MODEL FOR THE ESTIMATION OF THE SIZE OF A POPULATION BY USING CAPTURE-RECAPTURE METHOD

By KOITI TAKAHASI

(Received Dec. 23, 1960)

Summary

In the present paper we shall get an estimate, by using a multiple recapture method, of the total size of a population which consists of several classes, and consider some asymptotic properties of the estimate.

1. Introduction

In most of the models for the estimation of the total size of a population, it is usually assumed that any individual in the population has equal probability of being caught, but we may easily imagine a population where this assumption is not valid.

D. G. Chapman and C. O. Junge considered such a case [1]. In their model, the population is stratified and partial mixing takes place between strata. The individuals of different strata have different probabilities of being caught.

Now, our situation is as follows. Our experiment will consist of T repetitions of sampling. In i^{th} sampling, the individuals caught will be tagged with "i" and then released, where $i=1, 2, \cdots, T$. Therefore the data obtained by our experiment are the set of $n_{ij\cdots k}$'s, where $n_{ij\cdots k}$ denotes the number of individuals which are caught in the $i^{\text{th}}, j^{\text{th}}, \cdots$, and k^{th} samples. On the other hand, in Chapman's model, any individual caught is tagged with "s" when it is caught in s^{th} stratum. In our model, it will be assumed that a population consists of several classes and each class consists of the individuals of the population and the individuals of different classes have different probabilities of being caught, while each individual of the same class has the equal probability of being caught. Furthermore, it should be noticed that we can not know which class the individuals caught belong to.

Practically, our results may be applied to the estimation of the size of an animal or fish population where the individuals are moving in several groups within a bounded area. We notice that the procedures of our experiment are just the same as those of the experiment for the estimation of the size of a population where any individual has equal probability of being caught, when we disregard the restriction below, on the number of samples taken. Therefore, if the number of samples taken is large enough for our model, our model may be applicable to the case where the experiment is carried out under the assumption that all individuals of the population have the equal probability of being caught, and yet we can not obtain valid results, but the obtained data satisfy the conditions of our model for an appropriate number of classes. In this case, the number of classes may be estimated by using the idea of the latent structure analysis [cf. [2] Chapter 10, 11].

In order to get our estimate we shall use the notion of the latent structure analysis [see [2] for example]. The obtained data in our model correspond to the response pattern for a sequence of items in the usual latent structure analysis; tags correspond to positive answers, the difference among samples to that among the items, and our classes to the latent classes. However, there is a remarkable difference between our model and the usual latent class model; in our model, the total size of a population is unknown, so that the number of individuals which are not caught at all is unknown. Since it is very difficult to get the maximum likelihood estimate in our model, we use relations of the expectations of the observable random variables. Furthermore, we notice that to our method information about the frequencies of third or higher order may not be useful for obtaining an estimate of the total size of the population. But they will be used for estimation of the asymptotic variance of our estimate.

2. Notation and assumptions

N = the total number of individuals in the population (unknown),

s = the number of classes (assumed to be known),

$N^{(r)}$ = the number of individuals in r^{th} class ($r=1, 2, \cdots, s$) (unknown),

T = the number of samples taken,

$p_j^{(r)}$ = the probability that any individual in r^{th} class is caught in j^{th} sample ($r=1, 2, \cdots, s$; $j=1, \cdots, T$) (unknown),

$n_{j_1\cdots j_t}$ = the number of individuals that are caught in the j_1^{th}, j_2^{th}, $\cdots$, and j_t^{th} samples (observable),

$n^{(r)}_{j_1\cdots j_t}$ = the number of individuals in r^{th} class that are caught in the $j_1^{\text{th}}, j_2^{\text{th}}, \cdots$, and j_t^{th} samples (unobservable).

Assumptions

(I) $n^{(1)}_{j_1\cdots j_t}, n^{(2)}_{j_1\cdots j_t}, \cdots$, and $n^{(s)}_{j_1\cdots j_t}$ are independent for all $j_1, \cdots, j_t$. $n_1^{(r)}, n_2^{(r)}, \cdots$, and $n_T^{(r)}$ are independent for $r=1, 2, \cdots, s$.

(II) $$\begin{vmatrix} p^{(1)}_{j_1} p^{(1)}_{j_2} \cdots p^{(1)}_{j_s} \\ p^{(2)}_{j_1} p^{(2)}_{j_2} \cdots p^{(2)}_{j_s} \\ \cdots\cdots\cdots \\ p^{(s)}_{j_1} p^{(s)}_{j_2} \cdots p^{(s)}_{j_s} \end{vmatrix} \neq 0 \qquad \text{for all } j_1, \cdots, j_s.$$

We shall later see that this assumption is essentially necessary.

(III) $$T \geqq 2s\ .$$

3. Estimation of N

The expectations of the $n_{j_1\cdots j_t}$'s are easily calculated to be

(1) $$En_{j_1\cdots j_t} = \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_1} \cdots p^{(r)}_{j_t} \qquad \text{for all } j_1, \cdots, j_t,$$

and we have

(2) $$N = \sum_{r=1}^{s} N^{(r)}\ .$$

Since it is generally impossible to repeat the experiment under the same conditions, we shall practically substitute the values obtained by only one experiment into the left hand sides of equations (1).

