確率的変換に基づくインターネット調査手法の解析

田上　敦士[†]　佐々木　力[†]　長谷川輝之[†]　阿野　茂浩[†]
冨浦　洋一[††]

Analysis of Answering Method with Probability Conversion for Internet Research

Atsushi TAGAMI[†], Chikara SASAKI[†], Teruyuki HASEGAWA[†], Shigehiro ANO[†], and Yoichi TOMIURA[††]

あらまし　インターネットの普及により，ネットワークを介した情報収集が広く行われている．インターネット調査と呼ばれる，インターネットを介したアンケート調査は市場調査や社会調査だけではなく，様々な領域で利用されている．しかしながら，これらの情報は多くの個人情報を含み，匿名性を保った状態で収集することが求められている．これに対し筆者らは，確率的変換に基づくインターネット調査手法を提案する．本手法は，二つの回答関数を用いて生成した乱数を回答の代わりに質問者に送信することにより，匿名性を保証する．本論文では，回答者数・調査結果の信頼性/精度が与えられたとき，これらの条件を満足する回答関数に対する制約はその分散のみであることを示す．更に，匿名度という新たな指標を提案し，分散を固定したとき，最良な回答関数を導出する．これらにより，提案手法の適用可能範囲を明確にする．

キーワード　匿名性，インターネット調査，確率的変換

1. まえがき

近年，インターネットの普及により，ネットワークを介して膨大な情報を収集し，各種の調査に利用するということが広く行われている．インターネット調査と呼ばれる，インターネットを介したアンケート調査は，市場調査や社会調査だけではなく，アミューズメント等 [1] 様々な領域で利用されている．また，IPTV における視聴率調査等ネットワークを介した情報収集は今後もその利用領域を広げることが予想される．しかしながら，これらの情報は多くの個人情報を含むため，安全性を確保した収集手法の確立が重要な課題である．

安全性は第三者による情報閲覧を困難にする “秘密性（Securecy）” と，更に，質問者にさえ回答者の回答を知ることを困難にする “匿名性（Anonymity）” の 2 段階に分けられる．秘密性を保証する技術としては，SSL 等公開鍵を利用した暗号化通信が存在する．暗号化通信では，回答者と質問者以外の第三者による情報閲覧を困難とするが，質問者は回答者の回答を知ることができる．インターネット調査で必要とされる安全性には秘密性だけでは不十分であり，匿名性が必要とされる場面が多い．

匿名性を保証する技術としては，電子投票が挙げられる [2], [3]．電子投票はゼロ知識証明を利用することにより，どの投票者が誰に投票したのかはだれにも分からないことを数学的に保証しつつ，投票結果を得ることができる．しかしながら，電子投票では，公開鍵の交換や複数の権限者とミクサを必要とし，処理が複雑になる．インターネット調査は，そもそも全数調査ではなく標本調査である．したがって，インターネット調査では，複雑な処理で厳密な集計結果と匿名性を保証することよりは，むしろ，厳密な集計結果は保証しないが，簡易な技術で匿名性を保証することの方が重要と考えられる．RR 法（Randomized response thechniques）[4] は，簡易に匿名性を提供可能な技術である．しかしながら，理論的な検証が不十分であり，適用可能範囲が明確にされていない．

これに対し筆者らは，確率的変換によるインターネット調査技術を提案する [5], [6]．本技術は，回答者の回答 $x \in \{0,1\}$ を確率的に変換した乱数 v を x の代わりに質問者に送信する．ただし，回答が 0 の場合

[†]（株）KDDI 研究所，ふじみ野市
KDDI R&D Laboratories Inc., Fujimino-shi, 356–8502 Japan
[††]九州大学，福岡市
Kyushu University, Fukuoka-shi, 819–0395 Japan

の送信される値の期待値は 0，回答が 1 の場合の送信される値の期待値は 1 となるように確率的な変換を施す．送信される値は乱数であるため，質問者は個々の回答者の回答を知ることは困難である．しかし，送信された値の標本平均は大数の定理より 1 と回答した回答者の割合（1 回答割合）に確率収束する．これにより，匿名性を確保しつつユーザ動向を簡易に収集することが可能となる．

本技術の適用に際しては，回答者の数が決まっており，必要とする信頼性と精度が与えられたとき，どのような種類の確率的変換が最も高い匿名性を得られるかを明らかにする必要がある．そこで本論文では，本技術の数学的モデル化を行い，回答者数が与えられたとき，1 回答割合の推定値に要求される信頼性・精度を満たす確率的変換に対する制約を明らかにする．**2.** では，安全性を確保するアンケート調査手法に関する関連研究に関して述べ，**3.** では，数学的モデルを示し，上記の確率的変換に対する制約は変換の分散に対する制約のみであることを示す．**4.** では，回答者の回答 x を返信 v から推測することの難しさに基づいて匿名性を測る尺度である匿名度を定義する．**5.** では，匿名度を基準として，確率的変換の分散が与えられたとき（つまり，回答者数及び 1 回答割合の推定値に要求される信頼度・制度が与えられたときの）の最良な回答関数を導出する．**6.** では，その結果から提案手法の適用範囲について考察する．

2. 関 連 研 究

投票者の匿名性を保ちつつ，公開の場において投票と集計の正当性が証明可能な技術として，無記名式電子投票方式（Electronic Voting Scheme）がある．電子投票技術は，準同型性（Homomorphism）[2] に基づいた方式と，MIX-net [3] に基づいた方式に大別できる．両方式では，投票者が不正投票を行っていないことや，集計サーバが不正な集計を行っていないことを保証しつつ，正確な投票内容が得られる．しかしながら，投票者が投票内容の正当性をゼロ知識証明等で示す必要や，投票締め切り後に全投票者の暗号文をシャッフルして復号して集計する必要がある等，投票者・収集者ともに処理が複雑であるという問題がある．

Item Count 法 [7] は，二つの回答者集団に対して真に質問を行いたい項目（キー項目）を含んだ質問リストと，キー項目を含まない質問リストをそれぞれに配布する手法である．回答者は該当する質問の項目数のみを質問者に返信することにより，匿名性を保持可能である．しかしながら，等質な二つの回答者集団を確保する必要があるため，精度において問題がある．また，視聴率調査など，質問数が決まっている場合には適用できないという問題がある．

RR 法（Randomized Response Thecnique）[4] では，質問者は，二つの質問を用意する．一つは調査対象の質問（真の質問）であり，もう一つは調査対象の質問と全く逆の意味をもつ質問（逆の質問）である．回答者は確率 q で逆の質問に yes/no で回答し，確率 $1-q$ で真の質問に yes/no で回答する．このようにして，回答者がどちらの質問に回答したかを質問者に分からないようにすることで，匿名性を保持する．本来の調査に yes と回答した者の割合（すなわち，全回答者のうち，真の質問に yes と回答した，あるいは逆の質問に no と回答した者の割合）は，本論文で提案する手法と同様の原理（**3.1** 参照）で推定することができる．RR 法はデータマイニングへの応用等が検討されている [8]．また，回答を行列で扱い，マルコフ遷移行列に基づいてデータに攪乱を与える PRAM（Post RAndomization Method）[9], [10] も質問者から見た場合，RR 法と同等と見ることができる．しかしながら，これらの方法では，各パラメータの決定手法について明確な匿名性の指標が定義されておらず，提案手法の適用範囲が明確化されていない．

