任意標本調査法（三）

関　弥三郎

三　集落抽出法
(1) 集落抽出法に於ける推定式
(2) 母分数の推定
(3) 集落抽出法の精度
四　多段抽出法
(1) 多段抽出法に於ける期望値計算
(2) 二段抽出法に於ける推定式
(3) 適正配分
(4) 母分散の推定
(5) 二段抽出法の精度
(6) 三段抽出法に於ける推定式

三　集落抽出法

(1) 集落抽出法に於ける推定式

集落抽出法は母集団の構成単位を若干個宛集めて集落を作り母集団を集落の集団となし、これより集落を抽出して標本を構成する方法であるため、集落を母集団及び標本の構成単位と見る限り、集落抽出法は単純抽出法の場合に帰着する事になる。

集落抽出法（及び多段抽出法）に於ては、集落に就ての値と調査単位に就ての値との区別を明確にする事が必要である。以下集落の値は大文字で、調査単位の値は小文字で表し

x_{jk}…j番集落内のk番調査単位の変量

$\bar{x}_j$…j番集落に於ける調査単位当り平均値

$\bar{x}$…母集団に於ける調査単位当り平均値

X_j…j番集落の量的大いさ（集落を構成する単位の変量の総和）

$\bar{\bar{X}}$…母集団に於ける集落当り平均値

の記号とする。そして標本として抽出された単位の変量及びこれ等の母数の標本推定値は、ダッシュを附けて母集団に於ける値と区別する事にする。

今 N 個の単位より成る母集団 π を M 個の集落の母集団 π' に変換し、各集落は N_j 個の単位を有しその量的大いさを $X_j=\sum_{k=1}^{Nj}x_{jk}$,（$j=1, 2, \cdots\cdots, M$）とすると、$\pi'$ の平均値（即ち集落の量的大いさの平均）$\bar{X}$ 及び分散（即ち集落の量的大いさの分散）σ_e^2 は

$$\bar{X}=\frac{1}{M}\sum_{j=1}^{M}X_j \quad , \quad \sigma_e{}^2=\frac{1}{M}\sum_{j=1}^{M}(X_j-\bar{X})^2 \qquad (3\cdot1)$$

である。この π' より m 個の集落を抽出して標本を構成し、それより母平均 $\bar{X}$ を推定する場合、その不偏推定値 $\bar{X}'$ 及び分散は、(2・2)(2・31)及び(2・33)式に於て $x_j'=X_j$, $\bar{x}'=\bar{X}'$, $\bar{x}=\bar{X}$, $N=M$, $n=m$, $\sigma^2=\sigma_e{}^2$ と置けばよい。

故に

$$\bar{X}'=\frac{1}{m}\sum_{j=1}^{m}X_j{}^{1)} \qquad (3\cdot2)$$

$$E(\bar{X}')=\bar{X} \qquad (3\cdot3)$$

$$V(\bar{X}')=\frac{M-m}{M-1}\frac{\sigma_e{}^2}{m} \qquad (3\cdot4)$$

である。

1) 茲に標本に於ける集落の量的大いさ X_j にダッシュを附けないのは、この X_j は標本に於ける値とはいえ集落の全部調査によって知り得た正確な値であるから、それの推定値と区別するためである。この事は多段抽出法に於て記号の意味を明瞭にする利便を有するのである。

しかしながら一般に母集団の構成単位を集落に分けるのは、調査上の便宜に過ぎないのであって、真の目的は元の母集団 π に於ける母平均 $\bar{x}$ の推定にある。この場合 $\bar{x}$ は π' に於ては

$$\bar{x}=\frac{1}{N}\sum_{j=1}^{N}x_j=\frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{Nj}x_{jk}$$

$$=\frac{1}{N}\sum_{j=1}^{M}X_j=\frac{M\bar{X}}{N}\ [(3\cdot1)により] \qquad (3\cdot5)$$

であるため、集落平均の推定値に M/N を乗ずる事によってそれを推定する事が出来る。故に $\bar{x}$ の推定値 $\bar{x}'$ は

$$\bar{x}'=\frac{M}{N}\bar{X}'=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}X_j \qquad (3\cdot6)$$

であり[1)]、

$$E(x') = \frac{M}{N}E(\bar{X}') = \frac{M}{N}\bar{X} = \bar{x} \tag{3·7}$$

なるため、この推定値は不偏である。そして分散は

$$V(x') = \left(\frac{M}{N}\right)^2 V(\bar{X}') = \left(\frac{M}{N}\right)^2 \frac{M-m}{M-1}\frac{\sigma_e^2}{m} \tag{3·8}$$

となる。

1) (3·6)式は母集団πに於ける単位総数Nが既知である事を必要とする。若しそれが判って居らない時はそれをも推定しなければならない。

(3·2)式に於てX_jをその集落内の単位の数N_jとすれば $\bar{N}' = \frac{1}{m}\sum_{j=1}^{m} N_j$ は集落当り平均単位数 $\bar{N} = \frac{1}{M}\sum_{j=1}^{M} N_j = \frac{N}{M}$ の不偏推定値であり、これをM倍する事によってNの不偏推定値N'が得られる。即ち

$$N' = \frac{M}{m}\sum_{j=1}^{m} N_j \tag{3·9}$$

かくて(3·6)式は次のようになる。

$$x'' = \frac{M\bar{X}'}{N'} = \frac{\sum_{j=1}^{m} X_j}{\sum_{j=1}^{m} N_j} \tag{3·10}$$

しかるに(3·10)式に於てはN_jがすべて同一でない限り、分子も分母も共に確率変数であるため、(3·6)式と標本理論上の性質を全く異にし、これは比推定の一種であって一般には不偏推定ではないのである。故に(3·10)式の分散は最早や(3·8)式ではなく、別に規定されねばならないのである。（比推定の項参照）

以上は平均値の推定であるが、母集団πに於ける標識αを有する単位の割合pを推定せんとする時は、α標識に1、α以外の標識に0なる値を与えると、平均値は比例数を表す事になるから、平均値の場合と同様にして推定する事が出来る。故にpの不偏推定値及びその分散は、(3·6)(3·8)より

$$p' = \frac{1}{N}\frac{M}{m}\sum_{j=1}^{m} X_j \tag{3·11}$$

$$V(p') = \left(\frac{M}{N}\right)^2 \frac{M-m}{M-1}\frac{\sigma_e^2}{m} \tag{3·12}$$

となる。但しこの場合$X_j = \sum_{k=1}^{N_j} x_{jk}$は$j$番集落に於ける$\alpha$標識の単位の数である。

集落の大いさが均一なる場合　尚集落の大いさがすべて均一$N_j = \bar{N}$,（$j=1,2,\ldots\ldots,M$）である場合は、不偏推定式及びその分散式は比較的簡単になり、且後述する如くその推定の精度も昂るのである。

茲で集落の大いさが均一である場合に、母集団に於ける

諸種の分散が如何に変形されるかを考察して置くのが便利である。

母集団πを集落より成る母集団π'に変換した場合、πの分散σ^2は $\sigma^2=\sigma_w^2+\sigma_b^2$ 即ち集落内分散と集落間分散とより合成されているのである。（第一講 p. 125参照）そして集落内分散σ_w^2は

$$\sigma_w^2=\frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{Nj}(x_{jk}-\bar{x}_j)^2$$

$$=\frac{1}{N}\sum_{j=1}^{M}\frac{N_j}{N_j}\sum_{k=1}^{Nj}(x_{jk}-\bar{x}_j)^2=\frac{1}{N}\sum_{j=1}^{M}N_j\sigma_j^2 \quad (3\cdot13)$$

但 $\sigma_j^2=\dfrac{1}{Nj}\sum_{k=1}^{Nj}(x_{jk}-\bar{x}_j)^2$……

……j 番集落に於ける調査単位の分散

と書替えられ、$N_j=\bar{N}$ なる時は

$$\sigma_w^2=\frac{\bar{N}}{N}\sum_{j=1}^{M}\sigma_j^2=\frac{1}{M}\sum_{j=1}^{M}\sigma_j^2 \quad (3\cdot14)$$

となる。又集落間分散σ_b^2は

$$\sigma_b^2=\frac{1}{N}\sum_{j=1}^{M}(\bar{x}_j-\bar{x})^2N_j$$

$$=\frac{\bar{N}}{N}\sum_{j=1}^{M}(\bar{x}_j-\bar{x})^2=\frac{1}{M}\sum_{j=1}^{M}(\bar{x}_j-\bar{x})^2 \quad (3\cdot15)$$

となる。

π'の分散σ_e^2は

$$\sigma_e^2=\frac{1}{M}\sum_{j=1}^{M}(X_j-\bar{X})^2=\frac{1}{M}\sum_{j=1}^{M}(N_j\bar{x}_j-\bar{N}\bar{x})^2 \quad (3\cdot16)$$

［$\because$ (3・5) より $\bar{X}=\dfrac{N}{M}\bar{x}=\bar{N}\bar{x}$

j番集落の平均値：$\bar{x}_j=\dfrac{X_j}{N_j}$ より $X_j=N_j\bar{x}_j$］

であるから、$N_j=\bar{N}$ の時は

$$\sigma_e^2=\frac{\bar{N}^2}{M}\sum_{j=1}^{M}(\bar{x}_j-\bar{x})^2=\bar{N}^2\sigma_b^2 \quad [(3\cdot15)] \quad (3\cdot17)$$