For the purpose of obtaining an estimate of N, we consider algebraic relations between the right members of (1) and that of (2). For simplicity, the right hand sides of (1) are also denoted by the $n_{j_1\cdots j_t}$'s. We can use here the basic determinants in the latent structure analysis, that is, the determinants of the form

(3) $$\begin{vmatrix} N & n_{k_1} & \cdots & n_{k_t} \\ n_{j_1} & n_{j_1 k_1} & \cdots & n_{j_1 k_t} \\ \vdots & \vdots & & \vdots \\ n_{j_t} & n_{j_t k_1} & \cdots & n_{j_t k_t} \end{vmatrix}\ .$$

If $t=s$, (3) vanishes. For,

$$\text{the left hand side of (3)} = \begin{vmatrix} \sum_{r=1}^{s} N^{(r)} & \sum_{r=1}^{s} N^{(r)} p^{(r)}_{k_1} & \cdots & \sum_{r=1}^{s} N^{(r)} p^{(r)}_{k_s} \\ \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_1} & \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_1} p^{(r)}_{k_1} & \cdots & \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_1} p^{(r)}_{k_s} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_s} & \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_s} p^{(r)}_{k_1} & \cdots & \sum_{r=1}^{s} N^{(r)} p^{(r)}_{j_s} p^{(r)}_{k_s} \end{vmatrix}$$

$$
=\begin{vmatrix} 1 & \cdots & 1 & 0 \\ p_{j_1}^{(1)} & \cdots & p_{j_1}^{(s)} & 0 \\ \cdots & \cdots & \cdots & \\ p_{j_s}^{(1)} & \cdots & p_{j_s}^{(s)} & 0 \end{vmatrix} \begin{vmatrix} N^{(1)} & 0 & \cdots & 0 \\ 0 & N^{(2)} & \ddots & \vdots \\ \vdots & \ddots & N^{(s)} & 0 \\ 0 & \cdots & 0 & 0 \end{vmatrix} \begin{vmatrix} 1 & p_{k_1}^{(1)} & \cdots & p_{k_s}^{(1)} \\ \vdots & \vdots & & \vdots \\ 1 & p_{k_1}^{(s)} & \cdots & p_{k_s}^{(s)} \\ 0 & 0 & \cdots & 0 \end{vmatrix} .
$$

If

$$
(4) \qquad \begin{vmatrix} n_{j_1k_1} & \cdots & n_{j_1k_s} \\ \vdots & & \vdots \\ n_{j_sk_1} & \cdots & n_{j_sk_s} \end{vmatrix} \neq 0 ,
$$

we can solve (3) with respect to N, so that we have

$$
(5) \qquad N=-\frac{\begin{vmatrix} 0 & n_{k_1} & \cdots & n_{k_s} \\ n_{j_1} & n_{j_1k_1} & \cdots & n_{j_1k_s} \\ \vdots & \vdots & & \vdots \\ n_{j_s} & n_{j_sk_1} & \cdots & n_{j_sk_s} \end{vmatrix}}{\begin{vmatrix} n_{j_1k_1} & \cdots & n_{j_1k_s} \\ \vdots & & \vdots \\ n_{j_sk_1} & \cdots & n_{j_sk_s} \end{vmatrix}} .
$$

Thus we obtain an estimate, $\hat{N}$, of N when we regard the n_j's, n_{jk}'s as observable random variables;

$$
(6) \qquad \hat{N}=-\frac{\begin{vmatrix} 0 & n_{k_1} & \cdots & n_{k_s} \\ n_{j_1} & n_{j_1k_1} & \cdots & n_{j_1k_s} \\ \vdots & \vdots & & \vdots \\ n_{j_s} & n_{j_sk_1} & \cdots & n_{j_sk_s} \end{vmatrix}}{\begin{vmatrix} n_{j_1k_1} & \cdots & n_{j_1k_s} \\ \vdots & & \vdots \\ n_{j_sk_1} & \cdots & n_{j_sk_s} \end{vmatrix}} .
$$

The left hand side of (4) is

$$
(7) \qquad \begin{vmatrix} \sum N^{(r)} p_{j_1}^{(r)} p_{k_1}^{(r)} & \cdots & \sum N^{(r)} p_{j_1}^{(r)} p_{k_s}^{(r)} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \sum N^{(r)} p_{j_s}^{(r)} p_{k_1}^{(r)} & \cdots & \sum N^{(r)} p_{j_s}^{(r)} p_{k_s}^{(r)} \end{vmatrix}
$$

$$=\begin{vmatrix} p_{j_1}^{(1)} & \cdots & p_{j_1}^{(s)} \\ \vdots & & \vdots \\ p_{j_s}^{(1)} & \cdots & p_{j_s}^{(s)} \end{vmatrix} \begin{vmatrix} N^{(1)} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & N^{(r)} \end{vmatrix} \begin{vmatrix} p_{k_1}^{(1)} & \cdots & p_{k_s}^{(1)} \\ \vdots & & \vdots \\ p_{k_1}^{(s)} & \cdots & p_{k_s}^{(s)} \end{vmatrix} .$$

Therefore, (4) is satisfied if and only if

$$(8) \qquad \begin{vmatrix} p_{j_1}^{(1)} & \cdots & p_{j_1}^{(s)} \\ \vdots & & \vdots \\ p_{j_s}^{(1)} & \cdots & p_{j_s}^{(s)} \end{vmatrix} \neq 0 \quad \text{and} \quad \begin{vmatrix} p_{k_1}^{(1)} & \cdots & p_{k_1}^{(s)} \\ \vdots & & \vdots \\ p_{k_s}^{(1)} & \cdots & p_{k_s}^{(s)} \end{vmatrix} \neq 0 .$$

Our assumption (II) is thus seen to have its origin in (8).

