ランダム回答法における二，三の注意*

—クロス集計にもとずく推定の精度，偽答・D.K. の影響，補助質問使用の問題—

統計数理研究所　鈴木達三
高橋宏一
逆瀬川浩孝

(1976年8月 受付)

Some notes on randomized response techniques

Tatsuzo Suzuki, Koiti Takahasi and Hirotaka Sakasegawa

(The Institute of Statistical Mathematics)

Let us suppose that each individual belongs to either group A or Ā and also to either group B or B̄ and that for a random sample the randomized response approaches are applied to the estimations of the proportions of A and B in a population, separately. We first consider the precision of the estimator of the proportion A∩B based on the cross tabulation data.

Secondly, we consider the effects of untruthful reporting and no answer to the bias in the estimation.

Finally, we investigate the possibility of using auxiliary questions instead of randomizing devices and the bias introduced by the dependency of auxiliary questions on main questions.

§1. はじめに

母集団の各人はグループ A かグループ $\bar{A}$ のいずれか一方に属するものとする．またグループ B かグループ $\bar{B}$ のいずれか一方に属するものとする．2つの分類の組合せによって，母集団の各人は AB, $A\bar{B}$, $\bar{A}B$, $\bar{A}\bar{B}$ のいずれかに属することになる．母集団における A の比率や B の比率のほかに，AB, …, $\bar{A}\bar{B}$ の比率も推定したいときに，「あなたは A ですか？」という質問と，「あなたは B ですか？」という質問がなされていれば，AB, …, $\bar{A}\bar{B}$ の比率は2問のクロス集計の結果から推定される．A の母比率を π_A, B のそれを π_B, AB, …, $\bar{A}\bar{B}$ の母比率を π_{AB}, …, $\pi_{\bar{A}\bar{B}}$ と書くことにすると，母比率の推定量としての標本比率の分散は $\pi_A(1-\pi_A)/n$, $\pi_B(1-\pi_B)/n$, $\pi_{AB}(1-\pi_{AB})/n$, …, $\pi_{\bar{A}\bar{B}}(1-\pi_{\bar{A}\bar{B}})/n$ であり（簡単のため，以下復元無作為抽出を仮定する．n は標本の大きさ)，したがって，通常の調査法を用いている場合には，クロス集計にもとずく π_{AB}, …, $\pi_{\bar{A}\bar{B}}$ の推定において，推定の精度という点で特別な問題は生じない．

しかしながら，グループ A に属するかどうかということも，グループ B に属するかどうかということも共に答を知られたくないことであり，ウォーナ法 [7]，無関連質問法 [1]，モートン法 [1] などのランダム回答法をそれぞれの質問に適用した場合には，クロス集計の結果にもとずく π_{AB}, …, $\pi_{\bar{A}\bar{B}}$ の推定の精度は π_A や π_B の推定の精度と事情はかなりちがって

---

* この研究の一部は，文部省科学研究費補助金
昭和48年度一般研究（D）863012「回避的な質問項目の調査法」及び昭和51年度試験研究（2）183001「社会調査の標準化過程における回答誤差の研究」によるものである．

くる．このことについて注意を述べるのがこの論文の一つの目的である．

ランダム回答法は，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. への逃避をふせぐために本来考えられたのではあるが，質問の性質上，偽りの回答や D.K. への逃避がある程度存在し [3]，しかもそれが推定に偏りを惹き起こす恐れもある．実際の偽回答や D.K. の起こり方についてはまだ研究がなされていないので，いくつかの予想される起こり方について偏りへの影響を調べる．

ランダム回答法では，カードをはじめとして，質問を確率的に選択する，あるいは回答を確率的に変換するために何らかのランダム化器具を使用する．それらの器具のとり扱いに回答者が不慣れであるなどして，正しい使い方がなされないと推定の結果は信用できないものになってしまう．また，ランダム化器具を使用することは留置き調査法，電話調査法，郵便調査法などにとって困難である．そこでランダム化器具を使用するかわりに，補助質問を使用することが考えられる．このことに関連して，他の多くの質問と独立であるような質問の有無が問題になる．現在のところ，他の多くの質問と独立であるような質問を見つけることはかなり難しそうに思われる．しかも，独立性がくずれることは，推定に重大な偏りをもたらすことになる．

§2. クロス集計にもとずく推定の精度

グループ A に属する，属さないということも，グループ B に属する，属さないということもいずれも他人には知られたくないことがらであるとする．グループ A，グループ B およびグループ AB (A にも B にも属する人からなるグループ)，グループ $A\bar{B}$ (A に属し，B に属さない人からなるグループ)，グループ $\bar{A}B$，グループ $\bar{A}\bar{B}$ の母集団比率を推定したいものとする．そこで，直接に「あなたは A ですか？」，あるいは「あなたは B ですか？」ときくかわりに，それぞれに対してランダム回答法を適用した場合を考える．たとえば赤カードの割合を p_A，青カードの割合を $q_A\,(=1-p_A)$ とし，その中から1枚ぬき出して（調査員には見せない），回答者はそれが赤なら A に属するとき "1"，$\bar{A}$ に属するとき（A に属さないこと）"0" と答え，カードが青なら A に属するとき "0"，$\bar{A}$ に属するとき "1" と答える．このとき π_A の最尤推定量は，n 人の回答者中の "1" と答えた人数を x_A とすれば

$$\hat{\pi}_A = \frac{\dfrac{x_A}{n} - q_A}{p_A - q_A} \tag{1}$$

で与えられる．たゞし $p_A \neq 1/2$ とする．またその分散は

$$V(\hat{\pi}_A) = \frac{\pi_A(1-\pi_A)}{n} + \frac{p_A q_A}{n(p_A - q_A)^2} \tag{2}$$