そこで筆者らは，RR 法を包含した新たなインターネット調査手法を提案し，その適用範囲と最適なパラメータ設定手法について述べる．

3. 提 案 手 法

3.1 インターネット調査

まず，本論文が想定するインターネット調査について述べる．インターネット調査は複数の回答者と 1 人の質問者からなり，回答者は質問者に対して，0 か 1 で回答を返す．ただし，各回答者は質問に対して 1 回のみ回答することとする．質問者は，すべての回答者から回答を集め，1 と回答した人の割合（1 回答割合）r を求めることを目的とする．もし，2 個以上の選択肢がある場合においても，選択肢の数だけのビット列を用意し，回答に対応するビットのみを 1 にすることによって対応可能である．

しかし，回答者が 0 か 1 をそのまま質問者に送信したとすると，質問者はどの回答者が何と回答したか分かることになり，匿名性は保たれない．そこで，回答

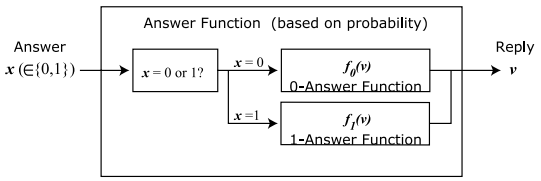

図 1　0/1 回答関数
Fig. 1　0/1-answer function.

者の回答 x ($\in \{0,1\}$) を確率的に変換した v を x の代わりに質問者に送信することを考える．返信 v と回答 x は 1 対 1 対応していないため，質問者が返信 v から回答者の回答 x を知ることは困難である．

図 1 に x から v の変換手順を示す．返信を表す確率変数を V とする．回答者は 0 と回答する場合は，V の確率関数（若しくは確率密度関数）$f_0(v)$ に従って V の実現値 v を発生させこれを質問者に送信する．また，1 と回答する場合は，V の確率関数（若しくは確率密度関数）$f_1(v)$ に従って V の実現値 v を発生させこれを送信する．本論文では $f_0(v)$ を 0 回答関数，$f_1(v)$ を 1 回答関数と呼ぶ．

ここで，0 回答関数に従う確率変数の期待値は 0，1 回答関数に従う確率変数の期待値は 1，両確率変数の分散は等しく σ^2 とする．すなわち，

$$\mathrm{E}[V; f_0(\cdot)] = 0 \tag{1}$$

$$\mathrm{E}[V; f_1(\cdot)] = 1 \tag{2}$$

$$\mathrm{Var}[V; f_0(\cdot)] = \mathrm{Var}[V; f_1(\cdot)] = \sigma^2 \tag{3}$$

を満たすとする．この三つの条件式を満たす 0/1 回答関数であれば，いかなる分布に従う確率関数（若しくは確率密度関数）であっても，提案手法における回答関数として利用できる．

ここで，$\overline{V}_n$ を

$$\overline{V}_n = \frac{1}{n}\sum_i V_i \tag{4}$$

と定義する．n は回答者数である．V_i は回答者 i が送信する値に対応する確率変数であり，V_i と V_j $(i \neq j)$ は独立である．

このとき，1 と回答した人の数を m とすると，一般性を損なわずに回答者 1〜m は 1 と回答し，$m+1$〜n は 0 と回答するものと考えることができる．すなわち，

$$\overline{V}_n = \frac{1}{n}\left(\sum_{i=1}^{m} V_i + \sum_{i=m+1}^{n} V_i\right) \tag{5}$$

となる．

ここで，$m/n = r$ を保った状態で，回答者数 n を大きくすることを考える．このとき，大数の法則及び式 (1)，(2) より，

$$\frac{1}{m}\sum_{i=1}^{m} V_i \xrightarrow{P} 1, \quad \frac{1}{n-m}\sum_{i=m+1}^{n} V_i \xrightarrow{P} 0 \tag{6}$$

が成立する．ただし，$\xrightarrow{P}$ は確率収束を表す．したがって，

$$\overline{V}_n \xrightarrow{P} \frac{1}{n}\{m\cdot 1 + (n-m)\cdot 0\} = \frac{m}{n} \tag{7}$$

となり，$\overline{V}_n$ が m/n つまり r の一致推定量であることが分かる．すなわち，n が十分大きいとき，1 と答えた人の割合 r は，$\overline{V}_n$ で近似できる．これにより，質問者は個々の回答者の回答を知ることなしにインターネット調査結果を取得することが可能となる．

3.2　σ^2 決定手法

本節では前節で提案したインターネット調査手法に関して，要求される信頼度と精度のインターネット調査結果を得るための 0/1 回答関数のパラメータ決定手法について述べる．

式 (1) (2) (3) と中心極限定理より，$\sum_{i=1}^{m} V_i$ は近似的に $N(m, m\cdot\sigma^2)$ に従い，$\sum_{i=m+1}^{n} V_i$ は近似的に $N(0, (n-m)\cdot\sigma^2)$ に従う．ただし，$N(\mu, \sigma^2)$ は平均 μ，分散 σ^2 の正規分布を表す．

したがって，$\overline{V}_n$ は近似的に，

$$N\left(\frac{1}{n}(m+0), \frac{1}{n^2}\left(m\cdot\sigma^2 + (n-m)\cdot\sigma^2\right)\right)$$

つまり，

$$N\left(r, \frac{\sigma^2}{n}\right) \tag{8}$$

に従う．

今，Z を標準正規分布に従う確率変数とし，$z_{\alpha/2}$ を

$$P(-z_{\alpha/2} \leq X \leq z_{\alpha/2}) = 1-\alpha \tag{9}$$

と満たす値とすると，

$$-z_{\alpha/2} \leq \frac{\overline{V}_n - r}{\sqrt{\sigma^2/n}} \leq z_{\alpha/2} \tag{10}$$

より，r の $100\cdot(1-\alpha)$%信頼区間は，

$$\left[\overline{V}_n - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \overline{V}_n + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right] \tag{11}$$

である．したがって，推定値と真値 (m/n) の誤差が $\pm\delta$ 以内であるためには σ^2 は

$$\sigma^2 = n\left(\frac{\delta}{z_{\alpha/2}}\right)^2 \tag{12}$$

となる．

これより，式 (12) を満たす σ^2 を設定したとき，集計結果 $\overline{V}_n$ は，$100\cdot(1-\alpha)\%$の信頼度で誤差 $\pm\delta$ の範囲内に含まれる．

4. 匿 名 度

4.1 定　　義

前章で述べたとおり，回答者数が与えられたとき，1 回答割合の推定値に要求される信頼度・精度を満たす 0/1 回答関数に対する制約は分散 σ^2 に対する制約のみである．本節では，式 (12) で定義された分散下で，最も回答者の回答を推測困難な 0/1 回答関数を導出するため，新しい評価値である匿名度を導入する．

匿名度は，回答者の回答を推定することの困難さとして定義する．具体的には，回答推定の難しさは返信値 v に依存するため，匿名度は回答者以外の者が回答者の返信値 v を知ったとき，v に基づいて回答者の回答 x を推定した場合に，推定が誤る確率（誤推定率）の平均と定義する．この値が 0.5 に近ければ匿名性が高く，0 の場合は匿名性がなく返信値 v により一意に回答 x が推定できることを意味する．

