MULTIPLEX SAMPLING

By David H. Evans[1]

Bell Telephone Laboratories, Whippany, New Jersey

0. Summary. The problem of estimating desired characteristics of a response X, where $X = h(y)$ and y is a vector random variable with statistically independent components from known distributions, may be handled by standard Monte Carlo techniques. We are interested in the generalization where several distributions are of interest, one at a time, for each component of y and the desired characteristics of X must be estimated for each combination of component distributions. A relatively small number of observations, compared to the total number required if each combination were posed as a separate Monte Carlo problem, may be used instead by sampling from a fictitious distribution, calculating an estimate by appropriately weighting the observations, and then reusing the sample. These techniques are standard; the contribution here is to find the fictitious distribution which is best for the characteristics desired, the distributions of interest, and allied considerations. Various concepts of the meaning for "best" are examined in the paper. Finally, a quantitative evaluation is made under restricted conditions.

1. Introduction.

1.1 *General discussion.* Multiplex sampling is a method for estimating characteristics of the response by a sampling technique when the response is a function of several independent variables and each of the several variables is available in a variety of completely known distributions. The need for such a method arose in the author's investigation into so-called statistical tolerancing [5]. In tolerancing each component is available in several known distributions, e.g., resistors come in distributions designated as 1%, 5%, 10% or 20% resistors. To choose a set of appropriate tolerances for the components it is necessary to know the statistical characteristics of the response engendered by each possible choice of component tolerances. To be sure, other data such as the desired response and the associated costs are also required, but we do not concern ourselves with these problems here. We are concerned with the problem of estimating the response characteristics in an efficient manner when a Monte Carlo approach is required and where the efficiency is measured in terms of the number of observations required to achieve a desired precision in the estimates.

1.2. *Statement of the problem.*

(1). Given the response X as a function of the y_j, $j = 1, 2, \cdots n$,

$$X = h(y) = h(y_1, y_2, \cdots, y_n), \tag{1.1}$$

Received January 16, 1961; revised May 23, 1963.

[1] Now at Research Laboratories, General Motors Technical Center, Warren, Michigan.

where h is continuous and may be known either analytically or experimentally and an exact numerical value of X is uniquely determined by a set of values of the y_j.

(2). Given for each $j = 1, 2, \cdots, n$, that y_j will be taken to have one of the $r(j)$ given density functions

$$d_j = \{w_j(1; y_j), w_j(2; y_j), \cdots, w_j(r(j); y_j)\} \tag{1.2}$$

(the jth component is available in corresponding distributions in the tolerancing problem), and that the $y_j, j = 1, 2, \cdots, n$, are statistically independent random variables.

(3). Given that a particular family of distributions is denoted by

$$\beta = \{i_1, i_2, \cdots i_n\}, \qquad 1 \leqq i_j \leqq r(j), \quad j = 1, 2, \cdots, n, \tag{1.3}$$

where y_j is from $w_j(i_j ; y_j)$, $j = 1, 2, \cdots, n$; the range of β is the Cartesian product of $\{1, 2, \cdots, r(j)\}, j = 1, 2, \cdots, n$; there are

$$r = \prod_j r(j) \tag{1.4}$$

different β's.

(4). Find the desired characteristics of the response X as a function of β by an efficient sampling procedure. The concept of an efficient sampling procedure will be made precise. The characteristics, or, as we will call them from now on, the parameters, which will be considered explicitly here are the average of X, the variance of X, and the probability that X is within a given set; other parameters could be considered.

Since we are only considering the case where the y_j are statistically independent random variables we use

$$w(\beta; y) = \prod_{j=1}^{n} w_j(i_j ; y_j), \qquad y = (y_1, y_2, \cdots, y_n) \tag{1.5}$$

and at times we take the liberty of expressing the range of β as $1 \leqq \beta \leqq r$. The y_j will be treated as real random variables for convenience; this convention does not rule out, however, the possibility that they may be vector quantities.

