性を失うことなく  $p_1 \leq p_2$  とできる。

(a) 最適な  $n_1, n_2$  の値

$$\frac{\partial V(\hat{\pi}_{pr})}{\partial n_i} = 0 \quad (i = 1, 2)$$

を満たす  $n_1, n_2$  を、それぞれ  $n_1^*, n_2^*$  とすると、任意

の  $p_1, p_2$  に対して、

$$n_1^* = \frac{\sqrt{A_1}}{\sqrt{A_1} + \sqrt{A_2}} n, \quad n_2^* = \frac{\sqrt{A_2}}{\sqrt{A_1} + \sqrt{A_2}} n$$

が成り立つ。

(b) 最適な  $p_1, p_2$  の値

最適な値  $n_1^*, n_2^*$  の下での  $V(\hat{\pi}_{pr})$  を  $V_1$  とす

ると、 $V_1$  は  $r_1, r_2$  ( $r_i = 1 - 2p_i$ ) の関数となるの

で、いまこれを  $V_1(r_1, r_2)$  と表すと、

$$V_1(r_1, r_2) = (\sqrt{A_1} + \sqrt{A_2})^2 / n$$

となる。

$V_1(r_1, r_2)$  については、以下の性質をもつことが示される。

[性質1] 任意に固定された  $r_2$  ( $p_2$ ) に対して、

$V_1$  は、 $r_1$  ( $p_1$ ) の単調減少 (増加) 関数である。

[性質2] 任意に固定された  $r_1$  に対して、 $V_1$  は、

$r_2$  の定義域内のある値  $r_2^*$  において、唯一の最小値をとる。

[性質1] から、 $p_1$  の最適な値  $p_1^*$  は  $p_{10}$  にとればよく、

[性質2] から、 $p_2$  の最適な値  $p_2^*$  は  $(1 - r_2^*)/2$  にと

ればよいことがわかる。

(c) その他の性質

$p_1, p_2, \pi$  を任意に固定したとき、 $V_1$  は  $a$  に関して単調減少であり、また  $p_1, p_2, a$  を任意に固定したとき、

$V_1$  は  $\pi$  に関して単調増加である。

#### 4. $\hat{\pi}_{pr}$ と $\hat{\pi}_w$ の数値的比較と検討

##### 4. 1 Warnerのモデルにおけるパラメータの最適な値

2. 3節で紹介した Warnerの推定量  $\hat{\pi}_w$  は、偽りの回答が存在する場合には、

$$E(\hat{\pi}_w) = \pi(1 - \gamma)$$

となり、不偏でなくなる。したがって、分散  $V(\hat{\pi}_{pr})$  との比較には、 $MSE(\hat{\pi}_w)$  を用いなければならない。さらに、 $\hat{\pi}_{pr}$  との同等な条件での比較を行うため、次の最適化法により、パラメータの最適な値を決定しておく。

条件:  $p_0 \leq p$  ( $p_0 = \min[p_{10}, p_{20}]$ )

の下で、 $MSE(\hat{\pi}_w)$  を最小にする  $p$  の値  $p^*$  を求める。

ただし、 $p_{10}, p_{20}$  は、3. 3節で与えたものである。

##### 4. 2 $\hat{\pi}_{pr}$ と $\hat{\pi}_w$ の数値的比較

3. 3節, 4. 1節での最適化法により選ばれたパラメータの最適値(表2では、 $p_0 = p_{01} = 0.1$ )を用いて計算した分散と平均二乗誤差をそれぞれ、 $V^*(\hat{\pi}_{pr})$ ,  $MSE^*(\hat{\pi}_w)$  とする。標本の大きさ  $n$  に関して、分散は、 $1/n$  のオーダーとなるから、 $nV^*(\hat{\pi}_{pr})$ ,  $nV^*(\hat{\pi}_w)$  は標本数に依存しない量である。しかし、これらの量は、未知のパラメータ  $a$  と  $\pi$  に依存するので、表2には、 $a$  と  $\pi$  のいくつかの値の組

合せについて、これらの値を計算したものが示されている。  
 ここで、 $a$ の値については、実際問題においてはラフな推定すら出来ないので、0と1の間の全ての可能な範囲で動かす、 $\pi$ については、ランダム回答法の実際の適用例（例えば、Chi, Chow and Rider[1972]参照）などから、現実的に意味のある値の範囲で動かした。また、 $n_0$ は、われわれの提案した推定量 $\hat{\pi}_{pr}$ の方がWarnerの推定量 $\hat{\pi}_w$ より良くなるためには、この $n_0$ の値より大きな標本数が必要であるということの意味する数値である。