となる。

次にα標識の割合の推定の場合は、α標識に1、其他の標識に0を与える事によって平均値は比例数を表すため、(3・13)—(3・15)式は次のようになる。

$$\sigma_w^2=\frac{1}{N}\sum_{j=1}^{M}N_jp_jq_j \quad (3\cdot18)$$

但 $\sigma_j^2=p_jq_j$,[1] （$p_j+q_j=1$, p_jは j 番集落に於けるα標識の単位の割合）

$N_j=\bar{N}$ ならば

$$\sigma_w^2=\frac{1}{M}\sum_{j=1}^{M}p_jq_j \quad (3\cdot19)$$

$$\sigma_b^2=\frac{1}{M}\sum_{j=1}^{M}(p_j-p)^2 \quad (3\cdot20)$$

1) 第二講 p. 68 参照。

さて母平均$\bar{x}$の推定の場合は(3・6)(3・8)式に於て$N_j=\bar{N}=\dfrac{N}{M}$及び(3・17)を代入すると

$$\bar{x}'=\frac{1}{m\bar{N}}\sum_{j=1}^{m}X_j \quad (3\cdot21)$$

$$V(\bar{x}')=\left(\frac{1}{\bar{N}}\right)^2\frac{M-m}{M-1}\frac{\bar{N}^2\sigma_b^{\,2}}{m}$$

$$=\frac{M-m}{M-1}\frac{\sigma_b^{\,2}}{m}, \quad 但 \quad \sigma_b^{\,2}=\frac{1}{M}\sum_{j=1}^{M}(\bar{x}_j-\bar{x})^2 \qquad (3\cdot22)$$

となる。

又母比例数pの推定の場合は、推定式、分数は(3・21)(3・22)と同じであり、只(3・22)に於て$\sigma_b^{\,2}$が(3・20)式となるのみである。

(2) 母分散の推定

標本推定値の誤差の規定に必要な母分散——$\sigma_e^{\,2}$又は$\sigma_b^{\,2}$——の値を標本より推定する場合、先づ$\sigma_e^{\,2}$は集落を単位とした時の単位間の分散であるため、それの不偏推定値$\hat{\sigma}_e^{\,2}$は(2・47)式より直ちに導く事が出来る。今m個の単位（集落）より成る標本の平均値及び分散を

$$\bar{X}'=\frac{1}{m}\sum_{j=1}^{m}X_j, \quad s_e^{\,2}=\frac{1}{m}\sum_{j=1}^{m}(X_j-\bar{X}')^2$$

とすると、(2・47)より

$$\hat{\sigma}_e^{\,2}=\frac{M-1}{M}\frac{m}{m-1}s_e^{\,2} \qquad (3\cdot23)$$

となる。

次に$N_j=\bar{N}$の場合は標本の平均値及び分散は次のようになる。

$$\bar{X}'=\frac{\bar{N}}{m\bar{N}}\sum_{j=1}^{m}X_j=\bar{N}\bar{x}' \quad [(3\cdot21)]$$

$$s_e=\frac{1}{m}\sum_{j=1}^{m}(\bar{N}\bar{x}_j-\bar{N}\bar{x}')^2=\frac{\bar{N}^2}{m}\sum_{j=1}^{m}(\bar{x}_j-\bar{x}')^2$$

$$=\bar{N}^2s_b^{\,2}, \quad s_b^{\,2}=\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j-\bar{x}')^2$$

従って(3・23)より

$$E\left(\frac{M-1}{M}\frac{m}{m-1}\bar{N}^2s_b^{\,2}\right)=\sigma_e^{\,2}=\bar{N}^2\sigma_b^{\,2} \quad [(3\cdot17)]$$

$$\therefore \quad E\left(\frac{M-1}{M}\frac{m}{m-1}s_b^{\,2}\right)=\sigma_b^{\,2}$$

即ち$\sigma_b^{\,2}$の不偏推定値$\hat{\sigma}_b^{\,2}$は

$$\hat{\sigma}_b^{\,2}=\frac{M-1}{M}\frac{m}{m-1}s_b^{\,2}, \quad 但 \quad s_b^{\,2}=\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j-\bar{x}')^2 \qquad (3\cdot24)$$

である。

(3) 集落抽出法の精度

(3・8)及び(3・12)式より明らかな如く、集落抽出法の精度は$\sigma_e^{\,2}$即ちX_jの分散によって支配されるのであって、これが小さい程精度が高くなるのである。しかるにX_jは集落内の単位の変量の総和又はα標識の単位の数であ

るため、それの分散σ_e^2は集落内の単位数N_jが均一乃至は略同一である場合は、N_jが集落によって大きく異る場合に比して小くなるのが普通であり、故に一般に集落の大いさの変動が少い程標本推定の精度は向上するのである。

そして集落の大いさが均一なる場合は、(3・22)式より明らかな如く、推定の精度はσ_b^2即ち集落間分散の値によって支配されるのであり、それは集落相互間が同質的（故に集落の内部は異質的）である程小となるため、[1] 母集団単位を集落に分ける場合可能な限り異質的な単位を集めて集落を作る事によって、標本推定の精度を高める事が出来るのである。

次に集落抽出法の精度と単純抽出法の精度とを、母平均の推定の場合に就て比較して見よう。但し比較の便宜上集落抽出法は集落の大いさが均一な場合とする。今M個の集落より成る母集団よりm個の集落を抽出して標本を構成した場合の分散をV_2とすると、(3・22)式より

$$V_2=\frac{M-m}{M-1}\frac{\sigma_b^2}{m}\doteqdot\frac{M-m}{M}\frac{\sigma_b^2}{m}\quad(\because\ M-1\doteqdot M)$$

である。この集落抽出法による標本には$m\bar{N}$個の調査単位が含まれているから、これと同個数の調査単位の単純抽出による標本の分散をV_1とすると、(2・33)式より

$$V_1=\frac{N-m\bar{N}}{N-1}\frac{\sigma^2}{m\bar{N}}\doteqdot\frac{N-m\bar{N}}{N}\frac{\sigma^2}{m\bar{N}}\quad(\because\ N-1\doteqdot N)$$

$$=\frac{M-m}{M}\frac{\sigma^2}{m\bar{N}}\quad(\because\ N=M\bar{N})$$

である。そこでV_1のV_2に対する比をとると

$$\frac{V_1}{V_2}\doteqdot\frac{\frac{M-m}{M}\frac{\sigma^2}{m\bar{N}}}{\frac{M-m}{M}\frac{\sigma_b^2}{m}}=\frac{\sigma^2}{\bar{N}\sigma_b^2}=\frac{1}{\bar{N}}\frac{\sigma_w^2+\sigma_b^2}{\sigma_b^2}\quad(\because\ \sigma^2=\sigma_w^2+\sigma_b^2)$$

$$=\frac{1}{\bar{N}}\left(1+\frac{\sigma_w^2}{\sigma_b^2}\right)\qquad(3・25)$$

であり、従って

$$\frac{1}{\bar{N}}\left(1+\frac{\sigma_w^2}{\sigma_b^2}\right)\gtreqless 1\quad なる時に\quad V_1\gtreqless V_2$$

となる。この関係は条件式を次のように書替るとよ

明瞭となる。

$$\frac{\sigma_w^2}{\sigma_b^2}\gtreqless\bar{N}-1\doteqdot\bar{N}\qquad\therefore\ \frac{\sigma_b^2}{\sigma_w^2}\lesseqgtr\frac{1}{\bar{N}}$$

なるため

$$\left.\begin{array}{l}\frac{\sigma_b^2}{\sigma_w^2} < \frac{1}{N} \\ 〃 = 〃 \\ 〃 > 〃\end{array}\right\} \text{ならば} \left\{\begin{array}{ll}V_1 > V_2 & \text{(i)} \\ 〃 = 〃 & \text{(ii)} \\ 〃 < 〃 & \text{(iii)}\end{array}\right.$$

となるのである。これより明らかな如く σ_b^2 が極めて小であり σ_w^2 が大である時は、即ち集落の内部が相当異質的な単位より成って居り従って集落相互間は同質的である場合は、[2] (i)が成立し従って集落抽出法の精度の方が単純抽出法の精度よりも高くなるのであるが、σ_b^2 が大となり σ_w^2 が小となるに従って、即ち集落内が同質的となり故に集落相互間が異質的となる時は、(iii)が成立し従って集落抽出法の精度は単純抽出法の精度よりも低くなるのである。かくて集落内が異質的であり集落相互間が同質的となる程 $\frac{\sigma_b^2}{\sigma_w^2}$ の値が小さくなり、集落抽出法の精度が高くなって行くのである。しかしながら実際上は集落内異質的、集落間同質的なるように集落を構成する事は多くの場合困難であり、殊に地域的区割を集落となす地域抽出法の場合は、社会的、自然的条件が同一地域内に類似の単位を集合せしめる傾向を有するため、（例えば住宅地域、商店街、工場地帯等）むしろ集落内同質的、集落間異質的となると考えられるため、一般には $\frac{\sigma_b^2}{\sigma_w^2}$ は $\frac{1}{\bar{N}}$ よりも大であり、従って集落抽出法は単純抽出法よりも精度は落ちるのである。

又(3・25)式より明らかな如く集落抽出法の精度は集落の大いさ $\bar{N}$ によっても影響を受けるのである。普通は $\frac{V_1}{V_2} < 1$ なるため、$\bar{N}$ が小なる程この比は1に近づき、従って集落抽出法の精度は単純抽出法の精度に近くなるのである。

かくて集落抽出法は同じ規模の（即ち調査単位の数を同じくする）単純抽出法よりもその精度は一般に低いのであり、集落抽出法の精度を高めるためには、集落の大いさを小さくし且出来る丈均一ならしめ、又集落内の単位を可能な限り異質的にする事が必要である。

1) 2) 第一講 P. 126 参照。

四　多段抽出法

(1) 多段抽出法に於ける期望値計算

二段抽出法はM個の集落より成る母集団よりm個の集落を抽出し、抽出されたm個の集落の各々の内で更にn_j個宛（$j=1,2,\ldots\ldots,m$）の調査単位（又は集落）を抽出して標本を構成する方法であり、三段抽出法は第二次に抽出された集落の各々より更にd_{jk}個宛（$j=1,2,\ldots\ldots,m$　$k=1,2,\ldots\ldots,n_j$）の調査単位（又は集落）を抽出して標本を構成する方法である。このような多段抽出法に於ける不偏推定及びその分散を導くためには、先づこの抽出法に於ける確率変数の期望値、分散及びそれに就て成立する諸関係を知らねばならぬ。