1.3. *Relation to usual and complementary methods.* In order to measure the efficiency of multiplex sampling we must compare it with a more usual procedure. For the purposes of the paper we define the usual method (perhaps somewhat unfairly) as follows: First choose a particular β, draw a random sample of size M_β where the size depends on the desired precision, and then estimate the desired parameters. To obtain estimates for all β, follow the same procedure r times. Note that we allow the use of standard Monte Carlo variance reducing techniques so long as they are also applicable in multiplex sampling; the important difference is that there be r samples. Then, the total number of observations required is

$$T_{\text{usual}} = \sum_\beta M_\beta . \tag{1.6}$$

Multiplex sampling is a method for estimating the same parameters with at least the desired precision by drawing only a single random sample of size N. It is accomplished by sampling from a fictitious distribution which is best under criterions to be adduced and then using weighted sampling formulas for all β of interest. The efficiency of multiplex sampling is measured by comparing N to T.

As noted, in multiplex sampling weighted sampling formulas are used to calculate the parameters; these essentially make the fictitious distribution from which the sample is drawn look like any another distribution of interest [8], [9]. The one sample is used to make the calculations for several distributions of interest [9]. Of course, these techniques are standard in Monte Carlo work as is the selection of the optimum fictitious distribution for the case of a single β [8]. The contribution here is the selection of this distribution when all β are of interest. In finding this distribution the standard Monte Carlo variance reducing device of introducing a tractable function $Z = g(y)$, which approximates the function $X = h(y)$, in addition to h is used [8]; there is a reason for explicitly considering it, see Section 3.5 and also [6]. As implied above, however, other well known Monte Carlo devices [2] (especially [8]) may be used in addition; their applicability will depend on the specific problem.

A fundamental assumption underlying the applicability of multiplex sampling is that the functions involved are either unknown analytically, or so intractable, or the like, that an experimental sampling approach is dictated. If the function h is tractable, one can use Tukey's nonlinear propagation of error formulas [11], [12] and [13]; of course, in the simplest case when h is at least approximately linear the classical propagation of error formulas are adequate.

2. Response statistics.

2.1. *Estimators for parameters of the response.* We require weighted sampling formulas; we derive them in order to have them in the desired forms. Define a fictitious distribution with known probability density function $p(y)$; for now p is arbitrary, and its efficient choice will be taken up in Section 3. Draw a random sample of size N from this distribution; let y^i denote the ith observation.[2] For each y^i there is an X_i by (1.1). Further calculate $w(\beta; y^i)$ and $p(y^i)$ for the ith observation and define the weights

$$q_i(\beta) = w(\beta; y^i)/p(y^i). \tag{2.1}$$

These depend on both the value of y^i and β.

As noted in Section 1.3 we find it especially useful to use an approximating function. Let us choose a function

$$Z = g(y) \tag{2.2}$$

which approximates the function $X = h(y)$ in the range of interest; g does not

[2] A superscript must be used to denote the ith observation of y since the subscript has been preempted; subscripts are used for this purpose elsewhere in this paper.

replace h, they are both used. Assume that g is continuous, and, for simplicity of exposition, that we have an analytic expression for it. We require that the chosen g be sufficiently tractable so that the mean and variance of Z can be computed relatively easily for each $w(\beta; y)$; let $a_g(\beta)$ and $\sigma_g^2(\beta)$ be the mean and variance for each β, respectively. For each y^i calculate the Z_i from (2.2).

Using the above, estimators of the mean and variance of X when y is distributed according to $w(\beta; y)$ are

Mean

$$m(\beta) = (1/N)\{\sum_i (X_i - Z_i)q_i(\beta)\} + a_g(\beta). \tag{2.3}$$

Variance

$$m_2(\beta) = [1/2N(N-1)]\{\sum_i \sum_j [(X_i - X_j)^2 - (Z_i - Z_j)^2]q_i(\beta)q_j(\beta)\} + \sigma_g^2(\beta), \tag{2.4}$$

$$= [1/N(N-1)]\{\sum_j q_j(\beta)\sum_i [X_i^2 - Z_i^2]q_i(\beta) - [\sum_i X_i q_i(\beta)]^2 + [\sum_i Z_i q_i(\beta)]^2\} + \sigma_g^2(\beta), \tag{2.5}$$

where all sums are from 1 to N. The second expression for $m_2(\beta)$, (2.5), is only an algebraic transformation of the first; the second is the form used for numerical computations.

