The Annals of Statistics
1979, Vol. 7, No. 2, 395–399

A MINIMAX APPROACH TO RANDOMIZATION AND ESTIMATION IN SURVEY SAMPLING

BY H. STENGER

University of Mannheim

We consider a finite set of units, a population. With each unit is associated a real value (unknown to us) and a label (identifying the unit). Based on the labels we may select a sample, i.e., a subset of the population, to estimate the mean of the real values. In simple random sampling (not necessarily of fixed size) the selection probabilities of all samples are not affected by a permutation of the labels.

It is assumed that we have to choose both a sampling design and a linearly invariant estimator, i.e., a linear function of the observed values with the property: equality of the observed values implies that the estimate is equal to this common value. Under these conditions we should use simple random sampling together with the sample mean as an estimator. This follows from the minimax criterion.

1. Introduction. The discussion of Bayesian ideas and superpopulation models has brought the role of randomization in survey design into question. "The only formal justification of a random design that we know that does not depend on the introduction of some form of unbiasedness is that of Blackwell and Girshick (1954, page 229) based on the minimax criterion", state Scott and Smith (1975, page 353). They show by generalizing at least partially the result of Blackwell and Girshick that, once having accepted a certain standard estimator, sampling with unequal probabilities is a minimax decision. Aggarwal (1959) and Royall (1970) describe the complementary approach, first accepting a design and then looking for a minimax estimator. However, a statistician usually has to determine a design and an estimator at the same time. The minimax principle allows us to justify simultaneously a standard design and a standard estimator, as will be shown in this paper.

2. Definitions and notation. The population to be considered consists of N distinct units. With each unit is associated a real number, called a variate value, and an integer, called a label. Labels of distinct units are assumed to be distinct. Without loss of generality we may assume that $\{1, 2, \cdots, N\}$ is the set of labels. The variate value of the unit with label i is denoted by x_i. Then $x = (x_1, x_2, \cdots, x_N)$ is called the parameter vector. X denotes the whole parameter space.

Let S be the collection of all samples, i.e., of all nonempty subsets of $\{1, 2, \cdots, N\}$. For $s \in S$ we denote by $n(s)$ the number of elements in s. Any probability function $p(s)$ is said to be a (sampling) design. If $p(s) > 0$, $p(s') > 0$ implies $n(s) = n(s')$ the design p has a fixed sample size.

Received May 1977; revised January 1978.

AMS 1970 *subject classifications.* Primary 65D05; secondary 62K05.

Key words and phrases. Finite populations, simple random sampling and symmetric estimation, modified minimax principle of Wesler, simultaneous application to randomization and estimation.

Any function $t(s, x)$ depending on x only through the coordinates x_i, $i \in s$ is called an estimator. The sample mean

$$t_0(s, x) = \Sigma_{i \in s} x_i / n(s)$$

is an estimator of special importance. A strategy (p, t) consists of a design p and an estimator t.

Let $L(x, a)$ be a given nonnegative function, which is convex in the real variable a for each x. Then L is called a loss function and associates a risk function

$$R(x; p, t) = \Sigma_{s \in S} p(s)[L(x, t(s, x)) + cn(s)]$$

with each strategy (p, t). In this formula c is a positive constant denoting the cost of drawing one unit from the population. When the population mean $\bar{x} = \Sigma x_i / N$ is to be estimated, we use the loss function

$$L(x, a) = l(\bar{x}, a)$$

where l is convex in a for each $\bar{x}$. It is common to consider linear estimators, where linearity of an estimator $t(s, x)$ means that functions $t_i(s)$ exist satisfying

$$t(s, x) = \Sigma_i t_i(s) x_i \qquad \text{for all} \quad s \in S, x \in X.$$

The definition of an estimator implies that $t_i(s) = 0$ for $i \notin s$.

A linear estimator $t(s, x)$ is said to be linearly invariant (see Roy and Chakravarti, 1960) if

$$\Sigma_i t_i(s) = 1 \qquad \text{for all} \quad s \in S.$$

(The term linearly invariant should be understood as an abbreviation of: linear and invariant with respect to translations.) Clearly, a linear estimator is linearly invariant if and only if equality of all the observed variate values implies that the estimate equals this common value. The sample mean t_0 is linearly invariant.

3. Relabelling and symmetry of sampling strategies. Let $\pi 1, \pi 2, \cdots, \pi N$ be a permutation of the integers $1, 2, \cdots, N$. We relabel the units associating the label πi with the unit originally labelled by i. This change of labels does not affect the variate value of a unit, i.e., with label πi is now associated the same variate value which was associated originally with label i. We denote by πx the parameter vector we obtain by relabelling; then we have

$$(\pi x)_{\pi i} = x_i \qquad \text{(equivalently, } (\pi x)_i = x_{\pi^{-1} i})$$

i.e., the πith component of πx is identical with the ith component of x. By relabelling $s \in S$ is transformed into $\pi s = \{\pi i : i \in s\} \in S$.

