Kontakt/Contact

Digizeitschriften e.V.
SUB Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen

info@digizeitschriften.de

ČASOPIS PRO PĚSTOVÁNÍ MATEMATIKY
Vydává Matematický ústav ČSAV, Praha
SVAZEK 84 * PRAHA, 29. X. 1959 * ČÍSLO 4

OPTIMUM STRATEGY AND OTHER PROBLEMS IN PROBABILITY SAMPLING

JAROSLAV HÁJEK, Praha

(Received August 4, 1958) DT: 519.271.3

The statistician's strategy in probability sampling consists in the choice of the sampling design (plan) and of the estimation method (procedure). A strategy may be called optimum if it solves the conflict between cost and accuracy in the best way. In this paper Bayes approach is accepted, i. e. the accuracy is measured by the expected variance with respect to a certain a priori distribution of ascertained values. A general solution of the problem is derived for a rather wide class of admissible sampling designs, estimators, cost functions, and for the following two most important assumptions concerning the a priori distribution: (a) The ascertained values are realizations of non-correlated random variables. (b) The ascertained values are realizations of a random sequence with stationary convex correlation function and stationary coefficients of variations.

In the introductory sections the conceptions of "sample" and "estimate" are defined, and a general formula for the variance and estimated variance of linear estimates is derived; furthermore, a method of improving estimates based on sufficient statistics is presented, and two sampling designs with varying probabilities are discussed.

A. INTRODUCTORY SECTIONS (1—5)

1. Definitions

Let us have a population S consisting of N elements of arbitrary nature, so that they may be represented by integers $1, \ldots, N$, $S = \{1, \ldots, N\}$. From S we select a subset s in such a way that any subset $s \subset S$ possesses a probability $\mathsf{P}(s)$ of being selected. The selected subset s will be called the *sample*.

Let us denote by $y_1, \ldots, y_N$ values of a certain variable associated with elements $1, \ldots, N$, respectively. We try to estimate the total

$$Y = \sum_{i=1}^{N} y_i \tag{1.1}$$

by an estimate $\hat{Y}$ having the form

$$\hat{Y} = \sum_{i \in s} y_i w_i(s) \,, \tag{1.2}$$

where $w_i(s)$ are arbitrary weights, $i \in s$, $s \subset S$, and $\sum_{i \in s}$ extends over the elements i included in the sample s. The estimate (1.2) will be called *a linear estimate.*

The necessary and sufficient condition for the estimate (1.2) to be *unbiassed* is, obviously, that

$$\sum_{s \ni i} w_i(s) \, \mathsf{P}(s) = 1 \,, \quad i = 1, \ldots, N \,, \tag{1.3}$$

where the sum $\sum_{s \ni i}$ extends over all samples containing the element i.

The probability of selecting a sample s which contains the element i, say π_i, equals

$$\pi_i = \sum_{s \ni i} \mathsf{P}(s) \,. \tag{1.4}$$

Similarly, the probability of selecting a sample which contains both elements i and j, say π_{ij}, equals

$$\pi_{ij} = \sum_{\substack{s \ni i \\ s \ni j}} \mathsf{P}(s) \,, \quad i, j = 1, \ldots, N \,. \tag{1.5}$$

If we put $w_i(s) = \dfrac{1}{\pi_i}$, $i \in s$, $s \subset S$, we get the *simple linear estimate*:

$$\hat{Y} = \sum_{i \in s} \frac{y_i}{\pi_i} \,. \tag{1.6}$$

It is easily seen that the simple linear estimate is unbiassed, i. e. that (1.3) holds.

We have defined the "sample" as a *subset* of the population. However, one could think of a more detailed specification of the "sample". For example, it is possible to define it as *an ordered subset* of the population, or, still more distinctively, as *a sequence*, all members of which belong to the population. For brevity, let us use the following symbols:

$$\text{subset } s \,, \quad \text{ordered subset } s' \,, \quad \text{sequence } s'' \,. \tag{1.7}$$

The sample s only tells us what elements have been selected, while s' and s'' comprise further information. The sample s' fully describes the element-by-element sampling without replacement, and s'' fully describes the element-by-element sampling with replacement.

If we delete in the sequence s'' all members which appear in some of the preceding places we get on ordered set s', $s' = s'(s'')$. If we dispense with the ordering in s', we get a set s, $s = s(s')$. Symbolically,

$$s = s(s') = s(s'(s'')) \,, \tag{1.8}$$

i. e. s' is an abstract function of s'' and s is an abstract function of s', and, naturally, of s'' too.

The collection (1.7) of possible definitions of the sample is naturally not exhaustive. For example, we shall use, in section 3, the sample s^* which tells how many times each element has been included in the sample. Clearly, if we dispense with the ordering in s'', we get s^*. Thus it holds that

$$s^* = s^*(s'') \,, \quad s = s(s^*) \,. \tag{1.9}$$

If we deal with double sampling, we may define the sample as a couple of subsets (s_1, s), where s_1 is the "larger" sample and s is the ultimately selected sample. As we can see, the possible definitions of the sample might be continued as long as we wished.

Now, let us define the *observation* and the *estimate* in probability sampling. The observation, say (s, y), (s', y), (s'', y), etc., involves a knowledge of the sample s, s', s'', etc. and of the values y_i associated with elements in s, s', s'', etc., respectively. The estimate t is any function of (s, y), (s', y), (s'', y), etc.:

$$t = t(s, y) \,, \quad t = t(s', y) \,, \quad t = t(s'', y) \,, \quad \text{etc.} \tag{1.10}$$

2. Estimating sampling error of linear estimates

We begin with the **Definition 2.1.** Any estimate $t_i(s)$, which equals 0 if s does not contain the element i, will be called an (i)-estimate; any estimate $t_{ij}(s)$, which equals 0 if s does not contain the element i or j (or both), will be called an (i, j)-estimate.

If we complete the definition of $w_i(s)$, as function of s, putting for s not containing the element i

$$w_i(s) = 0 \,, \quad s \text{ non} \ni i \,, \tag{2.1}$$

then $w_i(s)$ becomes an (i)-estimate of 1 and (1.2) may be rewritten in the following form:

$$\hat{Y} = \sum_{i=1}^{N} y_i w_i(s) \,. \tag{2.2}$$

Theorem 2.1. *If for the values $z_1, \ldots, z_N$ the equation*

$$\sum_{i=1}^{N} z_i w_i(s) = \sum_{i=1}^{N} z_i \quad (\text{i. e. } \hat{Z} = Z) \tag{2.3}$$

holds with probability 1, *then the mean-square error of the estimate* (2.2) *equals*

$$\mathsf{M}(\hat{Y} - Y)^2 = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 z_i z_j (\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j) \quad (2.4)$$

where $\mathsf{M}t = \sum t(s)\,P(s)$ *denotes the mean value over all possible* s. *In the unbiassed case* ($\mathsf{M}w_i = 1,\ i = 1, \ldots, N$) *we have*

$$\mathsf{M}(\hat{Y} - Y)^2 = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 z_i z_j (1 - \mathsf{M}w_i w_j)\,. \quad (2.5)$$

The sum of the weights standing in (2.4) *at the terms* $\frac{1}{2}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2$, $i \neq j$, *equals*

$$\sum_{i \neq j}\sum z_i z_j (\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j) = \sum_{i=1}^{N} z_i^2 \mathsf{M}(w_i - 1)^2\,. \quad (2.6)$$

Proof is based on the following easy identities:

$$\mathsf{M}(\hat{Y} - Y)^2 = \mathsf{M}[\sum y_i(w_i - 1)]^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \mathsf{M}(w_i - 1)(w_j - 1) =$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{N}\frac{y_i}{z_i}\frac{y_j}{z_j} z_i z_j \mathsf{M}(w_i - 1)(w_j - 1) =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\left[-\frac{1}{2}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 + \frac{1}{2}\left(\frac{y_i}{z_i}\right)^2 + \frac{1}{2}\left(\frac{y_j}{z_j}\right)^2\right] z_i z_j \mathsf{M}(w_i - 1)(w_j - 1) =$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j] +$$

$$+ \sum_{i=1}^{N}\left(\frac{y_i}{z_i}\right)^2 \mathsf{M}\left[z_i(w_i - 1)\sum_{j=1}^{N} z_j(w_j - 1)\right].$$

The last term, however, vanishes, since, according to (2.3), $\sum_{j=1}^{N} z_i(w_i - 1) = 0$ with probability 1. The identity (2.6) is implied by the same fact: We get

$$0 = -\mathsf{M}(\sum z_i(w_i - 1))^2 = \sum_{i \neq j}\sum z_i z_j [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j] -$$

$$- \sum_{i=1}^{N} z_i^2 \mathsf{M}(w_i - 1)^2\,, \quad \text{which is equivalent to (2.6).}$$

Formula (2.5) for simple estimate was given in [5].

As regards the estimated mean-square error, we shall use the following

Theorem 2.2. *On replacing* 1, $\mathsf{M}w_i$, $\mathsf{M}w_i w_j$ *in* (2.4) *or* (2.5) *by any their* (i, j)-*estimates, we get an estimated means square error of the estimate* (2.2). *Ones of the possible unbiassed* (i, j)-*estimates of* 1, $\mathsf{M}w_i$ *and* $\mathsf{M}w_i w_j$ *are the following*:

Estimate of $1 =$

$$= 1 + \frac{1}{n-1} \sum_{k \text{ non} \in s} \frac{\mathsf{P}(s - \{j\} + \{k\})}{\mathsf{P}(s)} + \frac{1}{n-1} \sum_{k \text{ non} \in s} \frac{\mathsf{P}(s - \{i\} + \{k\})}{\mathsf{P}(s)} +$$

$$+ \frac{1}{n(n-1)} \sum_{k \text{ non} \in s} \sum_{\substack{h \text{ non} \in s \\ k \neq h}} \frac{\mathsf{P}(s - \{i, j\} + \{k, h\})}{\mathsf{P}(s)}, \quad i, j \in s. \tag{2.7}$$

Estimate of $\mathsf{M}w_i =$

$$= w_i(s) + \frac{1}{n-1} \sum_{k \text{ non} \in s} w_i(s - \{j\} + \{k\}) \frac{\mathsf{P}(s - \{j\} + \{k\})}{\mathsf{P}(s)}, \quad i, j \in s. \tag{2.8}$$

$$\textit{Estimate of } \mathsf{M}w_i w_j = w_i(s)\, w_j(s), \quad i, j \in s, \tag{2.9}$$

where $s - \{j\} + \{k\}$ *and* $s - \{i, j\} + \{k, h\}$ *denotes the subset got from the subset* s *by replacing the element* j *by the element* k *or the elements* $\{i, j\}$ *by the elements* $\{k, h\}$, *respectively, and* $\sum_{k \text{ non} \in s}$ *denotes the sumation over all elements* k *not contained in* s.

Proof. The first assertion of the Theorem only tells that on replacing, 1, $\mathsf{M}w_i$ and $\mathsf{M}w_i w_j$ in (2.4) or (2.5) by any their (i, j)-estimates, we get a function of (s, y), i. e. an estimate (see Section 1).

Now, in view of (2.1), we have

$$\sum_{s \ni i,j} w_i(s)\, w_j(s)\, \mathsf{P}(s) = \sum_{s \subset S} w_i(s)\, w_j(s)\, \mathsf{P}(s) = \mathsf{M}w_i w_j,$$

where $\sum_{s \ni i,j}$ denotes the sum extended over the subsets s containing the both elements i and j, by which it is shown that $w_i w_j$ is an unbiassed (i, j)-estimate of $\mathsf{M}w_i w_j$.