である [7]．

π_B も同じ方法で推定するものとする．カード中の赤の割合は p_A と同じとは限らないのでそれを $p_B\,(\neq 1/2)$ とし，n 人の回答者中の "1" と答えた人数を x_B とすれば，(1)，(2) に対応して

$$\hat{\pi}_B = \frac{\dfrac{x_B}{n} - q_B}{p_B - q_B} \tag{3}$$

$$V(\hat{\pi}_B) = \frac{\pi_B(1-\pi_B)}{n} + \frac{p_B q_B}{n(p_B - q_B)^2} \tag{4}$$

となる．

次にπ_{AB}, $\pi_{A\bar{B}}$, $\pi_{\bar{A}B}$, $\pi_{\bar{A}\bar{B}}$ の推定を考える．A に対するランダム回答法で "i"，B に関するランダム回答法で "j" (i, j=0, 1) と答えた人数を y_{ij} とする．$n=y_{11}+y_{10}+y_{01}+y_{00}$ である．π_{AB}, $\pi_{A\bar{B}}$, $\pi_{\bar{A}B}$, $\pi_{\bar{A}\bar{B}}$ の最尤推定量を $\hat{\pi}_{AB}$, $\hat{\pi}_{A\bar{B}}$, $\hat{\pi}_{\bar{A}B}$, $\hat{\pi}_{\bar{A}\bar{B}}$ とすると，これらは次の連立１次方程式の解として与えられることは明きらかである．

$$\begin{pmatrix} q_Aq_B & q_Ap_B & p_Aq_B & p_Ap_B \\ q_Ap_B & q_Aq_B & p_Ap_B & p_Aq_B \\ p_Aq_B & p_Ap_B & q_Aq_B & q_Ap_B \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\pi}_{AB} \\ \hat{\pi}_{A\bar{B}} \\ \hat{\pi}_{\bar{A}B} \\ \hat{\pi}_{\bar{A}\bar{B}} \end{pmatrix} = \begin{pmatrix} y_{00}/n \\ y_{01}/n \\ y_{10}/n \\ 1 \end{pmatrix} \tag{5}$$

ここで，

$$\varDelta = \frac{\dfrac{y_{11}}{n} - \dfrac{y_{1\cdot}}{n} \cdot \dfrac{y_{\cdot 1}}{n}}{(p_A - q_A)\,(p_B - q_B)}, \tag{6}$$

ただし，$y_{1\cdot} = y_{10} + y_{11}$, $y_{\cdot 1} = y_{01} + y_{11}$, とおくと

$$\begin{aligned} \hat{\pi}_{AB} &= \hat{\pi}_A \cdot \hat{\pi}_B + \varDelta, \\ \hat{\pi}_{A\bar{B}} &= \hat{\pi}_A \cdot (1 - \hat{\pi}_B) - \varDelta, \\ \hat{\pi}_{\bar{A}B} &= (1 - \hat{\pi}_A) \cdot \hat{\pi}_B - \varDelta, \\ \hat{\pi}_{\bar{A}\bar{B}} &= (1 - \hat{\pi}_A)\ (1 - \hat{\pi}_B) + \varDelta \end{aligned} \tag{7}$$

と表すことができる．

推定量の分散は，たとえば $\hat{\pi}_{AB}$ では

$$\begin{aligned} V\,(\hat{\pi}_{AB}) = {} & \frac{\pi_{AB}(1-\pi_{AB})}{n} + \frac{p_Aq_A}{n\,(p_A - q_A)^2}\pi_B \\ & + \frac{p_Bq_B}{n\,(p_B - q_B)^2}\pi_A + \frac{p_Aq_A\,p_Bq_B}{n\,(p_A - q_A)^2\,(p_B - q_B)^2} \end{aligned} \tag{8}$$

となる．$\hat{\pi}_{A\bar{B}}$, $\hat{\pi}_{\bar{A}B}$, $\hat{\pi}_{\bar{A}\bar{B}}$ の分散は (8) の π_{AB}, π_A, π_B をそれぞれ $\pi_{A\bar{B}}$, π_A, $1-\pi_B$; $\pi_{\bar{A}B}$, $1-\pi_A$, π_B; $\pi_{\bar{A}\bar{B}}$, $1-\pi_A$, $1-\pi_B$ でおきかえればよい．

とくに，$p_A=p_B=p=1-q$ とすると

$$V\,(\hat{\pi}_{AB}) = \frac{\pi_{AB}(1-\pi_{AB})}{n} + \frac{pq}{n\,(p-q)^2}\,(\pi_A + \pi_B) + \frac{p^2q^2}{n\,(p-q)^4} \tag{9}$$

