BAYES AND MINIMAX PROCEDURES IN SAMPLING FROM FINITE AND INFINITE POPULATIONS[1]—I

By Om P. Aggarwal

Purdue University[2] and *Stanford University*

Summary. Some of the sampling methods and the methods of estimation usually employed in sample surveys are considered in terms of loss and risk functions. The loss function is taken as the sum of two components, one proportional to the square of the error of the estimate and the other proportional to the cost of obtaining the sample. Consideration is given to the problem of the allocation of the total sample size and only non-sequential estimates are discussed. As the loss function is convex and of finite expectation in each case, only non-randomized estimates are considered, since Hodges and Lehmann [5] have shown that under these conditions the class of non-randomized estimates is essentially complete. Only simple random sampling and stratified sampling methods are discussed in this part, the ratio, regression and sub-sampling methods will be discussed in subsequent parts.

1. Introduction. In the current practice of conducting sample surveys, the statisticians have adopted one of the following two procedures (see, e.g., [2], [3], [4], [7], or [8]): (i) to get an estimate of maximum precision for a given total cost of the survey, or (ii) to get an estimate of given precision for a minimum total cost of the survey. The allocation of the resources for a given survey is usually carried out, keeping in mind one or the other of the above two aims. It is possible, however, to consider jointly the losses resulting from the errors in the estimates and from the cost of sampling, and to employ such sampling and estimation procedures as will, in some sense, "minimize" the total expected loss. Accordingly we shall take as loss function the sum of two components, one proportional to the square of the error of the estimate and the other proportional to the cost of obtaining the sample. The problems generally met in sampling surveys will be formulated in terms of the decision theory using this loss function and it will be seen that their solutions are the classical results in estimation and design. This appears to be a preliminary step toward further research in this field.

2. Bayes and minimax estimates. The estimation problem with a fixed sample size has the following structure. We are given a sample space X, a space of probability distributions on X, $P_\Omega = \{p_\omega : \omega \in \Omega\}$, where Ω is an index set (generally called parameter space), and a numerical-valued function g defined on

Received April 29, 1958.

[1] Work done under the sponsorship of the Office of Naval Research while the author was at Stanford University.

[2] On leave from Purdue University during 1958–59 as Statistician with the F.A.O. of the United Nations on their Technical Assistance Mission to Chile.

Ω whose value $g(\omega)$ the statistician wishes to estimate on the basis of the outcome of an experiment, say $x \varepsilon X$. A non-randomized decision function for the statistician, usually called an estimate[3], is a numerical function δ defined on X, specifying for each x the number $a \varepsilon A$ which will be chosen to estimate $g(\omega)$ when that x is observed. The space of actions A is here the real line. The loss function L defined on $\Omega \times A$ is non-negative and is the loss incurred when $g(\omega)$ is estimated by a. The risk function R is defined by

$$R(\omega, \delta) = E_\omega L(\omega, \delta). \tag{2.1}$$

The subscript ω appended to the symbol E for expectation indicates that ω is to be regarded as fixed when expectation is taken.

If it is assumed that ω is obtained by nature as the value of a random variable having a probability distribution λ, a Bayes estimate with respect to the a priori distribution λ is defined as an estimate δ which minimizes the average risk $\int R(\omega, \delta)\, d\lambda(\omega)$. If the statistician knew λ, he would choose this estimate as his best action. But in the absence of any knowledge of λ the statistician may decide to use what is called in the terminology of two person zero-sum game a minimax strategy. A minimax estimate is defined as an estimate δ which minimizes the "maximum" risk, $\sup_{\omega \varepsilon \Omega} R(\omega, \delta)$. In the same terminology, a least favorable distribution or "maximin" strategy is defined as a distribution λ such that it maximizes the "minimum" risk $\inf_\delta \int R(\omega, \delta)\, d\lambda(\omega)$.

The following theorem [6] gives in many cases a minimax estimate as well as a least favorable distribution whenever the latter exists.

THEOREM 2.1. *If a Bayes estimate δ_λ has constant (independent of ω) risk $R(\omega, \delta_\lambda) = r$, then δ_λ is minimax and λ is a least favorable distribution.*

The following theorem [6] will give in many cases a minimax estimate where no least favorable distribution exists.