先づ二段抽出の場合に就て説明しよう。二段無作為抽出操作の結果一定変域内の値をとる確率変数ξの確率分布の期望値、分散をそれぞれ$\underset{jk}{E}(\xi)$, $\underset{jk}{V}(\xi)$と書く事にする。茲に期望値、分散の記号の下に添数j、kを附けるのは、それが母集団よりの一次抽出単位の抽出を可能なあらゆる場合丈変え、更に又抽出された一次抽出単位内での二次抽出単位の抽出を、可能なあらゆる場合丈変える時のξの期望値、分散なる事を表すためであって、添数jは抽出する一次抽出単位を動かす事を示し、添数kは抽出された一次抽出単位内で二次抽出単位の抽出を変える事を示すのである。

この二段抽出法の期望値並に分散の計算に於て次の諸関係が成立する事は、一段抽出法の場合と同様にして証明する事が出来る。（但し証明は略する）

$$\underset{jk}{E}(\xi+\eta)=\underset{jk}{E}(\xi)+\underset{jk}{E}(\eta) \tag{4·1}$$

一般に

$$\underset{jk}{E}(\xi_1+\xi_2+\cdots\cdots+\xi_n)=\underset{jk}{E}(\xi_1)+\underset{jk}{E}(\xi_2)+\cdots\cdots+\underset{jk}{E}(\xi_n) \tag{4·2}$$

$$\underset{jk}{E}(a\xi)=a\underset{jk}{E}(\xi),\qquad a=\text{const.} \tag{4·3}$$

ξ、ηが相互に独立なる確率変数の場合は

$$\underset{jk}{E}(\xi\eta)=\underset{jk}{E}(\xi)\cdot\underset{jk}{E}(\eta) \tag{4·4}$$

一般に$\xi_1\xi_2\cdots\cdots\xi_n$が相互に独立である時は

$$E_{jk}(\xi_1\xi_2\cdots\cdots\xi_n)=E_{jk}(\xi_1)E_{jk}(\xi_2)\cdots\cdots E_{jk}(\xi_n) \quad (4\cdot5)$$

次にξの分散 $V_{jk}(\xi)$ は、分散の定義式(2・22)により

$$V_{jk}(\xi)=E_{jk}\{\xi-E_{jk}(\xi)\}^2 \quad (4\cdot6)$$

$$=E_{jk}(\xi^2)-(E_{jk}\xi)^2 \quad (4\cdot7)$$

$$V_{jk}(a\xi)=a^2V_{jk}(\xi),\quad a=\mathrm{const.} \quad (4\cdot8)$$

ξ、ηが独立なる確率変数である時は

$$V_{jk}(\xi+\eta)=V_{jk}(\xi)+V_{jk}(\eta) \quad (4\cdot9)$$

一般に$\xi_1\xi_2\cdots\cdots\xi_n$が相互に独立なる場合は

$$V_{jk}(\xi_1+\xi_2+\cdots\cdots+\xi_n)=V_{jk}(\xi_1)+V_{jk}(\xi_2)+\cdots+V_{jk}(\xi_n) \quad (4\cdot10)$$

そして二段抽出法の期望値、分散は、次の式により一段抽出法の期望値、分散の計算によって求める事が出来るのである。先づ期望値の計算式から説明しよう。今抽出された一次抽出単位（j番集落とする）を固定し、その中での二次抽出単位の抽出を可能なあらゆる場合に就て動かす時のξの期望値を$E_k^{(j)}(\xi)$で表すと、それはj番集落に於ては一定の値である。そこで今度は一次抽出単位の抽出をあらゆる場合に就て動かす事にすると$E_k^{(j)}(\xi)$も又確率変数となるため、その期望値は$E_j(E_k^{(j)}\xi)$と書かれる。すると

$$E_{jk}(\xi)=E_j(E_k^{(j)}\xi) \quad (4\cdot11)$$

なる関係が成立するのである。

この事は次のようにして証明される。j、kをあらゆる場合に就て動かす時に ξ のとり得る値を x_{jk} $\left(\begin{matrix} j=1,\ 2,\ \cdots\cdots \\ k=1,\ 2,\ \cdots\cdots \end{matrix}\right)$とし、ξが特定の$x_{jk}$をとる確率を$p_{jk}$とすれば、$E_{jk}(\xi)$は期望値の定義式(2・13)より

$$E_{jk}(\xi)=\sum_j\sum_k p_{jk}x_{jk}$$

である。そして第一次抽出に於て特定の（j番）一次抽出単位が抽出される確率をp_jとし、又その一次抽出単位の中から第二次抽出の結果、ξが特定のx_{jk}の値をとるが如き二次抽出単位（k番単位）の抽出さる確率を$p_k^{(j)}$とすると、確率の乗法定理により

$$p_{jk}=p_jp_k^{(j)}$$

である。故に

$$\underset{jk}{E}(\xi)=\sum_j\sum_k p_j p_k{}^{(j)}x_{jk}=\sum_j p_j\underbrace{\sum_k p_k{}^{(j)}x_{jk}}_{=\underset{k}{E}{}^{(j)}(\xi)}$$

$$=\underbrace{\sum_j p_j(\underset{k}{E}{}^{(j)}\xi)}_{=\underset{j}{E}(\cdots\cdots)}=\underset{j}{E}(\underset{k}{E}{}^{(j)}\xi)\qquad[(2\cdot13)]$$

〔証明終り〕

かくて二段抽出法の場合の確率変数ξの期望値は、先づj番一次抽出単位よりの第二次抽出に於ける期望値を計算し、次で母集団よりの第一次抽出に於ける期望値を計算する事によって容易に求める事が出来るのである。

次に分散の計算式を求めると、(4·7)より

$$\underset{jk}{V}(\xi)=\underset{jk}{E}(\xi^2)-(\underset{jk}{E}\xi)^2$$
$$=\underset{j}{E}\underset{k}{E}{}^{(j)}\xi^2-(\underset{j}{E}\underset{k}{E}{}^{(j)}\xi)^2\qquad[(4\cdot11)]$$

この式の右辺に $\underset{j}{E}(\underset{k}{E}{}^{(j)}\xi)^2$ を加減すると

$$=\underset{j}{E}\underset{k}{E}{}^{(j)}\xi^2-\underset{j}{E}(\underset{k}{E}{}^{(j)}\xi)^2+\underset{j}{E}(\underset{k}{E}{}^{(j)}\xi)^2-(\underset{j}{E}\underset{k}{E}{}^{(j)}\xi)^2$$
$$=\underset{j}{E}\underbrace{\{\underset{k}{E}{}^{(j)}\xi^2-(\underset{k}{E}{}^{(j)}\xi)^2\}}_{=\underset{k}{V}{}^{(j)}(\xi)}+\underbrace{\underset{j}{E}(\underset{k}{E}{}^{(j)}\xi)^2-\{\underset{j}{E}(\underset{k}{E}{}^{(j)}\xi)\}^2}_{=\underset{j}{V}(\underset{k}{E}{}^{(j)}\xi)}\qquad[(2\cdot23)]$$

故に

$$\underset{jk}{V}(\xi)=\underset{j}{V}(\underset{k}{E}{}^{(j)}\xi)+\underset{j}{E}\underset{k}{V}{}^{(j)}(\xi)\qquad(4\cdot12)$$

この(4·12)式によって二段抽出法の場合の分散の計算は、一段抽出法の分散の計算に帰着させる事が出来、又二段抽出法の確率変数の分散を、第一次抽出の分散と第二次抽出の分散とに分解し得るのである。

次に三段抽出法の場合の確率変数ξの期望値及び分散を $\underset{jkl}{E}(\xi)$, $\underset{jkl}{V}(\xi)$ と書く事にする。これ等の記号の添数j、k、lは、j、kに就ては二段抽出法の場合と同じであり、lは抽出された二次抽出単位内で三次抽出単位の抽出を可能な限り変える事を表するである。

三段抽出法は、二段抽出法の第二次抽出に於て、それを第一次の抽出として更に第二次の抽出を行うものと見る事によって、その期望値及び分散の計算式は(4·11)(4·12)式を用いて容易に導く事が出来る。先づ期望値の計算式は、最初の二段抽出のみに着目して

$$\underset{jkl}{E}(\xi)=\underset{j}{E}\underset{kl}{E}{}^{(j)}(\xi)\qquad[(4\cdot11)]$$

しかるに第二次の抽出は後の二段抽出の第一次抽出に

相当するのであるから

$$E^{(j)}_{kl}(\xi) = E^{(j)}_{k} E^{(j,k)}_{l}(\xi) \qquad [(4\cdot11)]$$

$$\therefore \quad E_{jkl}(\xi) = E_j E^{(j)}_k E^{(j,k)}_l(\xi) \qquad (4\cdot13)$$

かくて三段抽出法の場合の確率変数ξの期望値は、先づj番一次抽出単位内のk番二次抽出単位よりの第三次抽出に於ける期望値を計算し、次でj番一次抽出単位よりの第二次抽出の期望値を求め、最後に母集団からの第一次抽出に於ける期望値を計算する事によって得られるのである。

次に分散 $V_{jkl}(\xi)$ の計算式は、始めの二段抽出に関して

$$V_{jkl}(\xi) = V_j(E^{(j)}_{kl}\xi) + E_j V^{(j)}_{kl}(\xi) \qquad [(4\cdot12)]$$

