

標本比率の分散に関する一考察

鈴木啓祐

はしがき*

いま、母集団は、 Γ 個の赤球と Δ 個の白球（球の総個数は T 個、すなわち、 $\Gamma + \Delta = T$ とする）とから構成され、この母集団は、 $\tau_1, \tau_2, \dots, \tau_i, \dots, \tau_N$ 個の球からなる N 個の部分に分けられ、それぞれの部分は、赤球を $\gamma_1, \gamma_2, \dots, \gamma_i, \dots, \gamma_N$ 個含んでいると仮定する（図1 (Figure 1)）。この母集団から、 n 個の部分抽出し、その各部分の個数が、 $t_1, t_2, \dots, t_j, \dots, t_N$ であり、その各部分に含まれる赤球の個数が、 $c_1, c_2, \dots, c_j, \dots, c_n$ であるとき、この t_j ($j=1, 2, \dots, n$) と c_j ($j=1, 2, \dots, n$) とから得られる

$$p = \frac{\bar{c}}{\bar{t}} \tag{0.1}$$

で定義される p を母集団の赤球の構成比率 π 、すなわち、

$$\pi = \frac{\Gamma}{\Gamma + \Delta} (= \frac{\Gamma}{T}) \tag{0.2}$$

の推定量とすることができよう。ただし、

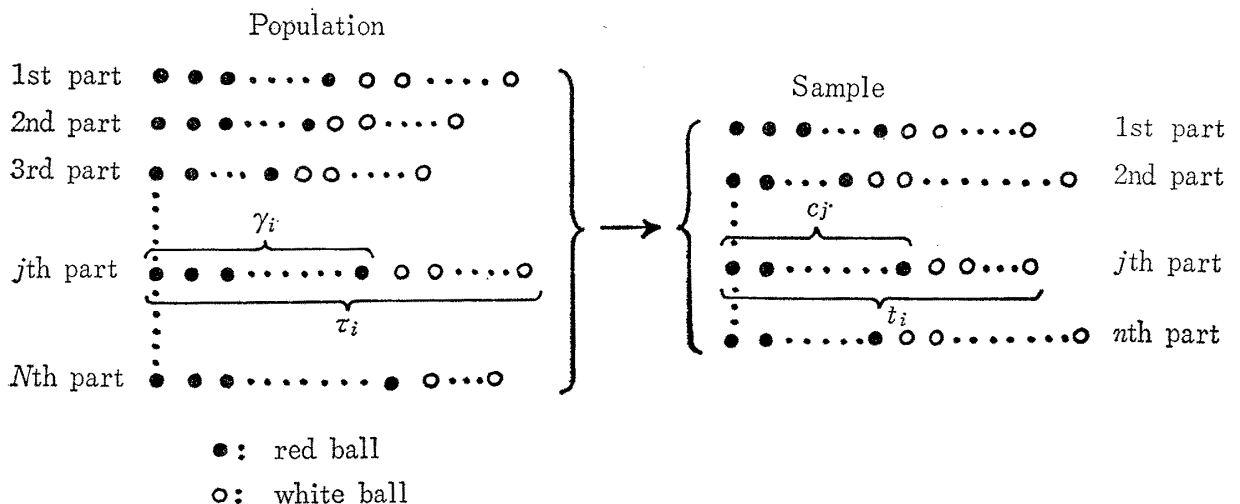
$$\bar{c} = \frac{1}{n} \sum_{j=1}^n c_j \tag{0.3.1}$$

$$\bar{t} = \frac{1}{n} \sum_{j=1}^n t_j \tag{0.3.2}$$

である。このとき、 p の分散 $V(p)$ が得られることが望ましいが、すでに、Cochran¹⁾ や多賀²⁾ が述べているように、この分散については、検討がおこなわれ、その近似的大きさが得られている。

ここでは、この p の分散に関して興味ある検

Figure 1 The structure of the population and the sample questioned



* この論文の作成にあたり、貴重なコメントを与えられた千葉大学の多賀保志教授に深く謝意を表す。

1) Cochran, William G.: *Sampling Techniques*, New York, John Wiley, 1963 (Second edition) pp.

2) 多賀保志:『サンプリング調査の理論』, 東京, サイエンス社, 1976年, 92頁.

