㉙ 観測測定値が確率変数と考へられる場合のSamplingについて

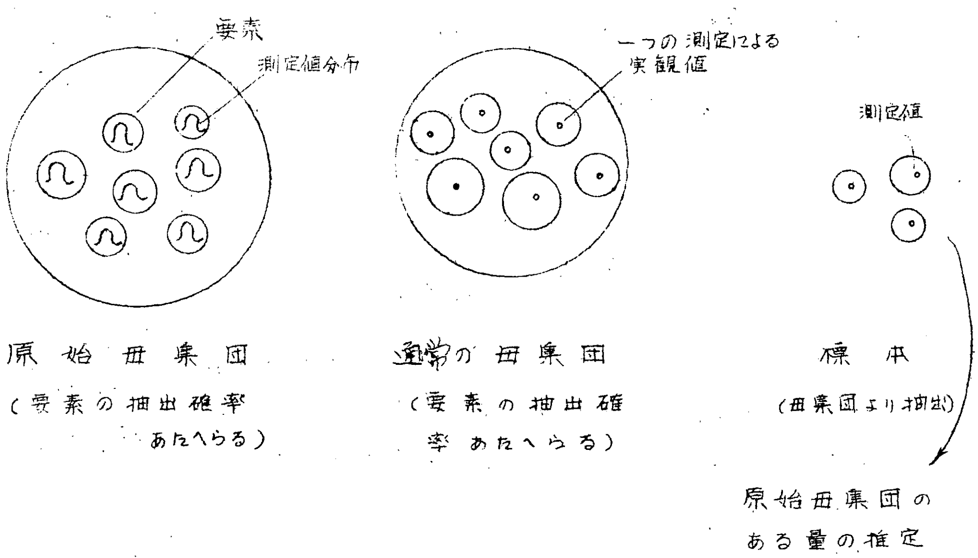
林　知己夫

我々は通常観測，測定値が一定であるとして Sampling の問題を考へて来た。 即ち調査対象の標識が観測測定に対して変化なく一定であるとして，此の各要素にある適当な抽出確率を與へて母集団を構成し此より Sampling を行つて調査しようとする母集団の或量を推定する場合を論じてゐたのである。

しかし，こゝで調査しようとするある量が測定ごとに異つてくる。（しかし此の量は測定に関してある確率法則に従つて分布してゐる。）場合を考へてみたい。

§1. 第一に我々の知りたいのは「Sample から得られた測定毎に異る量」からつくられた母集団の推定量ではなく「Sample から得られるべき（測定量の分布の平均値）」からつくられる母集団の推定量であるべき場合についてまづ考へることにする。つまり，知りたいのは通常の母集団の原母集団とも言ふべきものに対する推定量なのである。

原母集団は測定の分布を考慮に入れさらに調査対象の抽出確率を考へに加へてつくられた一つの母集団とも称すべきものである。



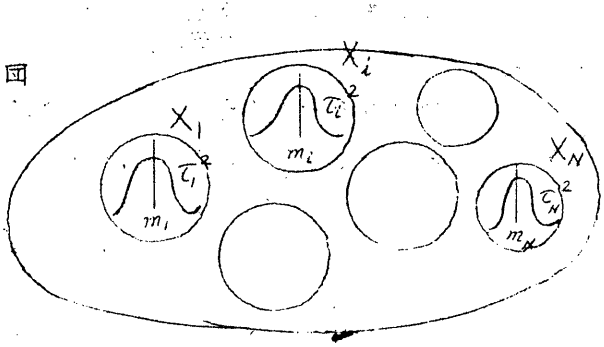
此を例についてのべてみよう。

母集団の大きさを N
各測定値を $X_1, \cdots\cdots, X_N$ （X_i は確率変数夫々平均を m_i, varianceを τ_i^2 とする）