In order to prove the assertion concerning (2.8), let us note, that any subset z containing the element i but not containing the element j may be converted in a subset s containing both elements i and j by omitting an element, say k, different from i, and by replacing it by j, and that this may be done in $n - 1$ ways. Consequently,

$$\sum_{s \ni i,j} \left[w_i(s) + \frac{1}{n-1} \sum_{k \text{ non} \in s} w_i(s - \{j\} + \{k\}) \frac{\mathsf{P}(s - \{j\} + \{k\})}{\mathsf{P}(s)} \right] \mathsf{P}(s) =$$

$$= \sum_{s \ni i,j} w_i(s)\, \mathsf{P}(s) + \sum_{s \ni i,j} \frac{1}{n-1} \sum_{k \text{ non} \in s} w_i(s - \{j\} + \{k\})\, (\mathsf{P}(s - \{j\} + \{k\}) =$$

$$= \sum_{s \ni i,j} w_i(s)\, \mathsf{P}(s) + \sum_{\substack{s \ni i \\ s \text{ non} \ni j}} \frac{1}{n-1} (n-1)\, w_i(z)\, \mathsf{P}(z) = \sum_{s \ni i} w_i(s)\, \mathsf{P}(s) = \mathsf{M}w_i.$$

Before proceeding to (2.7), let us note, that any subset z not containing either i or j may be converted in a subset s, which does contain both elements i and j, by omitting any two of his elements, say k and h, and by replacing them by i and j, and that this may be done in $n(n-1)$ ways. Consequently,

$$\sum_{s\ni i,j}\Bigg[1+\frac{1}{n-1}\sum_{k\,\mathrm{non}\in s}\frac{\mathsf{P}(s-\{j\}+\{k\})}{\mathsf{P}(s)}+\frac{1}{n-1}\sum_{k\,\mathrm{non}\in s}\frac{\mathsf{P}(s-\{i\}+\{k\})}{\mathsf{P}(s)}+$$

$$+\frac{1}{n(n-1)}\sum_{\substack{k\,\mathrm{non}\in s,\ h\,\mathrm{non}\in s\\ k\neq h}}\frac{\mathsf{P}(s-\{i,j\}+\{k,h\})}{\mathsf{P}(s)}\Bigg]\mathsf{P}(s)=$$

$$=\sum_{s\ni i,j}\mathsf{P}(s)+\sum_{\substack{s\ni i\\ s\,\mathrm{non}\ni j}}\frac{1}{n-1}(n-1)\,\mathsf{P}(s)+\sum_{\substack{s\ni j\\ s\,\mathrm{non}\ni i}}\frac{1}{n-1}(n-1)\,\mathsf{P}(s)+$$

$$+\sum_{\substack{s\,\mathrm{non}\ni i\\ s\,\mathrm{non}\ni j}}\frac{1}{n(n-1)}n(n-1)\,\mathsf{P}(s)=\sum_{s\subset S}\mathsf{P}(s)=1\,,$$

which accomplished the proof.

Remark 2.1. The (i, j)-estimates shown in the theorem 2.2 are of use in situations where $w_i(s)$ depends of s $(s \ni i)$ not much, and they will be used in section 4. Their scope might be widened by this device: We may, for each (i, j) separately, select several samples containing both elements i, j, say $s_1, \ldots, s_k$, and then replace the (i, j)-estimate, say $u_{ij}(s)$, by the arithmetic mean

$$u^*_{ij}(s_1, \ldots, s_k) = \frac{1}{k}\sum_{\nu=1}^{k} u_{ij}(s_\nu)\,. \tag{2.10}$$

If the probabilities π_{ij} of including both elements i and j and mean values $\mathsf{M}w_iw_j$ and $\mathsf{M}w_i$ are simple, we may use the following unbiassed (i, j)-estimates:

$$\textit{Estimate of}\,[\mathsf{M}w_i+\mathsf{M}w_j-1-\mathsf{M}w_iw_j]=\frac{\mathsf{M}w_i+\mathsf{M}w_j-1-\mathsf{M}w_iw_j}{\pi_{ij}}\ \textit{if}\ \{i,j\}\subset s,$$
$$=0\ \textit{in the other case}\,, \tag{2.11}$$

or, in the unbiassed case

$$\textit{Estimate of}\,[1-\mathsf{M}w_iw_j]=\frac{1-\mathsf{M}w_iw_j}{\pi_{ij}}\quad \textit{if}\quad \{i,j\}\subset s\,,$$
$$=0\quad \textit{in the other case}\,. \tag{2.12}$$

Remark 2.2. The device of ratio estimation may be useful also in estimating the mean square error. For example, in view of (2.6), we may hope that the estimated mean square error

$$\mathsf{m}(\hat{Y}-Y)^2=\frac{\frac{1}{2}\sum_{i\in s}\sum_{j\in s}\left(\frac{y_i}{z_i}-\frac{y_j}{z_j}\right)^2 z_iz_j[\mathsf{M}w_i+\mathsf{M}w_j-1-\mathsf{M}w_iw_j]_{i,j}}{\sum_{i\in s}\sum_{j\in s}z_iz_j[\mathsf{M}w_i+\mathsf{M}w_j-1-\mathsf{M}w_iw_j]_{i,j}}\sum_{i=1}^{N}z_i^2\mathsf{M}(w_i-1)^2, \tag{2.13}$$

where $[.]_{i,j}$ denotes a proper (i, j)-estimate of $[.]$, will be better than the estimated mean square error

$$\mathsf{m}(\hat{Y} - Y)^2 = \frac{1}{2} \sum_{i \in s} \sum_{j \in s} \left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 z_i z_j [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j]_{i,j} \,. \quad (2.14)$$

Remark 2.3. If the estimated mean square errors (2.13) and (2.14) are too laborious, we may select randomly with equal probabilities a subset M of k couples $\{i, j\} \subset s$ without common elements, when possible, and then replace (2.13) and (2.14) by

$$\mathsf{m}(\hat{Y} - Y)^2 = \frac{\frac{1}{2} \sum\sum_{\{i,j\} \in M} \left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 z_i z_j [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j]_{i,j}}{\sum\sum_{\{i,j\} \in M} z_i z_j [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j]_{i,j}} \sum_{i=1}^{N} z_i^2 \mathsf{M}(w_i - 1)^2 \quad (2.15)$$

and by

$$\mathsf{m}(\hat{Y} - Y)^2 = \frac{n(n-1)}{2k} \sum\sum_{\{i,j\} \in M} \left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 z_i z_j [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j]_{i,j}. \quad (2.16)$$

Remark 2.4. If not directly the values y_i but unbiassed estimates of them are at hand, say $\hat{y}_i$, e. g. in the case of subsampling, and if estimates $\hat{y}_i$ are mutually independent, then the formula (1.2) is changed into

$$\hat{\hat{Y}} = \sum_{i \in s} \hat{y}_i w_i(s) \,,$$

and the formula (2.3) into

$$\mathsf{M}(\hat{\hat{Y}} - Y)^2 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathsf{M}_2 \left[\left(\frac{\hat{y}_i}{z_i} - \frac{\hat{y}_j}{z_j}\right)^2\right] z_{ij} [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j] + \\ + \sum_{i=1}^{N} \mathsf{D}_2(\hat{y}_i) \,, \quad (2.17)$$

where $\mathsf{M}_2(.)$ and $\mathsf{D}_2(.)$ denote the mean value and the variance over the estimates $\hat{y}_i$ (for example, over the subsampling).

It means that the estimated mean square error has, generally, the form

$$\mathsf{m}(\hat{\hat{Y}} - Y)^2 = \frac{1}{2} \sum_{i \in s} \sum_{j \in s} \left(\frac{\hat{y}_i}{z_i} - \frac{\hat{y}_j}{z_j}\right)^2 [\mathsf{M}w_i + \mathsf{M}w_j - 1 - \mathsf{M}w_i w_j]_{i,j} + \hat{D}_2 \,, \quad (2.18)$$

where $\hat{D}_2$ is an estimate of $\sum_{i=1}^{N} \mathsf{D}_2(\hat{y}_i)$.

3. Improving estimates connected with element-by-element sampling with and without replacement

We have denoted by (s'', y) the observation consisting in a knowledge of s'' and of values y_i associated with elements appearing in s'', and a similar meaning has been ascribed to (s, y) and (s', y). It is easily seen that the knowledge of (s'', y) enables us to establish (z'', y) for all z'' such that

$$s(z'') = s(s'') , \tag{3.1}$$

or, similarly, knowledge of (s', y) enables us to establish (z', y) for all z' such that $s(z') = s(s')$. For example, if we know that on the sequence of elements $s'' = (2, 4, 3, 2, 4)$ there were ascertained values (13, 18, 15, 13, 18), respectively, we may infer that on the sequence $z'' = (3, 3, 2, 4)$ would be ascertained values (15, 15, 13, 18), respectively, because we simply know that $y_3 = 15$, $y_2 = 13$ and $y_4 = 18$. Conversely, knowing (z'', y) for any z'' such that (3.1) holds, we can establish (s'', y).

This trivial fact has an interesting application: Having an arbitrary estimate

$$t'' = t''(s'', y) , \tag{3.2}$$

we may replace it by the estimate

$$t = t(s, y) = \frac{\sum\limits_{[s]} t''(z'', y)\, \mathsf{P}(z'')}{\sum\limits_{[s]} \mathsf{P}(z'')} , \tag{3.3}$$

where the sum $\sum\limits_{[s]}$ extends over all z'' such that $s(z'') = s = s(s'')$. Indeed, if we know (s'', y), we can evaluate $t''(z'', y)$ for any z'' such that (3.1) holds, and hence we can evaluate (3.3).

A brief inspection of the equation (3.3) shows that t is a conditional mean value of t'' with respect to (s, y). Consequently

$$\mathsf{M}t = \mathsf{M}t'' \tag{3.4}$$

$$\mathsf{D}t = \mathsf{D}t'' - \mathsf{M}(t - t'')^2 . \tag{3.5}$$

In other words, t is at worst as equally good an estimate as t''. This means that the subset-definition of a sample is fully satisfactory, since any good estimate is a function of (s, y) only. (Of course, it may happen that t'' and the estimated variance of t'' are more easy to compute than t and the estimated variance of t.)

A similar conclusion may be drawn concerning the estimates $t' = t'(s', y)$, etc.

Now, let us show how the method works:

Example 3.1. Let us perform independent samples of one element with probabilities $\alpha_1, \ldots, \alpha_N$, $\sum_{i=1}^{N} \alpha_i = 1$, until n distinct elements have been selected. Let us choose any unbiassed estimator of the total $Y = \sum_{i=1}^{N} y_i$, for example,

$$\hat{Y} = \frac{y_{i_1}}{\alpha_{i_1}} \tag{3.6}$$

where i_1 denotes the element selected as the first. Let us take a conditional mean value of $\hat{Y}$ with respect to (s, y), where s is the set of n distinct elements included in the sample. Using the formula (3.3) we see that

$$\tilde{Y} = \mathsf{M}(\hat{Y}|(s, y)) = \sum_{i \in s} \frac{y_i}{\alpha_i} \mathsf{P}_i(s) , \tag{3.7}$$

where $\mathsf{P}_i(s)$ is the conditional probability that the element i will be included as the first in the sample under the condition that the distinct elements selected consist the set s. Probabilities $\mathsf{P}_i(s)$ are not easy to compute, except when the sampling is uniform, i. e. $\alpha_1 = \ldots = \alpha_N = \frac{1}{N}$, or when $n = 2$.