討結果が得られたので、その結果について報告したい。

なお、ここでとり扱われる理論は、一般に、比率の区間推定、たとえば「集落抽出法によって得られた製品中の不良品の数」から不良品の混入率を推定したり、あるいは「各世帯から得られた総収入と総消費支出の値」から全世帯の総所得中の消費支出の割合を推定したりする場合の「比率の区間推定」に有効な理論である。

I p の分散 $V(p)$ ($\tau_i \neq \bar{\tau}$ の場合)

p の分散 $V(\hat{p})$ は、

$$V(p) = E[(p - \hat{p})^2] \\ = \frac{1}{N^*} \sum_{k=1}^{N^*} (p_k - \hat{p})^2 \quad (1.1)$$

で定義される。ただし、 p_k とは、母集団に含まれる N 個の部分から n 個の部分抽出して得られるあらゆる標本のうちの第 k 番目の標本（第 k 標本）から得られる π の推定値 p であり、 \hat{p} は、 p_k の平均値、すなわち、

$$\hat{p} = \frac{1}{N^*} \sum_{k=1}^{N^*} p_k \quad (= E(p)) \quad (1.2)$$

である。また、 N^* は、母集団の N 個の部分から、 n 個の部分抽出して得られる標本（大きさ n の標本）の総数であり、 ${}_N C_n$ である。

多賀が示したように³⁾、母集団の第 i 部分に含まれる赤球の個数の理論値を $\pi\tau_i$ 、そして、これと実際の赤球の個数 γ_i との差を ε_i とすれば、 ε_i は、

$$\varepsilon_i = \gamma_i - \pi\tau_i \quad (1.3.1)$$

によって示される。また、標本の第 j 部分に含まれる赤球の個数の理論値を πt_j 、そして、これと実際の赤球の個数 c_j との差を e_j とすれば、 e_j は、

$$e_j = c_j - \pi t_j \quad (1.3.2)$$

によって示される。式 (1.3.2) から、

$$\sum_{j=1}^n e_j = \sum_{j=1}^n c_j - \pi \sum_{j=1}^n t_j \quad (1.4)$$

が得られる。したがって、この式から、

$$\frac{1}{n} \sum_{j=1}^n e_j = \frac{1}{n} \sum_{j=1}^n c_j - \pi \left(\frac{1}{n} \sum_{j=1}^n t_j \right) \quad (1.5)$$

が成立する。 $\sum_{j=1}^n e_j/n$ 、 $\sum_{j=1}^n c_j/n$ 、 $\sum_{j=1}^n t_j/n$ を、それぞれ、 \bar{e} 、 \bar{c} 、 \bar{t} とすれば、

$$\bar{e} = \bar{c} - \pi \bar{t} \quad (1.6.1)$$

したがって、

$$\frac{\bar{e}}{\bar{t}} = \frac{\bar{c}}{\bar{t}} - \pi \quad (1.6.2)$$

が得られる。

式 (0.1) によれば、 p は、

$$p = \frac{\bar{c}}{\bar{t}} \quad (0.1')$$

で与えられるが、ここで、この式の両辺から π を引けば、

$$p - \pi = \frac{\bar{c}}{\bar{t}} - \pi \quad (1.7)$$

が得られる。

式 (1.6) と式 (1.7) とから、

$$p - \pi = \frac{\bar{e}}{\bar{t}} \quad (1.8)$$

が成立する。この式の両辺に、 $\bar{\tau}/\bar{\tau}$ ($=1$) を乗じると、

$$p - \pi = \frac{\bar{e}}{\bar{\tau}} \cdot \frac{\bar{\tau}}{\bar{t}} \quad (1.9)$$

となる。ただし、 $\bar{\tau}$ は母集団の τ の平均値、すなわち、

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^N \tau_i \quad (1.10)$$

である。ここで、『 n が十分大きいとき、大数の法則 (law of large numbers)⁴⁾ により、

$$\bar{t} = \bar{\tau} \quad (1.11)$$

となるとみなし得る⁵⁾ から、 n が十分大きいとき、式 (1.9) と式 (1.11) から、

$$p - \pi = \frac{\bar{e}}{\bar{\tau}} \quad (1.12)$$

4) たとえば、宮沢光一：『近代数理統計学通論』、東京、共立出版、1954年、70-71頁。

5) Cochran, William G.: *op. cit.*, p. 30.

3) 多賀保志：前掲書、92頁。

となる。したがって、式 (1.12) の左辺の期待値 $E(p-\pi)$ は、式 (1.6.1) を用いて、

$$\begin{aligned} E(p-\pi) &= E\left(\frac{\bar{p}}{\bar{\tau}}\right) \\ &= E\left(\frac{\bar{c}-\pi\bar{t}}{\bar{\tau}}\right) \\ &= \frac{1}{\bar{\tau}}\{E(\bar{c})-E(\pi\bar{t})\} \\ &= \frac{1}{\bar{\tau}}(\bar{\gamma}-\pi\bar{\tau}) \end{aligned} \quad (1.13)$$

が得られる。

また、定義により、

$$\begin{aligned} \pi &= \frac{\sum_{i=1}^N \gamma_i}{\sum_{i=1}^N \tau_i} \\ &= \frac{\bar{\gamma}}{\bar{\tau}} \end{aligned} \quad (1.14)$$