そして第二次抽出は又次の二段抽出の第一次抽出であるから

$$E^{(j)}_{kl}(\xi) = E^{(j)}_k E^{(j,k)}_l(\xi) \qquad [(4\cdot11)]$$

$$V^{(j)}_{kl}(\xi) = V^{(j)}_k(E^{(j,k)}_l \xi) + E^{(j)}_k V^{(j,k)}_l(\xi) \qquad [(4\cdot12)]$$

$$\therefore \quad V_{jkl}(\xi) = V_j(E^{(j)}_k E^{(j,k)}_l \xi) + E_j V^{(j)}_k(E^{(j,k)}_l \xi) + E_j E^{(j)}_k V^{(j,k)}_l(\xi) \qquad (4\cdot14)$$

(4•14)式により三段抽出法の場合の分散の計算は、一段抽出法の分散の計算によって求められるのであり、又これより明らかな如く三段抽出法の確率変数の分散は、第三次抽出の分散の平均と第二次抽出の分散の平均及び第一次抽出の分散の総和である。

又、四段抽出法に於ける期望値、分散の計算式は、(4•11)(4•12)式を用いて三段抽出法の計算式(4•13)(4•14)より容易に求める事が出来る。かくて一般にn段抽出法に於ける計算式は(4•11)(4•12)式を用いて（$n-1$）段抽出法の計算式より誘導する事が出来るのである。

そして(4•1)—(4•10)式の期望値計算に於て成立する関係は多段抽出法一般に妥当するのである。

(2) 二段抽出法に於ける推定式

二段抽出法はM個の集落より成る母集団からm個の集落を抽出し、抽出されたm個の集落の各〻に於て更にn_j個（$j=1, 2, \ldots\ldots, m$）宛の調査単位(又は集落)を抽出して標本を構成するのである。従って集落抽出法の場合は第一次抽出によって抽出された集落の量的

大いさを正確に知り得るのに対して、二段抽出法の場合はそれを推定しなければならないのである。故に集落抽出法の場合の推定式に於て、集落の量的大いさX_jにそれの推定値X_j'を代入する事によって、二段抽出法の場合の推定式が得られるであろう。

さてj番集落に於てN_j個の単位よりn_j個抽出するのであるから、[1] 集落の量的大いさX_jの不偏推定値X_j'は、(2•34)式より

$$X_j'=\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}',\quad \underset{k}{E^{(j)}}(X_j')=X_j \qquad (4•15)$$

である。従ってこれを(3•6)式に代入する事によって母平均$\bar{x}$の推定式が得られる。

$$\bar{x}'=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}\left(\frac{N_j}{n_j}\sum_{j=1}^{n_j}x'_{jk}\right)^{2)} \qquad (4•16)$$

これは不偏推定値である。即ち

$$\underset{jk}{E}(\bar{x}')=\underset{j\ k}{EE^{(j)}}(\bar{x}')$$

$$\underset{k}{E^{(j)}}(\bar{x}')=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}\underbrace{\underset{k}{E^{(j)}}(X_j')}_{=X_j\ [(4•15)]} \quad [(2•20)(2•17)]$$

$$=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}X_j=\bar{x}' \qquad [(3•6)(4•17)]$$

$$\therefore\ \underset{jk}{E}(\bar{x}')=\underset{j}{E}(\bar{x}')=\bar{x} \qquad [(3•7)]\quad (4•18)$$

そしてこの分散は次の如くである。

$$\underset{jk}{V}(\bar{x}')=\underset{j}{V}(\underset{k}{E^{(j)}}\bar{x}')+\underset{j\ k}{EV^{(j)}}(\bar{x}') \qquad [(4•12)]$$

$$\underset{j}{V}(\underset{k}{E^{(j)}}\bar{x}')=\underset{j}{V}(\bar{x}') \qquad [(4•17)]$$

$$=\left(\frac{M}{N}\right)^2\frac{M-m}{M-1}\frac{\sigma_e^2}{m} \qquad [(3•8)(4•19)]$$

$$\underset{k}{V^{(j)}}(\bar{x}')=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{j=1}^{m}\underbrace{\underset{k}{V^{(j)}}(X_j')}_{=N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}\equiv A_j\ [(2•35)]} \qquad [(2•25)(2•27)^{3)}]$$

$$\therefore\ \underset{j\ k}{EV^{(j)}}(\bar{x}')=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{l=1}^{m}\underset{j}{E}(A_j)$$

しかるにA_jは集落毎に一定の値であるため、$\underset{j}{E}(A_j)$は母集団π'より一個の集落を抽出して得られるA_jの値の期望値である。故に

$$\underset{j}{E}(A_j)=\frac{1}{M}\sum_{j=1}^{M}A_j \qquad [(2•29)] \qquad (4•20)$$

である。故に

$$\underset{j\ k}{EV^{(j)}}(\bar{x}')=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{j=1}^{m}\frac{1}{M}\sum_{j=1}^{M}A_j$$

$$=\left(\frac{1}{N}\right)^2\frac{M}{m}\sum_{j=1}^{M}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \qquad (4•21)$$

かくて

$$\underset{jk}{V}(\bar{x}') = \frac{1}{N^2}\left\{ M^2\frac{M-m}{M-1}\frac{\sigma_e^2}{m} + \frac{M}{m}\sum_{j=1}^{m} N_j^2\,\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \right\} \qquad (4\cdot22)$$

(4·22)式の第一項はπ'のM個の集落よりm個の集落を抽出して推定する事による誤差分散であり、第二項は抽出された集落のN_j個の調査単位よりn_j個の調査単位を抽出して推定する事による誤差分散である。

1) 第二次抽出単位の抽出数n_jは、集落の大いさN_jに比例する。$\frac{n_j}{N_j}=g$，$(j=1,2,\cdots\cdots,m)$のが普通である。その時は(4·16)(4·22)式に於て$\frac{n_j}{N_j}=g=\text{const.}$を代入すればよい。

2) (4·16)式は母集団の単位総数Nが判って居らねばならず、それが不明の時はNをも推定しなければならない。抽出されたj番集落の単位数は、調査単位を第二次抽出の単位となす限り第二次抽出の枠の作成に於て判っているため、Nの推定値は(3·9)式によって得られる。従って(4·16)式は

$$\bar{x}'' = \frac{\sum_{j=1}^{m}\frac{N_j}{n_j}\sum_{k=1}^{n_j} x_{jk}'}{\sum_{j=1}^{m} N_j} \qquad (4\cdot23)$$

の比推定となる。

3) 二段抽出法に於て第二次抽出は各集落毎に別々に行われるため、集落の量的大いさの推定値X_j'は相互に独立なる確率変数である。

次に母比例数pの推定の場合は、割合を知らんとする標識に1、その他の標識に0なる値を与える事によって$\bar{x}$はpを表すため、その推定式及び分散は(4·16)(4·22)と同じである。但しこの場合は(4·22)に於て

$$\sigma_j^2 = p_j q_j$$

である。

集落の大いさが均一なる場合　一次抽出単位たる集落の大いさが均一　$N_j=\bar{N}$，($j=1,2,\cdots\cdots,M$) なる場合は、(従って二次抽出単位の抽出数も同一　$n_j=\bar{n}$，($j=1,2,\cdots\cdots,m$) とする) 推定式及び分散は簡単となるのである。

今(4·16)(4·22)に於て　$N_j=\bar{N}$，$n_j=\bar{n}$　と置けば、$N=M\bar{N}$なるため母平均の推定式は

$$\bar{x}' = \frac{1}{N}\frac{M\bar{N}}{m\bar{n}}\sum_{j=1}^{m}\sum_{k=1}^{\bar{n}} x_{jk}' = \frac{1}{m\bar{n}}\sum_{j=1}^{m}\sum_{k=1}^{\bar{n}} x_{jk}' \qquad (4\cdot24)$$

となり、分散は

$$\underset{jk}{V}(\bar{x}')=\frac{1}{\bar{N}^2}\left\{M^2\frac{M-m}{M-1}\frac{\sigma_e{}^2}{m}+\frac{M\bar{N}^2}{m\bar{n}}\frac{\bar{N}-\bar{n}}{\bar{N}-1}\sum_{j=1}^{M}\sigma_j{}^2\right\}$$

$$=\frac{1}{\bar{N}^2}\left\{M^2\bar{N}^2\frac{M-m}{M-1}\frac{\sigma_b{}^2}{m}+M^2\bar{N}^2\frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{\sigma_w{}^2}{m\bar{n}}\right\}$$

［(3・17)(3・14)より］

$$=\frac{M-m}{M-1}\frac{\sigma_b{}^2}{m}+\frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{\sigma_w{}^2}{m\bar{n}} \qquad (4\cdot25)$$

若しM、$\bar{N}$がm、$\bar{n}$に比して極めて大であり、有限母集団修正を無視し得る時は

$$\underset{jk}{V}(\bar{x}')\doteqdot\frac{\sigma_b{}^2}{m}+\frac{\sigma_w{}^2}{m\bar{n}} \qquad (4\cdot26)$$

となる。そして母平均推定値$\bar{x}'$の変異係数を$C_{\bar{x}}'$で表し、母集団に於ける集落内変異係数を $c_w=\frac{\sigma_w}{\bar{x}}$ 集落間変異係数を $c_b=\frac{\sigma_b}{\bar{x}}$ と置けば

$$C^2_{\bar{x}}{}'=\frac{M-m}{M-1}\frac{c_b{}^2}{m}+\frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{c_w{}^2}{m\bar{n}} \qquad 4\cdot27)$$

$$\doteqdot\frac{c_b{}^2}{m}+\frac{c_w{}^2}{m\bar{n}} \qquad (4\cdot28)$$

となる。

比例数の推定の場合は推定式、分散は(4・24)―(4・28)と同じであるが、唯$\sigma_w{}^2$、$\sigma_b{}^2$が(3・19)(3・20)式となるのみである。