In the case when the sampling is uniform, we clearly get $\mathsf{P}_i(s) = \frac{1}{n}$, so that

$$\tilde{Y} = \frac{N}{n} \sum_{i \in s} y_i .$$

Now, let us consider the case $n = 2$, denoting the two distinct elements included in the sample by i and j. The probability that the element i was selected first and the element j second, equals

$$P_{ij} = \mathsf{P}\{s' = [i, j]\} = \frac{\alpha_i \alpha_j}{1 - \alpha_i} . \tag{3.8}$$

Similarly, the probability that the element j was selected first and the element i second, equals

$$P_{ji} = \mathsf{P}\{s' = [j, i]\} = \frac{\alpha_i \alpha_j}{1 - \alpha_j} . \tag{3.9}$$

This means that the conditional probability that the element i has been selected first equals

$$\mathsf{P}_i(s) = \frac{1 - \alpha_j}{2 - \alpha_i - \alpha_j} . \tag{3.10}$$

When substituting into (3.7) we get

$$\tilde{Y} = \frac{(1 - \alpha_j) \frac{y_i}{\alpha_i} + (1 - \alpha_i) \frac{y_j}{\alpha_j}}{2 - \alpha_i - \alpha_j} . \tag{3.11}$$

As regards the variance of $\tilde{Y}$, we may use the formula (2.5) since (2.3) is clearly satisfied for $z_i = \alpha_i$. It means that

$$\mathsf{M}(\tilde{Y} - Y)^2 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right) \alpha_i \alpha_j [1 - \mathsf{M} w_i w_j] \,. \tag{3.12}$$

In order to find $\mathsf{M}w_i w_j$ let us first compute π_{ij}:

$$\pi_{ij} = \mathsf{P}\{i, j \in s\} = \mathsf{P}\{s = \{i, j\}\} = P_{ij} + P_{ji} =$$
$$= \frac{\alpha_i \alpha_j}{1 - \alpha_i} + \frac{\alpha_j \alpha_i}{1 - \alpha_j} = \frac{\alpha_i \alpha_j (2 - \alpha_i - \alpha_j)}{(1 - \alpha_i)(1 - \alpha_j)} \,. \tag{3.13}$$

Now, in view of (3.11),

$$w_i(s) = \frac{1 - \alpha_i}{2 - \alpha_i - \alpha_j} \frac{1}{\alpha_i} \quad \text{for} \quad s = \{i, j\}$$

and, consequently,

$$\mathsf{M} w_i w_j = \sum_{s \ni i,j} w_i(s)\, w_j(s)\, \mathsf{P}(s) = w_i w_j \pi_{ij} =$$
$$= \frac{\alpha_i \alpha_j (2 - \alpha_i - \alpha_j)}{(1 - \alpha_i)(1 - \alpha_j)} \frac{(1 - \alpha_j)(1 - \alpha_i)}{(2 - \alpha_i - \alpha_j)^2} \frac{1}{\alpha_i \alpha_j} = \frac{1}{2 - \alpha_i - \alpha_j} \,.$$

On substituting (3.13) into (3.12) we get that

$$\mathsf{M}(\tilde{Y} - Y)^2 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right)^2 \alpha_i \alpha_j \frac{1 - \alpha_i - \alpha_j}{2 - \alpha_i - \alpha_j} \,. \tag{3.14}$$

Finally, we may use the (i, j)-estimates (2.12), where π_{ij} are given by (3.13), and get the following unbiassed estimated variance:

$$\mathsf{d}\tilde{Y} = \mathsf{m}(\tilde{Y} - Y)^2 = \left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right)^2 \frac{(1 - \alpha_i)(1 - \alpha_j)(1 - \alpha_i - \alpha_j)}{(2 - \alpha_i - \alpha_j)^2} \,. \tag{3.15}$$

As $(1 - \alpha_i)(1 - \alpha_j) \leqq \left(1 - \dfrac{\alpha_i + \alpha_j}{2}\right)^2$, we have that

$$\mathsf{d}\tilde{Y} \leqq \frac{1}{4} \left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right)^2 (1 - \alpha_i - \alpha_j) \,. \tag{3.16}$$

If the ordered-set sample s' consists of elements $s' = [i_1, \ldots, i_n]$, then there are the following unbiassed estimates of Y:

$$\hat{Y}_1 = \frac{y_{i_1}}{\alpha_{i_1}} \,,$$
$$\hat{Y}_2 = y_{i_1} + \frac{y_{i_2}}{\alpha_{i_2}} (1 - \alpha_{i_1}) \,, \tag{3.17}$$
$$\vdots$$
$$\hat{Y}_n = y_{i_1} + \ldots + y_{i_{n-1}} + \frac{y_{i_n}}{\alpha_{i_n}} (1 - \alpha_{i_1} - \ldots - \alpha_{i_{n-1}}) \,,$$

where i_k is the element included in the sample as the k-th. These estimates are non-correlated, and, moreover, the conditional mean value of $\hat{Y}_k$, given $\hat{Y}_1, \ldots, \hat{Y}_{k-1}$, equals Y. Consequently, any constants $c_1, \ldots, c_n$, $\sum_{i=1}^{n} c_i = 1$, generate an unbiassed estimate $\hat{Y} = \sum_{i=1}^{n} c_i \hat{Y}_i$ such that

$$\mathsf{D}\left(\sum_{i=1}^{n} c_i \hat{Y}_i\right) = \sum_{i=1}^{n} c_i^2 \mathsf{D} \hat{Y}_i \,. \tag{3.18}$$

A further unbiassed estimate might be

$$\frac{1}{v} \sum_{s''} \frac{y_i}{\alpha_i}, \tag{3.19}$$

where the sum $\sum_{s''}$ extends over all elements of the sequence s'' and v denotes the number of members of s'' (i. e. the number of independent selections of one element until n distinct elements have been selected).

The estimate (3.19) may be identified as a conditional mean value of (3.1) with respect to (s^*, y), where s^* is the orderless-sequence sample defined in Section 1:

$$\frac{1}{v} \sum_{s''} \frac{y_i}{\alpha_i} = \mathsf{M}\left(\frac{y_{i_1}}{\alpha_{i_1}} \middle| (s^*, y)\right). \tag{3.20}$$

This means, in view of (1.9), that the conditional mean of (3.19) with respect to (s, y) also equals (3.7). In addition, it is thereby proved that (3.19) is an unbiassed estimate.

Example 3.2. Let us perform a fixed number, say m, of independent selections of one element always with probabilities $\alpha_1, \ldots, \alpha_N$ and consider the well-known estimate

$$\hat{Y} = \frac{1}{m} \sum_{s''} \frac{y_i}{\alpha_i} \tag{3.21}$$

where the sum $\sum_{s''}$ extends over the selected sequence $s'' = (i_1, \ldots, i_m)$. It may be easily shown that the estimate (3.21) is a conditional mean value of (3.1) with respect to (s^*, y), where s^* is again the orderless-sequence sample.

If the sampling is uniform, i. e. $\alpha_1 = \ldots = \alpha_N = \frac{1}{N}$, then the conditional mean value of (3.21) with respect to (s, y) is easily seen to be

$$\tilde{Y} = \mathsf{M}(\hat{Y} \mid (s, y)) = \frac{N}{d} \sum_{i \in s} y_i \,, \tag{3.22}$$

where d is number of distinct elements in the sample and s is the set of the distinct elements in the sample.

If the probabilities α_i are varying we may get $\tilde{Y}$ as follows: First let us rewrite $\hat{Y}$ in the form

$$\hat{Y} = \frac{1}{m} \sum_{i \in s} \frac{y_i}{\alpha_i} + \frac{1}{m} \sum_{i \in s} (k_i - 1) \frac{y_i}{\alpha_i} \tag{3.23}$$

where k_i is the number of times the element i has been selected. The conditional probabilities of events $\{k_i = k_i^0, i \in s\}$ obviously equal

$$\text{const} \prod_{i \in s} \frac{\alpha_i^{k^0_i - 1}}{k_i^0!}, \quad k_i^0 \geqq 1 , \quad \sum_{i \in s} k_i^0 = m . \tag{3.24}$$

If all elements in s'' are distinct then $\tilde{Y} = \hat{Y}$. If there are $m - 1$ or $m - 2$ distinct elements in the sample s'', then (3.24) generates the following conditional distribution of $\sum_{i \in s} (k_i - 1) \frac{y_i}{\alpha_i}$:

$$\mathsf{P} \left\{ \sum_{i \in s} (k_i - 1) \frac{y_i}{\alpha_i} = \frac{y_a}{\alpha_a} \middle| (s, y) \right\} = c\alpha_a , \quad a \in s , \quad \begin{bmatrix} m - 1 \text{ distinct} \\ \text{elements} \end{bmatrix} \tag{3.25}$$

$$\left. \begin{aligned} &\mathsf{P} \left\{ \sum_{i \in s} (k_i - 1) \frac{y_i}{\alpha_i} = \frac{y_a}{\alpha_a} + \frac{y_b}{\alpha_b} \middle| (s, y) \right\} = c \frac{\alpha_a \alpha_b}{2!2!} \\ &\mathsf{P} \left\{ \sum_{i \in s} (k_i - 1) \frac{y_i}{\alpha_i} = 2 \frac{y_a}{\alpha_a} \middle| (s, y) \right\} = c \frac{\alpha_a^2}{3!} \end{aligned} \right\} a, b \in s, \begin{bmatrix} m - 2 \text{ distinct} \\ \text{elements} \end{bmatrix}, \tag{3.26}$$

where c is a constant. From (3.25) and (3.26), after some computations, we get that

$$\tilde{Y} = \frac{1}{m} \left[\sum_{i \in s} \frac{y_i}{\alpha_i} + \frac{\sum_{i \in s} y_i}{\sum_{i \in s} \alpha_i} \right], \quad \begin{bmatrix} m - 1 \text{ distinct} \\ \text{elements} \end{bmatrix}, \tag{3.27}$$

$$\tilde{Y} = \frac{1}{m} \left[\sum_{i \in s} \frac{y_i}{\alpha_i} + 2 \frac{\sum_{i \in s} y_i \left(1 + \frac{1}{3} \frac{\alpha_i}{\sum_{j \in s} \alpha_j} \right)}{\sum_{i \in s} \alpha_i \left(1 + \frac{1}{3} \frac{\alpha_i}{\sum_{j \in s} \alpha_j} \right)} \right], \quad \begin{bmatrix} m - 2 \text{ distinct} \\ \text{elements} \end{bmatrix}. \tag{3.28}$$

If there are less then $m - 2$ distinct elements, then precise evaluation of $\tilde{Y}$ becomes too complicated. We may use, however, the approximation

$$\tilde{Y} = \frac{1}{m} \left[\sum_{i \in s} \frac{y_i}{\alpha_i} + r \frac{\sum_{i \in s} y_i}{\sum_{i \in s} \alpha_i} \right], \quad \begin{bmatrix} m - r \text{ distinct} \\ \text{elements} \end{bmatrix} \tag{3.29}$$

which seems to be a good one.