まず，v を知ったときの誤推定率について考える．$V=v$ であるときの回答 x は，ベイズ決定則 (Bayes decision rule) に従うと，

$$\begin{aligned} x &= \arg\max_{x\in\{0,1\}} f_{X|V}(x|v) \\ &= \arg\max_{x\in\{0,1\}} f_X(x) f_{V|X}(v|x) \end{aligned} \tag{13}$$

と推定できる．$f_{X|V}(x|v)$ は，$V=v$ が与えられたときの，$X=x$ である条件付確率，$f_{V|X}(v|x)$ は，$X=x$ が与えられたときの，$V=v$ である条件付確率（あるいは条件付確率密度）であり，$f_{V|X}(v|0)=f_0(v)$，$f_{V|X}(v|1)=f_1(v)$ である．また，$f_X(x)$ は X の確率関数（回答の事前確率）である．

ここで，三つの領域 D_0，D_1，D_e を

$$\begin{aligned} D_0 &= \{v\in\mathcal{V} \mid f_X(0)f_0(v) > f_X(1)f_1(v)\} \\ D_1 &= \{v\in\mathcal{V} \mid f_X(0)f_0(v) < f_X(1)f_1(v)\} \\ D_e &= \{v\in\mathcal{V} \mid f_X(0)f_0(v) = f_X(1)f_1(v)\} \end{aligned} \tag{14}$$

と定義する．ただし，$\mathcal{V}=\{v\in\mathbb{R} \mid f_V(v)>0)\}$ で，V のとり得る値の集合である（V が離散型確率変数の場合は $\mathcal{V}$ は可算集合）．式 (13) より回答者の回答 x は，$v\in D_0$ のときは 0，$v\in D_1$ のときは 1 と推定される．また，$v\in D_e$ のときは，一般化して確率 ξ $(0\le\xi\le 1)$ で 1 と推定すると考える．

このとき，$V=v$ を知ったときの誤推定率 $Error(v)$ は，

$$Error(v) = \begin{cases} \dfrac{f_X(1)\cdot f_1(v)}{f_V(v)} & : \quad v\in D_0 \\ \dfrac{f_X(0)\cdot f_0(v)}{f_V(v)} & : \quad v\in D_1 \\ \xi\dfrac{f_X(0)\cdot f_0(v)}{f_V(v)} + (1-\xi)\dfrac{f_X(1)\cdot f_1(v)}{f_V(v)} & : \quad v\in D_e \end{cases}$$

となる．ただし，

$$f_V(v) = f_X(1)\cdot f_1(v) + f_X(0)\cdot f_0(v). \tag{15}$$

一般に，インターネット調査をする以前では 1 と回答する回答者の割合（つまり，事前確率 $f_X(1)$）は不明であるから，$f_X(0)=f_X(1)=0.5$ として，送信された値 v から回答 x を推定すると仮定すると，

$$Error(v) = \frac{1}{2}\cdot\frac{\min\left(f_0(v), f_1(v)\right)}{f_V(v)} \tag{16}$$

となる．なお，事前確率 $f_X(1)$ が既知である場合については **6.3** で考察する．

匿名度 Anonymity は誤推定率の期待値，つまり $\mathrm{E}\left[Error(V)\right]$ で定義される．V が離散型確率変数の場合は，

$$\begin{aligned} &Anonymity \\ &= \mathrm{E}[Error(V)] \\ &= \frac{1}{2}\sum_v \frac{\min(f_0(v), f_1(v))}{f_V(v)} f_V(v) \\ &= \frac{1}{2}\left\{\sum_{v\in D_0} f_1(v) + \sum_{v\in D_1} f_0(v) + \sum_{v\in D_e} f_0(v)\right\} \end{aligned} \tag{17}$$

であり，V が連続型確率変数の場合は，

$$\begin{aligned} &Anonymity \\ &= \mathrm{E}[Error(V)] \end{aligned}$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \frac{\min(f_0(v), f_1(v))}{f_V(v)} f_V(v)\, dv$$
$$= \frac{1}{2}\left[\int_{D_0} f_1(v)dv + \int_{D_1} f_0(v)dv + \int_{D_e} f_0(v)dv\right] \quad (18)$$

である．

4.2 匿名度による回答関数の評価

本節では，回答関数の期待値条件 (1)，(2) と分散条件 (3) を満たす以下の二つの 0/1 回答関数を例として，匿名度を用いた評価を行う．

［正規分布 0/1 回答関数］

$$\begin{cases} f_0(v) & : & N(0,\sigma^2) \text{ の確率密度関数} \\ f_1(v) & : & N(1,\sigma^2) \text{ の確率密度関数} \end{cases}$$

［ベルヌーイ分布(注1) 0/1 回答関数］

$$f_0(v) = \begin{cases} 1-q & : v = a \\ q & : v = b \\ 0 & : \text{その他} \end{cases}$$

$$f_1(v) = \begin{cases} q & : v = a \\ 1-q & : v = b \\ 0 & : \text{その他} \end{cases}$$

ただし，$0 < q < 0.5$ (注2) とする．また，f_0 と f_1 がそれぞれ式 (1) (2) の条件を満たすように，a，b は，

$$a = -\frac{q}{1-2q}, \quad b = \frac{1-q}{1-2q}$$

と設定する．ここで，f_0 と f_1 の分散がともに σ^2 となるように，q は $0 < q < 0.5$ を満たす

$$\frac{q(1-q)}{(1-2q)^2} = \sigma^2 \quad (19)$$

の解として設定される．

RR 法と異なり，yes/no ではなく，a/b を送信するが，これは，本手法では返信される値の標本平均で 1 回答割合を推定できるようにしたからで，本質的には RR 法はベルヌーイ分布 0/1 回答関数を用いた場合の提案手法と同等である．

式 (18) より，正規分布 0/1 回答関数の匿名度は，

$$Anonymity = \int_{0.5}^{\infty} f_0(v)dv \quad (20)$$

となる．同様に式 (17) より，ベルヌーイ分布 0/1 回答関数の匿名度は，

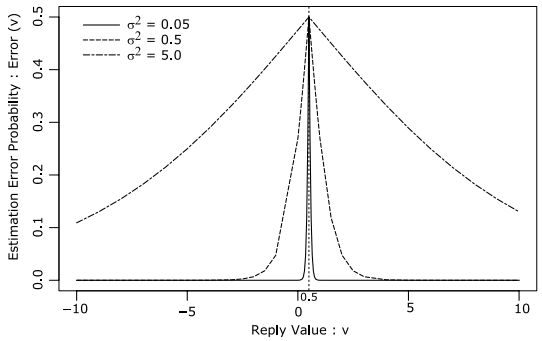

図 2　誤推定率の分布
Fig. 2　Error probability variance.

$$Anonymity = q \quad (21)$$

となる．例えば，$n = 10{,}000$，$\alpha = 0.05$，$\delta = 0.01$ とする．このとき，0/1 回答関数の分散 σ^2 は $z_{0.05/2} \simeq 1.96$ であるので，式 (12) より，$\sigma^2 \simeq 0.26$ でなければならない．したがって，正規分布 0/1 回答関数の場合の匿名度は，式 (20) より約 0.16，ベルヌーイ分布 0/1 回答関数の匿名度は式 (19) (21) より約 0.15 となる．わずかであるが正規分布 0/1 回答関数の方が匿名度が高い 0/1 回答関数といえる．