(3)　適　正　配　分

単純任意抽出法及び集落一段抽出法に於ては、標本

の精度及び調査の費用は抽出すべき調査単位又は集落の個数によって決るのであり、標本推定の精度を高めるためには調査単位又は集落の抽出数を増加しなければならず、しかしそれは調査費を増加せしめる事となるため、所要の精度の推定値を獲得するに要する最少抽出数の決定が、標本調査の設計に当って問題となるのであった。しかるに二段抽出法に於ては(4・22)式より明らかな如く、推定値の誤差は第一次抽出に於ける誤差と第二次抽出に於ける誤差とより成り、その各々は一次抽出単位の数及び二次抽出単位の数の増減によって減少又は増加せしめる事が出来るのであるため、同じ大いさの標本（即ち調査単位の数を同じくする標本）といえども、一次抽出単位の数mを多くし二次抽出単位の数n_jを少くするのと、逆にmを少くしてn_jを増すのとでは、その推定値の誤差は異るのである。そして第二次抽出の枠の作成費、調査単位の個別調査の費用等は後者の方が少くて済むであろうから、調査の費用に於ても同一ではないのである。

かくて二段抽出法に於ては標本の大いさのみならず、各抽出段階に於ける抽出個数の配分の如何によって、標本の精度及び調査の費用が変化するのであり、茲に於て最高の精度の標本推定を最小の費用で達成し得るような、又は与えられた費用で最高の精度の標本推定を獲得し得るような、各抽出段階に於ける抽出個数の配分即ち適正配分 optimum allocation の問題が生ずるのである。

このような各抽出段階に対する抽出個数の配分の如何によって、標本推定の精度及び調査費用が異る事は多段抽出法一般に就ていい得るため、多段抽出法に於ては適正配分の問題が存するのである。

次に二段抽出法の場合の適正配分を $N_j=\bar{N}, n_j=\bar{n}$ の場合に就て、且便宜上有限母集団修正が無視し得るものとして考察して見よう。それには先づ抽出単位数の変化による調査費用の変化を与える費用函数を決めねばならない。今抽出された集落一個当りの費用（例えば第二次抽出の枠の作成費、調査員の教育、監督費等）をk_1、調査単位一個当りの個別調査費をk_2とすると、二段抽出法による調査費用Tは

$$T=k_1m+k_2m\bar{n} \qquad (4\cdot29)$$

によって与えられる。

先づ調査費が与えられた場合の適正配分を考察しよう。この場合は(4・29)式のTが一定であるという条件の下に、(4・28)式の$C_{\bar{x}}^{2\prime}$を最小ならしめるm、$\bar{n}$の値を求めればよいのである。それはラグランジュ乗数λを用いて

$$f(m,\bar{n})=\frac{c_b^2}{m}+\frac{c_w^2}{m\bar{n}}+\lambda(k_1m+k_2m\bar{n}-T)$$

をm、$\bar{n}$でそれぞれ偏微分して0と置き、それ等を(4・29)式と連立せしめて解く事によって得られる。即ち、

$$\frac{\partial f}{\partial m}=-\frac{c_b^2}{m^2}-\frac{c_w^2}{m^2\bar{n}}+\lambda k_1+\lambda k_2\bar{n}=0 \qquad (\mathrm{i})$$

$$\frac{\partial f}{\partial \bar{n}}=-\frac{c_w^2}{m\bar{n}^2}+\lambda k_2m=0 \qquad (\mathrm{ii})$$

(ii)より　$m^2=\frac{c_w^2}{\lambda k_2\bar{n}^2}$

(i)に代入して　$-\frac{\lambda k_2\bar{n}^2}{c_w^2}\left(c_b^2+\frac{c_w^2}{\bar{n}}\right)+\lambda k_1+\lambda k_2\bar{n}=0$

$$\left(\frac{c_b}{c_w}\right)^2\lambda k_2\bar{n}^2+\lambda k_2\bar{n}=\lambda k_1+\lambda k_2\bar{n}$$

$$\bar{n}^2=\frac{k_1}{k_2}\left(\frac{c_w}{c_b}\right)^2 \quad \therefore \quad \bar{n}=\sqrt{\frac{k_1}{k_2}}\frac{c_w}{c_b} \qquad (4\cdot30)$$

(4・29)をmに就て解いて(4・30)を代入すると

$$m=\frac{T}{k_1+k_2\bar{n}}=\frac{T}{k_1+\sqrt{k_1k_2\frac{c_w}{c_b}}} \qquad (4\cdot31)$$

となる。

次に推定の精度が与えられた場合の適正配分は、(4・28)式の$C^2_{\bar{x}'}$が一定であるという条件の下に(4・29)式のTを最小ならしめるm、$\bar{n}$の値を求めればよいのであり、それは矢張りラグランジュ乗数λを用いて

$$f(m,\bar{n})=k_1m+k_2m\bar{n}+\lambda\left(\frac{c_b^2}{m}+\frac{c_w^2}{m\bar{n}}-C^2_{\bar{x}'}\right)$$

をm、$\bar{n}$で偏微分して0と置き(4・28)式と連立させて解く事によって得られる。即ち

$$\frac{\partial f}{\partial m}=k_1+k_2\bar{n}-\lambda\frac{c_b^2}{m^2}-\lambda\frac{c_w^2}{m^2\bar{n}}=0 \qquad (\text{i})$$

$$\frac{\partial f}{\partial \bar{n}}=k_2m-\lambda\frac{c_w^2}{m\bar{n}^2}=0 \qquad (\text{ii})$$

(ii)より　$m^2=\dfrac{\lambda}{k_2}\dfrac{c_w^2}{\bar{n}^2}$

(i)に代入して　$k_1+k_2\bar{n}-\lambda\dfrac{k_2}{\lambda}\dfrac{\bar{n}^2}{c_w^2}\left(c_b^2+\dfrac{c_w^2}{\bar{n}}\right)=0$

$$k_1+k_2\bar{n}=k_2\bar{n}^2\left(\frac{c_b}{c_w}\right)^2+k_2\bar{n}$$

$$\bar{n}^2=\frac{k_1}{k_2}\left(\frac{c_w}{c_b}\right)^2 \qquad \therefore\quad \bar{n}=\sqrt{\frac{k_1}{k_2}}\frac{c_w}{c_b}$$

これは(4・30)式と同じである。(4・28)をmに就て解いてこれを代入すると

$$m=\frac{c_b^2+\frac{c_w^2}{\bar{n}}}{C^2_{\bar{x}'}}=\frac{1}{C_{\bar{x}'}^2}\left(c_b^2+c_w^2\sqrt{\frac{k_2}{k_1}}\frac{c_b}{c_w}\right)=\frac{c_b^2+\sqrt{\frac{k_2}{k_1}}c_bc_w}{C^2_{\bar{x}'}} \qquad (4\cdot32)$$

となる。

1) (4・30)式によって与えられた$\bar{n}$の値が1より小である時は$\bar{n}=1$としなければならない。その時はこれに対するmは(4・31)又は(4・32)式によって与えられる値より若干小さくなるのである。

以上により明らかな如く適正分配に於ける二次抽出単位数$\bar{n}$は、費用が与えられた場合も精度が要求された場合も同じであって(4・30)式によって与えられ、それの決定には費用T又は精度$C^2_{\bar{x}'}$は関係しないのである。そして費用、精度の規定は一次抽出単位数mの値を決めるのみであって、費用の増大又は目標精度の向上（即ち$C^2_{\bar{x}}$の減少）はmを増加せしめる事となるのである。

二段抽出法に於ける最適抽出数m、$\bar{n}$の決定に見られるこの関係は、有限母集団修正を無視し得ない場合でも同様

にいゝ得るところであり、且このような費用、精度に関する要求は一次抽出単位数の決定のみに影響し、第二次以降の抽出段階の抽出数の規定には無関係である事は、多段抽出法一般の適正配分に見られる特徴である。

かくて二段抽出法の設計に於て最適抽出数の決定には、母分散σ_w^2、σ_b^2（又は母変異係数c_w、c_b）の値のみならず、費用函数の規定に必要な各抽出段階に於ける費用に就ての知識をも必要とするのである。

(4) 母分散の推定

次に達成精度の計算に必要な母分散の値を推定する式を求めよう。(4·22)式より明らかな如く二段抽出法に於ける誤差分散の計算に必要な母分散はσ_e^2、σ_j^2であって、それを標本より推定しなければならないのである。今標本に於ける同じ性質の値をs_e^2、s_j^2で表すと

$$s_e^2=\frac{1}{m}\sum_{j=1}^{m}(X_j'-\bar{X}')^2$$

$$\text{但}\quad X_j'=\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}',\quad \bar{X}'=\frac{1}{m}\sum_{j=1}^{m}X_j'$$

$$s_j^2=\frac{1}{n_j}\sum_{k=1}^{n_j}(x_{jk}'-\bar{x}_j')^2,\ \text{但}\quad \bar{x}_j'=\frac{1}{n_j}\sum_{k=1}^{n_j}x_{jk}'$$

である。

先づσ_j^2の推定より考察しよう。M個のσ_j^2の値を個々的に推定する事は、すべての集落より第二次抽出が行われるのでないため不可能である。しかし(4·22)式に於て必要なのは必ずしも個々の$\sigma^2{}_j$の値ではなく、それ等の加重された総和である

$$\sum_{j=1}^{M}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \tag{4·33}$$

である。そこでこれを推定するために、j番集落に於てそれの分散σ_j^2の不偏推定値$\hat{\sigma}_j^2$は、(2·47)式より

$$\hat{\sigma}_j^2=\frac{N_j-1}{N_j}\frac{n_j}{n_j-1}s_j,\qquad E_k^{(j)}(\hat{\sigma}_j^2)=\sigma_j^2 \tag{4·34}$$