であるから、

$$\bar{\gamma}-\pi\bar{\tau}=0 \quad (1.15)$$

となる。式 (1.13) と式 (1.15) とによって、

$$E(p-\pi)=0 \quad (1.16.1)$$

すなわち、

$$E(p)=\pi \quad (1.16.2)$$

が得られる。

式 (1.1) の \hat{p} は、式 (1.2) の定義から明らかのように、 $E(p)$ であるから、式 (1.1) は、式 (1.16.2) により、

$$V(p)=\frac{1}{N^*}\sum_{k=1}^{N^*}(p_k-\pi)^2 \quad (1.17)$$

と書かれ得る。

この式の p_k は、式 (0.1) の p に等しい（ここでは、第 k 標本から得られた式 (0.1) の p が p_k によって示されている）から、 $V(p)$ は、さらに、

$$V(p)=E\left[\left(\frac{\bar{c}}{\bar{t}}-\pi\right)^2\right] \quad (1.18)$$

によって示され得る。したがって、ここで、ふたたび、大数の法則による『 n が十分大きいと

き、

$$\bar{t} \doteq \bar{\tau} \quad (1.11')$$

となる』という性質を用いれば、

$$\begin{aligned} V(p) &= E\left[\left(\frac{\bar{c}}{\bar{\tau}}-\pi\right)^2\right] \\ &= \frac{1}{\bar{\tau}^2}E(\bar{c}^2)-\pi^2 \end{aligned} \quad (1.19)$$

を得る。

いま、ここで、 $E(\bar{c}^2)$ を計算すると、下記のようなになる。

母集団には、 N 個の部分があり、その各部には、赤球が、それぞれ、 $\gamma_1, \gamma_2, \dots, \gamma_i, \dots, \gamma_N$ 個含まれている。したがって、この N 個の部分からなる母集団から n 個の部分抽出することは、赤球の抽出によって表現するならば、「 $\gamma_1, \gamma_2, \dots, \gamma_N$ 個の赤球からなる n 個の赤球集団によって構成される母集団（大きさ N の母集団）から大きさ n の標本——この標本は、 c_1, c_2, \dots, c_n 個の赤球からなる n 個の赤球集団によって構成される。——を抽出すること」になる。

第 k 標本に見られる赤球の個数 $c_1, c_2, \dots, c_j, \dots, c_n$ の平均値を \bar{c}_k とすれば、

$$\begin{aligned} E(\bar{c}^2) &= \frac{1}{N^*}\sum_{k=1}^{N^*}\bar{c}_k^2 \\ &= \frac{1}{N^*}\sum_{k=1}^{N^*}\left\{\frac{1}{n}(c_{k1}+c_{k2}+\dots+c_{kn})\right\}^2 \\ &= \frac{1}{n^2N^*}\left\{\sum_{k=1}^{N^*}\sum_{j=1}^n c_{kj}^2 \right. \\ &\quad \left. + \sum_{k=1}^{N^*}\left(\sum_{j=1}^n \sum_{p=1, p \neq j}^n c_{kj}c_{kp}\right)\right\} \end{aligned} \quad (1.20)$$

となる。ただし、 c_{kj} とは、第 k 標本における第 j 番目の c の値 (c_j) である。

式 (1.20) の右辺の各項の値を γ_i を用いて表現すると、まず、

$$\sum_{k=1}^{N^*}\sum_{j=1}^n c_{kj}^2 = \frac{nN^*}{N}\sum_{i=1}^N \gamma_i^2 \quad (1.21.1)$$

となり、ついで、

$$\sum_{k=1}^{N^*} \left(\sum_{j=1}^n \sum_{p=1}^n c_{kj} c_{kp} \right) = \frac{n(n-1)}{N(N-1)} \sum_{i=1}^N \sum_{s=1}^N r_i r_s \quad (1.21.2)$$

となる。したがって、

$$E(\bar{c}^2) = \frac{1}{nN} \sum_{i=1}^N r_i^2 + \frac{n-1}{nN(N-1)} \sum_{i=1}^N \sum_{s=1}^N r_i r_s \quad (1.22)$$

となる。この結果を式 (1.19) に代入すれば、

$$V(p) = \frac{1}{\bar{\tau}^2} \left(\frac{1}{nN} \sum_{i=1}^N r_i^2 + \frac{n-1}{nN(N-1)} \sum_{i=1}^N \sum_{s=1}^N r_i r_s \right) - \pi^2 \quad (1.23)$$