式 (17) (18) において，匿名度を誤推定率の平均で定義した．しかしながら，確率分布によっては質問者が知り得た回答 v によって，誤推定率に大きな差が発生することが考えられる．

図 2 に正規分布に従う 0/1 回答関数において，式 (16) を用いて，各 v における誤推定率を求めた結果を示す．これより，$v = 0.5$ を頂点として 0.5 から遠ざかるほど，誤推定率が低下していることが分かる．これは，例えば，$v = 10$ をインターネット調査の返信として質問者に送信する確率は低いが，そのときは高い確

(注1)：ベルヌーイ分布の確率関数 $f(y)$ は

$$f(y) = \begin{cases} 1-p & : & y = 0 \\ p & : & y = 1 \\ 0 & : & \text{その他} \end{cases}$$

であるが，本論文では，確率関数が，

$$f(y) = \begin{cases} 1-p & : & y = a \\ p & : & y = b \\ 0 & : & \text{その他} \end{cases}$$

であるような，2 点（a，b）でのみ確率をもつ離散型分布もベルヌーイ分布と呼ぶ．

(注2)：これは本質的な制約ではない．実際，$q = q_0$ の場合の f_0 と，$q = 1 - q_0$ の場合の f_0 は同一になる．f_1 についても同様．

率で真の回答 x が推定されてしまうことを意味する．

5. 最良な回答関数

5.1 準　　備

続く **5.2**，**5.3** では，それぞれ，分散が σ^2 の 0/1 回答関数の中で，

- 誤推定率 $Error(v)$ が v によらず一定で匿名度を最大にする 0/1 回答関数，
- 誤推定率が与えられたある値 θ 以上で匿名度を最大にする 0/1 回答関数

を導出する．本節ではこのための準備として補題を与える．

f が実数の集合 $\mathbb{R}$ 上で定義された非負値関数とする．

$$D = \{y \in \mathbb{R} \mid f(y) > 0\}$$

が，可算集合であるとき，$\mathrm{E}_f[Y^k; D]$ を

$$\mathrm{E}_f[Y^k; D] = \sum_{y \in D} y^k f(y)$$

と定義する（注意：$\mathrm{E}_f[1; D] = \sum_{y \in D} f(y)$）．また，$f$ が非負値連続関数であるとき，$\mathrm{E}_f[Y^k; D]$（$D \subseteq \mathbb{R}$）を

$$\mathrm{E}_f[Y^k; D] = \int_D y^k f(y)\, dy$$

と定義する（注意：$\mathrm{E}_f[1; D] = \int_D f(y)\, dy$）．

［補題］ f を実数の集合 $\mathbb{R}$ 上で定義された非負値関数とする．$D \subseteq \mathbb{R}$ に対し，

$$\mathrm{E}_f[1; D] = S < \infty\,, \qquad \mathrm{E}_f[Y^2; D] < \infty$$

が成立するとき，ある実数 a が存在し，

$$\mathrm{E}_f[Y; D] = aS\,, \qquad \mathrm{E}_f[Y^2; D] \geq a^2 S$$

が成立する．ただし，後者の等号が成立するのは，

$$\{y \in \mathbb{R} \mid f(y) > 0\} = \{a\}$$

が成立するときのみである． □

本補題は，Jensen の不等式 [11] を利用して容易に証明することができる．

5.2 誤推定率一定の場合

［定理 1］ 分散が σ^2 の 0/1 回答関数の中で，誤推定率 $Error(v)$ が一定，つまり，

$$\exists\theta(\theta > 0)\, \forall v \in \mathcal{V}\ \ Error(v) = \theta \tag{22}$$

なる条件を満たし，匿名度を最大にする 0/1 回答関数は以下の (f_0, f_1) である．

$$f_0(v) = \begin{cases} 1 - \theta_m & : \quad v = -\dfrac{\theta_m}{1 - 2\theta_m} \\ \theta_m & : \quad v = \dfrac{1 - \theta_m}{1 - 2\theta_m} \\ 0 & : \quad \text{その他} \end{cases}$$

$$f_1(v) = f_0(1 - v)$$

ただし，

$$\theta_m = \frac{1}{2} - \frac{1}{2\sqrt{1 + 4\sigma^2}} \tag{23}$$

である．また，この回答関数を用いた場合の匿名度は，θ_m である．

（証明） 式 (15)，式 (16)，及び，仮定 $f_X(1) = f_X(0) = 0.5$ から，誤推定率 $Error(v)$ は，

$$Error(v) = \frac{\min(f_0(v), f_1(v))}{f_0(v) + f_1(v)} \tag{24}$$

と表せる．また，$f_X(1) = f_X(0) = 0.5$ の仮定のもとでは，

$$D_0 = \{v \in \mathcal{V} \mid f_0(v) > f_1(v)\}\,,$$
$$D_1 = \{v \in \mathcal{V} \mid f_0(v) < f_1(v)\}\,,$$
$$D_e = \{v \in \mathcal{V} \mid f_0(v) = f_1(v)\}$$

である．式 (24) より，

$$\begin{cases} v \in D_0 \text{ (つまり } , f_0(v) > f_1(v)) \text{ のとき} \\ \quad Error(v) = \dfrac{f_1(v)}{f_0(v) + f_1(v)} < \dfrac{1}{2}, \\ v \in D_1 \text{ (つまり } , f_0(v) < f_1(v)) \text{ のとき} \\ \quad Error(v) = \dfrac{f_0(v)}{f_0(v) + f_1(v)} < \dfrac{1}{2}, \\ v \in D_e \text{ (つまり } , f_0(v) = f_1(v)) \text{ のとき} \\ \quad Error(v) = \dfrac{1}{2} \end{cases} \tag{25}$$

である．$v \in D_e$ の場合と $v \in D_0 \cup D_1$ の場合とで，$Error(v)$ の値が異なるため，制約 (22) が成立するためには，$E_{f_1}[1; D_e] = E_{f_0}[1; D_e] = 0$ または $E_{f_1}[1; D_e] = E_{f_0}[1; D_e] = 1$ でなければならない．後者の場合は，

$$\forall v \in \mathcal{V} \quad f_0(v) = f_1(v)$$

となり，0/1 回答関数の条件 (1) (2) を満たさない．また，式 (25) より，例えば，$v \in D_0$ の場合，$Error(v)$

が一定値 θ であることから，

$$f_1(v) = \frac{\theta}{1-\theta} f_0(v)$$

を得る．したがって，制約 (22) は，仮定 $f_X(1) = f_X(0) = 0.5$ のもとでは，$0 < \theta < 0.5$ なる θ が存在して，

$$\begin{cases} f_1(v) = \dfrac{\theta}{1-\theta} f_0(v) & : \quad v \in D_0, \\ f_0(v) = \dfrac{\theta}{1-\theta} f_1(v) & : \quad v \in D_1 \end{cases} \tag{26}$$

かつ，

$$\mathrm{E}_{f_0}[1; D_e] = \mathrm{E}_{f_1}[1; D_e] = 0 \tag{27}$$

と等価である．式 (26) より，

$$\mathrm{E}_{f_1}[V^k; D_0] = \frac{\theta}{1-\theta}\mathrm{E}_{f_0}[V^k; D_0] \tag{28}$$

$$\mathrm{E}_{f_0}[V^k; D_1] = \frac{\theta}{1-\theta}\mathrm{E}_{f_1}[V^k; D_1] \tag{29}$$