Example 3.3. Let us consider a double sampling design and denote by s_1 and s $(s \subset s_1)$ the subsets of elements selected in the first and second phase, respectively. Let

$$\hat{Y} = \sum_{i \in s_1} y_i w_{i1}(s_1) \tag{3.30}$$

be an estimate based on the sample s_1. Now, we may judge the sample s_1 as a population with ascertained values $y_i w_i(s_1)$, $i \in s_1$, and construct a linear estimate

$$\hat{\hat{Y}} = \sum_{i \in s} y_i w_{i1}(s_1)\, w_{i2}(s_1, s),$$

whose weights depend on s_1. As the final observation (s, y) is, obviously, an abstract function of (s_1, s, y), any good estimate $\hat{\hat{Y}}$ must not depend on s_1, i. e. we must have

$$w_{i1}(s_1)\, w_{i2}(s_1, s) = w_i(s)\,, \quad s \subset s_1 \subset S\,. \tag{3.32}$$

If (3.32) does not hold, the estimate may be improved by the method we have used in Examples 3.1 and 3.2.

If $\hat{Y} = \underset{2}{\mathsf{M}}\hat{\hat{Y}}$, where $\underset{2}{\mathsf{M}}(.)$ denotes the mean value with respect to the second phase (or stage) of sampling, then

$$\mathsf{M}(\hat{\hat{Y}} - Y)^2 = \mathsf{M}(\hat{Y} - Y)^2 + \mathsf{M}(\hat{Y} - \hat{\hat{Y}})^2\,. \tag{3.33}$$

Example 3.4. The same considerations may be applied to sampling whose result is given by k interpenetrating samples $(s_1, \ldots, s_k)$. We come to the conclusion that the "good" estimates must not depend of how many times an element has been selected, i. e. ist must be a function the set s of elements contained in at least one set $s_1, \ldots, s_k$, and of the observations ascertained thereon.

Now we shall leave this topic, since, as will be shown in the section 4, there exists an exact theory of fixed-size sampling with varying probabilities without replacement.

Remark 3.1. The method could be formulated as an application of the well-known Rao-Blackwell theorem on improving estimates by taking their conditional mean value with respect to a sufficient statistic. In fact, in the space of all possible observations (s'', η) we many consider the system $\{\mathsf{P}_{y_1 \ldots y_N}(.)\}$ of admissible distributions generated by all possible sequences of values $y_1, \ldots, y_N$ in such a way that

$$\begin{aligned} \mathsf{P}_{y_1 \ldots y_N}(s'', \eta) &= \mathsf{P}(s'')\,, \quad \text{if} \quad (s'', \eta) = (s'', y)\,, \\ &= 0\,, \quad \text{otherwise}\,. \end{aligned}$$

Let us note that $(y_1, \ldots, y_N)$ plays a role of a parameter.

4. Rejective sampling

Let us perform n independent draws of one element always with probabilities $\alpha_1, \ldots, \alpha_N$ and accept or reject all the selected elements if or if not at no two draws the same elements has been selected, respectively. If the sample is rejected, let us repeat this procedure until we get an acceptable sample. In this well-known sampling scheme, we have, obviously,

$$\mathsf{P}(s) = \lambda \prod_{i \in s} \alpha_i \qquad \text{for any } s \text{ consisting of } n \text{ elements}\,, \tag{4.1}$$

where

$$\lambda = [\sum_{s \in V_n} \prod_{i \in s} \alpha_i]^{-1}\,,$$

where V_n denotes the class of all subsets of the population which consist of n elements. Our point is to show that the sampling design just described is capable of exact and easy treatment.

First, let us observe that any sample z not containing the element i may be converted in a sample which does contain the element i by omitting any one of its elements and replacing it by the element i. This may be done in n ways as z contains n elements. If the obtained sample is s, then, by (4.1),

$$\mathsf{P}(z) = \frac{\alpha_k}{\alpha_i}\,\mathsf{P}(s)\,,$$

where k is the element which has been omitted. This means that

$$\begin{aligned} 1 &= \sum_{s \ni i} \mathsf{P}(s) + \sum_{z \,\mathrm{non} \ni i} \mathsf{P}(z) = \sum_{s \ni i} \mathsf{P}(s) + \frac{1}{n} \sum_{s \ni i} \mathsf{P}(s) \sum_{k \,\mathrm{non} \in s} \frac{\alpha_k}{\alpha_i} = \\ &= \sum_{s \ni i} \mathsf{P}(s) \left[1 + \frac{1}{n\alpha_i} \sum_{k \,\mathrm{non} \in s} \alpha_k\right] = \sum_{s \ni i} \mathsf{P}(s)\, \frac{1 - \sum_{k \in s} \alpha_k + n\alpha_i}{n\alpha_i} \end{aligned} \tag{4.2}$$

where $\sum_{z \,\mathrm{non} \ni i}$ extends over all samples not containing the element i. Denoting

$$\alpha = \sum_{k \in s} \alpha_k\,, \quad w_i(s) = \frac{1 - \alpha + n\alpha_i}{n\alpha_i}\,, \quad i \in s\,, \tag{4.3}$$

we may rewrite (4.2) in the form

$$\sum_{s \ni i} w_i(s)\, \mathsf{P}(s) = 1\,. \tag{4.4}$$

The equation (4.4), however, means that $w_i(s)$ are weights of the unbiassed estimate

$$\hat{Y} = \sum_{i \in s} y_i \, \frac{1 - \alpha + n\alpha_i}{n\alpha_i} = \frac{1 - \alpha}{n} \sum_{i \in s} \frac{y_i}{\alpha_i} + \sum_{i \in s} y_i\,. \tag{4.5}$$

The estimate (4.5) equals Y identically when y_i are exactly proportional to numbers α_i, $i = 1, \ldots, N$, i. e. (2.3) is satisfied for $z_i = \alpha_i$. Consequently, according to (2.11), we have

$$\mathsf{D}\hat{Y} = \frac{1}{2}\sum_{i\in s}\sum_{j\in s}\left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right)^2 \alpha_i\alpha_j(1 - \mathsf{M}w_iw_j)\,. \tag{4.6}$$

The (i, j)-estimates of 1 and of $\mathsf{M}w_iw_j$ we shall seek in the form (2.7) and (2.9). Let us observe that, in view of (4.1),

$$\frac{\mathsf{P}(s - \{i\} + \{k\})}{\mathsf{P}(s)} = \frac{\alpha_k}{\alpha_i}\,,$$

$$\frac{\mathsf{P}(s - \{i, j\} + \{k, h\})}{\mathsf{P}(s)} = \frac{\alpha_k\alpha_h}{\alpha_i\alpha_j}\,,$$

and substitute these results into (2.7) and (2.9). We get:

$$\text{Estimate of } 1 = 1 + \frac{1}{n-1}\sum_{k\,\text{non}\in s}\frac{\alpha_k}{\alpha_j} + \frac{1}{n-1}\sum_{k\,\text{non}\in s}\frac{\alpha_k}{\alpha_i} +$$

$$+ \frac{1}{n(n-1)}\sum_{\substack{k\,\text{non}\in s,\ h\,\text{non}\in s\\ k\neq h}}\sum\frac{\alpha_k\alpha_h}{\alpha_i\alpha_j} =$$

$$= \frac{(1 - \alpha + n\alpha_i)(1 - \alpha + n\alpha_j) - n\alpha_i\alpha_j - \sum_{k\,\text{non}\in s}\alpha_k^2}{n(n-1)\,\alpha_i\alpha_j}\,, \quad i, j \in s\,.$$

$$\text{Estimate of } \mathsf{M}w_iw_j = w_iw_j = \frac{(1 - \alpha + n\alpha_i)(1 - \alpha + n\alpha_j)}{n^2\alpha_i\alpha_j}\,.$$

This gives:

$$\text{Estimate of } [1 - \mathsf{M}w_iw_j] =$$

$$= \frac{(1 - \alpha + n\alpha_i)(1 - \alpha + n\alpha_j)}{n^2(n-1)} - \frac{\alpha_i\alpha_j}{n-1} - \frac{\sum_{k\,\text{non}\in s}\alpha_k^2}{n(n-1)}\,. \tag{4.7}$$

On substituting (4.7) into (4.6) we obtain the following unbiassed estimated variance:

$$\mathsf{d}\hat{Y} = \frac{1}{2}\,\frac{1}{n-1}\sum_{i\in s}\sum_{j\in s}\left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right)^2\left[\frac{(1 - \alpha + n\alpha_i)}{n}\,\frac{(1 - \alpha + n\alpha_j)}{n} - \right.$$

$$\left. - \alpha_i\alpha_j - \frac{1}{n}\sum_{k\,\text{non}\in s}\alpha_k^2\right]. \tag{4.8}$$

Now, remembering that for any numbers p_i, $\sum_{i\in s} p_i = 1$, the identity

$$\frac{1}{2}\sum_{i\in s}\sum_{j\in s}\left(\frac{y_i}{\alpha_i} - \frac{y_j}{\alpha_j}\right)^2 p_ip_j = \sum_{i\in s}\left(\frac{y_i}{\alpha_i} - \sum_{i\in s}\frac{y_i}{\alpha_i}p_i\right)^2 p_i$$

holds, we can (4.8) rewrite in the following form:

$$\mathsf{d}\hat{Y} = \frac{1}{n-1}\left[\sum_{i\in s}\left(\frac{y_i}{\alpha_i} - \hat{Y}\right)^2 \frac{1-\alpha+n\alpha_i}{n} - n \sum_{k\,\mathrm{non}\in s} \alpha_k^2 \sum_{i\in s}\left(\frac{y_i}{\alpha_i} - \frac{1}{n}\sum_{i\in s}\frac{y_i}{\alpha_i}\right)^2 \frac{1}{n} - \right.$$
$$\left. - \alpha^2 \sum_{i\in s}\left(\frac{y_i}{\alpha_i} - \frac{\sum_{k\in s} y_k}{\sum_{k\in s}\alpha_k}\right)^2 \frac{\alpha_i}{\alpha}\right]. \tag{4.9}$$

The relative magnitudes of the terms on the right side are 1, $\frac{n}{N}\left(1-\frac{n}{N}\right)$ and $\left(\frac{n}{N}\right)^2$, respectively. If $\alpha_1 = \ldots = \alpha_N = \frac{1}{N}$, we get the well-known formula for simple random sampling.

If we wish to exploit the proportionality of values y_i to certain values x_i, we may use the ratio estimate

$$\check{Y} = \frac{\frac{1-\alpha}{n}\sum_{i\in s}\frac{y_i}{\alpha_i} + \sum_{i\in s} y_i}{\frac{1-\alpha}{n}\sum_{i\in s}\frac{x_i}{\alpha_i} + \sum_{i\in s} x_i} \sum_{i=1}^{N} x_i = fX\,. \tag{4.10}$$

Alternatively, we may also use the unbiassed ratio-type estimate

$$\check{Y} = \sum_{i\in s} y_i + \frac{1-\alpha}{n-1}\sum_{i\in s}\frac{y_i}{\alpha_i} + \frac{1}{n}\sum_{i\in s}\frac{y_i}{x_i}\left[X - \sum_{i\in s} x_i - \frac{1-\alpha}{n-1}\sum_{i\in s}\frac{x_i}{\alpha_i}\right] \tag{4.11}$$

which is a variant of the estimate introduced Goodman and Hartley in the case of simple random sampling (see [13]). To show the unbiassedness, let us rewrite (4.11) in the form

$$\check{Y} = \sum_{i\in s} y_i + \frac{1-\alpha}{n}\sum_{i\in s}\frac{y_i}{\alpha_i} +$$
$$+ \frac{1}{n}\sum_{i\in s}\frac{y_i}{x_i}\left[X - x_i - \sum_{j\in s-\{i\}} x_j - \frac{1-\alpha}{n-1}\sum_{j\in s-\{i\}}\frac{x_j}{\alpha_j}\right]. \tag{4.12}$$

The first two terms on the right side are nothing else but the unbiassed estimate (4.5) of Y. Consequently, if we show that the conditional mean value of

$$\sum_{j\in s-\{i\}} x_j + \frac{1-\alpha}{n-1}\sum_{j\in s-\{i\}}\frac{x_j}{\alpha_j} \tag{4.13}$$

under the condition $s \ni i$ equals $X - x_i$, our proof will be completed. However, from (4.1) it is easily seen that conditional probabilities of s under the condition $s \ni i$, say $P(s|i)$, equal

$$\mathsf{P}(s|i) = \lambda_i \prod_{j\in s-\{i\}} \alpha_j\,, \quad s \ni i\,,$$

i. e. conditional distribution has the same structure as the nonconditional one. Now the estimate (4.13) has the same form as the unbiassed estimate (4.5) except that n is replaced by $n - 1$, probabilities $\alpha_1, \ldots, \alpha_N$ are replaced by probabilities $\frac{\alpha_1}{1-\alpha_i}, \ldots, \frac{\alpha_{i-1}}{1-\alpha_i}, \frac{\alpha_{i+1}}{1-\alpha_i}, \ldots, \frac{\alpha_N}{1-\alpha_i}$, and the element i is omitted from the population. Consequently, the estimate (4.13) actually is an unbiassed estimate of $X - x_i$ under the condition $s \ni i$.