THEOREM 2.2. *If $\{\lambda_n\}$ is a sequence of a priori probability distributions, $\{r_n\}$ the sequence of associated Bayes risks, and if $r_n \to r$ as $n \to \infty$, and if there exists some estimate δ for which $R(\omega, \delta) \leqq r$ for all ω, then δ is a minimax estimate.*

We shall frequently need another theorem in the sequel. Using the term "minimax risk" for $\inf_\delta \sup_\omega R(\omega, \delta)$ we state and prove it as

THEOREM 2.3. *If δ, r are a minimax procedure and the minimax risk respectively, assuming that the observations X follow any probability distribution $\omega \varepsilon \Omega^*$, and if $\Omega \supset \Omega^*$ is a space of distributions for which the risk associated with δ does not exceed r, then δ is a minimax procedure and r the minimax risk for all distributions of X in Ω.*

PROOF. By hypothesis

$$r = \sup_{\omega \varepsilon \Omega^*} R(\omega, \delta) \leqq \sup_{\omega \varepsilon \Omega} R(\omega, \delta) \leqq r. \tag{2.2}$$

[3] It should be more properly called an estimator, to distinguish the function from a specific numerical value, but it is hoped no confusion will be caused by using the same term for both.

Hence equality holds. If d is any other procedure,

$$\sup_{\omega \varepsilon \Omega} R(\omega, \delta) = r \leqq \sup_{\omega \varepsilon \Omega^*} R(\omega, d) \leqq \sup_{\omega \varepsilon \Omega} R(\omega, d), \tag{2.3}$$

which shows that δ is a minimax procedure and r the minimax risk for all distributions in Ω.

3. Sampling from a finite population. In its simplest form, sampling from a finite population may be described in the following manner. We are given a sample space and nature (or a conscious being) performs a fixed sample size experiment and obtains a value of a random variable, which is a point $x = (x_1, x_2, \cdots, x_N)$ in the space R^N of ordered N-tuples of real numbers or vectors with real components. It should be mentioned, however, that all random variables may not be available to nature. The statistician has to select one out of a class A of possible actions in complete or partial ignorance of x and the particular probability distribution employed by nature to obtain x. He (the statistician) incurs a loss which is a bounded function of the selected action $a \varepsilon A$ and the point $x \varepsilon R^N$, but not of the underlying conceptual distribution. However, for a given a, the loss is assumed to be constant for all permutations of the coordinates of x. The statistician can obtain partial information on x by observing some fixed number of coordinates of x, say n. The problem is: If the cost of observing x_i, $i = 1, 2, \cdots, N$, is independent of i, how should the statistician select the sample and choose a?

In the case that $\Omega = R^N$ and the strategy for nature corresponding to ω is to choose $x = \omega$ with probability one, Blackwell and Girshick [1] have shown that the invariance and sufficiency principles require the sampling scheme to select each set of n distinct integers from 1 to N (without regard to order) with probability $\binom{N}{n}^{-1}$. This is the usual strategy of simple random sampling without replacement. The proof extends to the more complex situations discussed in this paper. Accordingly, we seek Bayes and minimax strategies using this sampling scheme.

4. Statement of the problem. We shall be estimating the mean of a finite population with our loss as squared error. If the variance is not restricted somehow, our risk may be arbitrarily large. Accordingly, we bound, under any strategy of nature for choosing the finite population, the expected value of the variance of the finite population. Another possibility would be to divide the loss by the expected variance. Our method actually shows that the sample mean is minimax for each expected variance.

Formally, the decision problem in which we are interested may be characterized as follows:

(a) The sample space is (X, Ω, p), where X is the N-dimensional Euclidean space R^N, Ω is the set of all distributions ω on hyperplanes in R^N of the form $x_1 + x_2 + \cdots + x_N =$ constant, say $N\mu_\omega$, and subject to the restriction that

$$E_\omega \sum_{i=1}^{N} (x_i - \mu_\omega)^2 = \int \cdots \int_{R^N} (x'x - N\mu_\omega^2)\, d\omega(x) \leqq (N-1)\sigma^2, \tag{4.1}$$

where x is the column vector with $x_1, x_2, \cdots, x_N$ as elements, σ is a given positive number, and $p_\omega = \omega$ for $\omega \varepsilon \Omega$.

(b) The action space A is the real line R^1.

(c) The loss function L, defined on $(\Omega \times A)$, is given by

$$L(\omega, a) = (a - \mu_\omega)^2. \tag{4.2}$$

(d) The space D of decision rules is the space of all ordered n-tuples of different integers from 1 through N together with all measurable mappings of n-space into A. If $\delta = (i_1, \cdots, i_n; f)$, then $\delta(x) = f(x_{i_1}, \cdots, x_{i_n})$.

The application of invariance and sufficiency principles, as in the problem stated in the last section, require the sampling scheme to select each set of ordered n-tuples of different integers from 1 through N with equal probability. Accordingly, it is sufficient to consider the following problem:

(a) The sample space (X, Ω, p), where X is the n-dimensional Euclidean space R^n, Ω is the same as before, and p_ω for $\omega \varepsilon \Omega$ is the distribution of a sample $x = (x_1, \cdots, x_n)$ obtained by simple random sampling without replacement from $x_1, x_2, \cdots, x_N$, which are distributed according to ω.