表2  $V^*(\hat{\pi}_{pr})$  と  $V^*(\hat{\pi}_w)$  の比較

$\pi$	$a$	0.10	0.30	0.50	0.70
		$p_1 = 0.1$			
0.10	varp	2.983	2.961	2.938	2.914
	varw	0.230	0.656	0.989	1.279
	$n_0$	6.9E4	2.8E4	1.7E4	1.1E4
0.20	varp	3.220	3.191	3.158	3.123
	varw	0.533	1.211	1.741	2.203
	$n_0$	4.3E4	1.5E4	7.7E3	4.0E3
0.30	varp	3.381	3.352	3.318	3.279
	varw	0.780	1.670	2.365	2.971
	$n_0$	3.2E4	9.8E3	4.0E3	1.1E3
0.40	varp	3.477	3.455	3.426	3.389
	varw	0.987	2.066	2.909	3.644
	$n_0$	2.5E4	6.7E3	1.8E3	—

上の表で  $\text{varp} = n V^*(\hat{\pi}_{pr})$ ,  $\text{varw} = n V^*(\hat{\pi}_w)$  である。

#### 4. 3 数値実験結果の考察

表 2 より、以下のようなことが読み取れる。

(a)  $V^*(\hat{\pi}_{pr})$  は、 $a$  に関して極めて頑健である。

(b)  $V^*(\hat{\pi}_{pr})$  は、 $\pi$  に関して  $MSE^*(\hat{\pi}_w)$  より頑健である。

(c)  $MSE^*(\hat{\pi}_w)$  は、 $a$  に関して sensitive である。

(d)  $V^*(\hat{\pi}_{pr}) < MSE^*(\hat{\pi}_w)$ 、すなわち推定量  $\hat{\pi}_{pr}$  の方が推定量  $\hat{\pi}_w$  より良くなるには、比較的大きな標本数が必要である。

#### 5. 頑健性の問題と $\hat{\pi}_{pr}$ による推定手順

##### 5. 1 パラメータに関する $V(\hat{\pi}_{pr})$ の頑健性

3. 3 節で考えたデザイン・パラメータの最適な値  $n_1^*$ ,  $p_1^*$  は、未知のパラメータ  $a$ ,  $\pi$  に依存する。従ってわれわれの提案する方法を現実の  $\pi$  の推定に適用する場合には、これらのデザイン・パラメータの最適値が  $a$  と  $\pi$  の値にどの程度影響を受けるかを検討しておかなければならない。3. 3 節で考えた最適化問題の結果から、指定された  $p_1^*$  の値（現実に意味のある  $p_1$  の下限値  $p_{10}$ ）の下で、 $p_2^*$  の値が決まったので、表 3 と 4 には、指定された  $p_1 (= p_1^*)$  の値の下で、 $a$  と  $\pi$  の値のいくつかの組合せに対して計算された  $n_1^*$  /

$n_2^*$ ,  $p_2^*$ の値が示されている。これらの値は指定された  $p_1$  の値の下で、 $a$  と  $\pi$  に関して頑健であることが読み取れる。このことから、 $p_1$  の値を指定したときの  $n_1^*/n_2^*$ ,  $p_2^*$  の値の  $a$  と  $\pi$  に関して平均をとった値  $n_1^{\circ}/n_2^{\circ}$ ,  $p_2^{\circ}$  を実際の調査において利用すべき値（推奨値）として与えた。