4. The asymptotic variance of $\hat{N}$

In this section, the term "asymptotic" will be used in the sense that the probabilities $p_j^{(r)}$'s are arbitrarily fixed and the $N^{(r)}$'s become sufficiently large. Furthermore, we shall assume that

$$(9) \qquad N^{(1)}, N^{(2)}, \cdots, N^{(r)}=O(N) \qquad \text{as } N\to\infty.$$

Now, (6) may be regarded as defining $\hat{N}$ as a function of $n_{j_1}, n_{j_2}, \cdots,$ and $n_{j_sk_s}$ and we may expand $\hat{N}$ about $\mathfrak{N}^\circ=(n_{j_1}^\circ, \cdots, n_{j_sk_s}^\circ)$, where $n_{j_1}^\circ, \cdots, n_{j_sk_s}^\circ$ denote $\sum_r N^{(r)}p_{j_1}^{(r)}, \cdots, \sum_r N^{(r)}p_{j_s}^{(r)}p_{k_s}^{(r)}$, respectively. (5) shows that $N=\hat{N}(\mathfrak{N}^\circ)$.

Hence, we have

$$(10) \qquad \hat{N}-N=\sum_{l=1}^{s}(n_{j_l}-n_{j_l}^\circ)\frac{\partial\hat{N}}{\partial n_{j_l}}\bigg|_{\mathfrak{n}^\circ}+\sum_{m=1}^{s}(n_{k_m}-n_{k_m}^\circ)\frac{\partial\hat{N}}{\partial n_{k_m}}\bigg|_{\mathfrak{n}^\circ}$$
$$+\sum_{l,m=1}^{s}(n_{j_lk_m}-n_{j_lk_m}^\circ)\frac{\partial\hat{N}}{\partial n_{j_lk_m}}\bigg|_{\mathfrak{n}^\circ}+\cdots .$$

It can readily be shown that

$$(11) \qquad \begin{cases} E(n_{j_l}-n_{j_l}^\circ)^2=\sum_{r=1}^{s}N^{(r)}p_{j_l}^{(r)}(1-p_{j_l}^{(r)}) , \\ E(n_{k_m}-n_{k_m}^\circ)^2=\sum_{r=1}^{s}N^{(r)}p_{k_m}^{(r)}(1-p_{k_m}^{(r)}) , \\ E(n_{j_lk_m}-n_{j_lk_m}^\circ)^2=\sum_{r=1}^{s}N^{(r)}p_{j_l}^{(r)}p_{k_m}^{(r)}(1-p_{j_l}^{(r)}p_{k_m}^{(r)}) , \\ E(n_{j_l}-n_{j_l}^\circ)(n_{j_lk_m}-n_{j_lk_m}^\circ)=\sum_{r=1}^{s}N^{(r)}p_{j_l}^{(r)}p_{k_m}^{(r)}(1-p_{j_l}^{(r)}) , \\ E(n_{k_m}-n_{k_m}^\circ)(n_{j_lk_m}-n_{j_lk_m}^\circ)=\sum_{r=1}^{s}N^{(r)}p_{j_l}^{(r)}p_{k_m}^{(r)}(1-p_{k_m}^{(r)}) , \\ E(n_{j_lk_m}-n_{j_lk_m}^\circ)(n_{j_lk_m}-n_{j_lk_m}^\circ)=\sum_{r=1}^{s}N^{(r)}p_{j_l}^{(r)}p_{k_m}^{(r)}p_{k_{m'}}^{(r)}(1-p_{j_l}^{(r)}) , \\ E(n_{j_lk_m}-n_{j_lk_m}^\circ)(n_{j_lk_m}-n_{j_{l'}k_m}^\circ)=\sum_{r=1}^{s}N^{(r)}p_{j_l}^{(r)}p_{j_{l'}}^{(r)}p_{k_m}^{(r)}(1-p_{k_m}^{(r)}) , \end{cases}$$

$$(11') \quad \begin{cases} E(n_{j_l}-n^\circ_{j_l})(n_{k_m}-n^\circ_{k_m})=0\,, & E(n_{j_l}-n^\circ_{j_l})(n_{j_{l'}}-n^\circ_{j_{l'}})=0\,, \\ E(n_{k_m}-n^\circ_{k_m})(n_{k_{m'}}-n^\circ_{k_{m'}})=0\,, & E(n_{j_l}-n^\circ_{j_l})(n_{j_{l'}k_m}-n^\circ_{j_{l'}k_m})=0\,, \\ E(n_{j_lk_m}-n^\circ_{j_lk_m})(n_{j_{l'}k_{m'}}-n^\circ_{j_{l'}k_{m'}})=0\,, & \text{where} \quad l\neq l',\ m\neq m'. \end{cases}$$