である．

f_0 が確率（密度）関数であることから，

$$1 = \mathrm{E}_{f_0}[1; D_0] + \mathrm{E}_{f_0}[1; D_1] + \mathrm{E}_{f_0}[1; D_e]$$

であり，これと，式 (27) (29) より

$$1 = \mathrm{E}_{f_0}[1; D_0] + \frac{\theta}{1-\theta}\mathrm{E}_{f_1}[1; D_1]$$

が成立する．同様に f_1 が確率（密度）関数であることと式 (27) (28) より，

$$1 = \frac{\theta}{1-\theta}\mathrm{E}_{f_0}[1; D_0] + \mathrm{E}_{f_1}[1; D_1]$$

が成立する．この二つの関係から，

$$\mathrm{E}_{f_0}[1; D_0] = \mathrm{E}_{f_1}[1; D_1] = 1 - \theta \tag{30}$$

を得る．

5. 1 で与えた表記を用いるならば，匿名度は

$$\frac{1}{2}\{\mathrm{E}_{f_1}[1; D_0] + \mathrm{E}_{f_0}[1; D_1] + \mathrm{E}_{f_0}[1; D_e]\}$$

となる．したがって，式 (27) (28) (29) (30) より，

$$\begin{aligned} &\textit{Anonimity} \\ &\quad = \frac{1}{2}\frac{\theta}{1-\theta}\{\mathrm{E}_{f_0}[1; D_0] + \mathrm{E}_{f_1}[1; D_1]\} = \theta \end{aligned}$$

である．つまり，回答関数の期待値条件，分散条件を満たす最大の θ を求めれば，これが，誤推定率一定の場合の最大の誤推定率であり，最大の匿名度となる．

f_0 の期待値条件 (1)，分散条件 (3) より，

$$\begin{aligned} 0 &= \mathrm{E}_{f_0}[V; D_0] + \mathrm{E}_{f_0}[V; D_1] + \mathrm{E}_{f_0}[V; D_e], \\ \sigma^2 &= \mathrm{E}_{f_0}[V^2; D_0] + \mathrm{E}_{f_0}[V^2; D_1] + \mathrm{E}_{f_0}[V^2; D_e] \end{aligned}$$

であり，これと，式 (27) (29) より，

$$0 = \mathrm{E}_{f_0}[V; D_0] + \frac{\theta}{1-\theta}\mathrm{E}_{f_1}[V; D_1], \tag{31}$$

$$\sigma^2 = \mathrm{E}_{f_0}[V^2; D_0] + \frac{\theta}{1-\theta}\mathrm{E}_{f_1}[V^2; D_1] \tag{32}$$

を得る．式 (31) (32) と，補題及び式 (30) より，ある定数 a_0，a_1 が存在し，

$$0 = a_0(1-\theta) + a_1\theta \tag{33}$$

$$\sigma^2 \geq {a_0}^2(1-\theta) + {a_1}^2\theta \tag{34}$$

を得る．また，f_1 の期待値条件 (2)，分散条件 (3) より，同様にして，

$$1 = a_0\theta + a_1(1-\theta) \tag{35}$$

$$\sigma^2 + 1 \geq {a_0}^2\theta + {a_1}^2(1-\theta) \tag{36}$$

を得る．

式 (33) (35) を a_0，a_1 について解くと，

$$a_0 = -\frac{\theta}{1-2\theta}, \quad a_1 = \frac{1-\theta}{1-2\theta} \tag{37}$$

が得られ，これを式 (34) (36) に代入して整理すると，ともに，

$$\theta - \theta^2 \leq (1-2\theta)^2\sigma^2$$

が得られる．$\theta < 1/2$ に注意してこれを解くと，

$$\theta \leq \frac{1}{2} - \frac{1}{2\sqrt{1+4\sigma^2}}$$

が得られる．

したがって，誤推定率一定の場合の誤推定率及び匿名度の最大値 θ_m は

$$\theta_m = \frac{1}{2} - \frac{1}{2\sqrt{1+4\sigma^2}} \tag{38}$$

である．これを実現する回答関数は式 (34) (36) で等号が成立するものであり，補題に示した等号の成立条件より，本定理の回答関数を得る． □

本定理が与える最良の回答関数は **4. 2** で例示したベルヌーイ分布 0/1 回答関数そのものである．

5.3 誤推定率に下限を与えた場合

［定理 2］ 分散が σ^2 で

$$\forall v\ Error(v) \geq \theta \tag{39}$$

なる制約を満たす 0/1 回答関数は,

$$\theta \leq \frac{1}{2} - \frac{1}{2\sqrt{1+4\sigma^2}}$$

のとき存在し，このうち匿名度を最大にする 0/1 回答関数は，以下の (f_0, f_1) である．

$$f_0(v) = \begin{cases} L & : \ v = 1/2 - \Delta \\ M & : \ v = 1/2 \\ L \cdot \theta/(1-\theta) & : \ v = 1/2 + \Delta \\ 0 & : \ \text{その他} \end{cases}$$

$$f_1(v) = f_0(1-v)$$

ただし,

$$L = \frac{1-\theta}{(1+4\sigma^2)(1-2\theta)^2}\ ,$$

$$M = 1 - \frac{1}{(1+4\sigma^2)(1-2\theta)^2}\ ,$$

$$\Delta = \frac{1}{2}(1+4\sigma^2)(1-2\theta)$$

である．またこの回答関数を用いた場合の匿名度は,

$$Anonymity = \frac{1}{1+4\sigma^2}\left(2\sigma^2 - \frac{\theta}{1-2\theta}\right) \tag{40}$$

である．

（証明） 定理 1 の証明とほぼ同様にして，仮定 $f_X(1) = f_X(0) = 0.5$ のもとでは，$\theta \geq 1/2$ のとき，制約 (39) を満たす 0/1 回答関数は存在せず，$0 < \theta < 1/2$ のとき，制約 (39) は,

$$f_1(v) \geq \frac{\theta}{1-\theta} f_0(v) \quad : \quad v \in D_0 \tag{41}$$

$$f_0(v) \geq \frac{\theta}{1-\theta} f_1(v) \quad : \quad v \in D_1 \tag{42}$$

と等価であることが導ける．

ここで,

$$g(v) = \begin{cases} \dfrac{1-\theta}{1-2\theta} f_1(v) - \dfrac{\theta}{1-2\theta} f_0(v) & : v \in D_0 \\ f_0(v) & : v \in D_e \\ \dfrac{1-\theta}{1-2\theta} f_0(v) - \dfrac{\theta}{1-2\theta} f_1(v) & : v \in D_1 \end{cases}$$

$$h(v) = \begin{cases} \dfrac{1-\theta}{1-2\theta}\left(f_0(v) - f_1(v)\right) & : \ v \in D_0 \\ 0 & : \ v \in D_e \\ \dfrac{1-\theta}{1-2\theta}\left(f_1(v) - f_0(v)\right) & : \ v \in D_1 \end{cases}$$