である．

(2) や (4) と同じように，$\hat{\pi}_{AB}$ の分散も，標本抽出による成分（(8) の右辺の第１項）と，ランダム回答法の使用による成分（(8) の右辺の第２項以下）の和にわかれる．また，グループ B に属するかどうかが他人に知られてもよいことであるなら，ランダム回答法を使用しないで直接質問するであろう．この場合は (8) において p_B=1, q_B=0 の場合に相当し，

$$V\,(\hat{\pi}_{AB}) = \frac{\pi_{AB}\,(1 - \pi_{AB})}{n} + \frac{p_Aq_A}{n\,(p_A - q_A)^2}\pi_B \tag{10}$$

となる．

ランダム回答法における推定量の分散は，(2)，(4)，(8) で見られるように，標本抽出による成分は殆んど影響力をもたず，それ以外の成分，すなわちランダム回答法使用に基因する成分が支配的である．この後者の成分は $\hat{\pi}_A$ では

$$\frac{p_Aq_A}{n\,(p_A - q_A)^2} \tag{11}$$

であり，π_A に無関係である．$\hat{\pi}_{AB}$ では

$$\frac{p_Aq_A}{n(p_A-q_A)^2}\pi_B+\frac{p_Bq_B}{n(p_B-q_B)^2}\pi_A+\frac{p_Aq_Ap_Bq_B}{n(p_A-q_A)^2(p_B-q_B)^2} \tag{12}$$

であり，π_A や π_B には関係するが π_{AB} には無関係であることがわかる．さらにこの成分の中で，p が 1/2 に近いときには (12) の右辺の最後の項が支配的になり，この項は π_A や π_B にも無関係である．

ここで大体のようすを知るために，$p_A=p_B=p=1-q$ の場合について数値例をあげてみよう．

$$f(p)=pq=p(1-p), \tag{13}$$

$$g(p)=pq/(p-q)^2=p(1-p)/(2p-1)^2 \tag{14}$$

第　1　表

p	$g(p)$	$g^2(p)$	$f(p)$
0.05	0.0586	0.0034	0.0475
0.10	0.1406	0.0198	0.0900
0.15	0.2602	0.0677	0.1275
0.20	0.4444	0.1975	0 1600
0.25	0.7500	0.5625	0.1875
$(5-\sqrt{5})/10(\fallingdotseq 0.2764)$	1.0000	1.0000	0.2000
0.30	1.3125	1.7227	0.2100
1/3	2.0000	4.0000	0.2222
0.35	2.5278	6.3897	0.2275
0.40	6.0000	36.0000	0.2400
0.45	24.7500	612.5625	0.2475

とおく．第1表を使えば必要な量は簡単に計算されるだろう．たとえば，$100\pi_A=20$(%)，$100\pi_B=30$(%)，$100\pi_{AB}=10$(%)，$n=1000$ として具体的に π_A, π_B, π_{AB}, $\pi_{\bar{A}B}$ $(=0.6)$ の推定量の標準誤差×2×100(%) を求めてみると，もし，普通の調査項目のように，直接質問をして正しい答が得られるものとすれば，π_A, π_B, π_{AB}, $\pi_{\bar{A}B}$ の推定量の分散は標本抽出による分散であるから，求める量はそれぞれ $200\sqrt{f(0.2)/1000}$, $200\sqrt{f(0.3)/1000}$, $200\sqrt{f(0.1)/1000}$, $200\sqrt{f(0.4)/1000}$, すなわち 2.5%，2.9%，1.9%，3.1% である．次に $p_A=p_B=p$ のランダム回答法を用いた場合は，求める量は $200\sqrt{(f(0.2)+g(p))/1000}$, $200\sqrt{(f(0.3)+g(p))/1000}$, $200\sqrt{\{f(0.1)+(0.2+0.3)g(p)+g^2(p)\}/1000}$, $200\sqrt{\{f(0.4)+(0.8+0.7)g(p)+g^2(p)\}/1000}$ となる．具体的に数値を求めて第2表にまとめてみる．

第2表を見ると，(8) から予想されるように，p が 0.5 に近くなると，π_A や π_B の推定の精度に比して，π_{AB} や $\pi_{\bar{A}B}$ の推定の精度はかなり悪くなる．第2表からすぐわかることであるが，たとえば $p=1/3$ のときに π_A の推定量の標準誤差の2倍を 5% にするためには $n=3456$ であるが，π_{AB} について同じ条件をみたすには $n=8144$ であり，$p=0.4$ のときには，同じ条件をみたすには π_A については $n=9856$ であるが，π_{AB} については $n=62544$ となる．このように，π_A や π_B の推定の精度を考えて標本の大きさを決めたとしても，p が 0.5 に近い場合（大まかにいって $0.3<p<0.7$）には，クロス集計にもとずく π_{AB} や $\pi_{\bar{A}B}$ の推定の精度は保障されているとはいえない．

第 2 表

		200×標準誤差（%），(n=1000)			
		π_A の推定	π_B の推定	π_{AB} の推定	$\pi_{\overline{AB}}$ の推定
直接質問		2.5	2.9	1.9	3.1
ランダム回答法	p=0.05	3.0	3.3	2.2	3.6
	0.10	3.5	3.7	2.7	4.3
	0.15	4.1	4.3	3.4	5.3
	0.20	4.9	5.1	4.5	6.6
	0.25	6.0	6.2	6.4	8.8
	0.30	7.7	7.8	9.9	12.5
	1/3	9.3	9.4	14.3	17.0
	0.35	10.4	10.5	17.6	20.4
	0.40	15.7	15.8	39.5	42.5
	0.45	31.6	31.6	158.1	161.2