(b) The action space A is the same as before.

(c) The loss function L is the same as before.

(d) The space D of decision rules is the set of all measurable mappings of X into A.

The problem of obtaining a minimax estimate of the mean μ_ω of the finite population $(x_1, \cdots, x_N)$ is solved in the following way. Consider nature's strategy as picking μ_ω from $N(0, \theta^2)$, a normal distribution with mean zero and variance θ^2, and given μ_ω, letting ω, with probability one, be singular N-variate normal with mean μ_ω and variance $\sigma^2(N - 1)/N$ for each component and covariance $-\sigma^2/N$ for each pair of components. A Bayes estimate is obtained with respect to this strategy of nature, regarded as a member of a sequence $\{\lambda_\theta\}$ of a priori distributions, and the limit, if any, of the corresponding sequence of Bayes risks $\{r_\theta\}$ as $\theta \to \infty$ is obtained, say r. Then if we can find some estimate δ for which the risk $R(\omega, \delta)$—without assuming normality of ω—does not exceed r, then by virtue of Theorems 2.2 and 2.3, δ is a minimax estimate.

The discussion so far is given in detail for the case when the sampling plan is simple random sampling without replacement. Under somewhat different situations different sampling plans will be required. We shall not discuss the derivation of the minimax strategies for the choice of sampling plans (it can be shown by an extension of Blackwell and Girshick's proof of the optimality of simple random sampling [1] that the sampling plans given are optimum under the circumstances) but will take the sampling plans as being given and confine our attention to the problem of estimating the mean of the populations by employing techniques similar to the one outlined above.

5. Bayes and minimax procedures for estimating the mean of a finite population with simple random sampling (without replacement). The average risk corresponding to an a priori distribution ξ for nature and a decision function δ

used for estimation of $g(\omega)$ by the statistician is obviously

$$R(\xi, \delta) = \underset{\Omega}{E} \underset{X}{E} [(\delta(x) - g(\omega))^2 \mid \omega], \tag{5.1}$$

where x stands for $(x_1, x_2, \cdots, x_n)$ and the symbol below E indicates the space over which the expectation is to be taken. For the sake of simplicity we shall not attempt to distinguish between a random variable and its observed value. Since the integrand is non-negative, we may change the order of taking expectations and write

$$R(\xi, \delta) = \underset{X}{E} \underset{\Omega}{E} [(\delta(x) - g(\omega))^2 \mid x]. \tag{5.2}$$

This is minimized by choosing, for each x, that number $\delta(x)$ which minimizes $\underset{\Omega}{E}[(\delta(x) - g(\omega))^2 \mid x]$ and the number $\delta(x)$ which does it is clearly

$$\delta_\xi(x) = E(g(\omega) \mid x). \tag{5.3}$$

This gives the minimum value of $\underset{\Omega}{E}[(\delta(x) - g(\omega))^2 \mid x]$ as $\sigma^2_{g(\omega)\mid x}$, the variance of the conditional distribution of $g(\omega)$ given x. Then, by (5.2), the Bayes risk r_ξ is given by

$$r_\xi = \underset{X}{E} \, \sigma^2_{g(\omega)\mid x}. \tag{5.4}$$

These results hold in general whenever the loss function for the estimation of $g(\omega)$ is of the form $L(\omega, a) = [a - g(\omega)]^2$.

In the problem under consideration, taking the a priori distribution for nature as mentioned in the last section, the distribution of the sample $(x_1, \cdots, x_n)$ given ω is n-variate normal with mean μ_ω and variance $\sigma^2(N - 1)/N$ for each x_i and covariance $-\sigma^2/N$ for each pair (x_i, x_j), $i \neq j$. Since it can be shown easily that the sample mean $\bar{x}$ is a sufficient statistic for μ_ω, we see from (5.3) and (5.4) respectively that the Bayes estimate $\delta_\theta(x) = E(\mu_\omega \mid x) = E(\mu_\omega \mid \bar{x})$, and the Bayes risk $r_\theta = E\sigma^2_{\mu_\omega \mid x} = E\sigma^2_{\mu_\omega \mid \bar{x}}$. Now μ_ω is $N(0, \theta^2)$ and, given μ_ω, $\bar{x}$ is $N(\mu_\omega, v)$ where $v = (n^{-1} - N^{-1})\sigma^2$, so μ_ω and $\bar{x}$ have a bivariate normal distribution. It is then easily seen that the conditional distribution of μ_ω given $\bar{x}$ is normal with mean $\theta^2\bar{x}(\theta^2 + v)^{-1}$ and variance $\theta^2 v(\theta^2 + v)^{-1}$. Since the variance is independent of x, these are respectively Bayes estimate $\delta_\theta(x)$ and Bayes risk r_θ.