と定義する．g，h を用いて f_0，f_1 を表現すると,

$$f_0(v) = \begin{cases} h(v) + g(v) & : \ v \in D_0 \\ g(v) & : \ v \in D_e \\ \dfrac{\theta}{1-\theta} h(v) + g(v) & : \ v \in D_1 \end{cases} \tag{43}$$

$$f_1(v) = \begin{cases} \dfrac{\theta}{1-\theta} h(v) + g(v) & : \ v \in D_0 \\ g(v) & : \ v \in D_e \\ h(v) + g(v) & : \ v \in D_1 \end{cases} \tag{44}$$

となる．また，これから,

$$\mathrm{E}_{f_0}[V^k; D_e] \ = \ \mathrm{E}_g[V^k; D_e]\,,$$

$$\mathrm{E}_{f_0}[V^k; D_1] \ = \ \frac{\theta}{1-\theta}\mathrm{E}_h[V^k; D_1] + \mathrm{E}_g[V^k; D_1]$$

などが得られる．g，h の定義，式 (41) (42) 及び $\theta < 1/2$ より,

$$\begin{cases} \forall v\ g(v) \geq 0\,, \\ \forall v \in (D_0 \cup D_1)\ h(v) > 0\,, \\ \forall v \in D_e\ h(v) = 0\,, \end{cases} \tag{45}$$

また，f_0，f_1 は確率（密度）関数であることから，式 (43) (44) より,

$$\mathrm{E}_h[1; D_0] + \frac{\theta}{1-\theta}\mathrm{E}_h[1; D_1] + \mathrm{E}_g[1; \mathcal{V}] = 1\,,$$

$$\frac{\theta}{1-\theta}\mathrm{E}_h[1; D_0] + \mathrm{E}_h[1; D_1] + \mathrm{E}_g[1; \mathcal{V}] = 1$$

を得る．これから

$$\mathrm{E}_h[1; D_0] = \mathrm{E}_h[1; D_1]$$

が得られ，上記二つの式はともに

$$\frac{L}{1-\theta} + M = 1 \tag{46}$$

と表せる．ただし,

$$\mathrm{E}_h[1; D_0] = \mathrm{E}_h[1; D_1] = L\,, \quad \mathrm{E}_g[1; \mathcal{V}] = M \tag{47}$$

である．また，$A = 2 \cdot Anonimity$ とおくと，式 (43)

(44) (47) より

$$\begin{aligned} A &= \mathrm{E}_{f_1}[1;D_0] + \mathrm{E}_{f_0}[1;D_1] + \mathrm{E}_{f_0}[1;D_e] \\ &= \frac{\theta}{1-\theta}\{\mathrm{E}_h[1;D_0] + \mathrm{E}_h[1;D_1]\} \\ &\quad + \mathrm{E}_g[1;D_0] + \mathrm{E}_g[1;D_1] + \mathrm{E}_g[1;D_e] \\ &= \frac{2\theta}{1-\theta}L + M \end{aligned} \tag{48}$$

を得る．式 (46) (48) を L，M について解くと，

$$L = \frac{(1-\theta)(1-A)}{1-2\theta} \qquad M = \frac{A-2\theta}{1-2\theta} \tag{49}$$

を得る．

f_0 の期待値条件 (1) と分散条件 (3) より，

$$\begin{aligned} 0 &= \mathrm{E}_{f_0}[V;D_0] + \mathrm{E}_{f_0}[V;D_1] + \mathrm{E}_{f_0}[V;D_e]\,, \\ \sigma^2 &= \mathrm{E}_{f_0}[V^2;D_0] + \mathrm{E}_{f_0}[V^2;D_1] + \mathrm{E}_{f_0}[V^2;D_e] \end{aligned}$$

であり，これと，式 (43) (44) より，

$$\begin{aligned} 0 &= \mathrm{E}_h[V;D_0] + \frac{\theta}{1-\theta}\mathrm{E}_h[V;D_1] + \mathrm{E}_g[V;\mathcal{V}]\,, \\ \sigma^2 &= \mathrm{E}_h[V^2;D_0] + \frac{\theta}{1-\theta}\mathrm{E}_h[V^2;D_1] + \mathrm{E}_g[V^2;\mathcal{V}] \end{aligned}$$

を得る．式 (45) より，g 及び h は領域 D_0，D_1，D_e で非負であるから，上式と式 (47)，補題より，実数 a_0，a_1，a_e が存在して，

$$0 = a_0 L + \frac{\theta}{1-\theta}a_1 L + a_e M \tag{50}$$

$$\sigma^2 \geq {a_0}^2 L + \frac{\theta}{1-\theta}{a_1}^2 L + {a_e}^2 M \tag{51}$$

が成立する．また，f_1 の期待値条件 (2) と分散条件 (3) より，同様にして，

$$1 = \frac{\theta}{1-\theta}\, a_0 L + a_1 L + a_e M \tag{52}$$

$$\sigma^2 + 1 \geq \frac{\theta}{1-\theta}\, {a_0}^2 L + {a_1}^2 L + {a_e}^2 M \tag{53}$$

を得る．

式 (50) (52) に (49) を代入して，a_0，a_1 について解くと，

$$a_0 = -\frac{A-2\theta}{1-A}\, a_e - \frac{\theta}{1-A} \tag{54}$$

$$a_1 = -\frac{A-2\theta}{1-A}\, a_e + \frac{1-\theta}{1-A} \tag{55}$$

を得る．

(54) (55) を式 (51) に代入し整理すると，

$$A \leq \frac{1}{{a_e}^2 + \sigma^2}\left(\sigma^2 + 2\theta {a_e}^2 - \frac{\theta(1-\theta)}{1-2\theta}\right) \tag{56}$$

を得る．上記の右辺を $A_0(a_e)$ と記す．(54) (55) を式 (53) に代入し整理すると，

$$\begin{aligned} A \leq{}& \frac{1}{(1-a_e)^2 + \sigma^2} \\ &\times \left(\sigma^2 + 2\theta(1-a_e)^2 - \frac{\theta(1-\theta)}{1-2\theta}\right) \end{aligned} \tag{57}$$

を得る．上記の右辺を $A_1(a_e)$ と記す．A の上限は，

$$\max_{a_e} \min(A_0(a_e), A_1(a_e))$$

である．

式 (45) と M，L の定義より，$M \geq 0$，$L > 0$ であるが，これを保証するには，(49) より，

$$2\theta \leq A < 1$$

でなければならない．この領域上に不等式 (56) (57) の解が存在する（積領域が空でない）ためには，明らかに，$A_0(a_e) < 1$，$A_1(a_e) < 1$ であるから，

$$A_0(a_e) \geq 2\theta\,, \quad A_1(a_e) \geq 2\theta$$

であればよい．両式からはともに

$$\theta - \theta^2 \leq (1-2\theta)^2\sigma^2$$

が得られ，これを，$0 < \theta < 1/2$ に注意して θ について解くと，

$$\theta \leq \frac{1}{2} - \frac{1}{2\sqrt{1+4\sigma^2}}$$

を得る．この θ の範囲に注意して，$A_0(a_e)$，$A_1(a_e)$ の a_e に関する増加減少を求め，更に，$A_0(a_e)$，$A_1(a_e)$ の対称性を考慮すると，

$$\begin{aligned} &\max_{a_e} \min(A_0(a_e), A_1(a_e)) \\ &\quad = A_0(1/2) \\ &\quad = \frac{1}{1+4\sigma^2}\left(4\sigma^2 - \frac{2\theta}{1-2\theta}\right) \end{aligned} \tag{58}$$

