NOTE ON SAMPLING FROM A SOCIOMETRIC PATTERN

By CHIKIO HAYASHI

(Received Aug. 2, 1957)

Usual sampling theory, treats mainly the universe in which the label of an element is a function of only that element itself, although the label is sometimes regarded as a random variable. Here in this note some remark will be made on this point.

We take up a universe consisting of people or things and of size N. Suppose that the relation between i element and j element ($i, j = 1, 2, \cdots, N$, $i \neq j$) is represented by e_{ij} (a numerical value) which is measured by a survey. We want to estimate some statistic made by e_{ij}'s in the universe, for example $I = \sum\sum_{i \neq j}^{N} \frac{e_{ij}}{N(N-1)}$, by the sampling method with the equal sampling probability $1/N$ for each element of the universe.

In those cases, it is not valid in the light of sociometry-technique to use this sampling method, where the value e_{ij} in the universe, which we use to make the required statistic, is a function of both i, j and of the universe, but e_{ij} in a probability sample is determined in the very sample set including fixed i and j, and so e_{ij} varies according to what elements else than i, j are sampled, that is to say, e_{ij} is a function of not only i and j but also of the sample set. In many socio-psychological surveys, we encount this situation.

However, this sampling method is effective in cases where e_{ij} is a functions of only i and j, and the size N is large. These conditions are fulfilled in some kinds of socio-psychological surveys and surveys of factual relations between elements.

In the present note, a problem of sampling from a sociometric pattern is treated as an example of some sampling estimation in a correlated pattern.

1. In-Group Choice.

Suppose each element of the universe concerned chooses positive, neutral or negative relation to all elements excluding itself, and element

i has a numerical value e_{ij} in relation to j ($j=1, 2, \cdots, N$, $i \neq j$) which describes the directive relation of i to j.

For example, let $e_{ij} > 0$ if i likes j and $e_{ji} < 0$ if j dislikes i. j does not always like i though i likes j. In this case we call the relation directive, because the relation of i to j is not same as that of j to i.

	1	2	3	…………………	N
1	×	e_{12}	e_{13}	…………………	e_{1N}
2	e_{21}	×	e_{23}	…………………	e_{2N}
3	e_{31}	e_{32}	×	…………………	e_{3N}
⋮	⋮	⋮	⋮		⋮
N	e_{N1}	e_{N2}	e_{N3}	…………………	×

Fig. 1

Generally $e_{ij} \neq e_{ji}$. We take a sample of size say, n, by equal probability sampling without replacement, in order to estimate the so-called index of cohesion in the universe $I = \sum\sum_{i \neq j}^{N} \frac{e_{ij}}{N(N-1)}$.

This idea leads to probability sampling of correlated lines in a grid.

In the first step, we assume that $e_{ij}=1$ or 0 and that $e_{ij}=1$ represents positive relation, $e_{ij}=0$ non-positive relation. In this case I represents the proportion of positive choices to total choices.

It is natural to take $\bar{x} = \frac{1}{n(n-1)} \sum\sum_{k \neq l}^{n} x_{kl}$ as an estimate of I, where x_{kl} is the label to describe the relation of k to l in a probability sample of size n ($k, l=1, 2, \cdots, n$, $k \neq l$). $\bar{x}$ is clearly an unbiased estimate of I. $\sigma_{\bar{x}}^2$, the variance of $\bar{x}$, is calculated as below.

$$\sigma_{\bar{x}}^2 = \frac{1}{n(n-1)} \Bigg[R + \frac{N-n}{N}(n-2) \Bigg\{ \frac{1}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} N_i^2$$

$$+ 2\frac{1}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} N_i M_i + \frac{1}{(N-1)(N-2)(N-3)} \sum\sum_{i \neq j}^{N} S_i$$

$$+ \frac{1}{(N-1)(N-2)} \left(\frac{1}{n-2} - \frac{1}{N-3} \right) \sum\sum_{i \neq j}^{N} \delta_{ij} \Bigg]$$

where

$$N_i = \sum_{j=1, j \neq i}^{N} e_{ij}$$

$$M_i = \sum_{j=1, j \neq i}^{N} e_{ji}$$

$$S_{ij}=\sum_{g=1,g\neq i,j}^{N} e_{ig}e_{jg}$$

$$\delta_{ij}=\sum\sum_{i\neq j}^{N} e_{ij}e_{ji}$$

$$R=\left[\frac{N-n}{N-2}I+N(N-1)\left\{\frac{(n-2)(n-3)}{(N-2)(N-3)}-\frac{n(n-1)}{N(N-1)}\right\}I^2\right]$$