and

$$(12) \quad \begin{cases} \left.\dfrac{\partial\hat{N}}{\partial n_{j_l}}\right|_{\mathfrak{n}^\circ}=(-1)^{l+1}\dfrac{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ p^{(1)}_{j_1} & p^{(2)}_{j_1} & \cdots & p^{(s)}_{j_1} \\ \cdots & \cdots & \cdots & \cdots \\ p^{(1)}_{j_{l-1}} & p^{(2)}_{j_{l-1}} & \cdots & p^{(s)}_{j_{l-1}} \\ p^{(1)}_{j_{l+1}} & p^{(2)}_{j_{l+1}} & \cdots & p^{(s)}_{j_{l+1}} \\ \cdots & \cdots & \cdots & \cdots \\ p^{(1)}_{j_s} & p^{(2)}_{j_s} & \cdots & p^{(s)}_{j_s} \end{vmatrix}}{\begin{vmatrix} p^{(1)}_{j_1} & \cdots & p^{(s)}_{j_1} \\ \vdots & & \vdots \\ p^{(1)}_{j_s} & \cdots & p^{(s)}_{j_s} \end{vmatrix}}\,, \\[2ex] \left.\dfrac{\partial\hat{N}}{\partial n_{k_m}}\right|_{\mathfrak{n}^\circ}=(-1)^{m+1}\dfrac{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ p^{(1)}_{k_1} & p^{(2)}_{k_1} & \cdots & p^{(s)}_{k_1} \\ \cdots & \cdots & \cdots & \cdots \\ p^{(1)}_{k_{m-1}} & p^{(2)}_{k_{m-1}} & \cdots & p^{(s)}_{k_{m-1}} \\ p^{(1)}_{k_{m+1}} & p^{(2)}_{k_{m+1}} & \cdots & p^{(s)}_{k_{m+1}} \\ \cdots & \cdots & \cdots & \cdots \\ p^{(1)}_{k_s} & p^{(2)}_{k_s} & \cdots & p^{(s)}_{k_s} \end{vmatrix}}{\begin{vmatrix} p^{(1)}_{k_1} & \cdots & p^{(s)}_{k_1} \\ \vdots & & \vdots \\ p^{(1)}_{k_s} & \cdots & p^{(s)}_{k_s} \end{vmatrix}}\,, \\[2ex] \left.\dfrac{\partial\hat{N}}{\partial n_{j_lk_m}}\right|_{\mathfrak{n}^\circ}=-\left.\dfrac{\partial\hat{N}}{\partial n_{j_l}}\right|_{\mathfrak{n}^\circ}\cdot\left.\dfrac{\partial\hat{N}}{\partial n_{k_m}}\right|_{\mathfrak{n}^\circ} \end{cases}$$

and generally

$$(13) \quad \begin{cases} E(n_{j_1}-n^\circ_{j_1})^{\nu_{j_1}}\cdots(n_{j_sk_s}-n^\circ_{j_sk_s})^{\nu_{j_sk_s}}=O(N^{[\nu/2]}) \\ \left.\dfrac{\partial^\nu\hat{N}}{\partial n^{\nu_1}_{j_1}\cdots\partial n^{\nu_{j_sk_s}}_{j_sk_s}}\right|_{\mathfrak{n}^\circ}=O(N^{-\nu+1})\,, \quad \text{where} \quad \nu=\nu_{j_1}+\cdots+\nu_{j_sk_s}\,. \end{cases}$$

From (10), (11′), and (13), we obtain

(14)

$$E(\hat{N}-N)^2\sim\sum_{\substack{i=j_1,\cdots,j_s\\k_1,\cdots,k_s}}E(n_i-n^\circ_i)^2\left.\frac{\partial\hat{N}}{\partial n_i}\right|^2_{\mathfrak{n}^\circ}+\sum_{\substack{l=1,\cdots,s\\m=1,\cdots,s}}E(n_{j_lk_m}-n^\circ_{j_lk_m})\left.\frac{\partial\hat{N}}{\partial n_{j_lk_m}}\right|^2_{\mathfrak{n}^\circ}$$
$$+2\sum_{\substack{l=1,\cdots,s\\m=1,\cdots,s}}E(n_{j_l}-n^\circ_{j_l})(n_{j_lk_m}-n^\circ_{j_lk_m})\left.\frac{\partial\hat{N}}{\partial n_{j_l}}\right|_{\mathfrak{n}^\circ}\cdot\left.\frac{\partial\hat{N}}{\partial n_{j_lk_m}}\right|_{\mathfrak{n}^\circ}$$

$$+2\sum_{\substack{l=1,\dots,s\\ m=1,\dots,s}} E(n_{k_m}-n^\circ_{k_m})(n_{j_lk_m}-n^\circ_{j_lk_m})\frac{\partial\hat N}{\partial n_{k_m}}\Big|_{\mathrm{n}^\circ}\cdot\frac{\partial\hat N}{\partial n_{j_lk_m}}\Big|_{\mathrm{n}^\circ}$$

$$+\sum_{\substack{l,m,m'=1,\dots,s\\ m\neq m'}} E(n_{j_lk_m}-n^\circ_{j_lk_m})(n_{j_lk_{m'}}-n^\circ_{j_lk_{m'}})\frac{\partial\hat N}{\partial n_{j_lk_m}}\Big|_{\mathrm{n}^\circ}\cdot\frac{\partial\hat N}{\partial n_{j_lk_{m'}}}\Big|_{\mathrm{n}^\circ}$$

$$+\sum_{\substack{l,l',m=1,\dots,s\\ l\neq l'}} E(n_{j_lk_m}-n^\circ_{j_lk_m})(n_{j_{l'}k_m}-n^\circ_{j_{l'}k_m})\frac{\partial\hat N}{\partial n_{j_lk_m}}\Big|_{\mathrm{n}^\circ}\cdot\frac{\partial\hat N}{\partial n_{j_{l'}k_m}}\Big|_{\mathrm{n}^\circ}\,.$$