2.2. *Properties of the estimators.* It can be shown easily that $m(\beta)$ and $m_2(\beta)$ are unbiased estimators. In doing so, one must remember that all averages must be taken with respect to p. The variances of the estimates are:

$$\text{var } m(\beta) = \frac{1}{N}\left\{\int [(h(y) - g(y)]^2 w^2(\beta; y)[p(y)]^{-1}\, dy - [a(\beta) - a_g(\beta)]^2\right\}, \tag{2.6}$$

$$\text{var } m_2(\beta) = \frac{1}{N}\left\{\iiint [|\Delta h(x, y)|^2 - |\Delta g(x, y)|^2] \cdot [|\Delta h(x, z)|^2 - |\Delta g(x, z)|^2]w^2(\beta; x)[p(x)]^{-1} \cdot w(\beta; y)w(\beta; z)\, dx\, dy\, dz - 4[\sigma^2(\beta) - \sigma_g^2(\beta)]^2\right\} + O(1/N^2), \tag{2.7}$$

where

$$\Delta h(x, y) \equiv h(x) - h(y), \qquad \Delta g(x, y) \equiv g(x) - g(y), \tag{2.8}$$

and $a(\beta)$ and $\sigma^2(\beta)$ are the mean and variance of X. These calculations are straightforward; the $O(1/N^2)$ term in (2.7) arises essentially from dropping an integral. All the above integrals will be assumed to exist.

From the simple version of the Central Limit Theorem (Theorem 17.4 in [4]) it is easily seen that $m(\beta)$ is asymptotically normally distributed for large N with unbiased mean and variance as above. The proof that $m_2(\beta)$ is asymptotically normally distributed is more difficult and one must appeal to the asymptotic normality of functions of moments (Theorem 28.4, [4]); the proof will not be reproduced here since it is only tangential to the main ideas of the paper.

2.3 *Mean value forms for the variances.* For use in Section 3, let us note that because of the assumed continuity of h and g, from the mean value theorem of integral calculus, it follows that

$$\text{var } m(\beta) = (1/N)\{|h - g|_\beta^2 Q_\beta - (\delta a_\beta)^2\}, \tag{2.9}$$

where

$$\delta a_\beta \equiv a(\beta) - a_g(\beta), \tag{2.10}$$

$$Q_\beta = \int w^2(\beta; y)[p(y)]^{-1}\, dy, \tag{2.11}$$

and

$$|h - g|_\beta^2 \equiv |h(y_\beta) - g(y_\beta)|^2 \tag{2.12}$$

where y_β is some point in the range of integration; y_β depends on $p(y)$ and on β thru $w(\beta; y)$. Similarly

$$\text{var } m_2(\beta) = \frac{1}{N}\{||\Delta h|^2 - |\Delta g|_\beta^2|_\beta^2 Q_\beta - 4(\delta\sigma_\beta^2)^2\}, \tag{2.13}$$

where

$$\delta\sigma_\beta^2 = \sigma^2(\beta) - \sigma_g^2(\beta), \tag{2.14}$$

and

$$||\Delta h|^2 - |\Delta g|^2|_\beta^2 \equiv ||\Delta h(x_\beta, y_\beta)|^2 - |\Delta g(x_\beta, y_\beta)|^2|^2, \tag{2.15}$$

and x_β, y_β are points in the range of integration and terms of $O(1/N^2)$ are dropped. Here, the simplified form of the mean value term (2.15) does not come directly from the mean value theorem, but can be obtained by exploiting the positivity of the variance and the required continuity of h and g. For the class of p functions we will be interested in, it can be shown that Q_β always exists.