If $N\gg n$, $N\gg 1$, $n\gg 1$, we obtain,

$$\sigma_{\bar{x}}^2 \doteqdot \frac{I}{n^2}+\frac{1}{n}\left\{\frac{1}{N(N-1)^2}\sum_{i=1}^{N} N_i^2+2\frac{1}{N(N-1)^2}\sum_{i=1}^{N} N_i M_i+\right.$$

$$\left.+\frac{1}{N(N-1)(N-2)}\sum\sum_{i\neq j}^{N} S_{ij}+\frac{1}{n}\frac{1}{N(N-1)}\sum\sum_{i\neq j}^{N}\delta_{ij}\right\}.$$

The estimates of various sociometric indices will be obtained by various definitions of e_{ij}. Suppose that element i has a numerical value e_{ij} in relation to j ($i=1, 2, \cdots, n$, $j=1, 2, \cdots, N$, $i\neq j$), i.e. element i in a probability sample is asked for the relations to all elements in the universe. In this case,

$$\bar{x}=\frac{\sum_{i=1}^{n}\sum_{j=1\, i\neq j}^{N} x_{ij}}{n(N-1)}$$

is an unbiased estimate of I. The variance of $\bar{x}$ can easily be obtained by the usual formulae of sampling theory.

2. Out-Group Choice.

Suppose that there are two groups A, B in the universe, the sizes of which are N_A, N_B, respectively, $N_A+N_B=N$. We take a sample of size n_α from the sub-universe α by equal probability sampling without replacement ($\alpha=A$, B). We want to estimate the statistics

$$O_{AB}=\frac{1}{N_A N_B}\sum_{i_A=1}^{N_A}\sum_{j_B=1}^{N_B} g_{i_A j_B} \quad \text{or} \quad O_{BA}=\frac{1}{N_A N_B}\sum_{j_B=1}^{N_B}\sum_{i_A=1}^{N_A} g_{j_B i_A},$$

by the sampling above mentioned, where $g_{i_A j_B}$ represents the relation of i_A element in group A to j_B element in group B, $g_{j_B i_A}$ vice versa, and generally $g_{i_A j_B}\neq g_{j_B i_A}$.

In sociometric research it is desirable in analyzing a group structure of the universe to use not only I_A, I_B (cohesion index in sub-universes A, B) but also these O_{AB} and O_{BA} simultaneously.

If we define $g_{i_A j_B}(g_{j_B i_A})=1$ or 0 as in the previous section, the following results are easily obtained by the idea of two stage sampling method. As to a further generalization it will be not necessary to describe it here.

We use $\bar{y}_{AB}=\frac{1}{n_A n_B}\sum_{k_A=1}^{n_A}\sum_{l_B=1}^{n_B} y_{k_A l_B}$ to estimate O_{AB}, where $y_{k_A l_B}$ is the label to describe the relation of k_A in a probability sample of A to l_B in that of B, $\bar{y}_{AB}$ is clearly an unbiased estimate of O_{AB}. The variance of $\bar{y}_{AB}$, $\sigma^2_{\bar{y}_{AB}}$, is

$$\frac{N_B-n_B}{N_B-1}\frac{\bar{\sigma}^2_{AB}}{n_A n_B}+\frac{N_A-n_A}{N_A-1}\frac{\sigma^2_{P_{AB}}}{n_A}+\frac{N_A-n_B}{N_B-1}\frac{1}{n_B}\frac{n_A-1}{n_A}\frac{1}{N_A(N_A-1)}\sum_{r_A}^{N_A}\sum_{s_A}^{N_A} C_{r_A s_A}$$

where

$$P_{kB}=\frac{1}{N_B}\sum_{l_B=1}^{N_B} g_{k l_B}\,, \qquad k=1, 2, \cdots, N_A$$

$$\bar{P}_{AB}=\frac{1}{N_A}\sum_{k=1}^{N_A} P_{kB}$$

$$\sigma^2_{P_{AB}}=\frac{1}{N_A}\sum_{k=1}^{N_A}(P_{kB}-\bar{P}_{AB})^2$$

$$\tau^2_{kB}=P_{kB}(1-P_{kB})\,, \qquad k=1, 2, \cdots, N_A$$

$$\bar{\sigma}^2_{AB}=\frac{1}{N_A}\sum_{k=1}^{N_A}\tau^2_{kB}$$

$$C_{r_A s_A}=\frac{1}{N_B}\sum_{j_B=1}^{N_B}(g_{r_A j_B}-P_{r_A B})(g_{s_A j_B}-P_{s_A B})\,.$$

It will be possible to design a sampling survey by using the ideas mentioned above.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Hayashi, C. and H. Akaike, On a Matching Problem, *Annals of the Institute of Statistical Mathematics*, Vol. IV, No. 2, 1953.

[2] Hayashi, C., Response Errors and Sampling Design, *Proceedings of the Institute of Statistical Mathematics*, Vol. 5, No. 1, 1957.