となる。

他方、 r の分散 $V(r)$ は、

$$\begin{aligned} V(r) &= E[(r - \bar{r})^2] \\ &= \frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2 \\ &= \frac{1}{N} \sum_{i=1}^N r_i^2 - \bar{r}^2 \end{aligned} \quad (1.24)$$

によって定義される。

$$\bar{r}^2 = \left\{ \frac{1}{N} (r_1 + r_2 + \dots + r_N) \right\}^2 \quad (1.25)$$

であるから、

$$N^2 \bar{r}^2 = \sum_{i=1}^N r_i^2 + \sum_{i=1}^N \sum_{s=1}^N r_i r_s \quad (1.26)$$

となる。式 (1.24) から、

$$NV(r) = N \sum_{i=1}^N r_i - N \bar{r}^2 \quad (1.27)$$

が、また、この関係から、

$$\sum_{i=1}^N r_i^2 = NV(r) + N \bar{r}^2 \quad (1.28)$$

が得られるから、式 (1.26) は、式 (1.28) を用いて、

$$N^2 \bar{r}^2 = NV(r) + N \bar{r}^2 + \sum_{i=1}^N \sum_{s=1}^N r_i r_s \quad (1.29)$$

となる。この関係を用いて、

$$\sum_{i=1}^N \sum_{s=1}^N r_i r_s = N(N-1) \bar{r}^2 - NV(r) \quad (1.30)$$

が得られる。

これらの結果のうち、式 (1.27) と式 (1.29) とを式 (1.23) に代入すれば、

$$\begin{aligned} V(p) &= \frac{1}{\bar{\tau}^2} \left[\frac{1}{nN} (NV(r) + N \bar{r}^2) \right. \\ &\quad \left. + \frac{n-1}{nN(N-1)} \{N(N-1) \bar{r}^2 - NV(r)\} \right] - \pi^2 \\ &= \frac{N-n}{n(N-1)} \cdot \frac{V(r)}{\bar{\tau}^2} \end{aligned} \quad (1.31)$$

となる。したがって、

$$V(p) = \frac{N-n}{N-1} \cdot \frac{V(r)}{n \bar{\tau}^2} \quad (1.32)$$

となる。

II p の分散 $V(p)$ ($\tau_i = \bar{\tau}$ の場合)

式 (1.32) によれば、 p の分散 $V(p)$ は、

$$V(p) = \frac{N-n}{N-1} \cdot \frac{V(r)}{n \bar{\tau}^2} \quad (2.1)$$

である。ところで、

$$V(\varepsilon) = E[(\varepsilon - \bar{\varepsilon})^2] \quad (2.2)$$

で定義される $\varepsilon (= \gamma - \pi\tau)$ の分散 $V(\varepsilon^2)$ は、

$$\begin{aligned} \bar{\varepsilon} &= E(\varepsilon) \\ &= E(\gamma - \pi\tau) \\ &= E(\gamma) - \pi E(\tau) \\ &= \bar{\gamma} - \pi \bar{\tau} \\ &= 0 \end{aligned} \quad (2.3)$$

であるから、

$$\begin{aligned} V(\varepsilon^2) &= E(\varepsilon^2) \\ &= E[(\gamma - \pi\tau)^2] \end{aligned} \quad (2.4)$$