The variances of (4.10) and (4.11) are naturally complicated but they may be estimated in lines with Theorem 2 . 2, since (2.3) is satisfied for $z_i = x_i$.

5. Permutation sampling

The rule of including and not-including the element i in the sample is the following: We take a random permutation $R_1, \ldots, R_N$ of numbers $1, \ldots, N$, all permutations having the same probability, and than include or not include the element i in the sample if or if not

$$R_i \leqq \pi_i N\,, \quad i = 1, \ldots, N\,. \tag{5.1}$$

If the numbers $\pi_i N$ are not integers, we may replace R_i by $R_i - \xi_i$, where ξ_i are independent random variables distributed uniformly over the interval (0,1).

It is easily seen that the numbers π_i used in (5.1) are directly probabilities of including the element in the sample. If $\pi_i \leqq \pi_j$, and the element i has been included in the sample, then, because $R_j \neq R_i$ and $R_i \leqq \pi_i N \leqq \pi_j N$, R_j may take on $\pi_j N - 1$ integers not greater than $\pi_j N$ and $N - \pi_j N$ integers greater then $\pi_j N$, each of them with the same probability $\frac{1}{N-1}$. Consequently, the probability of including both elements i and j in the sample equals

$$\pi_{ij} = \pi_i \frac{\pi_j N - 1}{N - 1}\,, \quad [\pi_i \leqq \pi_j]\,. \tag{5.2}$$

When we are using the above mentioned random variables ξ_i, the formula for π_{ij} must be slightly modified.

The permutation sampling is useful, for example, in connection with the ratio estimate

$$\check{Y} = \frac{\sum_{i \in s} \frac{y_i}{\pi_i}}{\sum_{i \in s} \frac{x_i}{\pi_i}} \sum_{i=1}^{N} x_i = f X\,. \tag{5.3}$$

Practical performation of permutation sampling goes as follows: We decide that π_i's should by proportional to numbers a_i and that the mean sample size should be $\bar{n}$. Then we compute the number

$$\psi = \frac{\sum_{i=1}^{N} a_i}{\bar{n}N} = \frac{\bar{a}}{\bar{n}}.$$

Now we select elements by simple random sampling element-by-element without replacement and the element selected as the k-th, say i_k, we accept or reject if or if not $a_{i_k} \geqq k\psi$. It is easily seen that, if $a_* = \min_{1 \leqq i \leqq N} a_i$ and $a^* = \max_{1 \leqq i \leqq N} a_i$, the first $k \leqq k_* = \frac{a_*}{\psi}$ elements are accepted certainly, and, on the other hand, the sampling is certainly finished when $k \geqq k^* = \frac{a^*}{\psi}$.

We may observe that in permutation sampling we need know the a_i's only for the first $k \leqq \frac{a^*}{\psi}$ selected elements and that no sums of a_i's are needed.

B. OPTIMUM STRATEGY. CASE I: THE ASCERTAINED VALUES ARE NON-CORRELATED

The assumption that the ascertained values are (a priori) non-correlated is of fundamental importance and, as the reader will see in section 8, of rather wide scope. For previous results see [6].

6. Bayes approach to the optimum strategy

Sampling-estimating strategy is defined by probabilities $\mathsf{P}(s)$, $s \subset S$, and by weights $w_i(s)$, $i \in s$, $s \subset S$. The quality of this strategy will be judged, on the one side, by the mean square error

$$\mathsf{M}(\hat{Y} - Y)^2 = \sum_{s \subset S} (\sum_{i \in s} y_i w_i(s) - \sum_{i=1}^{N} y_i)^2 \, \mathsf{P}(s) \,, \tag{6.1}$$

and, on the other side, by costs which have generally the form

$$C = \sum_{s \subset S} c(s, w) \, \mathsf{P}(s) \tag{6.2}$$

where $c(s, w)$, $s \subset S$, are costs of ascertaining the values y_i on s, and of computing $\hat{Y}$ with the weights $w_i(s)$, $i \in s$; perhaps, in $c(s, w)$ may be included costs of computing estimated sampling error, i. e. estimated square root of (6.1). The right side of (1.2) is called a cost function.

Sampling-estimating may be optimum in two senses: it either minimizes the mean square error (6.1) for given expected total cost (6.2), or conversely, it minimizes expected total cost (6.2) for given mean square error (6.1).

The notion of optimum strategy is useless when only one particular sequence $y_1, \ldots, y_N$ is considered, because the problem disappears, if we know it, and has no solution, if we does not know it. This difficulty may be overcome in various ways. In this paper we shall prefer the Bayes approach. It means we shall suppose that there is a certain probability distribution in the space of sequences $(y_1, \ldots, y_N)$, and our criterion of accuracy will be

$$\mathsf{EM}(\hat{Y} - Y)^2 \tag{6.3}$$

where E denotes the mean value over the random sequence $(y_1, \ldots, y_N)$ and M denotes the mean value over the samples s.

As can be seen from (6.1) $\mathsf{M}(\hat{Y} - Y)^2$ is a quadratic form in $y_j's$, so that $\mathsf{EM}(\hat{Y} - Y)^2$ depends only of the mean values, covariances and variances of $y_1, \ldots, y_N$. Consequently, specification of the distribution of $y_1, \ldots, y_N$ may only consist in the determination of the first and second-order moments:

$$\mathsf{E}y_i = \mu_i\,, \quad \mathsf{V}y_i = d_{ii}\,, \quad \text{Cov}\,(y_i, y_j) = d_{ij}\,, \quad i, j = 1, \ldots, N \tag{6.4}$$

where V denotes the variance in the $(y_1, \ldots, y_N)$-space.

Bayes approach seems to be reasonable on these grounds: (a) In most cases, we really have some knowledge of conditions producting values $y_1, \ldots, y_N$, and we can express them in the form (6.4). (b) Assumptions (6.4), if accepted, only influence our choice of sampling-estimating strategy and do not influence the validity of our estimated sampling errors, confidence intervals etc. Sampling error will be valid for any particular sequence $y_1, \ldots, y_N$ and consequently, any mistake in assumptions (6.4) will only cause the sampling errors to be on the average greater than they would be, if our assumptions were right.

There are, of course, cases, when the sample is selected in such a way that sampling error cannot be estimated on the basis of ascertained values $y_1, \ldots, y_N$ only (e. g. in systematic sampling). However, even in these cases, the Bayes approach is the only way of getting any estimated sampling error at all. We shall conclude with a few remarks.

Remark 6.1. The first to use the Bayes approach for the solution of sampling strategy was, according to the author's knowledge, W. G. Cochran, in the pioneering paper [1].

Remark 6.2. As emphasized by R. A. Fisher, any statistical work is a whole consisting of two aspects: *experimental design* and *statistical procedure*. This applies to probability sampling, too, the two aspects being the *sampling design* and the *estimation procedure*.

Remarks 6.3. Not only in probability sampling but in any statistical work there is a fundamental distinction between the case where the a priori distribution is an *organic part* of the statistical procedure (i. e. influences the validity of the probability statements), and the case where it only influences the *choice* of the experimental design and of the statistical procedure.

7. Sufficient conditions for the optimum strategy

We shall restrict ourselves to unbiassed linear estimators, i. e. $\sum_{s \subset S} w_i(s)\mathsf{P}(s) = 1$, $i = 1, \ldots, N$, and to the simple cost function

$$C = \sum_{i=1} c_i \pi_i \,. \tag{7.1}$$

(7.1) is based on the assumption that the cost associated with the element i equals c_i, i. e. does not depend on the estimation method and on which other elements were selected.

Supposing that the y_s's are non-correlated and that the estimator $\hat{Y}$ is unbiassed, let us evaluate (6.3). First, we change the order of the mean-value operators E and M, and then make use of the assumption that the y_s's are non-correlated:

$$\mathsf{EM}(\hat{Y} - Y)^2 = \mathsf{ME}(\hat{Y} - Y)^2 = \mathsf{ME}(\sum_{i \in s} y_i w_i(s) - \sum_{i=1}^{N} y_i)^2 =$$

$$= \mathsf{M}\{(\sum_{i \in s} \mu_i w_i(s) - \sum_{i=1}^{N} \mu_i)^2 + \sum_{i \in s} d_{ii}(w_i(s) - 1)^2 + \sum_{i \,\text{non} \in s} d_{ii}\} =$$

$$= \mathsf{M}(\sum_{i \in s} \mu_i w_i(s) - \sum_{i=1}^{N} \mu_i)^2 + \sum_{i=1}^{N} d_{ii}[\sum_{s \ni i} (w_i(s) - 1)^2 \,\mathsf{P}(s) + \sum_{s \,\text{non} \ni i} \mathsf{P}(s)] \tag{7.2}$$

where $\mu_i = \mathsf{E}y_i$, $d_{ii} = \mathsf{V}y_i$, and $s \in i$ and $i \in s$ denote that i is not contained in s. Bearing in mind (1.3) and (1.4) we get

$$\sum_{s \in i} (w_i(s) - 1)^2 \,\mathsf{P}(s) + \sum_{s \,\text{non} \ni i} \mathsf{P}(s) \geqq \left[\sum_{s \ni i} (w_i(s) - 1) \,\mathsf{P}(s)\right]^2 \frac{1}{\sum_{s \ni i} \mathsf{P}(s)} + \sum_{s \,\text{non} \ni i} \mathsf{P}(s) =$$

$$= \frac{1}{\pi_i} (1 - \pi_i)^2 + (1 - \pi_i) = \frac{1}{\pi_i} - 1 \,, \tag{7.3}$$

where the sign of equality holds if, and only if, $w_i(s)$ are independent of s, $s \ni i$. Independency of $w_i(s)$ of s, in connection with (1.3) and (1.4), implies that

$$w_i(s) = \frac{1}{\pi_i} \,, \quad i \in s \,, \quad s \subset S \,. \tag{7.4}$$

The result just obtained together with (7.2) gives that

$$\mathsf{E}\mathsf{M}(\hat{Y} - Y)^2 \geqq \sum_{i=1}^{N} d_{ii}\left(\frac{1}{\pi_i} - 1\right) \tag{7.5}$$

where the sign of equality holds if and only if, first, $\sum_{i \in s} \mu_i w_i(s) = \sum_{i=1}^{N} \mu_i$ for any s such that $\mathsf{P}(s) > 0$, and, second, (7.4) holds, i. e. $\hat{Y}$ is a simple linear estimate:

$$\hat{Y} = \sum_{i \in s} \frac{y_i}{\pi_i}\,. \tag{7.6}$$

Finally let us choose optimum $\pi_1, \ldots, \pi_N$ under the supposition that the expected total cost is given by (7.1). For brevity, let us denote the right side of (7.5) by

$$D = \sum_{i=1}^{N} d_{ii}\left(\frac{1}{\pi_i} - 1\right). \tag{7.7}$$

The use of Cauchy's inequality gives that

$$C\left[D + \sum_{i=1}^{N} d_{ii}\right] = \left(\sum_{i=1}^{N} c_i\pi_i\right)\left(\sum_{i=1}^{N} \frac{d_{ii}}{\pi_i}\right) \geqq \left(\sum_{i=1}^{N} \sqrt{c_i d_{ii}}\right)^2 \tag{7.8}$$

where the sign of equality holds if

$$c_i\pi_i = \lambda^2 \frac{d_{ii}}{\pi_i}, \quad i = 1, \ldots, N\,,$$

i. e.

$$\pi_i = \lambda\sqrt{\frac{d_{ii}}{c_i}}, \quad i = 1, \ldots, N\,, \tag{7.9}$$

where λ is a constant by which we may regulate the expected variance or cost when substituting into (7.7) or (7.1), respectively; of course, λ must be chosen so that $\pi_i \leqq 1$.