§3. 偽りの回答と D.K. の影響

直接質問したのでは，答を偽ったり，D. K.（わからない）と答えたりするおそれの多分にあるような場合に，そうした偽りの回答や D. K. をなくす目的でランダム回答法が考えられた．しかしながら，質問がプライヴエートなものであったり，デリケートなものであったりする関係上，ランダム回答法を使用した場合，原理は理解されたが，なおかつ偽りの回答や D. K. が起こることが考えられる（原理が理解できずにまったく出たら目に答えるといった種類の偽答はここでは考えない）．そこで，ランダム回答法における偽りの回答や D. K. が推定に及ぼす影響について調べてみる．

まだ実地調査結果にもとずく研究はほとんどなされていないので，偽回答や D. K. の起こり方について適当な仮定をたてて解析することにする．数学的には同等な方法であっても，心理的にちがった印象をもたらす方法では，偽回答や D. K. の起こり方は同じと考えることはできない．

回答者は赤，青のカードの割合が p: q $(=1-p)$ であるカードの中から1枚とり出し，

$\overline{A}$ で青の人は "1"，$\overline{A}$ で赤の人は "0"

A で赤の人は "1"，A で青の人は "0"

と答えるウォーナ法では，（赤，青のカードの割合に関する確率的判断を別にすると）回答者は "1" と "0" を回答する場合どちらかを特に嫌う根拠がない．この場合には，指示通りに答えれば "1" なのに偽って "0" と答える，あるいは逆に指示通りに答えれば "0" なのに "1" と答えるというような偽りの回答の危険はほとんどないものと考えられる（p を "0"，あるいは "1" にあまり近ずけすぎると偽りの回答をする可能性が無視できなくなるだろう）．確率的な判断までして偽るとすれば，たとえば A であると思われたくない場合に，$p>1/2$ のときには，"1" と答えるべき人が "0" と答える可能性があるが，その可能性は，極めて少いと考えられるので，ここでは，ウォーナ法における偽答による偏りはとり上げないことにする．

一方モートン法は赤，青，黄のカードの割合が p_r: p_b: p_y $(p_r+p_b+p_y=1)$ であるカードから1枚とり出し，

赤の人は A でも $\overline{A}$ でも "1"

青の人は A でも $\overline{A}$ でも "0"

黄の人は A なら "1"，$\overline{A}$ なら "0"

という方法である．これは $p_y=|2p-1|$，$p_r=p_b$ のとき，前述のウォーナ法と数学的モデルは同一である [5]．しかし，回答者の受ける心理的印象は必ずしも同じとは考えられない．上の指示を見ると，“0” という回答をすれば

黄で $\bar{A}$ か，青で A か $\bar{A}$ かは全く情報を与えていない

ということを意味し，一方，“1” という回答は

黄で A か，赤で A か $\bar{A}$ かは全く情報を与えていない

ということを意味する．$\bar{A}$ であることは人に知られてもかまわないが，A であることは知られたくないものとする．このような場合には1が A に対応し，0が $\bar{A}$ に対応するように受けとられる恐れがあり，確率的な判断以前に，心理的には “1” よりも “0” と回答することに安心感をもつことが考えられる（たとえば，$\bar{A}$ の人が赤カードをひいたとき，1と答えることに抵抗を感じるかもしれない）．

D. K. の起こり方についてはいろいろ考慮すべき点があり，簡単にその起こり方についてのモデルを設定することはできない．たとえば，カードをとり出し，質問文を読んだあとで D. K. と答えることは，意見に関する質問なら十分起こることが考えられるが，事実の有無などについての質問であると回答者としても途中から D. K. に移ることはしにくく，むしろ回答を偽る方向に進もうとするかもしれない．D. K. について，ここでは，その起こり方に実際的な意味づけをしないで，推定における偏りが D. K. の起こり方との間にどのような関係があるかを数値的に考察するにとどめることにする．

なお，偽りの回答や D. K. のほかに，“0” や “1” で答えるかわりに，直接 A であるか，$\bar{A}$ であるかを答える回答者が出てくることもあり得る．調査法を理解しないためのことではなくて，ここでいっているのは，調査法を理解した上で，自分が A であること，あるいは $\bar{A}$ であることをはっきり答えたいという理由で起こる場合である．この種の回答は，そこにまた偽りが入るようなときはとり扱いはきわめて困難であるし，またどの段階まで進んで直接 A，$\bar{A}$ で答えることを選んだかによって事情は異ってくる．たとえば，カードをとり出す前に A，$\bar{A}$ で答えるのと，カードをとり出したあとで A，$\bar{A}$ で答えるのとでは，とり扱いがちがってくる．前者の場合しか起こらないなら，問題は比較的簡単で，事後にランダムにカードを切って “0”，“1” に割り当てれば済むことである．しかし，後者の場合も含まれてくると，π_A の推定は不可能になる．この種の回答のとり扱いは，ここではとり上げない．