である。

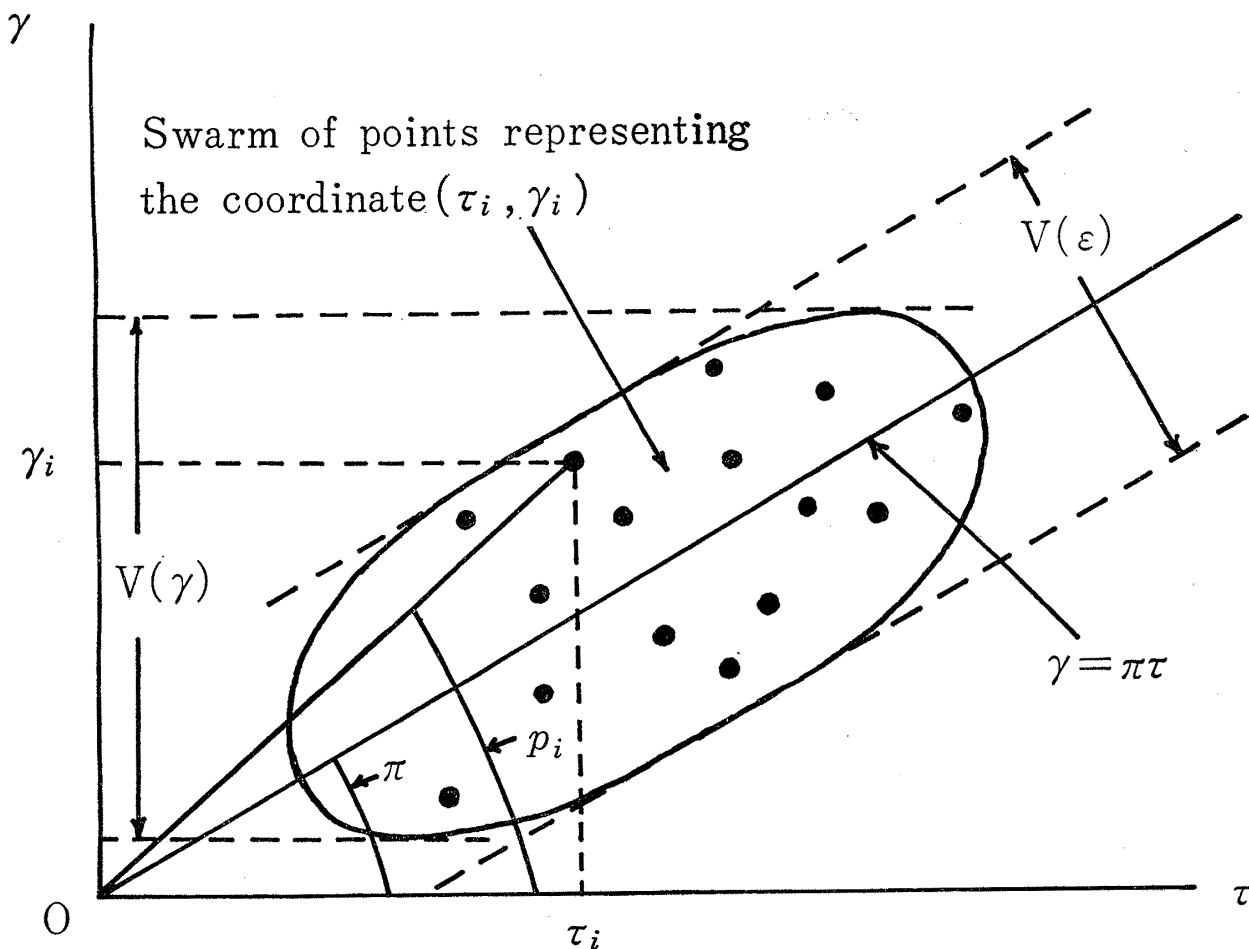
ここで、「母集団の各部分に含まれる赤球、白球の個数 τ_i がすべて等しく、 $\bar{\tau}$ である」という仮定を導入すれば、

$$\tau_i = \bar{\tau} \quad (2.5)$$

であるから、

$$\begin{aligned} V(\varepsilon^2) &= E[(\gamma - \pi\tau)^2] \\ &= E[(\gamma - \pi\bar{\tau})^2] \\ &= E[(\gamma - \bar{\gamma})^2] \end{aligned}$$

Figure 2 Distribution of γ_i and p_i in the graph which have τ - and γ -axes



$$= V(\gamma^2) \quad (2.6)$$

となる。すなわち、 $\tau_i = \bar{\tau}$ のとき、

$$V(\gamma) = V(\epsilon) \quad (2.7)$$

である。したがって、このとき、

$$V(p) = \frac{N-n}{N-1} \cdot \frac{V(\epsilon)}{n\bar{\tau}^2} \quad (2.8)$$

となる。これは、Cochran⁶⁾ または多賀⁷⁾ が示した $V(p)$ に一致する。

III 考 察

p の分散 $V(p)$ は、式 (1.31) によって示されるが、いま、この式を

$$V(p) = \frac{N-n}{(N-1)n\bar{\tau}^2} V(\gamma) \quad (3.1)$$

のように書きかえてみると、この式は、 $V(p)$ が $V(\gamma)$ に $(N-n)/\{(N-1)n\bar{\tau}^2\}$ 倍したものであることを示していることが明らかになる。したがって、 $V(p)$ の式をこのように変形した場合、 $V(p)$ の要素を図示することができる。

式 (3.1) を得る場合には、 γ_i の理論値を $\pi\tau_i$ とした。この理論値は、 γ_i が $\pi\tau_i$ の近くに現われることを前提として得られたものである。実際、式 (1.3.1) を変形すれば、