Since the $\pi_i' s$ satisfying (7.9) minimalize the product $C[D + \sum_{i=1}^{N} d_{ii}]$ for any λ, (7.9) solves simultaneously both problems: minimization of C for given D, and of D for given C. The minimum value of C or D for given D or C, respectively, may be obtained from the equation

$$C\left[D + \sum_{i=1}^{N} d_{ii}\right] = \left(\sum_{i=1}^{N} \sqrt{c_i d_i}\right)^2. \tag{7.10}$$

Let us summarize our results in a theorem:

Theorem 7.1. *Let us suppose that* $d_{ij} = 0,\ i \neq j,\ \sum_{s \subset S} w_i(s)\ \mathsf{P}(s) = 1,\ i = 1, .., N$ *and* $C = \sum_{i=1}^{N} c_i \pi_i$. *Then any strategy for which the conditions*

$$\pi_i = \lambda \sqrt{\frac{d_{ii}}{c_i}}, \quad i = 1, \ldots, N\,, \tag{7.11}$$

$$w_i(s) = \frac{1}{\pi_i}, \quad i \in s\,, \quad s \subset S\,, \tag{7.12}$$

$$\sum_{i \in s} \mu_i w_i(s) = \sum_{i=1}^{N} \mu_i\,, \quad P(s) > 0\,, \tag{7.13}$$

are fulfilled, is the optimum one.

The condition (7.11) determines the expected frequency π_i of the element i in the sample and will be called the condition of *optimum allocation*. The condition (7.12) will be called the condition of *constancy of weights*. The condition (7.13) means that for any sample s which can be really selected the sampling error vanishes as soon as y_i exactly equal their expected values μ_i, $i = 1, \ldots, N$. We can express this by words that the strategy is *representative* with respect to the values μ_i, $i = 1, \ldots, N$. In this way the vague notion of "representativity" becomes well-defined.

We can see that the condition (2.9) assumed in the section 2 was nothing else than the condition of representativity with respect to the values z_i, $i = 1, \ldots, N$.

From the course of the proof it is visible that small violations of the above conditions are immaterial. Most frequently it is the condition (7.12) which is not exactly fulfilled: the weights are not constant, and, moreover, they are very often such that the estimate is not unbiassed (e. g. in ratio estimation). Once we know that the above conditions are sufficient, it is inessential whether they are *approximately* fulfilled by an unbiassed or biassed estimate. Another question would be to choose an optimum solution within the whole class of linear estimates. However, this problem is too subtle and hardly fruitful.

Sometimes we may decide which of two possible systems of weights $w_i(s)$ and $w_i^*(s)$ is better in the following way: Suppose that the sampling design considered is such that (7.11) and (7.13) hold, then, according to (7.2),

$$\mathsf{EM}(\hat{Y} - Y)^2 = \mathsf{M}\{\sum_{i \in s} d_{ii}(w_i(s) - 1)^2 + \sum_{i\,\mathrm{non} \in s} d_{ii}\}\,. \tag{7.14}$$

If it happens that for any s with $\mathsf{P}(s) > 0$

$$\sum_{i \in s} d_{ii}(w_i(s) - 1)^2 \leqq \sum_{i \in s} d_{ii}(w_i^*(s) - 1)^2\,, \tag{7.15}$$

then, clearly, the weights $w\ (s)$ are the better of the two.

Example 7.1. Let us compare the estimates

$$\hat{Y}_1 = \sum_{i \in s} y_i \frac{1 - \alpha + n\alpha_i}{n\alpha_i}, \tag{7.16}$$

$$\hat{Y}_2 = \frac{1}{n} \sum_{i \in s} \frac{y_i}{\alpha_i}, \tag{7.17}$$

under the hypotheses that

$$d_{ii} = \lambda \alpha_i^\delta, \quad 0 \leqq \delta \leqq 2, \tag{7.18}$$

and that the sampling is rejective as described in section 4. The estimate (7.16), as we have seen, is unbiassed, and the estimate (7.17) is biassed. We cannot, however, a priori say which of them is better. Putting

$$w_i(s) = \frac{1 - \alpha + n\alpha_i}{n\alpha_i}, \quad w_i^*(s) = \frac{1}{n\alpha_i}, \quad i \in s,$$

we have

$$\sum_{i \in s} d_{ii}[(w_i(s) - 1)^2 - (w_i^*(s) - 1)^2] = \sum_{i \in s} d_{ii}(w_i(s) + w_i^*(s) - 2) \,.$$
$$. \,(w_i(s) - w_i^*(s)) = \lambda \sum_{i \in s} \alpha_i^{\delta-2}(2 - \alpha - n\alpha_i)(n\alpha_i - \alpha) \,.$$

The last expression, however, is non-positive since it equals the covariance of values $\lambda'\alpha_i^{\delta-2}(2 - \alpha - n\alpha_i)$ and values $n\alpha_i$ and since the relation

$$[n\alpha_i \leqq n\alpha_j] \Rightarrow [\alpha_i^{\delta-2}(2 - \alpha - n\alpha_i) \geqq \alpha_j^{\delta-2}(2 - \alpha - n\alpha_j)] \,, \tag{7.19}$$
$$0 \leqq \delta \leqq 2 \,, \quad i, j \in s$$

holds. Consequently, under the hypothesis (7.18), the estimate (7.16) is better than the estimate (7.17).

Unbiassed estimated mean square error of the estimate (7.17) can be derived in lines of the sections 2 and 4:

$$m(\hat{Y}_2 - Y)^2 = \frac{1}{n-1} \sum_{i \in s} \left(\frac{y_i}{\alpha_i} - \frac{1}{n} \sum_{i \in s} \frac{y_i}{\alpha_i} \right)^2 \frac{(1 - n\alpha_i)(1 - \alpha) + n \sum\limits_{k \text{ non} \in s} \alpha_k^2}{n} +$$
$$+ \alpha^2 \left(\frac{1}{n} \sum_{i \in s} \frac{y_i}{\alpha_i} - \frac{\sum_{i \in s} y_i}{\sum_{i \in s} \alpha_i} \right)^2 . \tag{7.20}$$

Now, we shall apply the theorem 7.1 to several typical examples.

Example 7.2. If $\mu_i = \text{const}$, $d_{ii} = \text{const}$, $c_i = \text{const}$, $i = 1, \ldots, N$, then all the requirements of the theorem 7.1 are fulfilled by simple random sampling together with the simple linear estimate $\frac{N}{n} \sum_{i \in s} y_i$. If μ_i's, d_{ii}'s and c_i's are constant within some parts (strata) of the population, and when these strata are large enough, then the requirements of the optimum strategy are satisfied by the optimumly allocated sampling together with the simple linear estimate. If the stratum S_h contains N_h elements from which we select n_h elements, then for an element i, belonging to the stratum S_h, we have

$$\pi_i = \frac{n_h}{N_h}, \quad i \in S_h . \tag{7.21}$$

According to (7.11) the n_h's must be chosen so that

$$\frac{n_h}{N_h} = \lambda \sqrt{\frac{d_{hh}}{c_h}}, \quad h = 1, \ldots, H, \tag{7.22}$$

where d_{hh} and c_h are common values of d_{ii}'s and c_i's within the stratum S_h. However, from (7.22) it is seen that whenever the numbers $\lambda N_h \sqrt{\frac{d_{hh}}{c_h}}$ are too small, then (7.22) may not be fulfilled by any *integers* n_h with a reasonable degree of precision. In such situations is stratified sampling ineffective.

Example 7.3. Very often we may suppose that μ_i's are proportional to known numbers x_i:

$$\mu_i = \lambda x_i , \quad i = 1, \ldots, N . \tag{7.23}$$

In such situations the ratio estimates (5.3), (4.10) or (4.11) meet the condition of representativity with respect to (7.23). The condition of constancy of weights, however, may not be fulfilled. When using the estimate (5.3), (4.10) or (4.11) we have weights

$$w_i^*(s) = \frac{1}{\pi_i} \frac{\sum_{i=1}^{N} x_i}{\sum_{i \in s} \frac{x_i}{\pi_i}}, \tag{7.24}$$

$$w_i^{**}(s) = \frac{1 - \alpha + n\alpha_i}{n\alpha_i} \frac{\sum_{i=1}^{N} x_i}{\sum_{i \in s} x_i \frac{1 - \alpha + n\alpha_i}{n\alpha_i}}, \tag{7.25}$$

$$w_i^{***}(s) = x_i \left[1 + \frac{1 - \alpha}{n\alpha_i}\right] + \frac{1}{nx_i} \left[X - \sum_{i \in s} x_i + \frac{1 - \alpha}{n - 1} \sum_{i \in s} \frac{x_i}{\alpha_i}\right], \tag{7.26}$$

respectively. The variability of the weights depends on the sampling design and on the relation between the values x_i and π_i or α_i. Generally, the weights (7.25) connected with the rejective sampling design described in Section 4 may be expected to vary least.