Substituting (11) and (12) into (14), we obtain the representation of the asymptotic mean square error of $\hat N$ in terms of the population parameters. Furthermore, we can see that the asymptotic bias of $\hat N$ has order O(1);

$$E(\hat N-N)\sim\sum_{l,m=1}^{s}\Big\{E(n_{j_l}-n^\circ_{j_l})(n_{j_lk_m}-n^\circ_{j_lk_m})\frac{\partial^2\hat N}{\partial n_{j_l}\partial n_{j_lk_m}}\Big|_{\mathrm{n}^\circ}$$
$$+E(n_{k_m}-n^\circ_{k_m})(n_{j_lk_m}-n^\circ_{j_lk_m})\frac{\partial^2\hat N}{\partial n_{k_m}\partial n_{j_lk_m}}\Big|_{\mathrm{n}^\circ}$$
$$+\frac{1}{2}E(n_{j_lk_m}-n^\circ_{j_lk_m})^2\frac{\partial^2\hat N}{\partial n^2_{j_lk_m}}\Big|_{\mathrm{n}^\circ}\Big\}$$
$$+\sum_{\substack{l,m,m'=1\\ m\neq m'}}^{s} E(n_{j_lk_m}-n^\circ_{j_lk_m})(n_{j_lk_{m'}}-n^\circ_{j_lk_{m'}})\frac{\partial^2\hat N}{\partial n_{j_lk_m}\partial n_{j_lk_{m'}}}\Big|_{\mathrm{n}^\circ}$$
$$+\sum_{\substack{l,l',m=1\\ l\neq l'}}^{s} E(n_{j_lk_m}-n^\circ_{j_lk_m})(n_{j_{l'}k_m}-n^\circ_{j_{l'}k_m})\frac{\partial^2\hat N}{\partial n_{j_lk_m}\partial n_{j_{l'}k_m}}\Big|_{\mathrm{n}^\circ}, \tag{15}$$

and, for example,

$$E(n_{j_l}-n^\circ_{j_l})(n_{j_lk_m}-n^\circ_{j_lk_m})\frac{\partial^2\hat N}{\partial n_{j_l}\partial n_{j_lk_m}}\Big|_{\mathrm{n}^\circ}$$

$$=-\frac{\left\{\sum_{r=1}^{s}N^{(r)}p^{(r)}_{j_l}p^{(r)}_{k_m}(1-p^{(r)}_{j_l})\right\}}{\left(\prod_{r=1}^{s}N^{(r)}\right)\begin{vmatrix}p^{(1)}_{j_1}&\cdots&p^{(s)}_{j_1}\\ \vdots&&\vdots\\ p^{(1)}_{j_s}&\cdots&p^{(s)}_{j_s}\end{vmatrix}^2\cdot\begin{vmatrix}p^{(1)}_{k_1}&\cdots&p^{(1)}_{k_s}\\ \vdots&&\vdots\\ p^{(s)}_{k_1}&\cdots&p^{(s)}_{k_s}\end{vmatrix}}\cdot\left\{\begin{vmatrix}1&\cdots&1\\ p^{(1)}_{j_1}&\cdots&p^{(s)}_{j_1}\\ \cdots&\cdots&\cdots\\ p^{(1)}_{j_{l-1}}&\cdots&p^{(s)}_{j_{l-1}}\\ p^{(1)}_{j_{l+1}}&\cdots&p^{(s)}_{j_{l+1}}\\ \cdots&\cdots&\cdots\\ p^{(1)}_{j_s}&\cdots&p^{(s)}_{j_s}\end{vmatrix}\right.$$

$$\times\left.\begin{vmatrix}\sum N^{(r)}p^{(r)}_{j_1}p^{(r)}_{k_1}&\cdots&\sum N^{(r)}p^{(r)}_{j_1}p^{(r)}_{k_{m-1}}&\sum N^{(r)}p^{(r)}_{j_1}p^{(r)}_{k_{m+1}}&\cdots&\sum N^{(r)}p^{(r)}_{j_1}p^{(r)}_{k_s}\\ \cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ \sum N^{(r)}p^{(r)}_{j_{m-1}}p^{(r)}_{k_1}&\cdots&\cdots&\cdots&\cdots&\cdots\\ \sum N^{(r)}p^{(r)}_{j_{m+1}}p^{(r)}_{k_1}&\cdots&\cdots&\cdots&\cdots&\cdots\\ \cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ \sum N^{(r)}p^{(r)}_{j_s}p^{(r)}_{k_1}&\cdots&\cdots&\cdots&\cdots&\sum N^{(r)}p^{(r)}_{j_s}p^{(r)}_{k_s}\end{vmatrix}\right\}$$

$$=\mathrm{O}(1)\,.$$

Therefore, it makes no difference whether we speak of asymptotic mean square error or of asymptotic variance of $\hat{N}$.

Now, it is very difficult to estimate the asymptotic variance from the observed data. We shall restrict our consideration to the following two cases.

(I) First we assume that $s=2$. In this case, we can write the asymptotic variance in terms of the n_i°'s, n_{ij}°'s, n_{ijk}°'s, n_{ijkl}°'s, and N etc. It is easily verified that