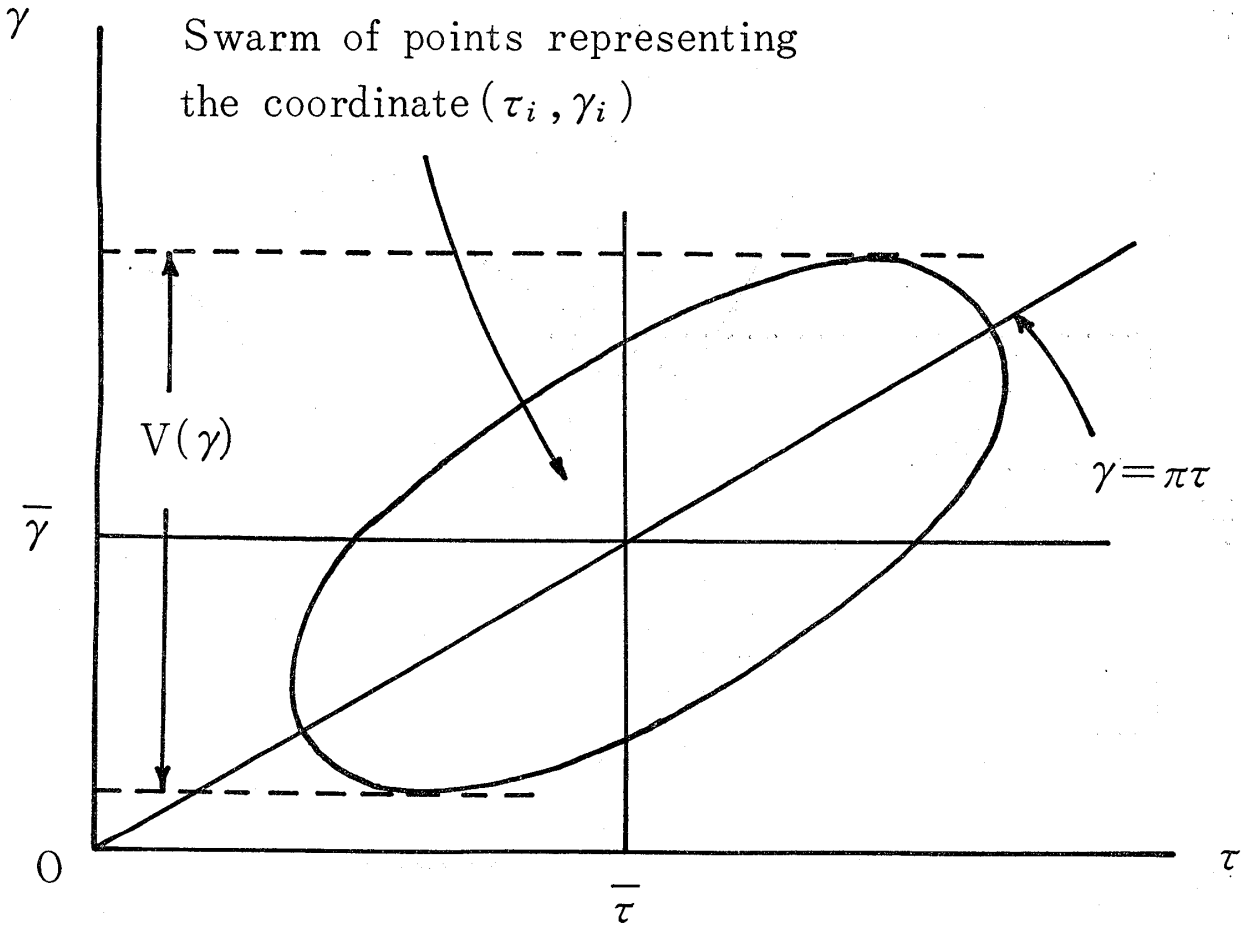
$$\gamma_i = \pi\tau_i + \epsilon_i \quad (3.2)$$

となる。この式で示される γ_i の分布を、 τ_i と関連させながら図示すれば、いいかえれば、横軸に τ の値 τ_i を、縦軸に γ の値 γ_i を示すような図に示せば、図 2 (Figure 2) のようになる。この図に示されるように、 γ_i を示す点は、回帰直線、

6) Cochran, William G.: *op. cit.*, p. 31.

7) 多賀保志: 前掲書, 92頁.

Figure 3 Graphical representation of variance $V(\gamma)$ ($\tau_i \neq \bar{\tau}$)



$$\gamma = \pi\tau \quad (3.3)$$

の周囲に現われ、 $p_i (= \gamma_i / \tau_i)$ は、原点と τ_i とを結ぶ直線と横軸とのなす角によって示される。

$\tau_i \neq \bar{\tau}$ の場合の $V(p)$ は、

$$\frac{N-n}{(N-n)n\bar{\tau}^2} = \omega \quad (3.4)$$

としたとき、

$$V(p) = \omega V(\gamma) \quad (3.5)$$

となるが、このときの $V(\gamma)$ は、図2の τ_i の γ 軸（縦軸）方向の分散（図3 (Figure 3)) であり、したがって、 $V(p)$ は、図3の τ_i の γ 軸方向の分散 $V(\gamma)$ に ω を乗じたものであるといえる。