表 3  $n_1^*/n_2^*$ ,  $n_1^{\circ}/n_2^{\circ}$  の値とそのロス

$p_1$	$\pi$	$a$			$n_1^{\circ}/n_2^{\circ}$ Mean Ratio	$L_3[n_1^{\circ}/n_2^{\circ}]$
		0.0	0.5	0.9		
0.10	0.05	0.119	0.121	0.117	0.120	0.027
	0.10	0.122	0.122	0.115		0.044
	0.20	0.127	0.119	0.116		-0.058
	0.30	0.127	0.119	0.115		-0.058
0.15	0.05	0.120	0.126	0.124	0.122	-0.029
	0.10	0.126	0.122	0.119		-0.031
	0.20	0.126	0.122	0.117		0.039
	0.30	0.123	0.119	0.122		0.026
0.20	0.05	0.123	0.122	0.122	0.123	0.012
	0.10	0.127	0.126	0.124		-0.034
	0.20	0.123	0.121	0.119		0.032
	0.30	0.127	0.125	0.122		-0.033
0.25	0.05	0.123	0.123	0.122	0.125	0.021
	0.10	0.126	0.125	0.124		-0.005
	0.20	0.130	0.129	0.128		-0.035
	0.30	0.120	0.119	0.130		0.049
0.30	0.05	0.119	0.118	0.118	0.121	0.023
	0.10	0.120	0.120	0.119		0.013
	0.20	0.122	0.122	0.122		-0.010
	0.30	0.124	0.123	0.123		-0.023
0.35	0.05	0.290	0.290	0.290	0.292	0.007
	0.10	0.291	0.291	0.291		0.003
	0.20	0.294	0.293	0.293		-0.005
	0.30	0.295	0.295	0.295		-0.011

表 4  $p_2^*$ ,  $p_2^\circ$  の値とそのロス

$p_1$	$\pi$	a			$p_2^\circ$	$L_4[p_2^\circ]$
		0.0	0.5	0.9		
0.10	0.05	0.284	0.280	0.280	0.288	0.003899
	0.10	0.288	0.284	0.284		0.001424
	0.20	0.292	0.292	0.288		0.001867
	0.30	0.296	0.296	0.292		0.005650
0.15	0.05	0.316	0.312	0.312	0.318	0.001732
	0.10	0.316	0.316	0.316		0.000592
	0.20	0.320	0.320	0.320		0.000621
	0.30	0.324	0.324	0.320		0.002312
0.20	0.05	0.344	0.344	0.344	0.346	0.000795
	0.10	0.344	0.344	0.344		0.000245
	0.20	0.348	0.348	0.348		0.000269
	0.30	0.348	0.348	0.348		0.000893
0.25	0.05	0.372	0.372	0.372	0.373	0.000378
	0.10	0.372	0.372	0.372		0.000151
	0.20	0.372	0.372	0.372		0.000351
	0.30	0.376	0.376	0.372		0.000503
0.30	0.05	0.400	0.400	0.400	0.400	0.0
	0.10	0.400	0.400	0.400		0.0
	0.20	0.400	0.400	0.400		0.0
	0.30	0.400	0.400	0.400		0.0

5. 2  $\hat{\pi}_{pr}$  による  $\pi$  の推定手順

5. 1 節における頑健性の検討の結果に基づき、現実の調査における  $\hat{\pi}_{pr}$  による  $\pi$  の推定手順をまとめておく。

- 1° 標本の大きさ  $n$  と確率  $p_1$  を決める。
- 2° 与えられた表を用いて、1° で決めた  $p_1$  に対応する推奨値  $n_1^\circ / n_2^\circ$ ,  $p_2^\circ$  から、標本配分数  $n_i$ , 確率  $p_2$  を決める。
- 3° 単純ランダム・サンプリングにより選ばれた  $n$  人の回答者にランダム回答法を適用し、 $\pi$  の推定値  $\hat{\pi}_{pr}$

を求める。

- 4° 3° で求めた推定値  $\hat{\pi}_{pr}$  を使い、 $a$  の推定値  $\hat{a}$  を計算し、 $\hat{\pi}_{pr}$  と  $\hat{a}$  を  $V(\hat{\pi}_{pr})$  の式に代入し、 $\hat{\pi}_{pr}$  の標準誤差を計算する。