To find a minimax estimate for μ_ω, we consider if the sequence $\{r_\theta\}$ tends to a limit as $\theta \to \infty$. It is seen that it does and the limit r is given by

$$r = \lim_{\theta \to \infty} r_\theta = v = \frac{N - n}{Nn} \sigma^2. \tag{5.5}$$

By Theorem (2.2) if we can find some estimate δ for which the risk does not exceed r, then that δ is a minimax estimate. Trying $\delta(x) = \bar{x}$ $(= \lim_{\theta \to \infty} \delta_\theta(x))$, we see that the risk corresponding to δ is given by

$$\begin{aligned} R(\omega, \delta) = E_\omega(\bar{x} - \mu_\omega)^2 &= E_\omega E[(\bar{x} - \mu_\omega)^2 \mid x_1, \cdots x_N] \\ &= \frac{N - n}{nN(N - 1)} E_\omega \sum_{i=1}^{N} (x_i - \mu_\omega)^2 \\ &\leqq \frac{N - n}{Nn} \sigma^2, \qquad \text{by (4.1)} \\ &= r. \end{aligned}$$

Hence $\bar{x}$ is a minimax estimate and the risk corresponding to this estimate (minimax risk) does not exceed $(N - n)\sigma^2/nN$.

6. Bayes and minimax procedures with stratified sampling plan. In survey designs stratification is a procedure whereby the entire population is divided into a number of strata and sampling is carried out independently in each stratum. Let c_i be the known cost of sampling per unit in the ith stratum, μ_i and σ_i^2 the unknown mean and the known variance of the population in the ith stratum, and let k be the number of strata. Then, for a given cost $C = \sum_{i=1}^k c_i n_i$, where n_i is the number of observations sampled from the ith stratum, it is well known (see e.g. [2], [3], [4], or [8]) that the procedure which estimates $\sum_{i=1}^k N_i\mu_i$ with minimum variance is to choose n_i proportional to $N_i\sigma_i/\sqrt{c_i}$ and to use $\sum_{i=1}^k N_i\bar{X}_i$ as an estimate, where N_i is the size of the ith stratum and $\bar{X}_i$ the sample mean of the n_i observations selected at random from it. The same values for n_i are obtained when for a given variance of the estimate, the object is to minimize the total cost of sampling. In this section we shall investigate for this problem some Bayes and minimax procedures, first for an infinite and then for finite populations.

A. *Infinite populations.* Suppose that the ith stratum consists of an infinite population with unknown mean μ_i and known upper bound σ_i^2 for the variance, $i = 1, 2, \cdots, k$, and that we have to estimate a linear function of the μ_i, say $U = \sum_{i=1}^k a_i\mu_i$ where the a_i are some given real numbers. Without loss of generality we may take $\sum a_i = 1$. For the sake of simplicity we shall assume that none of the a_i is zero. The loss function L is given by

$$L(U, \delta) = (\delta - U)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.1}$$

where $n_i(>0)$ is the size of the sample chosen from the ith stratum, c_i the sampling cost per unit in that stratum, and δ is a function of the sample $\{X_{ij}; i = 1, 2, \cdots k; j = 1, 2, \cdots, n_i\}$, where X_{ij} is the jth observation from the ith stratum. For the sake of simplicity of notation, as before, we shall not attempt to distinguish between a random variable and its observed value.

It may be noted that a slightly more realistic loss function would be

$$L(U, \delta) = \alpha(\delta - U)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.2}$$

where α is some constant depending upon the desired relative accuracy of the results and the cost of experimentation in a given situation. But it is easily seen that any procedure corresponding to this loss function can be obtained from the corresponding procedure when the loss function is (6.1) simply by substituting c_i/α for c_i. The risk associated with it will be simply α times the corresponding risk when the loss function is (6.1). For this reason also the condition $\sum a_i = 1$ in this section does not detract from the generality of the a_i.

Assume at first that, given ω, the distribution in each stratum is normal, with variance σ_i^2 in the ith stratum, $i = 1, 2, \cdots, k$. We may conjecture that for this problem there is no least favorable distribution of nature, since as U could have any real value, what we would expect it to be is a uniform distribution over the real line, but this is not a distribution.[4] We shall assume that the μ_i are normally and independently distributed each with mean zero and variance θ^2, and find Bayes solutions corresponding to the sequence $\{\lambda_\theta\}$ of a priori distributions of U resulting from the distribution of μ_i. Let δ_θ denote a corresponding Bayes estimate of U. If the Bayes risks r_θ corresponding to δ_θ tend to a limiting value r when θ tends to infinity, then any estimate which has its risk less than or equal to r will be a minimax estimate by Theorem 2.2 under the normality assumption of the observations. This assumption may then be removed easily with the help of Theorem 2.3.