元に戻って，まず偽りの回答の影響を調べよう．無関連質問法の場合については [1] に述べられている．ウォーナ法は，すでに述べたように，偽回答は起こりにくいと考えられるので，モートン法について述べる．上述の説明で “0” と回答すべき人はそのまゝ “0” と回答し，“1” と回答すべき人については赤をとり出したため “1” と答えるべき人についてはその中の γ の割合の人が，黄をとり出し A であるために “1” と答えるべき人についてはその中の δ の割合の人が偽って “0” と回答するものとする．標本の大きさを n，“1” という回答数を y とする．母集団における A に属する人の割合 π の推定量としては，偽回答がない場合の最尤推定量

$$\hat{\pi}=\frac{\dfrac{y}{n}-p_r}{p_y} \tag{15}$$

が用いられる．偽回答がない場合には勿論 $\hat{\pi}$ は π の不偏推定量である．上述のような偽回答がある場合には

$$E(\hat{\pi})=\frac{p_y\pi(1-\delta)+p_r(1-\gamma)-p_r}{p_y}=\pi-\left(\pi\delta+\frac{p_r\gamma}{p_y}\right) \tag{16}$$

となり，偏り $\beta(\pi, \delta, \gamma)$ は

$$\beta(\pi, \delta, \gamma) = -\left(\pi\delta + \frac{p_r\gamma}{p_y}\right) \tag{17}$$

である．これから直ちにわかることは，上述の偽回答の起こり方に関する仮定から予想されるように偏りはマイナスであること，p_y は大きく p_r は小さいことが偏りを減少させることがわかる．しかし，p_r の小さいことも，p_y の大きいことも，いずれも回答者の協力を得やすい方向とは反対である．$\delta=\gamma$，$p_r=p_b=(1-p_y)/2=1-p=q$ $(p>1/2)$ とおくと

$$\beta(\pi, \gamma, \gamma) = -\gamma\left(\pi + \frac{q}{p-q}\right) \tag{18}$$

である．$h(p)=q/(p-q)=(1-p)/(2p-1)$ とおく．$h(p)$ の値は次表のようになる．

第　3　表

p	0.55	0.6	2/3	0.7	0.75	0.8	0.85	0.9	0.95
$h(p)$	4.500	2.000	1.000	0.750	0.500	0.333	0.214	0.125	0.056

たとえば，γ が 0.1 以下，π は 0.3 以下として，偏りが 0.1 以下におさまるためには，p は大体 0.7 以上でなければならないことがわかる．デリケートな質問であるからといって，p を 0.5 に近ずける場合には，もし偽回答があると偏りへの影響は重大である．デリケートな質問ほど偽回答のおそれがあることを考えると，モートン法における偽回答の有無を実際の場面で検討することをゆるがせにはできないし，場合によってはウォーナ法を採用する方が無難かもしれない．

次に，D.K. の影響を調べよう．前述したように，現在のところ D.K. の起こり方について適当なモデルをつくることはできないので，D.K. の割合を ε として，偏りの大きいのはどんなときか，小さいのはどんなときかを調べる．偽りの回答はないものとして，ウォーナ法の場合について考察する．一般的にして，D.K. は

赤で A の人の $100a$%，赤で $\bar{A}$ の人の $100b$%，

青で A の人の $100c$%，青で $\bar{A}$ の人の $100d$%

という割合で起こるものとしよう．D.K. の割合が ε ということは，$\bar{\pi}=1-\pi$ とおいて

$$p\pi a + p\bar{\pi}b + q\pi c + q\bar{\pi}d = \varepsilon \tag{19}$$

と表わせる．標本の大きさを n，"1" と答えた人数を y，D.K. の人数を z とする．π の推定量としては

$$\hat{\pi} = \frac{\dfrac{y}{n-z} - q}{p-q} \tag{20}$$

を用いるものとする．$\varepsilon=0$ なら，これは π の不偏推定量である．n は大きいものとして

$$E\left(\frac{y}{n-z}\right) \fallingdotseq \frac{E(y)}{E(n-z)}$$

という近似を用いると

$$E(\hat{\pi}) = \frac{\dfrac{p\pi(1-a) + q(1-\pi)(1-d)}{1-\varepsilon} - q}{p-q} \tag{21}$$

となり，偏り $\beta(a, b, c, d)$ は

$$\beta(a, b, c, d)=\frac{\varepsilon(p\pi+q\bar{\pi})-(p\pi a+q\bar{\pi}d)}{(1-\varepsilon)(p-q)} \tag{22}$$

となる．$a=b=c=d$ とすると $\varepsilon=a$ であり

$$\beta(a, a, a, a)=0 \tag{23}$$

である．このことは，D.K. が A か $\bar{A}$ かということにも，どのカードをひいたかということにも無関係に起こるなら，偏りは生じないことを示している．

次に $a=b$, $c=d$ とすると，$\varepsilon=pa+qc$ であり，

$$\beta(a, a, c, c)=\frac{pq(c-a)(\pi-\bar{\pi})}{(1-\varepsilon)(p-q)} \tag{24}$$