『注意』

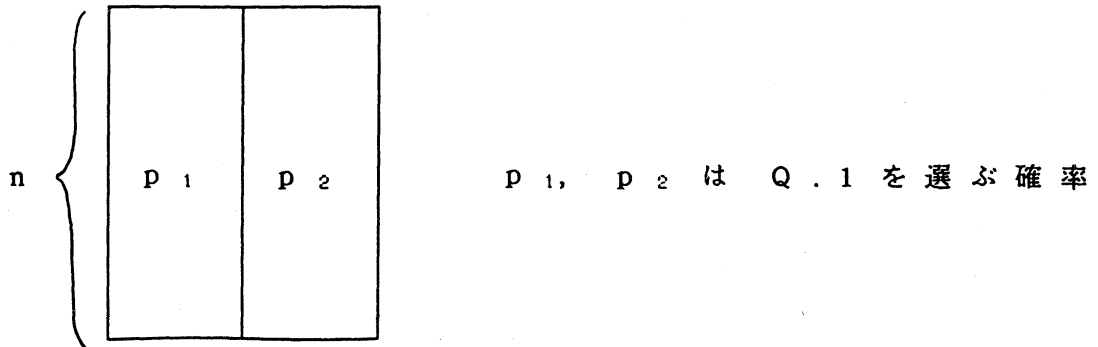
上の推定手順を遡ることにより、要求される  $\hat{\pi}_{pr}$  の推定精度から、必要な標本の大きさ  $n$  のおおよその値を決めることができる。一方  $\hat{\pi}_w$  の場合は 4. 1 節の最適化法によって決まる  $p^*$  は 4. 3 節の (b), (c) の考察により、 $a$  と  $\pi$  に関して頑健ではなく、さらに標本の大きさ  $n$  にも依存するので、われわれの考えている状況においては、 $\hat{\pi}_w$  によって  $\pi$  を推定することは現実には出来ない。

## 6. 提案する第 2 の推定量 $\hat{\pi}_{tt}$

### 6. 1 ランダム回答法による繰り返し質問

4. 3 節の結論から、推定量  $\hat{\pi}_{pr}$  は、平均二乗誤差の意味では、 $\pi_w$  に比べあまり良い推定量ではなかった。その一つの理由として、与えられた標本を 2 つに分けていることが考えられる。このため、第 2 のモデルでは、同じ回答者に繰り返し質問をする状況を考える。ただしここでは、繰り返し質問による回答者への負担による影響は考えないものとする。





回答パターンを模式的に表すと、次のようになる。

Trials		$A(\pi)$						$A_{(1-\pi)}^c$			
		$T_1$	$T_2$	$T_1$	$F_2$	$F_1$	$T_2$	$F_1$	$F_2$	$T_1$	$T_2$
1	2	$1-\gamma_1$	$1-\gamma_2$	$1-\gamma_1$	$\gamma_2$	$\gamma_1$	$1-\gamma_2$	$\gamma_1$	$\gamma_2$		
		$\eta_{11}$		$\eta_{10}$		$\eta_{01}$		$\eta_{00}$			
$Q_1$	$Q_1$	Y	Y	Y	N	N	Y	N	N	N	N
$p_1$	$p_2$										
$Q_1$	$Q_2$	Y	N	Y	Y	N	N	N	Y	N	Y
$p_1$	$1-p_2$										
$Q_2$	$Q_1$	N	Y	N	N	Y	Y	Y	N	Y	N
$1-p_1$	$p_2$										
$Q_2$	$Q_2$	N	N	N	Y	Y	N	Y	Y	Y	Y
$1-p_1$	$1-p_2$										

Y : Yes , N : No

3. 2節で  $\hat{\pi}_{pr}$  を導いたのと同様の方法によって  $\pi$  の推定量  $\hat{\pi}_{tt}$  を求めると、

$$\hat{\pi}_{tt} = \frac{r_2^3 z_1 - r_1^3 z_2}{r_1 r_2 (r_1^2 - r_2^2)},$$

$$z_i = \hat{\lambda}_i - (1 - p_i) \quad (i = 1, 2).$$

推定量の形は 3. 2節で得えられた推定量  $\hat{\pi}_{pr}$  と同じである。

## 6. 2 推定量 $\hat{\pi}_{tt}$ の性質

(a) 2 回の回答の間の相関

同じ回答者が繰り返し質問をうけた場合、2つの回答の間には相関が生ずる。この相関を定式化するために、次の確率変数  $U_i$  ( $i = 1, 2$ ) を導入する。

$$U_i = \begin{cases} 1 & i \text{ 回目の質問に正直に答える。} \\ 0 & i \text{ 回目の質問に偽りの回答をする。} \end{cases}$$