2.4 *Other parameters.* Other parameters besides the mean and variance of X can be used with multiplex sampling. In particular, one which has been considered is $P_S(\beta)$, the probability that X is in the set S. The estimator is the same as (2.3) with X_i and Z_i replaced by $\varphi(X_i)$ and $\varphi(Z_i)$ where φ is the characteristic function of the set S, and $a_g(\beta)$ replaced by the probability that Z is in the set S. Because of the close functional relationship of the two estimators almost identical results hold—including the results to be obtained—for the two statistics; the

reasoning needs to be changed to obtain a form equivalent to (2.9) because of the discontinuous nature of φ.

3. The efficient choice of p.

3.1. *The effect of p.* Throughout the preceding Section the function p was an arbitrary probability density function except for the mild constraint imposed by requiring that certain integrals converge. Assuming that the parameter under consideration exists, this constraint only requires essentially that $p(y)$ not be zero when $w(\beta; y) > 0$, all β, and that p go to zero for large $|y|$ no faster than the slowest of the w's. In dictating our approach, these are minor considerations.

The major considerations influencing the choice of p are: (1) the ease of producing random variables according to the distribution defined by p, (2) the size of the random sample, N, required to achieve a desired precision. As will be seen, these usually work against each other; that is, the most efficient p under (2) may necessitate such a complicated sampling procedure that the extra work involved in sampling more than offsets the work saved by minimizing the sample size. The balancing of (1) and (2) is important, but because (1) is not amenable to a precise mathematical definition, the optimum p depends on the interpretation selected. In the following sections the conditions on p, i.e., the conditions under which a minimization is carried out, will be changed. The first will be for optimum p from the viewpoint of minimizing N (for prescribed precision of the estimate) without regard to how difficult it is to construct the associated distribution. In successive steps the conditions will be changed with the idea of making it easier to realize p as a distribution at the expense of increasing N.

Two criterions are used to choose the optimum p; for the various relaxations of the requirements they must be adapted somewhat but essentially they are the following (stated only for the mean $m(\beta)$).

Criterion A: Given $\gamma_1, \gamma_2, \cdots, \gamma_r$, where $\gamma_i \geqq 0$, $\sum_i \gamma_i > 0$, choose p such that $\sum_\beta \gamma_\beta$ var $m(\beta)$ = minimum.

Criterion B (minimax criterion): Given $k_1, k_2, \cdots, k_r$, where $k_i > 0$, choose p such that $\max_\beta [k_\beta \text{ var } m(\beta)]$ = minimum. The k's in Criterion B define *interrelationships* among the var $m(\beta)$ in a precise manner; thus, if $\max_\beta [k_\beta \text{ var } m(\beta)] = B/N$ for a particular p and N, then for every β, var $m(\beta) \leqq B/Nk_\beta$. B is made as small as possible by choosing an optimum p; then at least the desired precision is realized for every β by making N sufficiently large. It will be seen that the γ's in Criterion A also define *interrelationships* among the var $m(\beta)$, but that the interrelationships are only qualitatively stated, i.e., for the optimum p it turns out that the larger γ_β is relative to the other γ_i, the smaller var $m(\beta)$ is relative to the other var $m(i)$. Again N may be adjusted separately to realize the desired precision. On the other hand, it will be seen that the implementation of Criterion A is relatively easy while the implementation of Criterion B is, in general, very difficult. One of the important results of this section will be to show that both criterions, through all their adaptations, lead to the same functional form for p.