また、 $\tau_i \neq \bar{\tau}$ の場合の $V(p)$ は、

$$V(p) = \omega V(\varepsilon) \quad (3.6)$$

となるが、このときの $V(\varepsilon)$ は、やはり、 τ_i の γ 軸方向の分散（図4 (Figure 4)) であり、——しかも、 τ_i は $\tau = \bar{\tau}$ で示される直線上にのみ現われるから、—— $V(\varepsilon)$ は、この場合の $V(\gamma)$ に等しい。——したがって、 $V(p)$ は、図4の τ_i の γ 軸方向の分散 $V(\varepsilon)$ に ω を乗じたものである。

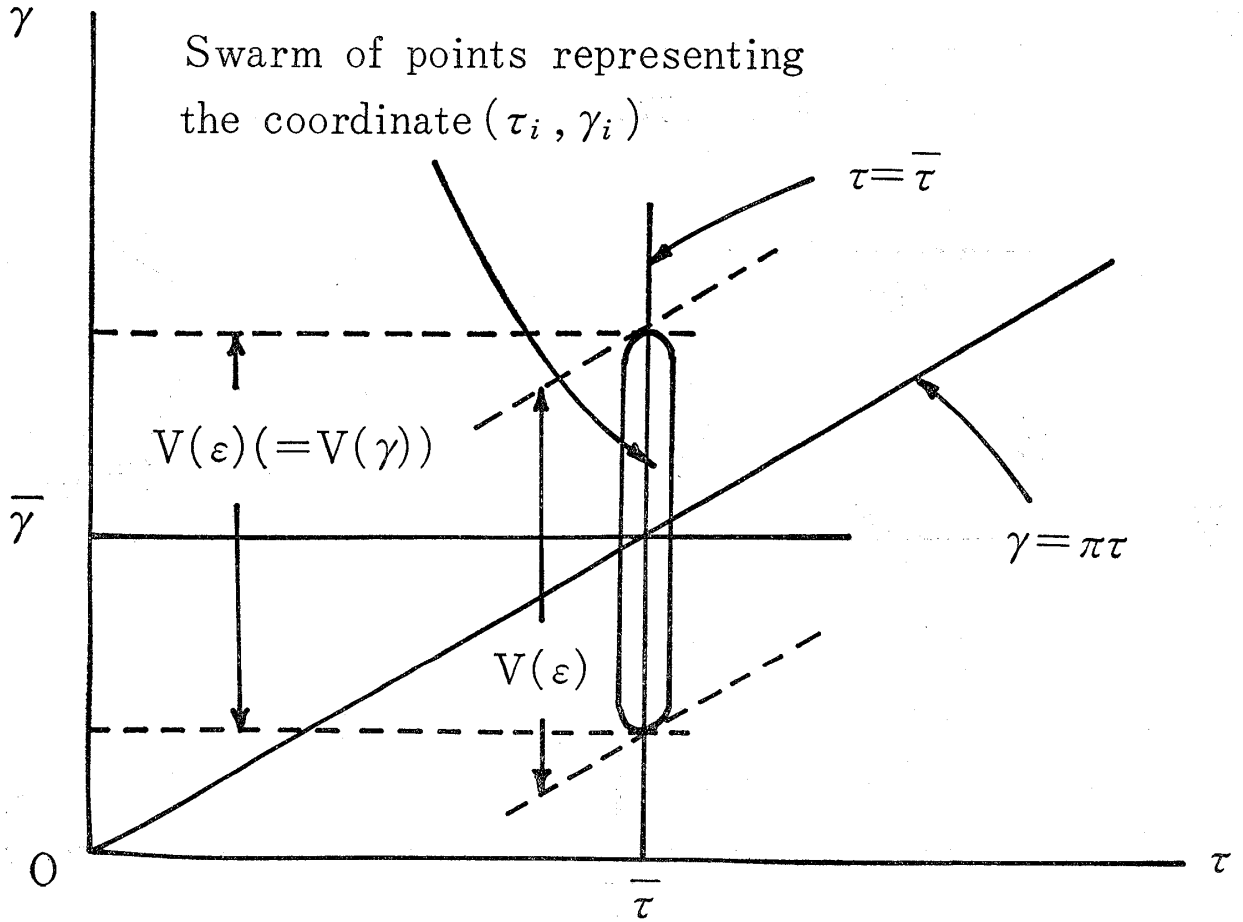
IV 結 語

T 個の球からなる母集団があり、その T 個のうち、 Γ 個は赤球であり、 A 個は白球であるとする。

また、この母集団が、 $\tau_1, \tau_2, \dots, \tau_N$ 個の球からなる N 個の部分に分けられ、それぞれの部分は、 $\gamma_1, \gamma_2, \dots, \gamma_N$ 個の赤球を含んでいるとする。