( $U_1, U_2$ ) の同時分布を次のように与えるとき、

$$P(U_1 = 1, U_2 = 1) = \eta_{11},$$

$$P(U_1 = 1, U_2 = 0) = \eta_{10},$$

$$P(U_1 = 0, U_2 = 1) = \eta_{01},$$

$$P(U_1 = 0, U_2 = 0) = \eta_{00}.$$

$U_1$  と  $U_2$  の相関係数  $\rho(U_1, U_2)$  は、

以下の式を満たすことがわかる。

$$\eta_{00} = \gamma_1 \gamma_2 + \delta,$$

$$\eta_{11} = (1 - \gamma_1)(1 - \gamma_2) + \delta,$$

$$\eta_{10} = (1 - \gamma_1)\gamma_2 - \delta,$$

$$\eta_{01} = \gamma_1(1 - \gamma_2) - \delta,$$

$$\delta = \rho(U_1, U_2) \sqrt{\gamma_1 \gamma_2 (1 - \gamma_1)(1 - \gamma_2)}.$$

(b)  $E(\hat{\pi}_{tt}) = \pi$  (不偏推定量),

$$\begin{aligned}
(c) \quad V(\hat{\pi}_{tt}) &= B_1^2 V(z_1) + B_2^2 V(z_2) \\
&\quad - 2 B_1 B_2 \text{Cov}(z_1, z_2) \\
&= \frac{1}{n} \left[ B_1^2 \left[ \pi r_1^2 (1 - \gamma_1) \{ 1 - \pi (1 - \gamma_1) \} + \right. \right. \\
&\quad \left. \left. (1 - r_1^2) / 4 \right] + \left[ \pi r_2^2 (1 - \gamma_2) \{ 1 - \pi (1 - \gamma_2) \} + (1 - r_2^2) / 4 \right] \right. \\
&\quad \left. - 2 B_1 B_2 \pi \left[ (1 - \pi) \gamma_1 \gamma_2 (1 - \gamma_1) (1 - \gamma_2) + a r_1 r_2 \right. \right. \\
&\quad \left. \left. \times \rho(U_1, U_2) \sqrt{(1 - \gamma_1)(1 - \gamma_2)} \right] \right],
\end{aligned}$$

ここで、

$$B_1 = \frac{r_2^2}{r_1(r_1^2 - r_2^2)}, \quad B_2 = \frac{r_1^2}{r_2(r_1^2 - r_2^2)}.$$

## 7. $\hat{\pi}_{tt}$ と Warnerの推定量との比較

### 7. 1 繰り返し質問を行った場合の Warnerの推定量

$i$  回目の質問における  $y e s$  の数を  $m_i$  とし、 $\hat{\lambda}_i = m_i / n_i$  とすると、 $i$  回目の質問から得られる Warnerの推定量は

$$\hat{\pi}_{wti} = \frac{\hat{\lambda}_i - (1 - p_i)}{2 p_i - 1}, \quad (i = 1, 2).$$

ここで  $p_i$  は、 $i$  回目の質問において、 $Q. 1$  を選ぶ確率で、

$p_i \neq 1/2$ ,  $p_1 \neq p_2$  を満たすものとする。

このとき  $\hat{\pi}_{tt}$  と比較の対照となる自然な推定量は、次式で与えられる推定量  $\hat{\pi}_{wt}$  である。

$$\hat{\pi}_{wt} = (\hat{\pi}_{wt1} + \hat{\pi}_{wt2}) / 2$$

このとき  $\hat{\pi}_{wt}$  の分散と偏りは次式で与えられる。

$$V(\hat{\pi}_{wt}) = [V(\hat{\pi}_{wt1}) + V(\hat{\pi}_{wt2}) + 2 \text{COV}(\hat{\pi}_{wt1}, \hat{\pi}_{wt2})] / 4,$$

$$B(\hat{\pi}_{wt}) = \pi(\gamma_1 + \gamma_2 - 1).$$

### 7. 2 $\hat{\pi}_{tt}$ と $\hat{\pi}_{wt}$ との数値的比較

4. 2 節同様、最適なパラメータが選ばれた条件のもとでの  $\hat{\pi}_{tt}$  の分散  $V^*(\hat{\pi}_{tt})$ 、 $\hat{\pi}_{wt}$  の平均二乗誤差  $MSE^*(\hat{\pi}_{wt})$  の数値による比較を行う。繰り返し質問の場合、標本の大きさ  $n$  は一定だから、以下の最適化問題を考えればよい。

まず、 $V(\hat{\pi}_{tt})$  については、

条件： $p_{i0} \leq p_i$  ( $0 < p_{i0} < 1/2$ ;  $i = 1, 2$ ) の下で、

$V(\hat{\pi}_{tt})$  を最小にするように  $p_i$  を決める。

次に、 $MSE(\hat{\pi}_{wt})$  に関しては、2 回の質問は独立に行えるから、 $p_1 = p_2 = p$  とおき、次の最適化法を行った。