3.2. *The most efficient choice of* p. Here let us consider only $m(\beta)$ and $g \equiv 0$, for simplicity, then from (2.6)

$$\text{var } m(\beta) = (1/N)[Q_\beta - a^2(\beta)], \qquad \beta = 1, 2, \cdots, r, \tag{3.1}$$

where

$$Q_\beta = \int h^2(\zeta) w^2(\beta; \zeta)[p(\zeta)]^{-1}\, d\zeta. \tag{3.2}$$[3]

Applying Criterion A gives a simple isoperimetric variational problem. Define

$$\begin{aligned} G(p) &= \sum_\beta \gamma_\beta \text{ var } m(\beta) + \nu \int p(\zeta)\, d\zeta \\ &= \sum_\beta \gamma_\beta [Q_\beta - a^2(\beta)] + \nu \int p(\zeta)\, d\zeta, \end{aligned} \tag{3.3}$$

where ν is a Lagrange multiplier since p must be a probability density function. Taking the first variation and setting it equal to zero gives

$$\delta G(p) = 0 = -\sum_i \gamma_i \int h^2(\zeta) w^2(i; \zeta) p^{-2}(\zeta) \delta p\, d\zeta + \nu \int \delta p\, d\zeta. \tag{3.4}$$

Then

$$p(y) = (1/\mu)|h(y)| \cdot [\sum_i \gamma_i w^2(i; y)]^{\frac{1}{2}}, \tag{3.5}$$

where the normalizing constant μ is

$$\mu = \mu(\gamma) = \int |h(\zeta)| [\sum_i \gamma_i\, w^2(i; \zeta)]^{\frac{1}{2}}\, d\zeta. \tag{3.6}$$

Let us denote the Q_β of (3.2) with $p(y)$ as given by (3.5) and (3.6) as $\bar{Q}_\beta$, i.e.,

$$\begin{aligned} \bar{Q}_\beta = \bar{Q}_\beta(\gamma) = \int \mid h(\eta) \mid w^2(\beta; \eta)\, [\sum_j \gamma_j\, w^2(j; \eta)]^{-1/2}\, d\eta \\ \cdot \int \mid h(\zeta) \mid [\sum_i \gamma_i\, w^2(i; \zeta)]^{1/2}\, d\zeta. \end{aligned} \tag{3.7}$$

A relation of some importance which follows from the above is

$$\mu^2 = \min_p [\sum_i \gamma_i Q_i] = \sum_i \gamma_i \bar{Q}_i\,. \tag{3.8}$$

[3] The letter Q will be used in a somewhat ambiguous manner throughout this section; that is, Q will be used to denote functionals which are not exactly the same functionals. However, all the functionals denoted by Q's have common properties and the notation is used to emphasize these properties.

Next let us consider the behavior of $\bar{Q}_j$ as γ_j is changed. One finds

$$\frac{\partial \bar{Q}_j}{\partial \gamma_j} = \frac{1}{2}\left\{\left[\int \frac{|h(\zeta)|w^2(j;\zeta)}{[\sum_i \gamma_i w^2(i;\zeta)]^{\frac{1}{2}}}\, d\zeta\right]^2\right.$$

(3.9)

$$\left. - \left(\left[\int |h(\xi)|\,[\sum_i \gamma_i w^2(i;\xi)]^{\frac{1}{2}}\, d\xi\right]\left[\int \frac{|h(\eta)|w^4(j;\eta)}{[\sum_l \gamma_l w^2(l;\eta)]^{\frac{3}{2}}}\, d\eta\right)\right\}.$$

By the Schwarz inequality the expression inside the brackets is negative.[4] Hence

$$\partial \bar{Q}_j/\partial \gamma_j < 0, \qquad 0 \leqq \gamma_j \tag{3.10}$$

and $\bar{Q}_j$ is a strictly monotone decreasing function of γ_j; obviously, so is var $m(\beta)$. The minimum value of $\bar{Q}_j$ is

$$\min \bar{Q}_j = [\alpha(j)]^2, \qquad \gamma_j > 0, \qquad \gamma_k = 0, \qquad k \neq j, \tag{3.11}$$

where $\alpha(j)$ is the first absolute moment of $h(x)$ when x is from distribution $w(j;\, x)$. The maximum value of $\bar{Q}_j$, considered only as a function of γ_j, is attained for $\gamma_j = 0$.