The weights (7.24) can be used in connection with the permutation sampling described in Section 5 with π_i satisfying the condition of optimum allocation

(7.11). If the weights (7.25) are used, in connection with the rejective sampling described in Section 4, the $\alpha_i's$ should be chosen so that the $\pi_i's$ again satisfy (7.11). Now from (1.4) and (4.3) it follows that

$$\frac{1}{\pi_i} = \frac{\sum_{i \in s} \frac{1 - \alpha + n\alpha_i}{n\alpha_i} P(s)}{\sum_{i \in s} P(s)} \tag{7.27}$$

whence it is easily seen that

$$\frac{n\alpha_i}{1 + (n-1)\max_j (\alpha_j - \alpha_i)} \leqq \pi_i \leqq b \frac{n\alpha_i}{1 - (n-1)\max_j (\alpha_j - \alpha_i)} . \tag{7.28}$$

Consequently, the equation (7.11) will be nearly fulfilled if simply

$$\alpha_i = \lambda'' \sqrt{\frac{d_{ii}}{c_i}}, \quad i = 1, \ldots, N . \tag{7.29}$$

If the $\pi_i's$ are not smaller than, say, 0,1, we may use the approximate relation

$$\pi_i \cong \frac{n\alpha_i}{1 + \frac{n-1}{N-1} N \left(\alpha_i - \frac{1}{N}\right)}, \tag{7.30}$$

i. e.

$$\alpha_i \cong \frac{\pi_i}{n} \frac{1 - \frac{n-1}{N-1}}{1 - \frac{(n-1)N}{(N-1)n} \pi_i} . \tag{7.31}$$

It may be of interest to modify the sampling design in order that the weights (7.24) vary as little as possible. Let us describe such a design for the uniform (self-weighting) case ($\pi_1 = \ldots = \pi_N = \pi$). In this case we try to select elements in such a way that $\pi_i = \pi$ and that the sum $\sum_{i \in s} x_i$ is nearly constant. Let us suppose that the $x_i's$ are integers (which causes no loss of generality) and consider the cyclical sequence

$$1, 2, \ldots, x_1 + \ldots + x_N , \quad 1, 2, \ldots, \omega .$$

Now we associate with the element i the segment of the sequence containing numbers $\{x_1 + \ldots + x_{i-1} + 1, \ldots, x_1 + \ldots + x_i\}$, $i = 1, \ldots, N$. Finally, we select an integer r with equal probabilities in the range $1 \leqq r \leqq x_1 + \ldots + x_N$, and include the element i in the sample with a probability $\frac{m_i(r)}{\omega}$, where $m_i(r)$ is the number of integers belonging simultaneously to the both segments $\{r, \ldots, r + \omega - 1\}$ and $\{x_1 + \ldots + x_{i-1} + 1, \ldots, x_1 + \ldots + x_i\}$. The probability of including an element in the sample is the same for all elements, namely

$$\pi_i = \frac{\omega}{x_1 + \ldots + x_N} . \tag{7.32}$$

This sampling may be preceeded by any, random or non-random, ordering of elements.

Example 7.4. If the a priori hypothesis is expressed by a regression relation

$$\mu_i = \sum_{j=1}^{k} \beta_j x_i^{(j)} + \beta_0 , \quad i = 1, \ldots, N , \tag{7.33}$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are unknown constants, then the fulfillment of the condition of representativity is not easy. We may use some type of regression estimates, which are, however, difficult to compute, or, when the regression is passing through the origin (i. e. $\beta_0 = 0$), some type of ratio estimates. Finally, we may use the so called acceptance-inspection method. This method consists in rejecting samples which give a bad result in estimating totals X_j of controlled (concomitant) variables used in the relation (7.31). For example, we repeat selections of a sample s until we get a sample for which

$$|\hat{X}_j - X_j| \leqq \varepsilon_j \sqrt{\mathsf{D}\hat{X}_j} , \quad j = 1, \ldots k , \tag{7.34}$$

where ε_j are properly chosen constants, say $\varepsilon_j = 0{,}5$. This method of attaining to the representativity emphasizes the sampling-design aspect, and this is right in situations, where we are dealing with extensive and fresh a priori data and ascertain a number of variables $y_i', y_i'', \ldots$, each of them is related to variable $x_i^{(1)}, \ldots, x_i^{(k)}$. The theory of the acceptance-inspection method is based on the supposition that the random vector $(\hat{Y}, \hat{X}_1, \ldots, \hat{X}_k)$ has a $(k + 1)$-dimensional normal distribution. However, no theoretical argument for this assumption is at our disposal in this time. Acceptance control is discussed in the paper [8].

8. Some remarks on the applications

8.1. Uniform sampling. The cost function (7.1) does not reflect the jump in cost which arises when the sampling is uniform (self-weighting) so that the simple linear estimates are reduced to arithmetic sums, i. e. $\hat{Y} = \frac{1}{\pi} \sum_{i \in s} y_i$. Therefore, an optimum non-uniform sampling-estimating strategy is really acceptable only if it is better than the uniform sampling connected with the simple linear estimate or with the simple ratio estimate, etc. This device, consisting in dividing the possible strategies into classes within which the problem has a simple mathematical formulation, is of wide use.

8.2. Subsampling. The flexibility of the previous theory is well-illustrated by the subsampling problem. Let us suppose that the elements are primary sampling units consisting of N_i secondary units. We wish select in any primary unit n_i secondary units by simple random sampling, where n_i is ascertained from the relation

$$\pi_i \frac{n_i}{N_i} = \tau , \quad i = 1, \ldots, M , \tag{8.1}$$

where π_i is the probability of including the primary unit in the sample and τ is the uniform overall probability of including a secondary unit in the sample.

A comparison of $\mathsf{EM}(\hat{\hat{Y}} - Y)^2$, where $\hat{\hat{Y}}$ is given by (2.30), with (7.2) shows that the effect of subsampling is the same as the effect of increasing the variances d_{ii} by the second-stage variance of $\hat{y}_i$, i. e. by

$$\frac{N_i^2}{n_i} \sigma_i^2 \left(1 - \frac{n_i}{N_i}\right) , \tag{8.2}$$

where σ_i^2 is the expected variance within the element (primary sampling unit) i. On the other hand, the costs associated with the element i is decreased by subsampling by

$$(N_i - n_i)\, e_i , \tag{8.3}$$

where e_i is the cost associated with a secondary unit within the element i.

If we insert the changed a priori variances and costs into (7.1) and (7.7), we get ($N \equiv M$)

$$C^* = \sum_{i=1}^{M} \pi_i [c_i - (N_i - n_i)\, e_i] , \tag{8.4}$$

$$D^* = \sum_{i=1}^{M} \left[d_{ii} + \frac{N_i^2}{n_i} \sigma_i^2 \left(1 - \left(\frac{n_i}{N_i}\right)\right]\left(\frac{1}{\pi_i} - 1\right)\right) \tag{8.5}$$

which by means of (8.1) gives

$$C^* = \sum_{i=1}^{M} \pi_i (c_i - N_i e_i) + \tau \sum_{i=1}^{M} e_i N_i , \tag{8.6}$$

$$D^* = \sum_{i=1}^{M} \frac{d_{ii} - N_i \sigma_i^2}{\pi_i} + \frac{1}{\tau} \sum_{i=1}^{M} N_i \sigma_i^2 - \sum_{i=1}^{M} (d_{ii} - N_i \sigma_i^2) . \tag{8.7}$$

It means that, if $d_{ii} - N_i\sigma_i^2 \geqq 0$, $c_i - N_i e_i \geqq 0$, we have

$$C^*[D^* + \sum_{i=1}^{M} (d_{ii} - N_i\sigma_i^2)] \geqq \left(\sum_{i=1}^{M} \sqrt{d_{ii} - N_i\sigma_i^2)(c_i - N_i e_i)} + \sqrt{\sum_{i=1}^{M} N_i\sigma_i^2 \sum_{i=1}^{N} N_i e_i}\right)^2,$$

where the sign of equality holds if

$$\frac{d_{ii} - N_i\sigma_i^2}{\pi_i} = \lambda^2\pi_i(c_i - N_ie_i) \quad \text{and} \quad \frac{\sum_{i=1}^{M} N_i\sigma_i^2}{\tau} = \lambda^2\tau \sum_{i=1}^{M} N_ie_i ,$$

i. e.

$$\pi_i = \lambda \sqrt{\frac{d_{ii} - N_i\sigma_i^2}{c_i - N_ie_i}} , \quad \tau = \lambda \sqrt{\frac{\sum_{i=1}^{M} N_i\sigma_i^2}{\sum_{i=1}^{M} N_ie_i}} . \tag{8.8}$$

See also [10], § 4.

8.3. Relative emphasis on the sampling aspect. The conditions stated in the Theorem 7.1 can be approximately fulfilled in many ways. When choosing a particular way, we may utilize some information which has not been included either in the assumptions about the a priori mean values and variances or in the cost function. A problem of this kind is how to distribute the effort between the sampling design and the estimation method. On some occasions it is better to pay attention to the sampling design (e. g. see Example 7.3) and on other occasions it is better to choose carefully the estimation method.

8.4. Connection between the a priori mean values and the assumption of non-correlation. The assumption of non-correlation of ascertained values is the more realistic, the more specific are the a priori mean values. For example, in a population of areas on which total yields of cereal-crops are ascertained, we may put

$$\mu_i = \mu \tag{8.9}$$

or

$$\mu_i = \mu x_i \tag{8.10}$$

or, finally,

$$\mu_i = bt_ix_i + ax_i \tag{8.11}$$

where x_i is the size of the area and t_i is an eye-estimate of the yield per unit area. The assumption (8.9) means that all influences are understood as random factors. The correlation of all these factors will cause the correlation of the values y_i. For example, in some regions there is a greater average size of areas than in other and, therefore, a greater average total yield on there areas. If we know the sizes x_i and take the model (8.10), then this source of correlation will disappear. In model (8.11), there is excluded a further source of correlation, namely the correlation of fertility on neighbourhooding areas. When using this last model we may hope that the yields are a priori approximately non-correlated.

8.5. Sampling designs with maximum entropy. A simple way of avoiding unfavourable consequences of the possible violation of the hypothesis of the non-correlation is to choose a sampling designs which distributes the probabilities $P(s)$ as uniformly as possible. This requirement may be formulated in the form that we wish to maximize the entropy

$$E = - \sum_{s \subset S} \mathsf{P}(s) \log \mathsf{P}(s) \, . \tag{8.12}$$

If the probabilities π_i, see (1.4), are fixed, the sampling which maximizes (8.12) is the Poisson sampling described in [10]. In fact, any sampling may be understood as N experiments having two possible outcomes: including or not including the element i in the sample. When the probability of including the element i in the sample, namely π_i, is given for each of these experiments, then the entropy will be maximum, as is well-known, if all these experiments are independent. By the last requirement just the Poisson sampling is defined.

Now we shall show that the rejective sampling (Section 4) with fixed $\alpha_1, \ldots, \alpha_N$ has the greatest entropy in the class of sampling design with the same $\pi_1 \ldots, \pi_N$ and with fixed sample size, n. In fact, bearing in mind (1.3) and using the usual method of Lagrange multiplicators λ_i, we get for $\mathsf{P}(s)$ and λ_i the following equations

$$\frac{\partial}{\partial P(s)} \left[\sum_{s \in V_n} - \mathsf{P}(s) \log \mathsf{P}(s) + \sum_{i=1}^{N} \lambda_i \sum_{s \ni i} \mathsf{P}(s) \right] = - \log \mathsf{P}(s) - 1 + \sum_{i \in s} \lambda_i = 0 \, , \quad s \in V_n \, , \tag{8.13}$$

$$\sum_{s \ni i} \mathsf{P}(s) = \pi_i \, , \quad i = 1, \ldots, N \, , \tag{8.14}$$

where V_n is the set of all samples with sample size n. Since the function $- x \log x$ is strictly concave, there is a single maximum. If it happens that the solution is such $0 \leqq \mathsf{P}(s) \leqq 1$, then it clearly coincides with the solution of the problem restricted to the domain $0 \leqq \mathsf{P}(s) \leqq 1$, $s \in V_n$. Now, it is easily seen that (8.13) is satisfied by the probabilities (4.1) and by

$$\lambda_i = \log \alpha_i + \frac{1}{n} (1 - \log \sum_{s \in V_n} \prod_{i \in s} \alpha_i) \, . \tag{8.15}$$

As regards (8.14), it is fulfilled automatically, since we considered just those $\pi_1, \ldots, \pi_N$ which are yielded by the given $\alpha_1, \ldots, \alpha_N$.