条件： $p_0 \leq p$  ( $0 < p_0 < 1/2$ ) の下で、

$MSE(\hat{\pi}_{wt})$  を最小にするように  $p$  を決める。

上の最適化法において、 $p_{i0}$ 、 $p_0$  は、それぞれ現実に使用可能な  $p_i$ 、 $p$  の下限値である。

表 5 には実際の数値実験結果が示されている。\* は 4. 2

節同様、最適なパラメータが選ばれた条件のもとで、計算した数値であるごとを示す。

表5  $V \cdot (\hat{\pi}_{tt})$ ,  $V \cdot (\hat{\pi}_{wt})$  と  $B \cdot (\hat{\pi}_{wt})$  の値

$p_1$	$\rho$	$\pi$		$a$		
				0.3	0.5	0.7
0.1	0.1	0.1	nvarp	2.1853	2.1949	2.2029
			nvarw	0.0567	0.0382	0.0228
			$n(Bias)^2$	0.8928	2.4801	4.8609
		0.2	nvarp	2.2727	2.2919	2.3079
			nvarw	0.1031	0.0708	0.0433
			$n(Bias)^2$	3.5713	9.9202	19.444
		0.3	nvarp	2.3401	2.3681	2.3907
			nvarw	0.1397	0.0984	0.0619
			$n(Bias)^2$	8.0354	22.321	43.748
	0.5	0.1	nvarp	2.1644	2.1644	2.1659
			nvarw	0.0627	0.0413	0.0367
			$n(Bias)^2$	0.8928	2.4801	4.8609
		0.2	nvarp	2.2322	2.2322	2.2352
			nvarw	0.1151	0.0907	0.0711
			$n(Bias)^2$	3.5713	9.9202	19.444
		0.3	nvarp	2.2799	2.2800	2.2844
			nvarw	0.1576	0.1283	0.1037
			$n(Bias)^2$	8.0354	22.321	43.748
0.9	0.1	nvarp	2.1434	2.1331	2.1275	
		nvarw	0.0687	0.0581	0.0507	
		$n(Bias)^2$	0.8928	2.4801	4.8609	
	0.2	nvarp	2.1897	2.1679	2.1560	
		nvarw	0.1270	0.1106	0.0990	
		$n(Bias)^2$	3.5713	9.9202	19.444	
0.3	nvarp	2.2149	2.1820	2.1629		
	nvarw	0.1755	0.1582	0.1455		
	$n(Bias)^2$	8.0354	22.321	43.748		

上の表は  $n = 1000$  の場合について計算されたものである。

また、 $nvarp = n V \cdot (\hat{\pi}_{tt})$ ,  $nvarw = n V \cdot (\hat{\pi}_{wt})$ ,

$n(Bias)^2 = n \{ B \cdot (\hat{\pi}_{wt}) \}^2$  である。

### 7. 3 数値実験結果の考察

表 5 より、以下のことがわかる。

(a) ほとんどの場合において、 $V^*(\hat{\pi}_{tt})$  は、

$V^*(\hat{\pi}_{wt})$  より大きいが、 $n$ ,  $\pi$ ,  $a$ , の値すべてが小さい場合を除いて、 $B^*(\hat{\pi}_{wt})$  は  $V^*(\hat{\pi}_{tt})$  に比べてかなり大きいため、 $V^*(\hat{\pi}_{tt})$  と  $MSE^*(\hat{\pi}_{wt})$  の比較を行うと、ほとんどの場合において、前者の方が、後者より小さい。

(b)  $V^*(\hat{\pi}_{tt})$  は、 $\rho$ ,  $\pi$  に関してそれぞれ、増加、減少関数であり、 $a$  に関しては、下に凸な関数である（定義域内で高々 1 つの最小値をもつ）が、これらのパラメータの値の変化に対して、極めて頑健である。

(c)  $MSE^*(\hat{\pi}_{wt})$  は、 $\pi$ ,  $a$ ,  $\rho$  に関して sensitive である。

## 8. 頑健性の問題と $\hat{\pi}_{tt}$ による推定手順

### 8. 1 パラメータに関する頑健性

表 6 には、 $p_1$  と  $\rho$  の値を固定し、 $\pi$  と  $a$  の値を動かしたときの、 $p_2$  の最適値  $p_2^*$  が与えられている。これらの値を見ると、 $p_1$  の値が固定された下では極めて頑健であることが判る。こ