Next consider the minimax criterion—Criterion B. Primarily, we want to demonstrate that

$$\inf_p \{\max_{1\leqq i\leqq r} k_i[Q_i - a^2(i)]\} = \min_\gamma \{\max_{1\leqq i\leqq r} k_i[\bar{Q}_i(\gamma) - a^2(i)]\}, \tag{3.12}$$

where on the left the Q_i are as in (3.2) and the minimization is over all probability density functions, and on the right the $\bar{Q}_i(\gamma)$ are as in (3.7) and the minimization is over all non-zero vectors $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_r)$ with non-negative components. There are two initial observations: Since p of (3.5) is a homogeneous function of γ we may, for convenience and with no loss, consider only γ satisfying

$$\gamma_i \geqq 0, \qquad \sum_i \gamma_i/k_i = 1, \qquad i = 1, 2, \cdots, r. \tag{3.13}$$

Second, it is obvious that the expression on the left in (3.12) is less than or equal to the expression on the right; hence only the converse conditional inequality need be demonstrated. By Theorem 1, [7], for any γ satisfying (3.13)

$$\max_i k_i[Q_i - a^2(i)] \geqq \sum_i \gamma_i[Q_i - a^2(i)]. \tag{3.14}$$

The minimization over p of the right hand side of (3.14) for any γ satisfying (3.13) is the same isoperimetric variational problem considered above, (3.3)

[4] It can be zero only if $w^2(j;\, y) = c\sum_i \gamma_i w^2(j;\, y)$ where c is a constant of proportionality—see Theorem 18, p. 132, [7]. This is a degenerate condition and hence has been excluded from consideration.

through (3.7). Hence,

$$\begin{aligned}\inf_p \{\max_i k_i[Q_i - a^2(i)]\} &\geqq \inf_p \{\max_\gamma \sum_i \gamma_i[Q_i - a^2(i)]\} \\ &= \max_\gamma \sum_i \gamma_i[\bar{Q}_i(\gamma) - a^2(i)].\end{aligned} \tag{3.15}$$

To complete the proof, it must be shown that

$$\max_\gamma \{\sum_i \gamma_i[\bar{Q}_i(\gamma) - a^2(i)]\} \geqq \min_\gamma \{\max_i k_i[\bar{Q}_i(\gamma) - a^2(i)]\}. \tag{3.16}$$

First, from (3.6) and (3.8)

$$M(\gamma) = \mu^2(\gamma) - \sum_i \gamma_i a^2(i) = \sum_i \gamma_i[\bar{Q}_i(\gamma) - a^2(i)]. \tag{3.17}$$

Adapting a theorem in nonlinear programming, [10], pp. 249f: Any γ^o that maximizes $M(\gamma)$ subject to (3.13) must satisfy

$$B \geqq k_i \frac{\partial M}{\partial \gamma_i}\bigg|_{\gamma^o} = k_i[\bar{Q}_i(\gamma^o) - a^2(i)], \qquad i = 1, 2, \cdots, r, \tag{3.18}$$

and

$$B = \sum_i \gamma_i^o \frac{\partial M}{\partial \gamma_i}\bigg|_{\gamma^o} = M(\gamma^o) \tag{3.19}$$

for some B. Using (3.18) and (3.19), in (3.16) the term inside the brackets on the right when evaluated at γ^o is equal to the left hand side; hence (3.16) is true.