このとき、 n 個の部分抽出すると（いいか

Figure 4 Graphical representation of variance $V(\gamma)$ ($\tau_i = \bar{\tau}$)



えれば、一種の集落抽出をすると), t_1, t_2, \dots, t_n 個の球からなる n 個の部分から, c_1, c_2, \dots, c_n の赤球が得られる。

これらの t_j と c_j ($j=1, 2, \dots, n$) とから得られる,

$$p = \frac{\sum_{j=1}^n c_j}{\sum_{j=1}^n t_j} \quad (4.1)$$

で定義される比率 p は,

$$\pi = \frac{\Gamma}{\Gamma + \Delta} \left(= \frac{\Gamma}{T} \right) \quad (4.2)$$

で定義される母集団比率 (母集団の赤球の構成比率) π の不偏推定量となる。そして、その推定量から得られる p の分散 $V(p)$ は、 ω を

$$\omega = \frac{N-n}{(N-n)n\bar{\tau}^2} \quad (4.3)$$

と定義したとき、 $\tau_i = \bar{\tau}$ の場合、

$$V(p) = \omega V(\gamma) \quad (4.4)$$

であり、また、 $\tau_i = \bar{\tau}$ の場合、

$$V(p) = \omega V(\varepsilon) \quad (4.5)$$

であることが明らかになった。式 (4.5) の $V(p)$ は、Cochran や多賀によって示されている $V(p)$ と一致する。

また、 $V(p)$ の決定要素の一つになっている $V(\gamma)$ 、あるいは、 $V(\varepsilon)$ は、横軸を τ 、縦軸を γ としてえがいた γ_i の分布を示すグラフにおいては、「いずれも、 γ_i の縦軸 (γ 軸) 方向の分布の分散である」ことが明らかにされた。

Synopsis

SUZUKI, KEISUKE: On the Variance of

the Ratio Obtained by Sampling, *The Journal of Ryūtsū Keizai University (Ryūtsū Keizai Daigaku Ronshū)* Vol. 15, No. 4, 1981/3, pp. 28-35.

It is supposed that a population contains Γ red balls and Δ white balls and it is divided into N parts. And it is also supposed that the i th part of population contains τ_i balls, and γ_j red balls (Figure 1).

If the n parts in this population are chosen, the sample ratio p which is defined by

$$p = \frac{\sum_{j=1}^n c_j}{\sum_{j=1}^n t_j}$$

is the unbiased estimator of the population ratio π which is written by

$$\pi = \frac{\Gamma}{\Gamma + \Delta} = \frac{\Gamma}{T}$$

where t_j is the number of balls contained in the j th part of sample, c_j is the number of red balls contained in j th part of the sample, and T is the total number of balls contained in the population.

When $\tau_i \neq \bar{\tau}$, the variance of p , $V(p)$ is approximately expressed by

$$V(p) = \omega V(\gamma),$$

where

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^N \tau_i$$

$$V(\gamma) = \frac{1}{N} \sum_{i=1}^N r_i$$

$$\omega = \frac{N-n}{(N-n)n\bar{\tau}^2}$$

And, when $\tau_i = \bar{\tau}$, $V(p)$ is approximately ex-

pressed by

$$V(p) = \omega V(\epsilon)$$

which is exactly equal to the result found in Cochran's book,* where

$$V(\epsilon) = E[(\gamma_i - \pi\tau_i)^2]$$

It is very interesting that if we represent γ_i in the graph which has coordinate axes, τ -axis and γ -axis as shown in Figure 2, the variances $V(\gamma)$ and $V(\epsilon)$ can be expressed clearly in the graph. Both the variances $V(\gamma)$ and $V(\epsilon)$ are regarded as the indicator of the dispersion of distribution of γ in the direction of γ -axis (Figure 2).

When $\tau_i \neq \bar{\tau}$, the points which show the γ_i ($i=1, 2, \dots, N$) distribute along the regression line :

$$\gamma = \pi\tau$$

and in this case, the variance $V(\gamma)$ is the indicator of the dispersion of the distribution of γ in the direction of γ -axis (Figure 3).

When $\tau_i = \bar{\tau}$, the points which show the τ_i 's ($i=1, 2, \dots, N$) distribute along the line $\tau = \bar{\tau}$, since τ_i 's are always $\bar{\tau}$, and in this case, the variance $V(\epsilon)$ is the indicator of the dispersion of distribution of γ in the direction of γ -axis, and the variance $V(\epsilon)$ becomes exactly equal to the variance $V(\gamma)$ (Figure 4).

* Cochran, William G.: *Sampling Techniques*, New York, John Wiley, 1963 (Second edition), p. 31.