を得る．

式 (58) の右辺を A_m とおく．$A = A_m$，$a_e = 1/2$ を式 (49) (54) (55) に代入し，

$$L = \frac{1-\theta}{(1+4\sigma^2)(1-2\theta)^2}$$

$$M = 1 - \frac{1}{(1+4\sigma^2)(1-2\theta)^2}$$
$$a_0 = \frac{1}{2} - \frac{1}{2}(1+4\sigma^2)(1-2\theta)$$
$$a_1 = \frac{1}{2} + \frac{1}{2}(1+4\sigma^2)(1-2\theta)$$

を得る．$A = A_m$ を実現する g，h は，式 (51) (53) で等号が成立するものであり，補題より，$a_0 \in D_0$，$a_1 \in D_1$，$a_e = 1/2 \in D_e$ であり，

$$g(v) = \begin{cases} M & : \quad v = 1/2, \\ 0 & : \quad \text{その他} \end{cases}$$

$$h(v) = \begin{cases} L & : \quad v = a_0, \\ L & : \quad v = a_1, \\ 0 & : \quad \text{その他} \end{cases}$$

であるから，式 (43) (44) より，本定理を得る． □

6. 考　　察

6.1 適用範囲

本節では，提案技術の適用範囲に関して考察する．図 3 に，定理 1 で与えられる誤推定率一定の場合の最良の 0/1 回答関数を用いた場合の回答者数と匿名度の関係を示す．これは信頼度 95%（$\alpha = 0.05$），誤差 $\delta = 0.01, 0.05$ として，式 (12) (23) より求めたものである．

匿名度の許容範囲は適用例によって様々であるが，高い匿名性が求められる場合においても，0.4 以上で十分であると仮定すると，本手法を利用する場合には $\delta = 0.05$ の場合で，10,000 人程度の回答者が必要となる．また，信頼度を一定としたとき，匿名度とインターネット調査結果の誤差はトレードオフの関係にあり，インターネット調査結果の誤差を大きくすることにより，匿名度を高めることが可能となる．

6.2 誤推定率の下限と匿名度の関係

5.3 で述べたように，誤推定率 Error(v) の下限によって，匿名度の値は変化する．本節では，その関係について考察する．

図 4 に，誤推定率の下限 θ と匿名度 Anonymity の関係を式 (12) (40) より求めた結果を図示する．回答者数 $n = 10^4, 5.0\times10^4, 10^5, 10^6$，信頼度 95%（$\alpha = 0.05$），誤差 $\delta = 0.01, 0.05$ とした．

定理 2 で与えられる制約を満たす回答関数が存在する最大の誤推定率に θ を設定すると，定理 2 で与えられる 0/1 回答関数は定理 1 で与えられる 0/1 回答関数に一致する．また，式 (40) より，誤推定率の下限 θ が小さくなるほど，匿名度が高くなることが分かる．これらのことは，図 4 からも読み取れる．更に，その上昇幅は回答者数が少ないほど大きく，回答者数が十分であるときには影響が小さいことが分かる．これより，十分な回答者数が確保できない場合には，誤推定率 Error(v) が v に依存する回答関数を用いることにより，収束度を変えずに匿名性を上げることが可能となる．

6.3 事前確率が既知である場合

4.1 において，回答 x の事前確率が不明，つまり $f_X(0) = f_X(1) = 0.5$ と仮定して，匿名度を定義した．しかしながら，‘1’ と回答する割合がある程度予測でき，これを利用して回答 x を推定する場合や，悪意のある質問者が集計結果から得られる ‘1’ と回答された割合の推定値を利用して，回答 x を推定する場合も考えられる．そこで，本節では事前確率が既知である場合について考察する．

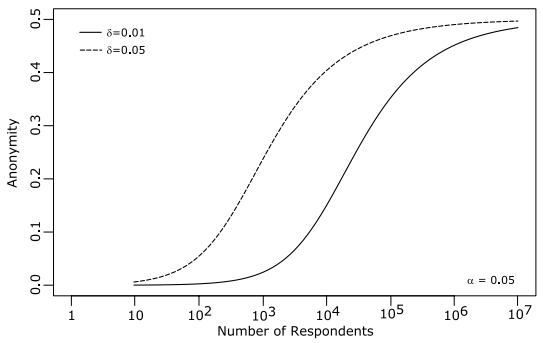

図 3　匿名度と回答者数の関係
Fig. 3　Anonymity v.s. number of sample.

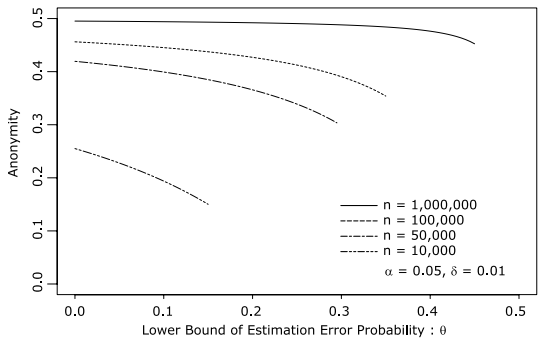

図 4　誤推定率の下限と匿名度の関係
Fig. 4　Anonymity v.s. lower bound of error(v).

事前確率を利用して，返信 v から回答 x を推定する場合，誤推定率は，$\min(f_X(0), f_X(1))$ 以下である．例えば $x = 1$ の事前確率が 0.1 である場合，返信値 v を利用しないで回答を推定する場合でさえ，誤推定率及び匿名度はともに 0.1 である．したがって，どのようなインターネット調査手法を採用したとしても誤推定率及び匿名度は 0.1 以下である．重要なことは，返信 v を利用して回答を推定した場合に，v を利用しない場合と比較して誤推定率や匿名度がどの程度低下するかである．もちろん，v を利用しない場合の誤推定率や匿名度に近い方が良い調査方式といえる．このような観点に立つならば，匿名性を（極力）保証するという意味で最良の回答関数は，やはり，誤推定率が返信値によらず一定で匿名度が最大，あるいは，誤推定率がある値 θ 以上で匿名度が最大の回答関数であり，そのような回答関数は **5.** と同様の手法で求めることができる．例えば，誤推定率が返信値によらず一定である場合，期待値条件 (1)，(2) と分散条件 (3) を満たす回答関数の中で，匿名度を最大とする 0/1 回答関数には以下の定理が成り立つ（以下に結果のみを示す）．

［定理 3］ 分散が σ^2 の 0/1 回答関数の中で，誤推定率 $Error(v)$ が一定，つまり，

$$\exists\theta(\theta > 0)\forall v \in \mathcal{V}\ \ Error(v) = \theta$$

なる条件を満たし，事前確率 $p_1 = f_X(1)$ が既知である場合，匿名度を最大にする 0/1 回答関数は以下の (f_0, f_1) である．

$$f_0(v) = \begin{cases} \dfrac{(1-\theta)(p_0-\theta)}{(1-2\theta)p_0} & : \quad v = -\dfrac{\theta p_1}{p_0-\theta} \\ \dfrac{\theta(p_1-\theta)}{(1-2\theta)p_0} & : \quad v = \dfrac{(1-\theta)p_1}{p_1-\theta} \\ 0 & : \quad \text{otherwise} \end{cases}$$

$$f_1(v) = \begin{cases} \dfrac{\theta(p_0-\theta)}{(1-2\theta)p_1} & : \quad v = -\dfrac{\theta p_1}{p_0-\theta} \\ \dfrac{(1-\theta)(p_1-\theta)}{(1-2\theta)p_1} & : \quad v = \dfrac{(1-\theta)p_1}{p_1-\theta} \\ 0 & : \quad \text{otherwise} \end{cases}$$