We may regard the n_i as fixed for the purpose of finding the estimates. Letting δ^* be a minimax estimate for given n_i, we shall choose the n_i so as to minimize the risk,

$$R(U, \delta^*) = E(\delta^* - U)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.3}$$

as a function of the n_i.

Since we are working with fixed n_i, we shall omit the $\sum c_i n_i$ term. The loss function is now simply the square of the difference between the estimate and the quantity U being estimated and, as in the last section, Bayes estimate δ_θ and Bayes risk r_θ are given respectively by the mean and the expectation of the variance of the conditional distribution of U, given the sample x. However, since the stratum sample means are jointly sufficient for $\mu_1, \cdots, \mu_k$, we may replace the sample x in the last sentence by $\bar{X}_1, \cdots, \bar{X}_k$, where $\bar{X}_i$ is the sample mean from the ith stratum.

Now, the μ_i are independently and normally distributed with means zero and variance θ^2, and given μ_i, the $\bar{X}_i$ are independently and normally distributed with mean μ_i and variance σ_i^2/n_i, so $\mu_1, \cdots, \mu_k$ and $\bar{X}_1, \cdots, \bar{X}_k$ have a joint $2k$-variate normal distribution. It is then easily seen that the conditional distribution of μ_i, given $\bar{X}_1, \cdots, \bar{X}_k$, is normal with mean

$$y_i = \frac{n_i \theta^2 \bar{X}_i}{\sigma_i^2 + n_i \theta^2}, \tag{6.4}$$

[4] I understand from an oral communication from H. Rubin that a proof of this conjecture has been given by M. A. Girshick.

and variance

$$v_i = \frac{\theta^2 \sigma_i^2}{\sigma_i^2 + n_i \theta^2}, \tag{6.5}$$

and that $\mu_1, \cdots, \mu_k$ are mutually independent given $\bar{X}_1, \cdots, \bar{X}_k$. Thus, for given $\bar{X}_1, \cdots, \bar{X}_k$, the distribution of U is normal with mean $\sum a_i y_i$ and variance $\sum a_i^2 v_i$. We thus conclude that the Bayes estimate $\delta_\theta(x) = \delta_\theta(\bar{X}_1, \cdots, \bar{X}_k) = \sum_{i=1}^k a_i y_i$, and since the variance of the conditional distribution of U is independent of $\bar{X}_1, \cdots, \bar{X}_k$, the Bayes risk $r_\theta = \sum_{i=1}^k a_i^2 v_i$.

Minimax estimate for given n_i. Letting $\theta \to \infty$, we see that $r_\theta \to r$, where $r = \sum_{i=1}^k a_i^2 \sigma_i^2 / n_i$. Thus, if we can find some estimate δ^* with risk $\leqq r$, then δ^* will be a minimax estimate by virtue of Theorem 2.2. Let us try the limiting Bayes estimate,

$$\lim_{\theta \to \infty} \delta_\theta(x) = \sum_{i=1}^k a_i \bar{X}_i = \delta^*(x), \quad \text{say}. \tag{6.6}$$

Since the $\bar{X}_i$ are normal and independent with means μ_i and variances σ_i^2/n_i, $\sum a_i \bar{X}_i$ is normal with mean $\sum a_i \mu_i = U$ and variance $\sum a_i^2 \sigma_i^2 / n_i$. Hence the risk corresponding to the estimate $\delta^*(x) = \sum a_i \bar{X}_i$ is equal to r which proves that $\sum a_i \bar{X}_i$ is a minimax estimate of U for given n_i.

It may be of interest to point out that although the Bayes estimates $\delta_\theta = \sum a_i y_i$, where y_i is given by (6.4), being unique (for given θ) are admissible, we cannot conclude from this the admissibility of δ^* because of the limiting process. The same remark applies to the other Bayes and minimax estimates obtained later, but we shall not go into the question of admissibility in this paper.

Removal of the normality assumption. Let us now do away with the assumption of normality of the distribution of X_{ij}. Suppose that whatever the joint distributions of the X_{ij}, the distribution in the ith stratum has an unknown mean μ_i, and the sample mean $\bar{X}_i$, for any sample size n_i, has a variance not exceeding a known positive number σ_i^2/n_i for $i = 1, 2, \cdots, k$, and that the $\bar{X}_i$ are uncorrelated. This is somewhat more general than the usual assumption of X_{ij} being independent with mean μ_i (unknown) and variance σ_i^2 (known) in the ith stratum, in which case the stratified sampling procedure is generally used. Let us calculate the risk R corresponding to the minimax estimate $\delta^*(x) = \sum a_i \bar{X}_i$ obtained under the assumption of the normal distribution of the observations in each stratum. It is easily verified that under these general assumptions, for given n_i,

$$\begin{aligned} R &= E\left(\sum_{i=1}^k a_i \bar{X}_i - U\right)^2 \\ &= E\left[\sum_{i=1}^k a_i(\bar{X}_i - \mu_i)\right]^2 \leqq a_i^2 \sigma_i^2 / n_i = r. \end{aligned} \tag{6.7}$$

Applying Theorem 2.3 now, we conclude that the minimax estimate $\sum a_i \bar{X}_i$ obtained under the assumption of normality of observations is still a minimax estimate for given n_i under the general assumptions given in the beginning of this paragraph.