我々は $\dfrac{\sum m_i}{N}$ を推定したいとする。

原始母集団

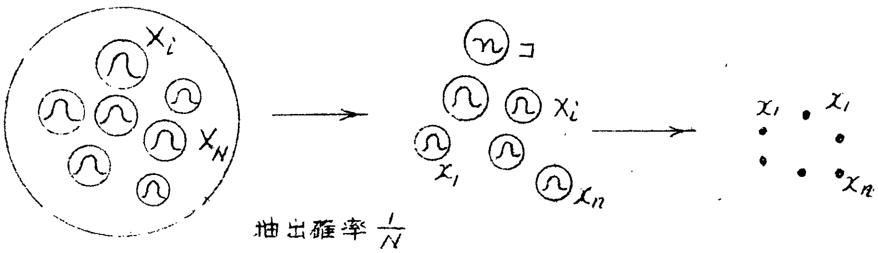
X_i の抽出確率は夫々 $\frac{1}{N}$ とする



我々は，此から n 個の Sample $x_1 \cdots\cdots x_n$ を抽出，それを測定して $y_1, \cdots\cdots, y_n$ を求め，此より $\frac{\sum y_n}{n}$ をつくり，原始母集団の平均 $\frac{\sum m_i}{N}$ の推定値としたいのである。 此の時の Variance は如何なるか。

此の Variance は Sub Sampling の場合の Variance と同様に考へればよいのである。

$\sum_N^1 \left(m_i - \frac{\sum m_i}{N} \right)^2 = \sigma_b^2$ を以て between の Variance τ^2 を以て within の Variance と考へ，primary Sampling Unit を $x_1, \cdots\cdots, x_n$（此等は確率変数ある分布法則をもつ無限母集団と考へる）とし此から，各一つの element を抽出調査するものと考へればよいのである。



抽出確率 $\frac{1}{N}$

したがって Sampling による Variance S^2

$$S^2 = \frac{N-n}{N-1} \cdot \frac{\sigma_b^2}{n} + \sigma_w^2 \cdot \frac{1}{n}$$

$$\sigma_w^2 = \frac{1}{N} \sum \tau_i^2$$

となる。

特に $\tau_i^2 = \tau^2 \ (i = 1 \cdots\cdots N)$ ならば $\sigma_w^2 = \tau^2$

となり

$$s^2 = \frac{N-n}{N-1}\frac{\sigma_b^2}{n} + \frac{\tau^2}{n}$$

従つて，精度は普通の場合に比して第二項の分丈おちることになる。であるから Sampling design にあたつては第一項丈によつて Sample を決定すべきではなく，第二項の影響を考慮に入れて決定してゆかねばならない。ために σ_w^2 が大きいときには，σ_b^2 が小であつても Sample 数を大にせねばならぬであらう。

通常，測定によつて値がことなりうると考へられる場合，所定の精度をうるためには第一項丈でもとめた Sample 数をある程度増加せしめる心掛が必要である事を注意しておかう。

又，測定を一つの Sample x_i に対して反覆行へる様な場合にあつては（実際的には random，independent の條件内ではさう反覆することはできない。反覆するとき明らかに非独立性があらはれ所要の測定のいみを失ふ）一つの primary unit からいくつかの Sample を抽出する場合と考へられるから，此の場合の variance は

$$S^2 = \frac{N-n}{N-1}\frac{\sigma_b^2}{n} + \sigma_w^2\frac{1}{mn} \qquad \text{となる}$$

m は測定回数

Sampling design にあたつては n の決定と共に σ_w^2 の estimate，m の指定は重大な意味をもつものと思ふ。

§2. 次に，基本の考へは §1 と同様なのであるが，$x_1, \cdots x_N$ の分布が従属であるべき（此にともない其他いささかの差異はあるが）時の実例について考へよう。

今，M なる Call（呼出し）をとらへる。Operater（Op とかく）が N_A ゐたとする。

今此のN人の中n人を調査しMを推定する様な場合を考へる。各N人が一つのCallをCatchする確率は$1/N$であると假定する。此の場合N人のOpがMなるCallをCatchしてゐるpattern $M_1, \cdots\cdots, M_N$は多項分布$P(M_1 \cdots\cdots M_N)$

$$= \frac{M!}{\prod_{i=1}^{N} M_i!} \left(\frac{1}{N} \right)^M$$

をなしてゐることになる。

さて我々は各~~N人の~~Opの抽出確率は$\frac{1}{N}$と假定する。N人中n人を調査しMの推定Call

$$x = \frac{\sum_{i=1}^{n} x_i}{n} \cdot N \qquad (x_i \text{ は } M_j \ j=1 \cdots N \text{ の中の何れかを示す})$$