C. OPTIMUM STRATEGY. CASE II: THE ASCERTAINED VALUES FORM A RANDOM SERIES WITH A STATIONARY CONVEX CORRELATION FUNCTION AND STATIONARY COEFFICIENTS OF VARIATIONS (9—10)

If the population is meaningfully ordered (in time or space), the ascertained values are very often governed by a convex correlation function. An empirical correlation function of this kind is shown in the Fig. 1, where also the estimated correlation function is drawn. The data refer to a population of plots where the area of forest land has been measured.[1])

Convex correlation function are so often met with that they deserve a detailed study. This has been undertaken in the paper [1], and in subsequent papers [3] and [7], where, under the supposition of convexity, it is proved that systematic sampling is superior to simple random sampling, stratified sampling, and several independent systematic samplings. In the present paper, using quite a different method, we shall derive a general result that systematic sampling is better than any other sampling design under consideration. Moreover, we shall replace the supposition that $\mathsf{E}y_i$ and $\mathsf{V}y_i$ are stationary by a more general supposition that only the coefficients of variation, $\sqrt{\mathsf{V}y_i}/\mathsf{E}y_i$, are so. As a solution we shall obtain systematic sampling with generally unequal probabilities.

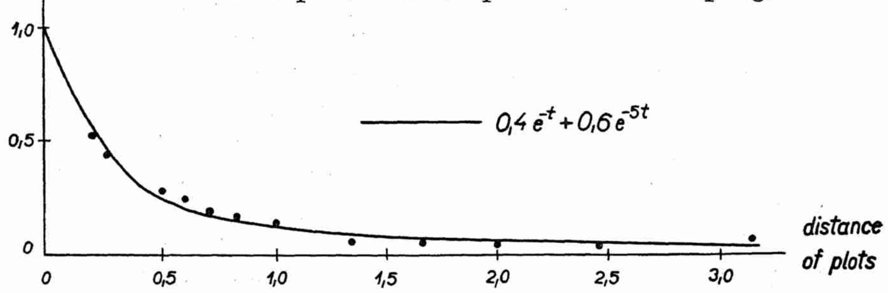

Fig. 1.

9. Preliminaries

Lemma 9.1. *Among all integral-valued random variables x possessing a given mean value ξ the random variable which takes on only the values $[\xi]$ and $[\xi] + 1$ is of least variance. Here $[\xi]$ denotes the greatest integral number not exceeding ξ.*

Proof. Putting $a = [\xi] + \frac{1}{2}$, we see that

$$\mathsf{D}x = \sum_{k=-\infty}^{\infty} (k - a)^2\, \mathsf{P}\{x = k\} - (\xi - a)^2 ,$$

where

$$(k - a)^2 = \tfrac{1}{4}\,, \quad \text{if} \quad k = [\xi] \quad \text{or} \quad [\xi] + 1\,,$$
$$> \tfrac{1}{4}\,, \quad \text{otherwise}\,.$$

This accomplishes our lemma.

[1]) The Figure 1 is borrowed from the paper [2] with author's kind permission.

Definition. A correlation function $R(i-j)$, $i, j = 1, \ldots, N$ that satisfies the condition

$$R(u) - 2R\left(\frac{u+v}{2}\right) + R(v) \geqq 0\,, \quad 0 \leqq u, v \leqq N\,, \tag{9.1}$$

is called convex.

Lemma 9.2. *Any convex correlation function $R(i-j)$, $i, j = 1, \ldots, N$, of of a discrete stationary random process may be expressed in the form*

$$R(i-j) = \sum_{\lambda=1}^{\infty} \sum_{\omega=1}^{N+\lambda-1} [R(\lambda+1) - 2R(\lambda) + R(\lambda-1)]\, q_{i\lambda\omega} q_{j\lambda\omega}\,, \tag{9.2}$$
$$1 \leqq i, j \leqq N$$

where

$$q_{i\lambda\omega} = 1\,, \quad \text{if} \quad \omega - \lambda < i \leqq \omega\,, \tag{9.3}$$
$$= 0\,, \quad \text{otherwise}\,.$$

and

$$R(\lambda+1) - 2R(\lambda) + R(\lambda-1) \geqq 0\,. \tag{9.4}$$

Proof. The inequality (9.4) follows from the assumption (9.1) when putting $u = \lambda + 1$, $v = \lambda - 1$. The equation (9.2) is a consequence of the following identities:

$$\sum_{\omega=1}^{N+\lambda-1} q_{i\lambda\omega} q_{j\lambda\omega} = \lambda - |i-j|\,, \quad \text{if} \quad \lambda > |i-j|$$
$$= 0\,, \quad \text{otherwise}$$

and

$$\sum_{\lambda=|i-j|+1}^{\infty} (\lambda - |i-j|)[R(\lambda+1) - 2R(\lambda) + R(\lambda-1)] =$$
$$= \sum_{\lambda=|i-j|+1}^{\infty} \sum_{\mu=|i-j|+1}^{\lambda} [R(\lambda+1) - 2R(\lambda) + R(\lambda-1)] =$$
$$= \sum_{\mu=|i-j|+1}^{\infty} \sum_{\lambda=\mu}^{\infty} [R(\lambda+1) - 2R(\lambda) + R(\lambda-1)] =$$
$$= \sum_{\mu=|i-j|+1}^{\infty} [R(\mu-1) - R(\mu)] = R(|i-j|) = R(i-j)\,.$$

10. An optimum property of systematic sampling

We shall suppose that the random sequence $y_1, \ldots, y_N$ possesses stationary coefficients of variation and a stationary convex correlation function, i. e.

$$\mathsf{E} y_i = \mu x_i\,, \quad \mathsf{V} y_i = \sigma^2 x_i^2\,, \quad \text{Cov}\,(y_i, y_j) = \sigma^2 x_i x_j R(i-j)\,, \tag{10.1}$$
$$i, j, = 1, \ldots, N\,,$$

and shall choose a sampling design which

(a) yields samples of fixed size, n,

(b) gives the following probabilities π_i of including the element i in the sample

$$\pi_i = \frac{nx_i}{x_1 + \ldots + x_N}, \quad i = 1, \ldots, N \tag{10.2}$$

(c) minimizes the expected variance of the estimator

$$\hat{Y} = \frac{X}{n} \sum_{i \in s} \frac{y_i}{x_i}, \quad X = \sum_{i=1}^{N} x_i . \tag{10.3}$$

Remark 10.1. The correlation function

$$R(i - j) = 1 , \quad \text{if} \quad i = j ,$$
$$= 0 , \quad \text{otherwise} ,$$

is convex, and, consequently, the scheme (10.1) incorporates the case of non-correlated random variables with stationary coefficients of variation as a particular case. When supposing constant costs, i. e. $c_i = c$, $i = 1, \ldots, N$, then the condition (10.2) and the estimator (10.3) garantee that the sampling-estimating strategy, we are choosing, will be optimum for this particular case. In fact, then (10.2) is equivalent to (7.11) and the estimator (10.3) implies the fulfilment of (7.12) and (7.13).

Theorem 10.1. *The above problem is solved by systematic sampling defined as follows: The sample consists of those elements i for which the sum $x_1 + \ldots + x_i$ at first reaches or exceeds some of the numbers*

$$r, r + \frac{1}{n} X , \ldots, r + \frac{n-1}{n} X , \quad \left(X = \sum_{i=1}^{N} x_i\right) ,$$

where r is a random variable uniformly distributed over the interval $0 < r \leqq \leqq \frac{1}{n} X$.

Proof. Unbiassedness of the estimator (10.3) for any particular sequence $y_1, \ldots, y_N$ justifies the first of the following identities (the remaining ones are obvious):

$$\mathsf{EM}(\hat{Y} - Y)^2 = \mathsf{EM}(\hat{Y} - \sum_{i=1}^{N} \mu x_i)^2 - \mathsf{E}(Y - \sum_{i=1}^{N} \mu x_i)^2 =$$
$$= \mathsf{ME} \left(\frac{X}{n} \sum_{i \in s} \frac{y_i}{x_i} - \mu X\right)^2 - \mathsf{E}(Y - \mu X)^2 =$$
$$= \left(\frac{X}{n}\right)^2 \sigma^2 \mathsf{M} \left(\sum_{i \in s} \sum_{j \in s} R(i - j)\right) - \mathsf{E}(Y - \mu X)^2 .$$

Consequently, it suffices to minimize $\mathsf{M}(\sum_{i\in s}\sum_{j\in s} R(i-j))$. Using (9.2) we obtain

$$\mathsf{M}(\sum_{j\in s}\sum_{i\in s} R(i-j)) = \sum_{\lambda=1}^{\infty}\sum_{\omega=1}^{N-\lambda-1}[R(\lambda+1) - 2R(\lambda) + R(\lambda-1)]\,\mathsf{M}(\sum_{i\in s} q_{i\lambda\omega})^2 .$$

According to (9.4) it suffices to prove that $\mathsf{M}(\sum_{i\in s} q_{i\lambda\omega})^2$ is minimized for each λ and ω. The definition (9.3) of $q_{i\lambda\omega}$ implies that

$$\sum_{i\in s} q_{i\lambda\omega} = \sum_{i=\omega-\lambda+1}^{\omega} p_i(s) \tag{10.4}$$

where p_i are random variables such that

$$\begin{aligned} p_i(s) &= 1\,, \quad \text{if} \quad s \ni i\,,\\ &= 0\,, \quad \text{otherwise.} \end{aligned} \tag{10.5}$$

This means that

$$\mathsf{M}\left(\sum_{i\in s} q_{i\lambda\omega}\right) = \sum_{i=\omega-\lambda+1}^{\omega} \mathsf{M}[p_i(s)] = \sum_{i=\omega-\lambda+1}^{\omega} \pi_i = \frac{n}{X}\sum_{i=\omega-\lambda+1}^{\omega} x_i\,.$$

i. e., in accordance with the point (b), the mean value of $\sum_{i\in s} q_{i\lambda\omega}$ is constant for any considered sampling design. Thus minimizing $\mathsf{M}(\sum_{i\in s} q_{i\lambda\omega})^2$ is equivalent to minimizing $\mathsf{D}(\sum_{i\in s} q_{i\lambda\omega})$ for given mean value $\frac{n}{X}\sum_{i=\omega-\lambda+1}^{\infty} x_i$. As $\sum_{i\in s} q_{i\lambda\omega}$ is an integral-valued random variable, it suffices, according to lemma 9.1, to show that $\sum_{i\in s} q_{i\lambda\omega}$ only takes on values $\left[\frac{n}{X}\sum_{i=\omega-\lambda+1}^{\omega} x_i\right]$ and $\left[\frac{n}{X}\sum_{i=\omega-\lambda+1}^{\omega} x_i\right] + 1$, where $[\xi]$ again denotes the greatest integral number not exceeding ξ. This may be done as follows: $\sum_{i\in s} q_{i\lambda\omega}$ is the number of members of the set $\{\omega-\lambda+1, \ldots, \omega\}$ which have been included in the sample, and this number, according to the definition of the systematic sampling in Theorem 10.1, equals the number of integrals $k = 1, \ldots, n$ for which

$$x_1 + \ldots + X_{\omega-\lambda} < r + \frac{k-1}{n} X \leqq x_1 + \ldots + x_\omega$$

i. e.

$$\frac{n}{X}(x_1 + \ldots + x_{\omega-\lambda} - r) + 1 < k \leqq \frac{n}{X}(x_1 + \ldots + x_\omega - r) + 1\,. \tag{10.6}$$

The number of integrals lying in a interval of fixed length l, however, can only equal $[l]$ or $[l] + 1$. The proof may be thus completed by observing that the length of the interval (10.6) equals $\frac{n}{X}\sum_{i=\omega-\lambda+1}^{\omega} x_i$.