Minimax strategy for choosing the n_i. Restoring the term $\sum c_i n_i$ in the risk function, one can choose "optimum" n_i if the variances of the populations in different strata are known, rather than only the upper bounds, by minimizing the risk as a function of the n_i. However, if only the upper bounds and not the actual variances are assumed to be known, the optimum choice of the n_i against the largest allowed variances σ_i^2 may not be optimum for other values of the variances, but it will still be a "minimax" choice. In other words, a minimax strategy for the statistician is to choose the n_i to be "optimum" against the maximum allowed variances σ_i^2, and then estimate U using the δ^* for these n_i. This statement follows from the following theorem.

THEOREM 6.1. *Suppose the space of strategies for the statistician is a union of spaces, say* $D = \bigcup_c D_c$. *If* δ_c *is minimax in* D_c *against* Ω, *the space of nature's strategies, and if* $R(\omega, \delta_c)$ *is constant for each c, say* $R(\omega, \delta_c) = r_c$, *then the* δ_c *minimizing* r_c, *if it exists, is minimax in D.*

PROOF. Let $\delta \varepsilon D$ be any strategy for the statistician. Then $\delta \varepsilon D_c$ for some c. Since δ_c is minimax in D_c,

$$\max_\omega R(\omega, \delta) \geqq \max_\omega R(\omega, \delta_c) = r_c. \tag{6.8}$$

Let δ_{c^*} be the δ_c minimizing r_c and denote the risk corresponding to δ_{c^*} by r_{c^*}. Then

$$r_c \geqq r_{c^*} = \max_\omega R(\omega, \delta_{c^*}). \tag{6.9}$$

From (6.8) and (6.9), $\max_\omega R(\omega, \delta_{c^*}) \leqq \max_\omega R(\omega, \delta)$ for all $\delta \varepsilon D$, hence δ_{c^*} is minimax in D.

We, therefore, choose optimum n_i corresponding to the variances in the different strata as the σ_i^2. For given n_i, the risk corresponding to δ^* is given by

$$R(\omega, \delta^*) = \sum_{i=1}^{k} \left[\frac{a_i^2 \sigma_i^2}{n_i} + c_i n_i \right]. \tag{6.10}$$

Now we want to choose the n_i so that this risk is minimum under the restriction that the n_i are positive integers. Since the ith term on the right hand side of (6.10) depends on n_i alone, it is sufficient to minimize $a_i^2\sigma_i^2/n_i + c_i n_i$ subject to the restriction that n_i is a positive integer. Denoting this expression by $f(n_i)$, we see that

$$f(n_i + 1) - f(n_i) = c_i - \frac{a_i^2 \sigma_i^2}{n_i(n_i + 1)}. \tag{6.11}$$

To minimize $f(n_i)$, we choose the smallest positive integral value for n_i for which the difference (6.11) is positive; in other words, the smallest positive integer n_i

for which $(n_i + 1/2)^2$ exceeds $a_i^2 \sigma_i^2/c_i + 1/4$. This gives

$$n_i = \text{integer nearest to } \sqrt{\frac{a_i^2 \sigma_i^2}{c_i} + \frac{1}{4}}. \tag{6.12}$$

When $\sqrt{a_i^2 \sigma_i^2/c_i + 1/4}$ lies exactly between two integers, say m and $m + 1$, the risk is equal and minimum for both $n_i = m$ and $n_i = m + 1$, and it is immaterial which of the two nearest integers is chosen for n_i.

B. *Finite population.* Suppose that x_{ij}, $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, N_i)$ denotes some numerical characteristic of the jth unit in the ith stratum. Suppose further that the $N_i(>1)$ are known, the means u_i of the strata are unknown, the upper bounds of the variances of the populations in the strata are known, say $\sigma_i^2 \geqq 1/(N_i - 1)E\sum_{j=1}^{N_i} (x_{ij} - u_i)^2$, and that we are required to estimate a linear function,

$$T = \sum_{i=1}^{k} a_i u_i, \tag{6.13}$$

of the population means u_i, where a_i are arbitrary known real numbers with $\sum a_i = 1$, the loss function L being given by

$$L(T, \delta) = (\delta - T)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.14}$$

where δ is an estimate for T, and c_i, n_i denote the cost of sampling per unit and the number of units sampled in the ith stratum. The sampling plan given is to decide upon k positive integers n_i, $i = 1, 2, \cdots, k$, and then choose a sample of size n_i by simple random sampling without replacement from the ith stratum, thus obtaining a sample of total size $n = \sum_{i=1}^{k} n_i$. We shall first assume that the n_i are determined somehow and obtain Bayes and minimax procedures and the corresponding risks for given n_i. Later we shall see how to choose the n_i so that the risk obtained is minimized over the choice of n_i. As before, by Theorem 6.1, this choice of the "optimum" n_i will be a minimax strategy for the statistician.