のことから  $p_1$  と  $\rho$  の値が決められた下では、 $p_2$  の最適値として、 $p_2^*$  の  $a$  と  $\pi$  に関して平均をとった値  $p_2^{r^0}$  を使っても差し支えないといえる。さらに  $p_2^{r^0}$  の値は  $\rho$  の値の変化に対しても頑健である。従って  $p_2^{r^0}$  の値は  $a$ 、 $\pi$ 、 $\rho$  の値にほとんど依存しないので実際の調査において役に立つ値である。

### 8. 2 $\hat{\pi}_{tt}$ による $\pi$ の推定手順

8. 1 節における頑健性の検討の結果に基づいて、現実の調査における  $\hat{\pi}_{tt}$  による  $\pi$  の推定手順をまとめておく。

- 1° 標本の大きさ  $n$  と確率  $p_1$  を決める。
- 2° 与えられた表を用いて、1° で決めた  $p_1$  に対する推奨値  $p_2^{r^0}$  を決める。
- 3° 単純ランダム・サンプリングにより選ばれた  $n$  人の回答者にランダム回答法を適用し、 $\pi$  の推定値  $\hat{\pi}_{tt}$  を求める。
- 4° 3° で求めた推定値  $\hat{\pi}_{tt}$  を使い、 $a$  の推定値  $\hat{a}$  を計算し、 $\hat{\pi}_{tt}$  と  $\hat{a}$  を  $V(\hat{\pi}_{tt})$  の式に代入し、 $\hat{\pi}_{tt}$  の標準誤差を計算する。

#### 『注意』

5. 2 節の『注意』同様、上の推定手順を遡ることにより、要求される  $\hat{\pi}_{tt}$  の推定精度から、必要な標本の大きさ  $n$  の

おおよその値を決めることが出来る。また実際の調査における  $\hat{\pi}_{wt}$  による  $\pi$  の推定手順は提案出来ない。

表 6  $p_2^*$ ,  $p_2^{re}$  の値とそのロス

$\rho$	$p_1$	$\pi$	a			$p_2^{re}$	$L(p_2^{re})$		
			0.0	0.5	0.9				
0.1		0.10	0.252	0.256	0.256		0.0006		
		0.20	0.252	0.256	0.260		0.256	0.0006	
		0.30	0.252	0.260	0.260			0.0013	
	0.2		0.10	0.324	0.324	0.324		0.0000	
			0.20	0.324	0.324	0.324		0.324	0.0000
			0.30	0.324	0.324	0.324			0.0000
0.5	0.1	0.10	0.252	0.252	0.252		0.0000		
		0.20	0.252	0.252	0.252		0.252	0.0000	
		0.30	0.252	0.252	0.252			0.0000	
	0.2		0.10	0.324	0.324	0.324		0.0000	
			0.20	0.324	0.324	0.324		0.324	0.0000
			0.30	0.324	0.324	0.324			0.0000
0.9	0.1	0.10	0.252	0.256	0.256		0.0018		
		0.20	0.252	0.256	0.260		0.256	0.0040	
		0.30	0.252	0.260	0.260			0.0071	
	0.2		0.10	0.324	0.324	0.324		0.0000	
			0.20	0.324	0.324	0.324		0.324	0.0000
			0.30	0.324	0.324	0.324			0.0000



## 参 考 文 献

- Bourke, P. D., and Dalenius, T. (1974), "Randomized Response Models with lying", *Technical Report 71*, Institute of Statistics, University of Stockholm, Sweden.
- Chaudhuri, A., and Mukerjee, R. (1988), *Randomized Response: Theory and Techniques*, New York: Marcel Dekker.
- Chi, I. C., Chow, L. P., and Rider, R. V. (1972), "The Randomized Response Technique as used in the Taiwan Outcome of Pregnancy Study", *Studies in Family Planning*, Vol, 3, No. 11.
- Greenberg, B. G., Abul-Ela, Abdel-Latif A., Simmons, Walt R., and Horvitz, Daniel G. (1969), "The Unrelated Question Randomized Response Model Theoretical Framework," *Journal of the American Statistical Association*, 64, 529-539.
- Warner, S. L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, bf 60, 63-69.