である事に注意して

$$\frac{M}{m}\sum_{j=1}^{m}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j} \tag{4·35}$$

を考え、これの期待値を求めると

$$E_{jk}\left(\frac{M}{m}\sum_{j=1}^{m}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\right)$$

$$=E_j\left\{\frac{M}{m}\sum_{j=1}^{m}N_j^2\frac{N_j-n_j}{N_j-1}\frac{1}{n_j}\underbrace{E_k^{(j)}(\hat{\sigma}_j^2)}_{=\sigma_j^2\ [(4·34)]}\right\}$$

$$=\frac{M}{m}\sum_{j=1}^{m}E_j\underbrace{\left(N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}\right)}_{=A_j}$$

$$=\frac{M}{m}\sum_{j=1}^{m}\frac{1}{M}\sum_{j=1}^{M}A_j \qquad [(4\cdot20)により]$$

$$=\sum_{j=1}^{M}A_j$$

となり、これは(4・33)に外ならない。従って(4・35)は(4・33)の不偏推定値である。即ち

$$\frac{M}{m}\sum_{j=1}^{m}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\approx\sum_{j=1}^{M}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \tag{4・36}$$

$$但\quad \hat{\sigma}_j^2=\frac{N_j-1}{N_j}\frac{n_j}{n_j-1}s_j^2$$

次にσ^2_eの不偏推定値を求めるに当り、先づs_e^2の期望値を計算して見よう。

$$s_e^2=\frac{1}{m}\sum_{j=1}^{m}(X_j'-\bar{X}')^2$$

$$=\frac{1}{m}\sum_{j=1}^{m}\{(X_j'-\bar{X})-(\bar{X}'-\bar{X})\}^2$$

$$=\frac{1}{m}\Big\{\sum_{j=1}^{m}(X_j'-\bar{X})^2-2(\bar{X}'-\bar{X})\underbrace{\sum_{j=1}^{m}(X_j'-\bar{X})}_{=m(\bar{X}'-\bar{X})}+m(\bar{X}'-\bar{X})^2\Big\}$$

$$=\frac{1}{m}\sum_{j=1}^{m}(X_j'-\bar{X})^2-(\bar{X}'-\bar{X})^2$$

なるため

$$\underset{jk}{E}(s_e^2)=\underset{jk}{E}\Big\{\frac{1}{m}\sum_{j=1}^{m}(X_j'-\bar{X})^2-(\bar{X}'-\bar{X})^2\Big\}$$

$$=\underset{jk}{E}(X_j'-\bar{X})^2-\underset{jk}{E}(\bar{X}'-\bar{X})^2$$

しかるに

$$\underset{jk}{E}(X_j')=\underset{j}{E}\underset{k}{E^{(j)}}(X_j')=\underset{j}{E}(X_j) \qquad [(4\cdot15)]$$

$$=\frac{1}{M}\sum_{j=1}^{M}X_j=\bar{X}$$

$$\underset{jk}{E}(\bar{X}')=\underset{jk}{E}\Big(\frac{1}{m}\sum_{j=1}^{m}X_j'\Big)=\frac{1}{m}\sum_{j=1}^{m}\underbrace{\underset{jk}{E}(X_j')}_{=\bar{X}}=\bar{X}$$

であるから、分散の定義式(4・6)より

$$\underset{jk}{E}(X_j'-\bar{X})^2=\underset{jk}{V}(X_j') \tag{i}$$

$$\underset{jk}{E}(\bar{X}'-\bar{X})^2=\underset{kj}{V}(\bar{X}') \tag{ii}$$

と書ける。そして $X_j'=\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}'$ であるため(4・16)式に於て $m=1$ と置き N/M を乗じたものに相等しい。即ち $X_j'=\frac{N}{M}x'$， 但 $m=1$ である。故に(i)は(4・22)式より

$$\underset{jk}{V}(X_j')=\left(\frac{N}{M}\right)^2\underset{jk}{V}(x') \qquad 但\quad m=1$$

$$=\sigma_e^2+\frac{1}{M}\sum_{j=1}^{M}\underbrace{N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}}_{=A_j}$$

又 $\bar{X}'=\frac{1}{m}\sum_{j=1}^{m}\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}'$ なるため(4・16)式に於て $\frac{N}{M}$ 倍したものに等しい。即ち $\bar{X}'=\frac{N}{M}\bar{\bar{x}}'$ である。故に(ii)は(4・22)式より

$$\underset{jk}{V}(\bar{X}')=\left(\frac{N}{M}\right)^2\underset{jk}{V}(\bar{\bar{x}}')$$
$$=\frac{M-m}{M-1}\frac{\sigma_e^2}{m}+\frac{1}{mM}\sum_{j=1}^{M}\underbrace{N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}}_{=A_j}$$

かくて

$$\underset{jk}{E}(s_e^2)=\left(\sigma_e^2+\frac{1}{M}\sum_{j=1}^{M}A_j\right)-\left(\frac{M-m}{M-1}\frac{\sigma_e^2}{m}+\frac{1}{mM}\sum_{j=1}^{M}A_j\right)$$
$$=\frac{m-1}{m}\frac{M}{M-1}\sigma_e^2+\frac{m-1}{m}\frac{1}{M}\sum_{j=1}^{M}A_j$$
$$\therefore\ \underset{jk}{E}\left(\frac{m}{m-1}s_e^2\right)=\frac{M}{M-1}\sigma_e^2+\frac{1}{M}\sum_{j=1}^{M}A_j$$
$$\therefore\ \frac{M-1}{M}\left\{\underset{jk}{E}\left(\frac{m}{m-1}s_e^2\right)-\frac{1}{M}\sum_{j=1}^{M}A_j\right\}=\sigma_e^2$$

かくて左辺第二項の $\sum_{j=1}^{M}A_j$ にそれの不偏推定値(4・36)を代入した

$$\frac{M-1}{M}\left(\frac{m}{m-1}s_e^2-\frac{1}{m}\sum_{j=1}^{m}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\right)\approx\sigma_e^2 \tag{4・37}$$

は σ_e^2 の不偏推定値である。

集落の大いさが均一なる場合　$N_j=\bar{N}$, $n_j=\bar{n}$ なる場合は誤差分散の計算に必要な母分散は σ_w^2、σ_b^2 である。次にそれ等の推定式を求めよう。今標本に於ける集落内分散を s_w^2、集落間分散を s_b^2 で表すと

$$s_w^2=\frac{1}{m\bar{n}}\sum_{j=1}^{m}\sum_{k=1}^{\bar{n}}(x_{jk}'-\bar{x}_j')^2=\frac{1}{m}\sum_{j=1}^{m}s_j^2$$

但 $\bar{x}_j'=\frac{1}{\bar{n}}\sum_{k=1}^{\bar{n}}x_{jk}'$, $s_j^2=\frac{1}{\bar{n}}\sum_{k=1}^{\bar{n}}(x_{jk}'-\bar{x}_j')^2$

$$s_b^2=\frac{1}{m\bar{n}}\sum_{j=1}^{m}(\bar{x}_j'-\bar{\bar{x}}')^2\bar{n}=\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j'-\bar{\bar{x}}')^2$$

但 $\bar{\bar{x}}'=\frac{1}{m\bar{n}}\sum_{j=1}^{m}\sum_{k=1}^{\bar{n}}x_{jk}'=\frac{1}{m}\sum_{j=1}^{m}\bar{x}_j'$

である。

先づ σ_w^2 の不偏推定値を求めるために s_w^2 の期望値を計算すると

$$\underset{jk}{E}(s_w^2)=\underset{j}{E}\left\{\frac{1}{m}\sum_{j=1}^{m}\underbrace{\underset{k}{E(j)}(s_j^2)}_{=\frac{\bar{N}}{\bar{N}-1}\frac{\bar{n}-1}{\bar{n}}\sigma_j^2\quad[(2\cdot46)]}\right\}$$
$$=\frac{1}{m}\sum_{j=1}^{m}\frac{\bar{N}}{\bar{N}-1}\frac{\bar{n}-1}{\bar{n}}\underset{j}{E}(\sigma_j^2)$$
$$=\frac{\bar{N}}{\bar{N}-1}\frac{\bar{n}-1}{\bar{n}}\underbrace{\frac{1}{M}\sum_{j=1}^{M}\sigma_j^2}_{=\sigma_w^2}\qquad[(3\cdot14)]$$

$$=\frac{\bar{N}}{\bar{N}-1}\,\frac{\bar{n}-1}{\bar{n}}\,\sigma_w^2$$

$$\therefore\quad \underset{jk}{E}\left(\frac{\bar{N}-1}{\bar{N}}\,\frac{\bar{n}}{\bar{n}-1}\,s_w^2\right)=\sigma_w^2$$

かくてσ_w^2の不偏推定値は

$$\frac{\bar{N}-1}{\bar{N}}\,\frac{\bar{n}}{\bar{n}-1}\,s_w^2\approx\sigma_w^2 \qquad (4\cdot38)$$

である。

次にσ_b^2の不偏推定値を求めるためにs_b^2の期望値を計算しなければならない。

$$s_b^2=\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j'-\bar{x}')^2$$

$$=\frac{1}{m}\sum_{j=1}^{m}\{(\bar{x}_j'-\bar{x})-(\bar{x}'-\bar{x})\}^2$$

$$=\frac{1}{m}\left\{\sum_{j=1}^{m}(\bar{x}_j'-\bar{x})^2-2(\bar{x}'-\bar{x})\underbrace{\sum_{j=1}^{m}(\bar{x}_j'-\bar{x})}_{=m(\bar{x}'-\bar{x})}+m(\bar{x}'-\bar{x})^2\right\}$$

$$=\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j'-\bar{x})^2-(\bar{x}'-\bar{x})^2$$