Subsequently Π represents the set of all permutations of the integers $1, 2, \cdots, N$. We assume that the parameter space X is symmetric, i.e., $\pi x \in X$ for all $x \in X$, $\pi \in \Pi$. Then Π is a group of one-to-one mappings of X onto X.

A function $f(x)$ is called symmetric if $f(\pi x) = f(x)$ for all $x \in X$, $\pi \in \Pi$. In particular, the population mean $\bar{x} = \Sigma x_i / N$ is symmetric. For a sampling strategy

(p, t) we define

$$\bar{p}(s) = \Sigma_{\pi\in\Pi} p(\pi s)/N!,$$

$$\bar{t}(s, x) = \Sigma_{\pi\in\Pi} t(\pi s, \pi x)/N!.$$

It is easy to see that $(\bar{p}, \bar{t})$ is a sampling strategy. If t is linearly invariant, we have $\bar{t} = t_0$. A strategy (p, t) is said to be symmetric if p and t are both symmetric, i.e., $\bar{p} = p$, $\bar{t} = t$. Any symmetric strategy (p, t) is invariant with regard to relabelling, i.e.,

$$p(\pi s) = p(s) \qquad \text{and} \quad t(\pi s, \pi x) = t(s, x)$$

for all $s \in S$, $x \in X$, $\pi \in \Pi$.

REMARK 1. We have $\bar{p} = p$ if and only if $p(s) = p(s')$ for all $s, s' \in S$ with $n(s) = n(s')$, i.e., there exist $\alpha_1, \alpha_2, \cdots, \alpha_N \geqslant 0$ with $\Sigma_i \alpha_i = 1$ and

$$p(s) = \alpha_{n(s)} \Big/ \binom{N}{n(s)} \qquad \text{for all} \quad s \in S.$$

REMARK 2. For a linear estimator t we have $\bar{t} = t$ if and only if real numbers $\beta_1, \beta_2, \cdots, \beta_N$ exist with

$$t_i(s) = \beta_{n(s)} \qquad \text{for all} \quad i \in s, s \in S.$$

4. A minimax theorem for sampling strategies. Royall (1970) has shown that

$$\max_{\pi\in\Pi} R(\pi x; \bar{p}, \bar{t}) \leqslant \max_{\pi\in\Pi} R(\pi x; \bar{p}, t)$$

for all strategies (p, t). This inequality is a strong justification for the use of a symmetric estimator after a symmetric design has been accepted. (See Royall, 1970, page 1778.) Royall's result is complementary to the well-known result of Blackwell and Girshick (1954, pages 229–233) which states

$$\max_{\pi\in\Pi} R(\pi x; \bar{p}, \bar{t}) \leqslant \max_{\pi\in\Pi} R(\pi x; p, \bar{t})$$

for any design p of fixed sample size. (See Royall, 1970, page 1776.) From the result of Blackwell and Girshick we conclude that we can restrict ourselves to a symmetric design, whenever a symmetric estimator and a design of fixed sample size have to be applied. The restriction on symmetric strategies, however, cannot be justified by combining the results of Royall and of Blackwell and Girshick. Indeed, it is easy to see that

$$\max_{\pi\in\Pi} R(\pi x; \bar{p}, \bar{t}) \leqslant \max_{\pi\in\Pi} R(\pi x; p, t)$$

is not true for all strategies (p, t), not even for all strategies (p, t) with t linear. We have, however:

THEOREM. *Let $L(x, a) = l(\bar{x}, a)$ with l convex in a. Then*

$$\max_{\pi\in\Pi} R(\pi x; \bar{p}, t_0) \leqslant \max_{\pi\in\Pi} R(\pi x; p, t)$$

for all designs p and all linearly invariant estimators t.

PROOF. Suppose $s \in S$. We denote by $\Pi(s)$ the set of all $\pi \in \Pi$ with $\pi i = i$ for all $i \notin s$. If (p, t) is a sampling strategy with t linearly invariant, we have for any

$x \in X, s \in S$

$$(1) \qquad \frac{1}{[n(s)]!}\Sigma_{\varphi\in\Pi(s)}t_{\varphi i}(s) = \frac{1}{n(s)} \quad \text{if} \quad i \in s;$$
$$= 0 \quad \text{if} \quad i \notin s,$$

and for $\varphi \in \Pi(s)$

$$(2) \qquad \Sigma_i t_{\varphi i}(s)x_i = \Sigma_i t_i(s)(\varphi x)_i \left(= \Sigma_i t_i(s)x_{\varphi^{-1}i}\right).$$