In addition to demonstrating (3.12) under fairly general conditions, the above has also demonstrated a dual i.e., the maximum of $M(\gamma)$ under (3.13) is equal to the expressions in (3.12). The theorem used above, [10], also states that if $M(\gamma)$ is strictly concave, then at most one γ^o satisfies (3.13), (3.18) and (3.19). Let $\gamma' \neq \bar{\gamma}$ both satisfying (3.13) and set

$$\gamma = \theta\gamma' + (1 - \theta)\bar{\gamma}, \qquad 0 < \theta < 1, \tag{3.20}$$

which also satisfies (3.13). By Minkowski's inequality[5]

$$\begin{aligned}\mu^2(\gamma) &= \left(\int |h|\,[\theta \sum_i \gamma_i' w^2(i;\zeta) + (1-\theta)\sum_i \bar{\gamma}_i w^2(i;\zeta)]^{\frac{1}{2}}\,d\zeta\right)^2 \\ &> \theta\left(\int |h|\,[\sum_i \gamma_i' w^2(i;\zeta)]^{\frac{1}{2}}\,d\zeta\right)^2 \\ &\quad + (1-\theta)\left(\int |h|\,[\sum_i \bar{\gamma}_i w^2(i;\zeta)]^{\frac{1}{2}}\,d\zeta\right)^2 = \theta\mu^2(\gamma') + (1-\theta)\mu^2(\bar{\gamma}).\end{aligned} \tag{3.21}$$

Since M differs from μ^2 only by a linear term M is strictly concave, also.

[5] Theorem 198, p. 146, [7]. Equality is possible only if the w^2 are linearly dependent; we exclude this degenerate condition from consideration, as before.

The significance of the above, from the standpoint of finding the desired γ^o, is that if one can find a γ^* such that either all $k_i[\bar{Q}_i(\gamma^*) - a^2(i)]$ are equal to some number, or for each exceptional β then $k_\beta[\bar{Q}_\beta(\gamma^*) - a^2(\beta)]$ is less than that number and $\gamma^*_\beta = 0$, then γ^* is proportional to γ^o and the number is B. Further, γ^* is a minimizing γ in (3.12) and (3.12) is equal to B.

3.3. *The case where p is independent of h.* The results of the foregoing section are satisfactory on theoretical grounds; however, in the practical case they are generally useless. There are two objections:

Objection (1): $p(y)$ as given by (3.5) is useful for a single variable. If one replaces y by $(y_1, y_2, \cdots, y_n)$, for any but the most trivial case, it is obvious that it would be extremely difficult, if at all possible, to produce the required random sample.

Objection (2): $p(y)$ as given by (3.5) depends on the function $h(y)$ as a function of y (or if g were included it would depend on $[h(y) - g(y)]$). Since the formulas of Section 2 are predicated on the assumption that p is a probability density function it is necessary to integrate (3.6) to obtain the normalizing constant μ, but since the basic assumption is that a Monte Carlo approach is required, it would be a rare case indeed when it could be evaluated.

Objection (2) will be dealt with in the present section essentially as a prelude to dealing with Objection (1) in Section 3.4.

We use the mean value expressions for the variances given in Section 2.3. Now Criterion A cannot be applied directly to the sum of the variances when the variances are in the form (2.9), for in the expression

$$\sum_\beta \gamma_\beta \operatorname{var} m(\beta) = (1/N)\{\sum_\beta \gamma_\beta[|h - g|^2_\beta Q_\beta - (\delta a_\beta)^2]\} \tag{3.22}$$

the $|h - g|_\beta$ depend on p. However, since var $m(\beta)$ varies directly with Q_β, an adaptation of Criterion A can be used.

Criterion A′. Given $\lambda_1, \lambda_2, \cdots, \lambda_r$ where $\lambda_i \geqq 0$, $\sum\lambda_i > 0$, choose p such that $\sum_\beta \gamma_\beta Q_\beta$ = minimum. Using this criterion and applying the same variational technique[6] as before one finds

$$p(y) = (1/\mu)[\sum_\beta \gamma_\beta w^2(\beta; y)]^{\frac{1}{2}}, \tag{3.23}$$

where

$$\mu = \int [\sum_\beta \lambda_\beta w^2(\beta; \zeta)]^{\frac{1}{2}}\, d\zeta. \tag{3.24}$$

Now, what has been accomplished by the above? The principal benefit is that $p(y)$ as given by (3.23) is independent of $[h(y) - g(y)]$, as a function of y. In turn, the chance of performing the required integration (3.24) is much better and the distribution is easier to construct (although Objection (1) still holds). The disadvantage is, of course, that $\sum \gamma_\beta$ var $m(\beta)$ is generally greater than its

[6] There is a procedural hitch here which one must remove to be mathematically correct. The results, however, are unchanged, see Appendix I.

minimum value (obviously, it cannot be less than that given in Section 3.2), or, equivalently, to obtain the same precision of estimate N, the sample size, must be larger.