であるため

$$\underset{jk}{E}(s_b^2)=\underset{jk}{E}\left\{\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j'-\bar{x})^2-(\bar{x}'-\bar{x})^2\right\}$$

$$=\underset{jk}{E}(\bar{x}_j'-\bar{x})^2-\underset{jk}{E}(\bar{x}'-\bar{x})^2$$

しかるに

$$\underset{jk}{E}(\bar{x}_j')=\underset{j}{E}\left\{\frac{1}{\bar{n}}\sum_{k=1}^{\bar{n}}\underbrace{\underset{k}{E^{(j)}}(x_{jk}')}_{=\bar{x}_j}\right\} \qquad [(2\cdot29)]$$

$$=\underset{j}{E}(\bar{x}_j)=\frac{1}{M}\sum_{j=1}^{M}\bar{x}_j=\frac{1}{N}\sum_{j=1}^{M}\bar{N}\bar{x}_j=\bar{x}$$

$$\underset{jk}{E}(\bar{x}')=\underset{jk}{E}\left(\frac{1}{m}\sum_{j=1}^{m}\bar{x}_j'\right)=\frac{1}{m}\sum_{j=1}^{m}\underbrace{\underset{jk}{E}(\bar{x}_j')}_{=\bar{x}}=\bar{x}$$

であるから、分散の定義式(4·6)より

$$\underset{jk}{E}(\bar{x}_j'-\bar{x})=\underset{jk}{V}(\bar{x}_j') \qquad (\mathrm{i})$$

$$\underset{jk}{E}(\bar{x}'-\bar{x})=\underset{jk}{V}(\bar{x}') \qquad (\mathrm{ii})$$

と書ける。そして$\bar{x}_j'=\frac{1}{\bar{n}}\sum_{k=1}^{\bar{n}}x_{jk}'$、なるため(4·24)式に於て$m=1$と置いたものに等しい。故に(i)は(4·25)式より

$$\underset{jk}{V}(\bar{x}_j')=\sigma_b^2+\frac{\bar{N}-\bar{n}}{\bar{N}-1}\,\frac{\sigma_w^2}{\bar{n}}$$

又$\bar{x}'$は(4·24)式に外ならないため(ii)は(4·25)式に同じである。

故に

$$\underset{jk}{E}(s_b^2)=\left(\sigma_b^2+\frac{\bar{N}-\bar{n}}{\bar{N}-1}\,\frac{\sigma_w^2}{\bar{n}}\right)-\left(\frac{M-m}{M-1}\,\frac{\sigma_b^2}{m}+\frac{\bar{N}-\bar{n}}{\bar{N}-1}\,\frac{\sigma_w^2}{m\bar{n}}\right)$$

$$=\frac{m-1}{m}\,\frac{M}{M-1}\,\sigma_b^2+\frac{m-1}{m}\,\frac{\bar{N}-\bar{n}}{\bar{N}-1}\,\frac{\sigma_w^2}{\bar{n}}$$

$$\therefore\quad \underset{jk}{E}\left(\frac{m}{m-1}\,s_b^2\right)=\frac{M}{M-1}\,\sigma_b^2+\frac{\bar{N}-\bar{n}}{\bar{N}-1}\,\frac{\sigma_w^2}{\bar{n}}$$

$$\therefore\ \frac{M-1}{M}\left\{\sum_{jk} E\left(\frac{m}{m-1}s_b^2\right)-\frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{\sigma_w^2}{\bar{n}}\right\}=\sigma_b^2$$

かくて左辺第二項のσ_w^2にそれの不偏推定値(4・38)を代入した

$$\frac{M-1}{M}\left(\frac{m}{m-1}s_b^2-\frac{\bar{N}-\bar{n}}{\bar{N}}\frac{s_w^2}{\bar{n}-1}\right)\sigma\approx\sigma_b^2 \quad (4\cdot39)$$

はσ_b^2の不偏推定値である。

(5) 二段抽出法の精度

(4・22)と(3・8)又は(4・25)と(3・22)の比較より明らかな如く二段抽出法による推定値の誤差分散は、集落一段抽出法による推定値の誤差分散に第二次抽出による誤差分散が加ったものであり、従って同個数の集落(一次抽出単位)を抽出する限り[1]、二段抽出法の精度の方が集落一段抽出法の精度よりも落ちるのである。そして集落の内部が同質的でありσ_j^2 ($j=1, 2, \ldots\ldots, M$) 又はσ_w^2の値が小さい程(4・22)又は(4・25)の右辺第二項が小となり、従って二段抽出法の精度は集落一段抽出法の精度に接近し、そして集落内の単位が完全に同質である場合は$\sigma_j^2=0$, $\sigma_w^2=0$となるため両者は一致するのである。しかしながら集落の内部が同質的である時は、単純抽出法に比して集落一段抽出法の精度が低くなるため、二段抽出法の精度としては向上するも単純抽出法に比較する時はむしろ悪くなるであろう。

1) 二段抽出法に於ては抽出された集落より更に調査単位の抽出が行はれるため、標本の大いさは集落一段抽出法の方が大きい。従ってこれは当然の事柄である。唯集落の内部が完全に同質の単位より成る時は、集落内の単位の全部観察によっても一部観察によっても同一の結果が得られるため、二段抽出法も集落一段抽出法も精度が等しくなるのである。

しかし厳密な精度の比較は同じ大いさの標本に就てなされねばならない。次に集落の大いさが均一な場合に就て且便宜上有限母集団修正を無視し得るものとしてそれを考察しよう。今M個の集落より成る母集団からm個の集落を抽出し、その各々に於て$\bar{N}$個の調査単位から$\bar{n}$個の単位を抽出して標本を構成した場合の分散をV_3とすると、(4・26)式より

$$V_3 \doteqdot \frac{\sigma_b^2}{m}+\frac{\sigma_w^2}{m\bar{n}}$$

である。この二段抽出法による標本には $m\bar{n}$ 個の調査単位が含まれているから、これと同じ個数の調査単位の集落一段抽出法による標本の分散を V_2 とすると、(3•22) 式より

$$V_2 \doteqdot \frac{\dfrac{\sigma^2{}_b}{m\bar{n}}}{\bar{N}}$$

である。そこで両者の比をとると

$$\frac{V_3}{V_2} \doteqdot \frac{\dfrac{\sigma_b{}^2}{m}+\dfrac{\sigma_w{}^2}{m\bar{n}}}{\dfrac{\bar{N}\sigma_b{}^2}{m\bar{n}}} = \frac{\bar{n}\sigma_b{}^2+\sigma_w{}^2}{N\sigma_b{}^2}$$

$$= \frac{1}{\bar{N}}\left(\bar{n}+\frac{\sigma_w{}^2}{\sigma_b{}^2}\right) \tag{4•40}$$

であり、従って

$$\frac{1}{\bar{N}}\left(\bar{n}+\frac{\sigma_w{}^2}{\sigma_b{}^2}\right) \gtreqless 1 \quad \text{なる時に} \quad V_3 \gtreqless V_2$$

となる。この条件式は

$$\frac{\sigma_w{}^2}{\sigma_b{}^2} \gtreqless \bar{N}-\bar{n} \qquad \therefore \quad \frac{\sigma_b{}^2}{\sigma_w{}^2} \lesseqgtr \frac{1}{\bar{N}-\bar{n}}$$

であるため

$$\left.\begin{array}{l} \dfrac{\sigma_b{}^2}{\sigma_w{}^2} < \dfrac{1}{\bar{N}-\bar{n}} \\ \quad〃 = 〃 \\ \quad〃 > 〃 \end{array}\right\} \text{ならば} \left\{\begin{array}{ll} V_3 > V_2 & (\mathrm{i}) \\ 〃 = 〃 & (\mathrm{ii}) \\ 〃 < 〃 & (\mathrm{iii}) \end{array}\right.$$

となる。これより明らかな如く $\sigma_b{}^2$ が極めて小であり $\sigma_w{}^2$ が大なる時は、即ち集落内が可成り異質的であって集落相互間は同質的である時は、(i) が成立し二段抽出法の精度は集落一段抽出法の精度よりも低いのであるが、$\sigma_b{}^2$ が大となり、$\sigma_w{}^2$ が小となるに従って、即ち集落内同質的、集落間異質的となる時は、(iii) が成立し、二段抽出法の精度の方が高くなるのである。既に「集落抽出法の精度」に於て述べたように、集落の構成は多くの場合は集落内同質的、集落間異質的となるのが実際であるため、同じ大いさの標本による限り、二段抽出法は集落一段抽出法に比して必ずしも精度は低くないのである。[1)]

又 (4•40) 式に於て $\bar{N}$ が大なる程右辺の値は小、従って二段抽出法の精度は集落一段抽出法の精度に接近し乃至はより高くなるため、二段抽出法に於ては一次抽出単位を大きくする方が精度が高まる傾向を有するのである。[2)]

1)　これは、標本の大いさが同じである時は抽出される

集落の数が二段抽出法の場合は集落一段抽出法の時よりも多くなるため、第一次抽出による誤差分散が小さくなるからである。しかし集落が相互に同質的であり何れの集落をとっても余り大差がない時は、少数の集落の抽出によっても誤差は小であるため、その時はむしろ第二次抽出の誤差が加る二段抽出法の方が劣るのである。

2) 又(4・40)式に於て$\bar{n}$が小さい程右辺の値は小であるため、各集落に於ける二次抽出単位の抽出数が少い程二段抽出法の精度は高まるのである。それは標本の大いさが同じである限り、$\bar{n}$が小さくなる事は一次抽出単位の抽出数mを大ならしめる事となるからである。