As in the case of simple random sampling discussed in Section 4, which is a special case of this problem for $k = 1$, we are considering now a decision problem in which the distribution $\omega \varepsilon \Omega$ consists of the product of k independent distributions ω_i on hyperplanes in R^{N_i} of the form $x_{i1} + \cdots + x_{iN_i} =$ constant, say $N_i\mu_{\omega_i}$, and subject to the restrictions that

$$E_{\omega_i} \sum_{j=1}^{N_i} (x_{ij} - u_i)^2 \leqq (N_i - 1)\sigma_i^2, \qquad i = 1, 2, \cdots, k, \tag{6.15}$$

where the constant is denoted by $N_i\mu_{\omega_i}$ to make μ_{ω_i} the mean of $x_{i1}, x_{i2}, \cdots, x_{iN_i}$ and μ_{ω_i} itself is being written as u_i for the sake of convenience of notation, the σ_i are given positive numbers, and p_ω for $\omega \varepsilon \Omega$ is the distribution of k independent samples $x_i = (x_{i1}, \cdots, x_{in_i})$, the ith sample being obtained by simple random sampling without replacement from $x_{i1}, \cdots, x_{iN_i}$, distributed according to ω_i.

The problem of obtaining a minimax estimate of T for given n_i is solved as before. Consider nature's strategy as picking each $u_i(=\mu_{\omega_i})$ from $N(0, \theta^2)$ and given u_i, letting the distributions ω_i, with probability one, be singular N_i-variate normal with mean u_i and variance $\sigma_i^2(N_i - 1)/N_i$ for each component and covariance $-\sigma_i^2/N_i$ for each pair of components. A Bayes estimate of U is obtained with respect to this strategy of nature, which is regarded as a member of a sequence $\{\lambda_\theta\}$ of a priori distributions, and the limit, if any, of the corresponding sequence of Bayes risks $\{r_\theta\}$ as $\theta \to \infty$ is obtained, say r. Then an estimate δ for which the risk $R(\omega, \delta)$—without assuming normality of ω—does not exceed r is, by Theorems 2.2 and 2.3, a minimax estimate for given n_i.

With nature's strategy as explained in the last paragraph, the distribution of the sample $x = \{x_{ij}; j = 1, \cdots, n_i, i = 1, \cdots, k\}$ given ω is the product of k distributions, the ith being n_i-variate normal with mean u_i and variance $\sigma_i^2(N_i - 1)/N_i$ for each component and covariance $-\sigma_i^2/N_i$ for each pair of components. Again, since the set $(\bar{x}_1, \cdots, \bar{x}_k)$ of the sample means from the k strata is a sufficient statistic for the set $(u_1, \cdots, u_k)$, and hence for T, we may replace the sample x in (5.3) and (5.4) by the set $(\bar{x}_1, \cdots, \bar{x}_k)$. Now, the strata are independent, i.e. the k pairs $(u_i, \bar{x}_i)$ are independent, and it follows from the calculations in Section 5 that the conditional distribution of an individual u_i given $\bar{x}_i$ is normal with mean $y_i = \theta^2\bar{x}_i(\theta^2 + v_i)^{-1}$ and variance $\theta^2 v_i(\theta^2 + v_i)^{-1}$, where $v_i = (n_i^{-1} - N_i^{-1})\sigma_i^2$. Thus, for given $\bar{x}_1, \cdots \bar{x}_k$, the conditional distribution of $T = \sum a_i u_i$ is normal with mean $\sum a_i y_i$, and variance $\sum a_i^2\theta^2 v_i(\theta^2 + v_i)^{-1}$. We thus conclude that the Bayes estimate for T is

$$\delta_\theta(x) = \sum_{i=1}^{k} a_i\, y_i = \sum_{i=1}^{k} a_i\, \theta^2\bar{x}_i(\theta^2 + v_i)^{-1}, \tag{6.16}$$

and as the variance of the conditional distribution of T is independent of x, the Bayes risk is

$$r_\theta = \sum_{i=1}^{k} a_i^2\, \theta^2 v_i(\theta^2 + v_i)^{-1} + \sum_{i=1}^{k} c_i\, n_i . \tag{6.17}$$