となる．当然予想されるように，ε が一定のときには，A か $\bar{A}$ かということでD.K. の割合がきまるときには，c と a の差が大きい程偏りは大きいことがわかる．

p, ε が与えられたとき，(22) の絶対値は $a=d=\varepsilon$ のとき最小で0になること ($a=d=\varepsilon$ は $b=c=\varepsilon$ とおけば (19) を満たす) がわかる．ところで，

$$\min\{p, q\}\leq p\pi+q\bar{\pi},\ p\bar{\pi}+q\pi\leq\max\{p, q\} \tag{25}$$

であり，$\varepsilon<\min\{p, q\}$ であると考えてよい ($\min\{p, q\}$ は 0.2, 0.25, 0.3, 1/3, 0.4 などに定められるであろうが，0に近くできるということは質問がそれほどデリケートでないことを意味し，そのときは D.K. も少いと考えられる)．このとき，(22) に示されている偏りの絶対値の最大値は

$$|\beta(a, 0, 0, d)|=\frac{\varepsilon(p\bar{\pi}+q\pi)}{(1-\varepsilon)|p-q|} \tag{26}$$

と

$$|\beta(0, b, c, 0)|=\frac{\varepsilon(p\pi+q\bar{\pi})}{(1-\varepsilon)|p-q|} \tag{27}$$

のうちの大きい方である．すなわち，$p\pi+q\bar{\pi}\geq0.5$ なら (27)，$p\pi+q\bar{\pi}\leq0.5$ なら (26) である．(26) では $b=c=0$, すなわち D.K. はすべて "1" と答えるべき人によるものであり，(27) では $a=d=0$, すなわち D.K. はすべて "0" と答えるべき人によるものである．これも当然予想されたことであり，D.K. の影響は，D.K. が一方の答をすべき人に偏って起こるときに大きく，いずれの答をすべき人にも同じ割合で起こるときは小さい（推定の偏りのみを問題にしている）という結論になる．

$\pi=0.2$, $\varepsilon=0.05$, $p=0.25$ とすると

$$\beta(a, b, c, d)=\frac{0.0325-(0.05a+0.6d)}{0.95\times0.5} \tag{28}$$

であるが，$x=0.05a+0.6d$ とおくと，(19) より，$0.05a+0.2b+0.15c+0.6d=0.05$ であるから，$0\leq x\leq0.05$ であり，

$$\beta(a, b, c, d)=\frac{0.0325-x}{0.475} \tag{29}$$

と表せる．数値的には第4表のようになる．

同じ π, ε に対して，$p=1/3$, $p=0.4$ の場合を調べてみる．前者では $x=0.067a+0.533d$ とおいて

第 4 表

x	0.0	0.01	0.02	0.03	0.0325	0.04	0.05
$\beta(a, b, c, d)$	0.068	0.047	0.026	0.005	−0.000	−0.016	−0.037

$$\beta(a,\ b,\ c,\ d)=\frac{0.03-x}{0.3167} \tag{30}$$

後者では $x=0.08a+0.48d$ とおいて

$$\beta(a,\ b,\ c,\ d)=\frac{0.028-x}{0.19} \tag{31}$$

である．いずれの場合も $0\leq x\leq 0.05$ である．それぞれの数値は第5表，第6表のようになる．

第 5 表

x	0.00	0.01	0.02	0.03	0.04	0.05
β	−0.095	−0.063	−0.032	0.000	0.032	0.063

第 6 表

x	0.00	0.01	0.02	0.028	0.03	0.04	0.05
β	−0.147	−0.095	−0.042	0.000	0.011	0.063	0.116

これらの表を見ると，全体としてのD.K.が5％のとき，$p=0.25$ なら最悪のときでも偏りは6.8％以下であるが，p が1/3，0.4とだんだん0.5に近ずくと最悪の場合は9.5％，14.7％となる．勿論，すでに述べたように，ランダム回答法における D.K. は，ランダム化器具を使い，自分が“0”と答えるべきか，“1”と答えるべきかを知ってから D.K. と答えたり，回答を拒否したりすることは，とくにウォーナ法の場合，少いとも考えられる．D.K. があるとしても，その内容は上述の最悪の場合よりは，むしろランダム化器具を使って問題に入る以前に回答を拒否する場合が多いものと思われる．もしそうだとすると，偏りへの影響を調べるには(24)の方がよいかもしれない．(24)の絶対値の最大値は，$\varepsilon<p<0.5$ のときは，$a=\varepsilon/p$，$c=0$ のときに起こり

$$|\beta(\varepsilon/p,\ \varepsilon/p,\ 0,\ 0)|=\frac{pq\,(\varepsilon/p)\,|\pi-\bar{\pi}|}{(1-\varepsilon)\,|p-q|}=\frac{\varepsilon|\pi-\bar{\pi}|q}{(1-\varepsilon)\,|p-q|} \tag{32}$$

である．$\pi=0.2$，$\varepsilon=0.05$ とすると(32)は $0.0316q/|p-q|$ となる．これは $p=0.25$，1/3，0.4の場合に，それぞれ0.047，0.063，0.095である．

§4. 質問によるランダム化器具の代用

ランダム回答法では何らかのランダム化器具，たとえばカード，サイコロ，箱に入った色つきビーズ [2]，フラスコに入った色つきボール [3] などが用いられる．§1に述べたように，面接法以外への応用という面からはランダム化器具を使わないと便利である．上述のランダム化器具のほかに乱数表を使用させることが考えられる．しかし一般の人に正しい使い方を期待することはできないであろう．ここではランダム化器具の役目を補助質問によって果たすことの可能性を調べる．結論は否定的である．