By the convexity of l we derive from (1) and (2)

$$l(\bar{x}, t_0(s, x)) \leqslant \frac{1}{[n(s)]!}\Sigma_{\varphi\in\Pi(s)}l(\bar{x}, \Sigma_i t_i(s)(\varphi x)_i)$$

and have, therefore,

$$(3) \quad \Sigma_{\pi\in\Pi}R(\pi x; p, t_0)$$
$$\leqslant \Sigma_{s\in S}\frac{p(s)}{[n(s)]!}\Sigma_{\varphi\in\Pi(s)}\Sigma_{\pi\in\Pi}\left[l(\bar{x}, \Sigma_i t_i(s)(\varphi\pi x)_i) + cn(s)\right].$$

As

$$\Sigma_{\pi\in\Pi}\left[l(\bar{x}, \Sigma_i t_i(s)(\varphi\pi x)_i) + cn(s)\right]$$

is independent of φ we deduce from (3)

$$(4) \qquad \Sigma_{\pi\in\Pi}R(\pi x; p, t_0) \leqslant \Sigma_{\pi\in\Pi}R(\pi x; p, t).$$

Now

$$(5) \qquad \Sigma_{\pi\in\Pi}R(\pi x; \bar{p}, t_0) = \Sigma_{\pi\in\Pi}R(\pi x; p, t_0).$$

As $R(\pi x; \bar{p}, t_0)$ is independent of π, the theorem follows from (4) and (5).

REMARK 3. Using Wesler's (1959) terminology the theorem may be stated as follows: let $L(x, a) = l(\bar{x}, a)$ with l convex in a and let (p, t) be a strategy with t linearly invariant. Then, the strategy $(\bar{p}, t_0)$ is at least as good as the strategy (p, t) (in the modified minimax sense).

REMARK 4. Let p be any design. By $\overset{*}{p}$ we denote the uniquely determined symmetric design with

$$\Sigma_{s\in S}n(s)\overset{*}{p}(s) = \Sigma_{s\in S}n(s)p(s)$$
$$\overset{*}{p}(s) > 0, \overset{*}{p}(s') > 0 \quad \text{implies that} \quad |n(s) - n(s')| \leqslant 1.$$

If $L(x, a) = l_0(\bar{x})(\bar{x} - a)^2$ with l_0 strictly positive, we have (see Ramakrishnan, 1969, (12))

$$R(x; \overset{*}{p}, t_0) \leqslant R(x; \bar{p}, t_0).$$

Combining this result with our theorem gives us

$$\max_{\pi\in\Pi}R(\pi x; \overset{*}{p}, t_0) \leqslant \max_{\pi\in\Pi}R(\pi x; p, t)$$

for all strategies (p, t) with t linearly invariant, i.e., $(\overset{*}{p}, t_0)$ is at least as good (in the modified minimax sense) as (p, t) if t is linearly invariant.

REMARK 5. Let $\xi(x)$ be an exchangeable (i.e., symmetric) prior density. The Bayes risk

$$r_\xi(p, t) = \int R(x; p, t)\xi(x)\, dx$$

of the strategy (p, t) then is an increasing function of the average risk (see Royall, 1970)

$$\bar{R}(x; p, t) = \frac{1}{N!}\Sigma_{\pi\in\Pi}R(\pi x; p, t)$$

of this strategy. Now, from (4) and (5) in the proof of Theorem 1

$$\bar{R}(x; \bar{p}, t_0) \leqslant \bar{R}(x; p, t)$$

and, therefore,

$$r_\xi(\bar{p}, t_0) \leqslant r_\xi(p, t)$$

for all p and all t linearly invariant, i.e., $(\bar{p}, t_0)$ is at least as good as (p, t) in the Bayes sense as long as the prior density is exchangeable.

Acknowledgment. The author wishes to thank the referees for their valuable comments.

REFERENCES

[1] AGGARWAL, O. P. (1959). Bayes and minimax procedures in sampling from finite and infinite populations I. *Ann. Math. Statist.* **30** 206–218.

[2] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[3] GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. Ser. B* **17** 268–278.

[4] RAMAKRISHNAN, M. K. (1969). Some results on the comparison of sampling with and without replacement. *Sankhyā Ser. A* **31** 333–342.

[5] ROY, J. and CHAKRAVARTI, I. M. (1960). Estimating the mean of a finite population. *Ann. Math. Statist.* **31** 392–398.

[6] ROYALL, R. M. (1970). Finite population sampling—on labels in estimation. *Ann. Math. Statist.* **41** 1774–1779.

[7] SCOTT, A. J. and SMITH, T. M. F. (1975). Minimax designs for sample surveys. *Biometrika* **62** 353–357.

[8] WESLER, O. (1959). Invariance theory and a modified minimax principle. *Ann. Math. Statist.* **30** 1–20.

DEPARTMENT OF ECONOMICS AND STATISTICS
UNIVERSITY OF MANNHEIM
6800 MANNHEIM, WEST GERMANY