To find a minimax estimate for T now, we consider if the sequence $\{r_\theta\}$ tends to a limit as $\theta \to \infty$. It will be seen that it does, and the limit r is given by

$$\begin{aligned} r = \lim_{\theta\to\infty} r_\theta &= \sum_{i=1}^{k} a_i^2\, v_i + \sum_{i=1}^{k} c_i\, n_i \\ &= \sum_{i=1}^{k} a_i^2 \frac{N_i - n_i}{N_i\, n_i} \sigma_i^2 + \sum_{i=1}^{k} c_i\, n_i . \end{aligned} \tag{6.18}$$

All we have to do now is to find some estimate δ^* for which the risk does not exceed r, and by Theorem 2.2, if any such δ^* exists, it will be a minimax estimate for given n_i. Trying $\delta^*(x) = \sum a_i\bar{x}_i(=\lim_{\theta\to\infty} \delta_\theta(x))$, we see that the risk corresponding to δ^* is given by

$$(6.19)\qquad \begin{aligned} R(\omega, \delta^*) &= E_\omega(\delta^* - T)^2 + \sum_{i=1}^{k} c_i n_i \\ &= E_\omega\left[\sum_{i=1}^{k} a_i(\bar{x}_i - u_i)\right]^2 + \sum_{i=1}^{k} c_i n_i . \end{aligned}$$

Noting the fact that the strata are independent and utilizing the result (5.6) for a single stratum, it is seen at once that (6.19) reduces to

$$(6.20)\qquad R(\omega, \delta^*) \leqq \sum_{i=1}^{k} a_i^2 \frac{N_i - n_i}{N_i n_i} \sigma_i^2 + \sum_{i=1}^{k} c_i n_i = r.$$

Hence the usual estimate $\sum a_i\bar{x}_i$ is a minimax estimate for given n_i.

Minimax strategy for choosing the n_i. We now choose the n_i so that the minimax risk for given n_i and largest allowed variances in the strata is minimum under the restriction that the n_i are positive integers and $\leqq N_i$. This risk is given by

$$(6.21)\qquad r = \sum_{i=1}^{k}\left[a_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\sigma_i^2 + c_i n_i\right].$$

This expression differs from that in (6.10) by a quantity which is independent of $n_1, n_2, \cdots, n_k$, and hence is minimized by the same n_i as before provided $n_i \leqq N_i$, and otherwise by $n_i = N_i$.

As a special case, consider the problem of the estimation of the overall mean of the finite population, $\mu = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i} x_{ij}$, where $N = \sum_{i=1}^{k} N_i$. Choosing $a_i = N_i/N$, $T = \sum a_i u_i = \mu$, and we see that a minimax procedure of estimating the mean μ is to choose the n_i as

$$(6.22)\qquad n_i = \text{integer nearest to } \sqrt{\frac{N_i^2\sigma_i^2}{N^2 c_i} + \frac{1}{4}} \text{ and } \leqq N_i,$$

and then employ the usual estimate $N^{-1}\sum N_i\bar{x}_i$, for these n_i. This rule for finding the minimax n_i is more exact than the one commonly stated in the literature, namely the allocation of the total sample size in proportion to $N_i\sigma_i/\sqrt{c_i}$. This greater exactness may be quite useful in the case of high c_i's.

7. Acknowledgement. I wish to express my gratitude to the late Professor M. A. Girshick, who encouraged me to work on this problem and gave generous help and guidance. Thanks are also due to Professors David Blackwell and Herman Rubin for several helpful discussions during the course of this work. I also take this opportunity to record my appreciation of some discussions with Professor John Pratt which helped to clarify some of the concepts during the preparation of this paper.

REFERENCES

[1] David Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[2] W. G. COCHRAN, *Sampling Techniques*, John Wiley and Sons, New York, 1953.
[3] W. E. DEMING, *Some Theory of Sampling*, John Wiley and Sons, New York, 1950.
[4] M. H. HANSEN, W. N. HURWITZ, AND W. G. MADOW, *Sample Survey Methods and Theory*, Vol. II, John Wiley and Sons, New York, 1953.
[5] J. L. HODGES, JR., AND E. L. LEHMANN, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.
[6] E. L. LEHMANN, *Mimeographed notes on the Theory of Estimation*, University of California, Berkeley, Chap. IV, 1950.
[7] P. C. MAHALANOBIS, "On Large Scale Sample Surveys," *Phil. Trans. Roy. Soc. London, Ser. B*, Vol. 231 (1944), pp. 329–451.
[8] P. V. SUKHATME, *Sampling Theory of Surveys with Applications*, Iowa State College Press, Ames, 1954.