ランダム回答法に用いられるランダム化器具の主要な特徴は次の3つである：

(i)　ランダム化器具による試行の結果は回答者の如何なる属性とも，如何なる意見とも独立である．

(ii)　試行の結果に関する確率は既知であり，一般にコントロールできる（たとえば，赤カードと青カードの割合は既知であるし，適当にきめることができる）．

(iii)　とり扱いが簡単でなければならない．

もし，ある2項質問に対する答 C と $\bar{C}$ の選択が，上の性質 (i)，(ii)，(iii) をみたすなら，それで問題は解決である．すなわち，その2項質問に対する答が回答者のもつ属性，意見などと独立であり，C あるいは $\bar{C}$ の選ばれる確率が既知であり，質問の内容が簡単でわかりやすいものとすれば，

> あなたの補助質問に対する答が C ならば「あなたは A ですか？」という質問に“はい”か“いいえ”で答えて下さい．また，あなたの補助質問に対する答が $\bar{C}$ ならば「あなたは $\bar{A}$ ですか？」という質問に“はい”か“いいえ”で答えて下さい．

という要領でランダム回答法を行なうことができるわけである．勿論，C を選ぶ確率を p_c とするとき，$p_c \neq \frac{1}{2}$ でなければならない．また，カードの場合，赤カードと青カードの割合を調節することができたが，補助質問の場合には p_c は既知であっても変えるわけにはいかないから，希望する値をもつような補助質問を探さなければならない．この点は別にしても，p_c が既知という条件をみたすものも簡単には考えつかない．勿論，補助質問の答が C か $\bar{C}$ かを回答者は誰にも告げないことにより，回答の秘密を守っているのだから，個々の回答者について C か $\bar{C}$ かがわかってしまうものでは意味がない．「あなたは早うまれですか」という質問では p_c は既知と考えられるが，回答者が実際に早うまれかどうかはわかってしまうから，これを補助質問に使うことはできない．p_c が未知であるとしても，他の標本に補助質問だけを実施して p_c を推定することはできる．

問題は (i) に述べた独立性である．その質問に対する答が，回答者の属性や意見などと独立であるような質問が見つかるかどうかということが当面の関心事になる．この点についてしばらく考察しよう．

ある統計のクラスで，A，B，C，D，E，F の6色の中でどの色が一番好きかということと，日本とソ連の外交交渉で日本は譲歩しすぎると思うかということを調べクロス集計をしたものが第7表である．イェーツの修正をした χ^2 の値は 3.09 で，これは自由度1の χ^2 分布の上側 7.9% 点である．この例からは，どの色が好きかといった問題でも外交交渉についての意見などの問題と必らずしも独立とはいいきれないことが考えられる．あるいは，「国民性調査　1973年」の結果の中に次のような例が見られる．問1「もういちど生れかわるとしたら，あなたは男と女のどちらに，生れてきたい

第　7　表

好きな色＼意見	思　う	思わない	計
ABC	23	32	55
DEF	13	6	19
計	36	38	74

第　8　表

性別＼問1	男に生れたい	女に生れたい
男	599(428)	34(205)
女	371(542)	432(261)

第　9　表

性別＼問2	働　く	や め る
男	472(470)	164(166)
女	607(609)	216(214)

と思いますか？」，問2「もし，一生楽に生活できるだけのお金がたまったとしたら，あなたはずっと働きますか，それとも働くのをやめますか？」という2つの質問と性別の3つの項目の中の2つづつのクロス集計を第8，9，10表に示す(( )内は独立としたときの期待値．D.K. やその他が除かれている)．お金がたまっても働くかどうかという問2を中心に考えると，第9表から見る限り性別とはほとんど関係がない．一方．第10表から見る限り問2は問1と独立であるとはいえない($\chi^2=9.14$，この値は自由度1の χ^2 分布の上側0.25%点)．ところで第8表を見ると性別と問1の間には，かなりはっきりした相関が見られる．問1は普通に考えると性別と密接する質問群に入れられるであろう．すべての種類の属性や意見と独立な質問を見つけることは無理としても，ある種の質問群とは独立な質問なら見つけることができるであろうと考えるかもしれないが，上の例で見ると，簡単にそういえないことがわかる．相関の有無を予想することが難しい例として次のものをあげることができよう．同じ調査の問20「自然と人間との関係について，つぎのような意見があります．あなたがこのうち真実に近い（ほんとうに近い）と思うものを，ひとつだけえらんで下さい？　1.　人間が幸福になるためには，自然に従わなければならない．2.　人間が幸福になるためには，自然を利用しなければならない．3.　人間が幸福になるためには，自然を征服してゆかなければならない」と，性別，前出の問1，問21「つぎのような意見があります．あなたはどちらに賛成ですか．もちろん，場合により，また程度によって違うでしょうが，ひとくちでいうと，どちらを重視すべきでしょうか？　1.　個人の権利をみとめるためには，公共の利益が多少犠牲になることがあっても，しかたがない．2.　公共の利益のためには，個人の権利が多少犠牲になることがあっても，しかたがない」との関係がどうなっているかを予想しよう．次のクロス集計の結果と予想とが3つとも一致する人はそう多くないかもしれない．問20と性別は有意水準1%で