$$
(16)\quad \begin{cases}
E(n_i-n_i^\circ)=\dfrac{n_i^\circ(N-n_i^\circ)}{N}-\dfrac{1}{N}\dfrac{\begin{vmatrix} N & n_j^\circ \\ n_i^\circ & n_{ij}^\circ \end{vmatrix}\cdot\begin{vmatrix} N & n_k^\circ \\ n_i^\circ & n_{ik}^\circ \end{vmatrix}}{\begin{vmatrix} N & n_j^\circ \\ n_k^\circ & n_{jk}^\circ \end{vmatrix}}\,, \\[2ex]
E(n_{ij}-n_{ij}^\circ)=\dfrac{n_{ij}^\circ(N-n_{ij}^\circ)}{N}-\dfrac{1}{N}\dfrac{\begin{vmatrix} N & n_{ij}^\circ \\ n_{kl}^\circ & n_{ijkl}^\circ \end{vmatrix}\cdot\begin{vmatrix} N & n_{ij}^\circ \\ n_{hm}^\circ & n_{ijhm}^\circ \end{vmatrix}}{\begin{vmatrix} N & n_{kl}^\circ \\ n_{hm}^\circ & n_{hmkl}^\circ \end{vmatrix}}\,, \\[2ex]
E(n_{ij}-n_{ij}^\circ)(n_{ik}-n_{ik}^\circ)=\dfrac{n_{ijk}^\circ(N-n_i^\circ)}{N}-\dfrac{1}{N}\dfrac{\begin{vmatrix} N & n_i^\circ \\ n_{hmg}^\circ & n_{ihmg}^\circ \end{vmatrix}\cdot\begin{vmatrix} N & n_l^\circ \\ n_{ijk}^\circ & n_{ijkl}^\circ \end{vmatrix}}{\begin{vmatrix} N & n_l^\circ \\ n_{hmg}^\circ & n_{hmg}^\circ \end{vmatrix}}\,, \\[2ex]
E(n_i-n_i^\circ)(n_{ij}-n_{ij}^\circ)=\dfrac{n_{ij}^\circ(N-n_i^\circ)}{N}-\dfrac{1}{N}\dfrac{\begin{vmatrix} N & n_i^\circ \\ n_{hk}^\circ & n_{ikh} \end{vmatrix}\cdot\begin{vmatrix} N & n_l^\circ \\ n_{hk}^\circ & n_{hkl}^\circ \end{vmatrix}}{\begin{vmatrix} N & n_l^\circ \\ n_{hk}^\circ & n_{hkl}^\circ \end{vmatrix}}\,,
\end{cases}
$$

where the different letters of sub-indices take the different numbers. On the other hand, each of (12) can obviously be represented by $n_{j_1}^\circ,\cdots,n_{j_sk_s}^\circ$ and N. Thus, from (14), (16) and (12), if for simplicity we put $j_1=1$, $j_2=2$, $k_1=3$, $k_2=4$, we have

$$
(17)\quad E(\hat{N}-N)^2\sim\frac{1}{ND_0^2}\left[\sum_{i=1}^{4}\left\{n_1^\circ(N-n_1^\circ)-\frac{\begin{vmatrix} N & n_h^\circ \\ n_1^\circ & n_{1h}^\circ \end{vmatrix}\cdot\begin{vmatrix} N & n_g^\circ \\ n_1^\circ & n_{1g}^\circ \end{vmatrix}}{\begin{vmatrix} N & n_h^\circ \\ n_g^\circ & n_{hg}^\circ \end{vmatrix}}\right\}D_i^2\right]
$$

$$
+\frac{1}{ND_0^4}\left[\sum_{\substack{i=1,2\\ j=3,4}}\left\{n_{ij}^\circ(N-n_{ij}^\circ)-\frac{\begin{vmatrix} N & n_{ij}^\circ \\ n_{kl}^\circ & n_{ijkl}^\circ \end{vmatrix}\cdot\begin{vmatrix} N & n_{ij}^\circ \\ n_{hm}^\circ & n_{ijhm} \end{vmatrix}}{\begin{vmatrix} N & n_{kl}^\circ \\ n_{hm}^\circ & n_{hmkl}^\circ \end{vmatrix}}\right\}D_i^2D_j^2\right]
$$

$$+\frac{1}{ND_0^4}\left[\sum_{\substack{i=1,2\\k,j=3,4}}\left\{n_{ijk}^{\circ}(N-n_i^{\circ})-\frac{\begin{vmatrix}N & n_i^{\circ}\\ n_{hmg}^{\circ} & n_{ihmg}\end{vmatrix}\cdot\begin{vmatrix}N & n_l^{\circ}\\ n_{ijk}^{\circ} & n_{ijkl}^{\circ}\end{vmatrix}}{\begin{vmatrix}N & n_l^{\circ}\\ n_{hmg}^{\circ} & n_{lhmg}^{\circ}\end{vmatrix}}\right\}D_i^2D_jD_k\right]$$

$$+\frac{2}{ND_0^3}\left[\sum_{\substack{(i,j)=(1,3),(1,4)\\(2,3),(2,4)\\(3,1),(4,1)\\(3,2),(4,2)}}\left\{n_{ij}^{\circ}(N-n_i^{\circ})-\frac{\begin{vmatrix}N & n_i^{\circ}\\ n_{kh}^{\circ} & n_{ikh}^{\circ}\end{vmatrix}\cdot\begin{vmatrix}N & n_l^{\circ}\\ n_{ij}^{\circ} & n_{ijl}^{\circ}\end{vmatrix}}{\begin{vmatrix}N & n_l^{\circ}\\ n_{hk}^{\circ} & n_{hkl}^{\circ}\end{vmatrix}}\right\}D_i^2D_j\right],$$

where $D_0=\begin{vmatrix}n_{13} & n_{14}\\ n_{23} & n_{24}\end{vmatrix}$, $D_1=\begin{vmatrix}n_3 & n_4\\ n_{23} & n_{24}\end{vmatrix}$, $D_2=-\begin{vmatrix}n_3 & n_4\\ n_{13} & n_{14}\end{vmatrix}$, $D_3=\begin{vmatrix}n_1 & n_{14}\\ n_2 & n_{24}\end{vmatrix}$ and $D_4=-\begin{vmatrix}n_1 & n_{13}\\ n_2 & n_{23}\end{vmatrix}$ and the different letters of sub-indices take the different numbers in one and the same { }-bracket.