をつくる。此はMのUnbiased Estimateである。この時Mの推定量としてのxのVarianceは如何であるかまづ條件附き確率の考へで推してゆく

Callの一つのpattern $M_1 \cdots\cdots M_N$があたへられて居り，これからn人を抽出したとすればxのVariance S^2は

$$S^2 = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n} N^2$$

但し

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \left(M_i - \frac{M}{N} \right)^2$$

となる。

したがってPatternを対象としてのxのVariance（S^2の平均）V^2は

$$V^2 = E(S^2) = \sum_1 \sum \cdots\cdots \sum_N S^2 P(M_1 \cdots M_N)$$
$$= \frac{N-n}{N-1} \cdot \frac{N^2}{n} \cdot M \left(\frac{1}{N} \right) \left(1 - \frac{1}{N} \right)$$

$$= \frac{N-n}{n}\cdot M$$ となる。

varianceがSampling Ratioにのみより Sample Size によらぬのは興味深い

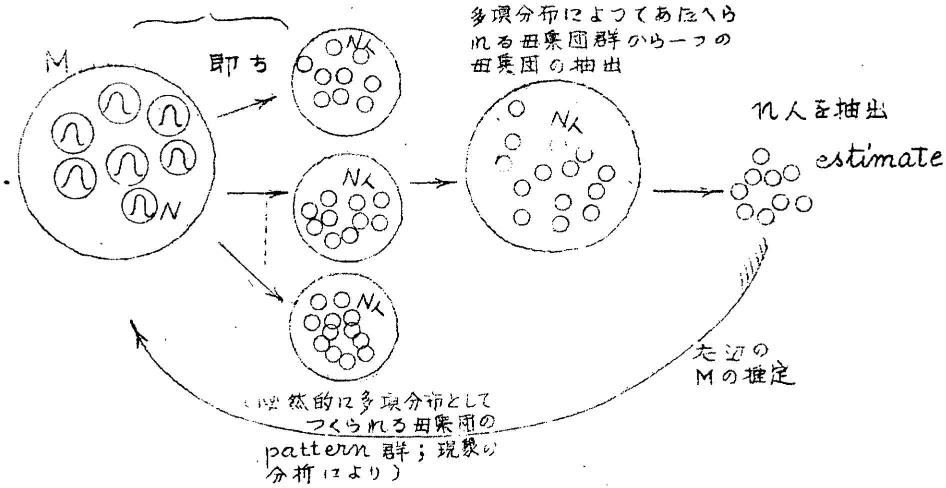
此の時の Coefficient of variation は

$$C.V = \sqrt{\left(\frac{N}{n}-1\right)}\cdot\frac{1}{\sqrt{n}}$$ となる。

例へば　$C.V = 0.02$　$M = 10,000$　$N = 200$
とすれば　$n = 40$　を得る。

C.VはMの増大と共に小さくなる（即ちσ^2が小さくなる）のは多項分布の立場から当然考へられる所であらう。

此の場合の考へは現象調査と言ふものが必然的に



の流れにしたがつて行はれてゐるものと仮定してゐることになるのである。分析的にみて此の様な図式によつて調査が行はれる場合も亦多いのではないかと思はれる。

さてStrataがL個ある場合を考へてみる。

Call	Op.	Sample
M_i	N_i	n_i

Estimate $\quad \Sigma M_i = M$

Sampleからの Estimate

$$\Sigma \overline{x}_i \cdot N_i$$

$\overline{x}_i$ は i 層の Sample 平均

此の時の Variance は

$$V^2 = \Sigma \frac{N_i - n_i}{n_i} M_i$$

$\Sigma n_i = n$ として各層への最も分散の小となる様な割当を求めると

($\Sigma \frac{N_i - n_i}{n_i} M_i + \lambda \Sigma n_i$ を n_i に関して最小にすると考へる。)

$$n_i = n \cdot \frac{\sqrt{M_i N_i}}{\Sigma \sqrt{M_i N_i}}$$

$$= n \cdot \frac{N_i a_i}{\Sigma N_i a_i}$$

$$a_i = \sqrt{\frac{M_i}{N_i}}$$

となる a_i を S.D と考へれば Neymanと云ふ学者のとなる割当法と同様な形をもつこととなる。

以上は単なる二つの例ではあるが，此の様な考へで Sampling design を実際に企画せねばならぬ場合もあり此の時のための一つの覚え書きとして記述してみたものである。