第　10　表

問1＼問2	働　く	やめる
男　に	709(686)	217(240)
女　に	306(329)	139(116)

第　11　表

問20＼性別	男	女
従　う	190(217)	297(270)
利　用	323(301)	352(374)
征　服	126(120)	144(150)

第　12　表

問20＼問1	男　に	女　に
従　う	308(312)	148(144)
利　用	442(436)	195(201)
征　服	172(174)	82 (80)

独立性は棄却される．第12表，第13表から明らかに，問20と問1，問20と問21は独立であることを棄却できない（χ^2 値はそれぞれ自由度2の χ^2 分布の下側22%点，下側12%点）．

第　13　表

問20＼問21	個人重視	公共重視
従　う	162(159)	270(273)
利　用	227(227)	391(391)
征　服	88 (91)	160(157)

これまで述べたことは，ランダム化器具を使う試行は，試行を行う人の属性や意見と独立であるが，これと同じように回答者の属性や意見と（少くともある種の属性や意見と）独立な質問を見つけることは困難であるということである．しかし，独立性が成立しないとしても，その程度が軽微であれば推定結果にそれほど影響をおよぼさないことも考えられる．この点について調べてみる．補助質問は「C か $\bar{C}$ か？」であり，本質問は「A か $\bar{A}$ か？」であるものとし，CA，$\bar{C}\bar{A}$ なら“はい”，$C\bar{A}$，$\bar{C}A$ なら“いいえ”と答えるものとする．n 人の回答者中 y 人が“はい”と答えたとする．C の母比率 p は既知

とする．A の母比率 π の推定量としては，2つの質問が独立としたときの不偏推定量

$$\hat{\pi}=\frac{\dfrac{y}{n}-q}{p-q} \tag{33}$$

を用いる．AC の母比率は，2つの質問が独立ならば勿論 $p\pi$ であるが，これが $p\pi+\varepsilon$ になったものとしよう．このとき $\hat{\pi}$ の期待値は

$$E(\hat{\pi})=\pi+\frac{2\varepsilon}{p-q} \tag{34}$$

である．したがって偏りは，

$$E(\hat{\pi})-\pi=\frac{2\varepsilon}{p-q}=k(p)\,\varepsilon \tag{35}$$

である．ただし，$k(p)=2/(p-q)$ とおいた．$k(p)$ は第14表に示されている．

第　14　表

p	0.1	0.2	0.25	0.3	1/3	0.4	0.45
$\lvert k(p)\rvert$	2.50	3.33	4	5	6	10	20

これをみると，たとえ ε が 1% であっても p が 0.4 のときは偏りは 10% に達する．したがって，この種の偏りも p が 0 や 1 に近い場合を除いて無視することはできない．

§5.　む　す　び

(i)　ランダム回答法の場合に，2問以上のクロス集計の結果を使って推定する場合には，各問ごとの周辺比率の推定より精度が一般に悪くなるから十分気をつけなければならない．とくに，ウォーナ法でいって，ランダム化器具によって出現させる確率 p が $1/3\leq p\leq 2/3$ というときには，クロス集計の結果を使う推定の精度は，通常の標本の大きさでは保障されないと考えるべきであろう．

(ii)　偽答や D.K. のもたらす推定への偏りは，偽答や D.K. の起こり方がランダムな起こり方からはなれる程重大になってくる．且つ p が 1/2 に近ずくほど重大になってくる．

(iii)　ランダム化器具を使う試行を，単純に補助質問の答の選択によって代用することは不可能なように思われる．ランダム化器具を用いずに，補助質問によってランダム回答法の主旨をいかす調査を行なうためには，補助質問と本質問との間に相関があってもかまわないような工夫をする必要がある．これについては [6] で論じる．

参　考　文　献

[1] Greenberg, Bernard G., Abul-Ela, Abdel-Latif A., Simmons, Walt R., and Horvitz, Daniel G., (1969). The unrelated question randomized response model: Theoretical framework, *J. Amer. Statist. Ass.*, 64, 520–539.

[2] Horvitz, Daniel G., Shah, B.V., and Simmons, Walt R. (1967). The unrelated question randomized response model, *Social Statistics Section, Proceedings of the American Statistical Association*, 65–72.

[3] Liu, P.T., Chow, L.P. and Mosley, W.H. (1975). Use of the randomized response technique with a new randomizing device, *J. Amer. Statist. Ass.*, 70, 329–332.

[4] Locander, W., Sudman, S. and Bradburn, N. (1976). An investigation of interview

method, threat and response distortion, *J. Amer. Statist. Ass.*, 71, 269–275.

[5] 高橋宏一 (1975). ランダム回答法，数理科学 (サイエンス社), No. 141, 34-38.

[6] Takahasi, K. and Sakasegawa, H. (1977). A randomized response technique without making use of any randomizing device, *Ann. Inst. Statist. Math.*, (to appear)

[7] Warner, Stanley L. (1965). Randomized response: A survey technique for eliminating evasive answer bias, *J. Amer. Statist. Ass.*, 60, 63–69.