ただし，

$$p_0 = 1 - p_1$$
$$\theta = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{p(1-p)\sigma^2}{\sigma^2 + p^2}}$$
$$p = \max(p_1, p_0).$$

また，このときの匿名性は

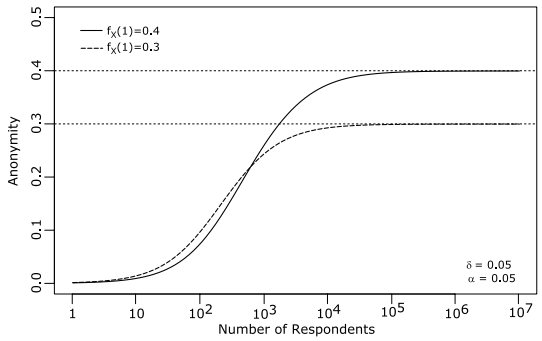

図 5　事前確率が既知である場合の匿名度と回答者数
Fig. 5　With known prior probability.

$$\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{p(1-p)\sigma^2}{\sigma^2 + p^2}} \tag{59}$$

となる． □

図 5 に，事前確率が既知である場合の回答者数と匿名度の関係を式 (59) から導出した結果を図示する．信頼度 95%（$\alpha = 0.05$），誤差 $\delta = 0.01$ とし，事前確率として $f_X(1) = 0.3, 0.4$ とした．これより，回答者数が大きくなる（すなわち，分散を大きくできる）と，匿名度は $\min(f_X(0), f_X(1))$ に近づくことが分かる．

また，事前確率が既知である場合，0/1 回答関数の分散をそれぞれ異なる値に設定する方が若干匿名度が向上することが考えられる．しかしながら，そのような回答関数は分散条件 (3) を満たさず，本論文の範疇を超えており，その解析は今後の課題とする．

7. むすび

本論文では，確率的変換を用いた匿名性保証技術について述べた．本手法は，回答者の回答を確率的に変換した乱数を，回答の代わりに質問者に送信することにより，匿名性を保つことを特徴とする．また，0/1 それぞれの回答に対応する確率的変換関数（回答関数）の分散を，所定の値に決定することにより，特定の精度で集計結果を得ることを可能とする．

調査数が与えられたとき，1 回答の推定値に要求される信頼度・精度を満たす 0/1 回答関数に対する制約は，その分散だけであることを示した．また，分散が σ^2 である 0/1 回答関数の中での最良性を評価するため，返信値から回答を推定する場合の困難さの尺度として，誤推定率及びその平均である匿名度を定義し，誤推定率・匿名度の観点から離散型の確率関数を用いた 0/1 回答関数が最良であることを示した．また，本

技術の適用範囲に関して解析を行った．その結果，本手法はサンプル数 n が 10^4 程度から適用可能であり，回答者数 $n = 10^4$ のとき，40%以上の匿名度を確保しつつ，信頼度 95%で誤差範囲 ± 0.05 の集計結果を得ることが可能であることを明らかにした．

本論文では確率的変換を用いた簡易な方法により，匿名性を保ちつつ，1 回答の割合を推定できることを示した．これにより，センサやユビキタス端末のようにリソースが限られた端末からも，匿名性を保ちつつ情報を収集することが可能となる．

文　献

[1] Everybody Votes Channel, http://www.nintendo.com/customer/wii/en_na/channelsEverybodyVotes.jsp

[2] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” Proc. EUROCRYPT’99, pp.223–238, Czech Republic, May 1999.

[3] J. Furukawa and K. Sako, “An efficient scheme for proving shuffle,” Crypto 2001, pp.368–387, California, Aug. 2001.

[4] S.L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” J. American Statistical Association, vol.60, no.309, pp.63–69, March 1965.

[5] A. Tagami, C. Sasaki, T. Hasegawa, S. Ano, and Y. Tomiura, “Analysis of answering method with probability conversion for Internet research,” Fifth Annual IEEE Consumer Communications & Networking Conference, NV, Jan. 2008.

[6] A. Tagami, C. Sasaki, T. Hasegawa, S. Ano, and Y. Tomiura, “Optimization of the answering method with probability conversion,” Workshop on Heuristic Methods for the Design, Deployment, and Reliability of Network and Network Applications, Finland, July 2008.

[7] J. Droitcour, R. Caspor, M. Hubbard, T. Parsley, W. Vissher, and T. Ezzati, “The item count technique as a method of indirect questioning: A review of its development and a case study application,” in Measurement Errors in Surveys, pp.185–210, John Wiley & Sons, New York, 1991.

[8] W. Du and Z. Zhan, “Using randomized response techniques for privacy-preserving data mining,” 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.505–510, Washington DC, Aug. 2003.

[9] P.L. Kooiman, L. Willenborg, and J. Gouweleeuw, “PRAM: A method for disclosure limitation of microdata,” Technical Report, Statistics Netherlands, 1997.

[10] J. Gouweleeuw, P. Kooiman, Willenborg, and P. Wolf, “Post randomization for statistical disclosure control: Theory and implementation,” J. Official Statistics, vol.14, pp.463–478, 1998.

[11] 野田一雄，宮岡悦良，数理統計学の基礎，共立出版，1992.

（平成 20 年 7 月 22 日受付，10 月 31 日再受付）

田上　敦士　（正員）

平 9 九州大学大学院システム情報科学研究科知能システム学専攻修士課程了．同年 KDD（株）入社．以来，研究所にて，高速通信プロトコル，オーバレイネットワークに関する研究に従事．現在，（株）KDDI 研究所 IP 品質制御システムグループ主任研究員．

佐々木　力　（正員）

平 16 東京工業大学大学院理工学研究科集積システム専攻修士課程了．同年 KDDI（株）入社．QoS，マルチキャストの研究に従事．現在，（株）KDDI 研究所 IP 品質制御システムグループ研究員．

長谷川輝之　（正員）

平 5 京都大学大学院修士課程了．同年 KDD（株）入社．以来，研究所にて，高速通信プロトコル，次世代インターネットの研究に従事．現在，（株）KDDI 研究所 IP 品質制御システムグループ主任研究員．博士（情報理工学）平 15 年度電波産業会電波功績賞受賞．

阿野　茂浩　（正員）

平元早稲田大学大学院修士課程了．同年 KDD（株）入社．以来，研究所にて，ATM 交換方式，IP ネットワーク管理・制御，次世代インターネットの研究に従事．現在，（株）KDDI 研究所 IP 品質制御システムグループリーダ．平 7 年度情報処理学会学術奨励賞受賞．

冨浦　洋一

昭 59 九大・工・電子卒，平元同大大学院工学研究科電子工学専攻博士課程単位取得退学．同年九州大学工学部助手，平 7 同助教授，現在，九州大学大学院システム情報科学研究院准教授．博士（工学）．統計的自然言語処理，計算言語学に関する研究に従事．平 3 年度情報処理学会研究賞．Pacling2005 Best Paper Award，FIT2006 論文賞受賞．