Therefore, we can get an estimate of the asymptotic variance of $\hat{N}$ by substituting the observed values n_j's, n_{jk}'s and $\hat{N}$ into n_j°'s, n_{jk}°'s and N, respectively. It should be noticed that the number of repetitions of sampling required for estimation of the asymptotic variance of $\hat{N}$ is at least seven while for estimation of $\hat{N}$ the required number is four. If we can estimate the asymptotic variance of $\hat{N}$ as in the above case, we may solve to some extent the problem of which combination of $(j_1, \cdots, j_s; k_1, \cdots, k_s)$ we should use for the estimation of N.

(II) It may be useful to consider the fluctuations of the asymptotic variance subject to that of parameters. However, since our model contains a very large number of parameters, general analysis will be complicated. Now, we shall consider the following simple model. Let us now assume $p_{j_l}^{(r)}=p_{k_l}^{(r)}=a$ for $r\neq l$ and $p_{j_r}^{(r)}=p_{k_r}^{(r)}=c$, where $r=1, \cdots, s$ and $l=1, \cdots, s$. Then we have

(18) $$\text{asymptotic variance of } \hat{N}\sim N\frac{\{(s-1)a(1-a)+c(1-c)\}^2}{\{(s-1)a+c\}^4}.$$

Let

(19) $$\varphi(a, c)=\frac{\{(s-1)a(1-a)+c(1-c)\}^2}{\{(s-1)a+c\}^4}.$$

We consider $\varphi(a, c)$ for fixed c. Differentiating $\varphi(a, c)$ with respect to a, we have

(20) $$\frac{\partial\varphi(a, c)}{\partial a}=\frac{(s-1)\{(s-1)+2c\}\cdot(a-a_1(c))(a-a_2(c))(a-a_3(c))}{\{(s-1)a+c\}^5},$$

where $$a_1(c)=\frac{(s-1)+\sqrt{(s-1)^2+(s-1)c(1-c)}}{2(s-1)},$$

$$a_2(c)=\frac{(s-1)-\sqrt{(s-1)^2+(s-1)(1-c)}}{2(s-1)}$$

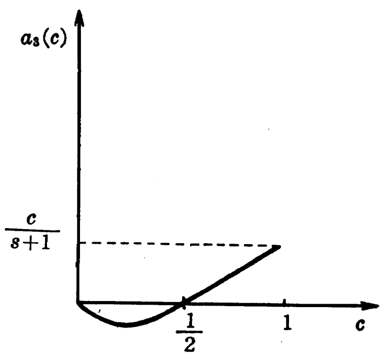
and
$$a_3(c)=\frac{c(2c-1)}{(s-1)+2c}\ .$$

It is easily shown that $a_1(c)\geqq 1$, $a_2(c)\leqq 0$, and also $a_3(c)<0$ when $0<c<1/2$, $a_3(c)=0$ when $c=0$ or $1/2$, $a_3(c)>0$ when $1/2<c\leqq 1$ [cf. Fig. 1]. Hence, if $0\leqq c\leqq 1/2$, $\varphi(a, c)$ is monotone decreasing with respect to a, and if $1/2<c\leqq 1$, it is increasing when $0<a<a_3(c)$ and decreasing when $a_3(c)<a<1$, so that $\varphi(a, c)$ takes the maximum value $\{s-(2c-1)^2\}^2/16s^2c^2$ at $a=a_3(c)$. As an example, we show the case $s=3$ in Fig. 2.

For reference, $\varphi(1, 0)=\varphi(1, 1)=0$, $\varphi\{1, (3/4)\}=0.0006$,
$\varphi\{1, (1/2)\}=0.0016$, $\varphi\{1, (1/4)\}=0.0014$.

It should be mentioned that $\varphi(c, c)$ cannot be defined if $s=1$. If $a=c$, s must be 1. However, the value $\varphi(c, c)$ computed from (19) is just equal to $(1/s^2)\cdot$[asymptotic variance of the estimate $\hat{N}$ when we take $s=1$].

In the case $s=1$, we have $\hat{N}=(n_i\cdot n_j)/n_{ij}$. This is the Peterson estimate. The aspects of $\varphi(a, c)$ as a function of c for fixed a are similar to those of $\varphi(a, c)$ as a function of a for fixed c. Even the simplest example mentioned above shows that it does not always give smaller variance of our estimate to increase the probabilities of being caught without any consideration of the situation.