次に二段抽出法の精度を単純抽出法の精度と比較して見よう。二段抽出法による標本は$m\bar{n}$個の調査単位が含まれているため、それと同じ大いさの単純抽出法による標本の分散V_1は、(2・33)式より

$$V_1 \doteqdot \frac{\sigma^2}{m\bar{n}}$$

である。V_3のV_1に対する比を求めると

$$\frac{V_3}{V_1} \doteqdot \frac{\dfrac{\sigma_b^2}{m}+\dfrac{\sigma_w^2}{m\bar{n}}}{\dfrac{\sigma^2}{m\bar{n}}} = \frac{\bar{n}\sigma_b^2+\sigma_w^2}{\sigma^2} = \frac{(\bar{n}-1)\sigma_b^2+\sigma^2}{\sigma^2}$$

$$\because \quad \sigma^2=\sigma_w^2+\sigma_b^2$$

$$=1+(\bar{n}-1)\frac{\sigma_b^2}{\sigma^2} \qquad (4・41)$$

(4・41)の右辺第二項は負数とならないため

$$\left.\begin{array}{l}(\bar{n}-1)\dfrac{\sigma_b^2}{\sigma^2}>0 \\ \qquad〃 \qquad =0\end{array}\right\} \text{ならば} \left\{\begin{array}{ll}V_3>V_1 & (\mathrm{i}) \\ 〃 = 〃 & (\mathrm{ii})\end{array}\right.$$

となる。$\bar{n}>1$とすると、[1] $\sigma_b^2=0$即ち集落は相互に全く同質であり集落内部が異質的である時は、(ii)が成立し二段抽出法の精度と単純抽出法の精度とは一致する[2]が、そうでない時は(i)が成立し、σ_b^2が大即ち集落間異質的、集落内同質的となる程、二段抽出法の精度は単純抽出法の精度に比してより低くなるのである。

1) 尚この事は$\bar{n}$の変化に関してもいう事が出来る。$\bar{n}=1$ならば(ii)が成立し、$\bar{n}>1$なる時は(i)が成立する。そして$\bar{n}$が大なる程二段抽出法の精度は低下するのである。それは標本の大いさが同じである限り$\bar{n}$の増大は第一次抽出単位の数mの減少となるからである。

2) この時は、既述の如く、集落一段抽出法の精度は単純抽出法の精度よりも高いのである。

かくて標本の大いさが同じである時は、集落の大い

さを大きくし、且集落の内部が異質的であるようにする[1]のが望まし事が判るのである。尚$\bar{n}$を小さくしてmを増すのが良いのであるが、$\bar{n}$、mの数の如何は費用に影響するため、それとの関連に於て適正配分として決定されるのである。

1）地域抽出による二段抽出法に於ては、集落の大いさを大ならしめる時は集落内は異質的となる傾向を有するため、この二つの条件は同時に満足され得るのである。

(6) 三段抽出法に於ける推定式

三段抽出法は二段抽出法の第二次に抽出された集落の各〻より、更にd_{jk}個$\left(\begin{matrix} j=1,2,\cdots\cdots,m \\ k=1,2,\cdots\cdots,n_j \end{matrix}\right)$宛の調査単位（又は集落）を抽出して標本を構成する方法であるから、二段抽出法の場合は第二次抽出によって抽出された集落の正確な量的大いさを知り得るのに対して、三段抽出法の場合はそれを推定しなければならないのであり、従って二段抽出法の推定式に於て、二次抽出単位の量的大いさにそれの推定値を代入する事によって、三段抽出法の推定式が得られるのである。

j番一次抽出単位中のk番二次抽出単位はD_{jk}個の調査単位より成って居り、これよりd_{jk}個の単位が抽出されるのであるから、jk番集落の量的大いさX_{jk}の不偏推定値X_{jk}'は、(2・34)式より

$$X_{jk}'=\frac{D_{jk}}{d_{jk}}\sum_{l=1}^{d_{jk}}x_{jkl}', \qquad \underset{l}{E}(j,k)(X_{jk}')=X_{jk} \qquad (4\cdot42)$$

であり、これを(4・16)式に代入して得られる[1]

$$\bar{x}'=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}\left\{\frac{N_j}{n_j}\sum_{k=1}^{n_j}\left(\frac{D_{jk}}{d_{jk}}\sum_{l=1}^{d_{jk}}x_{jkl}'\right)\right\} \qquad (4\cdot43)$$

は母平均$\bar{x}$の不偏推定値である。即ち

$$\underset{jkl}{E}(\bar{x}')=\underset{j}{E}\,\underset{k}{E}(j)\,\underset{l}{E}(j,k)(\bar{x}') \qquad [(4\cdot13)]$$

$$\underset{l}{E}(j,k)(\bar{x}')=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}\left\{\frac{N_j}{n_j}\sum_{k=1}^{n_j}\underbrace{\underset{l}{E}(j,k)(X_{jk}')}_{=X_{jk}\ [(4\cdot42)]}\right\}$$

$$=\frac{1}{N}\frac{M}{m}\sum_{j=1}^{m}\left(\frac{N_j}{n_j}\sum_{k=1}^{n_j}X_{jk}\right)=\bar{x}' \qquad [(4\cdot16)] \quad (4\cdot44)$$

$$\therefore\ \underset{jkl}{E}(\bar{x}')=\underset{j}{E}\,\underset{k}{E}(j)(\bar{x}')=\bar{x} \qquad [(4\cdot18)] \quad (4\cdot45)$$

1）二段抽出法に於ては一般に二次抽出単位は調査単位であるため、それの有する値はx_{jk}と小文字で表される

が、三段抽出法に於ては二次抽出単位は集落であるため、それの値は X_{jk} と大文字で表すのである。故に(4・16)式の x_{jk} はこの場合は X_{jk} と解さねばならない。

この推定式の分散は(4・14)より

$$\underset{jkl}{V}(\bar{x}')=\underset{j}{V}(\underset{k}{E}^{(j)}\underset{l}{E}^{(j,k)}\bar{x}')+\underset{j}{E}\underset{k}{V}^{(j)}(\underset{l}{E}^{(j,k)}\bar{x}')+\underset{j}{E}\underset{k}{E}^{(j)}\underset{l}{V}^{(j,k)}(\bar{x}')$$

しかるに(4・44)より $\underset{l}{E}^{(j,k)}(\bar{x}')=\bar{x}'$ なるため

$$=\underbrace{\underset{j}{V}(\underset{k}{E}^{(j)}\bar{x}')}_{\substack{\parallel\\(4\cdot19)}}+\underbrace{\underset{j}{E}\underset{k}{V}^{(j)}(\bar{x}')}_{\substack{\parallel\\(4\cdot21)}}+\underset{j}{E}\underset{k}{E}^{(j)}\underset{l}{V}^{(j,k)}(\bar{x}')$$

となり、右辺の最初の二項は二段抽出法に於ける分散式(4・22)に外ならない。そして第三項は

$$\underset{l}{V}^{(j,k)}(\bar{x}')=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{j=1}^{m}\left(\frac{N_j}{n_j}\right)^2\sum_{k=1}^{n_j}\underbrace{\underset{l}{V}^{(j,k)}(X_{jk}')}_{\substack{\parallel\\D^2_{jk}\frac{D_{jk}-d_{jk}}{D_{jk}-1}\frac{\sigma_{jk}^2}{d_{jk}}\equiv A_{jk}}}$$

[(2・35)]

$$\therefore\quad\underset{k}{E}^{(j)}\{\underset{l}{V}^{(j,k)}(\bar{x}')\}=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{j=1}^{m}\left(\frac{N_j}{n_j}\right)^2\sum_{k=1}^{n_j}\underset{k}{E}^{(j)}(A_{jk})$$

しかるに A_{jk} は二次抽出単位たるに集落毎に一定の値であり、$\underset{k}{E}^{(j)}(A_{jk})$ は j 番一次抽出単位に於て、それより一個の二次抽出単位を抽出して得られる A_{jk} の値の期望値であるため

$$\underset{k}{E}^{(j)}(A_{jk})=\frac{1}{N_j}\sum_{k=1}^{N_j}A_{jk}$$

故に

$$\underset{k}{E}^{(j)}\{\underset{l}{V}^{(j,k)}(\bar{x}')\}=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{j=1}^{m}\frac{N_j}{n_j}\sum_{k=1}^{N_j}A_{jk}$$

従って

$$\underset{j}{E}\{\underset{k}{E}^{(j)}\underset{l}{V}^{(j,k)}(\ ')\}=\left(\frac{1}{N}\frac{M}{m}\right)^2\sum_{j=1}^{m}\underbrace{\underset{j}{E}\left(\frac{N_j}{n_j}\sum_{k=1}^{N_j}A_{jk}\right)}_{\substack{\parallel\\\frac{1}{M}\sum_{j=1}^{M}\left(\frac{N_j}{n_j}\sum_{k=1}^{N_j}A_{jk}\right)}}$$

$$=\left(\frac{1}{N}\right)^2\frac{M}{m}\sum_{j=1}^{M}\frac{N_j}{n_j}\sum_{k=1}^{N_j}A_{jk}$$

かくて

$$\underset{jkl}{V}(\bar{x}')=\frac{1}{N^2}\left\{M^2\frac{M-m}{M-1}\frac{\sigma_e^2}{m}+\frac{M}{m}\sum_{j=1}^{M}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}+\frac{M}{m}\sum_{j=1}^{M}\frac{N_j}{n_j}\sum_{k=1}^{N_j}D_{jk}^2\frac{D_{jk}-d_{jk}}{D_{jk}-1}\frac{\sigma_{jk}^2}{d_{jk}}\right\}\qquad(4\cdot46)$$

即ち三段抽出法の推定値の誤差分散は二段抽出法の誤差分散に第三次抽出の誤差分散が加ったものである。

この場合も各抽出段階に於ける抽出単位数の如何によって標本推定の精度並に調査の費用が異り、適正配分が問題となるのであるが、その求め方は二段抽出法の場合と全く同じであるため省略する。