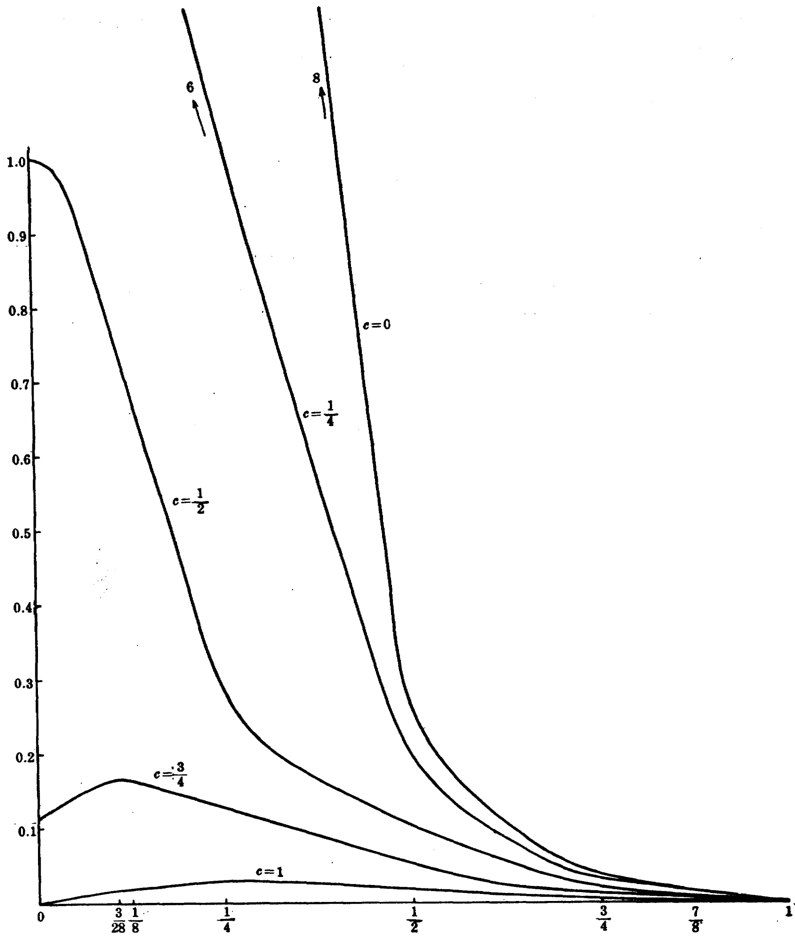


[Fig. 1]

A numerical example of asymptotic bias.

For simplicity, let us put $s=2$ and $N^{(1)}=N^{(2)}$. In this case, we have as an asymptotic bias of $\hat{N}$

$$\begin{aligned}
(21)\quad &\frac{1}{2\Theta^2}\Big[2\{p^{(1)}_{j_1}p^{(1)}_{k_1}(1-p^{(1)}_{j_1}p^{(1)}_{k_1})+p^{(2)}_{j_1}p^{(2)}_{k_1}(1-p^{(2)}_{j_1}p^{(2)}_{k_1})\}\\
&\qquad\times(p^{(1)}_{j_2}-p^{(2)}_{j_2})(p^{(1)}_{k_2}-p^{(2)}_{k_2})(p^{(1)}_{j_2}p^{(1)}_{k_2}+p^{(2)}_{j_2}p^{(2)}_{k_2})+\cdots\\
&\quad+\{p^{(1)}_{j_1}p^{(1)}_{k_1}p^{(1)}_{k_2}(1-p^{(1)}_{j_1})+p^{(2)}_{j_1}p^{(2)}_{k_1}p^{(2)}_{k_2}(1-p^{(2)}_{j_1})\}\\
&\qquad\times\{(p^{(1)}_{j_2}-p^{(2)}_{j_2})\{(p^{(1)}_{j_2}-p^{(2)}_{j_2})(p^{(1)}_{k_1}-p^{(2)}_{k_1})(p^{(1)}_{k_2}-p^{(2)}_{k_2})\\
&\qquad-(p^{(1)}_{j_2}+p^{(2)}_{j_2})(p^{(1)}_{k_1}p^{(1)}_{k_2}-p^{(2)}_{k_1}p^{(2)}_{k_2})\}\}+\cdots\\
&\quad+\{p^{(1)}_{j_1}p^{(1)}_{k_1}(1-p^{(1)}_{j_1})+p^{(2)}_{j_1}p^{(2)}_{k_1}(1-p^{(2)}_{j_1})\}\\
&\qquad\times(p^{(1)}_{j_2}-p^{(2)}_{j_2})(p^{(1)}_{k_1}p^{(2)}_{k_2}-p^{(1)}_{k_2}p^{(2)}_{k_1})(p^{(1)}_{j_2}p^{(1)}_{k_2}+p^{(2)}_{j_2}p^{(2)}_{k_2})+\cdots\Big]
\end{aligned}$$



[Fig. 2]

where $\Theta = p_{j_1}^{(1)} p_{j_2}^{(2)} p_{k_1}^{(1)} p_{k_2}^{(2)} + p_{j_1}^{(2)} p_{j_2}^{(1)} p_{k_1}^{(2)} p_{k_2}^{(1)} - p_{j_1}^{(1)} p_{j_2}^{(2)} p_{k_1}^{(2)} p_{k_2}^{(1)} - p_{j_1}^{(2)} p_{j_2}^{(1)} p_{k_1}^{(1)} p_{k_2}^{(2)}$.

From (21), we can roughly say that if we multiply each $p_i^{(r)}$ by constant c, then the asymptotic bias will be multiplied by about $1/c$.

For example, if we put

$$p_{j_1}^{(1)} = 0.1, \quad p_{j_2}^{(1)} = 0.04, \quad p_{k_1}^{(1)} = 0.05, \quad p_{k_2}^{(1)} = 0.125$$
$$p_{j_1}^{(2)} = 0.05, \quad p_{j_2}^{(2)} = 0.1, \quad p_{k_1}^{(2)} = 0.125, \quad p_{k_2}^{(2)} = 0.05$$

then the asymptotic bias of $\hat{N}$ is 18.7.

If

$$p_{j_1}^{(1)} = 0.1, \quad p_{j_2}^{(1)} = 0.05, \quad p_{k_1}^{(1)} = 0.125, \quad p_{k_2}^{(1)} = 0.04,$$
$$p_{j_1}^{(2)} = 0.05, \quad p_{j_2}^{(2)} = 0.125, \quad p_{k_1}^{(2)} = 0.05, \quad p_{k_2}^{(2)} = 0.1,$$

then the asymptotic bias of $\hat{N}$ is 17.5.

In such a case as the above example, for using our estimate it may be desirable that N is greater than about a thousand.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] D. G. Chapman and C. O. Junge, Jr., "The estimation of the size of a stratified animal population", *Ann. Math. Stat.*, Vol. 27 (1956), pp. 375-389.

[2] S. A. Stouffer and others, *Measurement and prediction*, Studies in Psychology in World War II, Vol. 4, Princeton University Press, 1950, pp. 362-472.