The correspondence between the γ_β of Criterion A and the λ_β of Criterion A′ is

$$\lambda_\beta = \gamma_\beta |h - g|_\beta^2, \tag{3.25}$$

and, thus,

$$\begin{aligned}\sum_\beta \gamma_\beta \operatorname{var} m(\beta) &= \sum_\beta (1/N)\{\gamma_\beta |h - g|_\beta^2 Q_\beta - \gamma_\beta(\delta a_\beta)^2\} \\ &= (1/N)\sum_\beta \lambda_\beta\{Q_\beta - (\delta a_\beta)^2/|h - g|_\beta^2\}.\end{aligned} \tag{3.26}$$

Hence Criterion A′ is a minimization[7] of the right hand side of (3.22) given the λ_β, and the γ_β are the resultant numbers. However, the desired procedure is to go from given γ_β to resultant λ_β. In the theoretical case obtaining λ_β from (3.25) given the desired γ_β is difficult because $|h - g|_\beta^2$ depends on λ_β through the function p. However, in the practical case it does not impose any additional difficulties since the $|h - g|_\beta^2$ can only be approximations and thus the λ_β can only be approximations. Since the whole procedure is aimed at overcoming a difficulty inherent in the practical case this is consistent.

Next, it is important to note that if one starts with the statistic $m_2(\beta)$, since the expression for var $m_2(\beta)$, (2.13), parallels precisely the formula for var $m(\beta)$, the results will be identical. Thus, one can perform the Monte Carlo to determine both $m(\beta)$ and $m_2(\beta)$ at the same time and from the same distribution $p(y)$, although one may have to compromise on the choice of the λ_β used in $p(y)$.

Now, let us consider the minimax criterion—Criterion B. The proper adaptation for this section is *Criterion* B′. Given $\kappa_1, \kappa_2, \cdots, \kappa_r$ where $\kappa_i > 0$ find p such that $\max_\beta[\kappa_\beta(Q_\beta - R_\beta)]$ = minimum, where R_β is a constant and $0 \leqq R_\beta \leqq 1$.

Completely analogous relationships exist between the Q_β of this section and the $\bar{Q}_\beta$ of Section 3.2—simply set $h \equiv 1$ in $\bar{Q}_\beta$. Hence the results carry over and p satisfying Criterion B′ has the functional expression (3.23). Note that $\min \bar{Q}_j = 1$ now for $\lambda_j > 0$, $\lambda_l = 0$, $l \neq j$; hence the restriction on R_β.

Criterion B′ is not as theoretically satisfying as desired, nor can it be easily made so. However, it is satisfactory for the practical case. The following is the reason behind this statement: Set

$$\kappa_\beta = k_\beta |h - g|_\beta^2, \tag{3.27}$$

then

$$\begin{aligned}k_\beta \operatorname{var} m(\beta) &= (1/N)k_\beta\{|h - g|_\beta^2 Q_\beta - (\delta a_\beta)^2\} \\ &= (1/N)\kappa_\beta\{Q_\beta - (\delta a_\beta)^2/|h - g|_\beta^2\}.\end{aligned} \tag{3.28}$$

[7] See Appendix I for the precise sense.

Thus, Criterion B′ is not theoretically satisfying because it sets

$$R_\beta = (\delta a_\beta)^2/|h - g|_\beta^2 = \text{constant}. \tag{3.29}$$

Obviously (3.29) is not true in general, but, also, carrying out the minimization over the λ's when R_β is not constant would be extremely difficult.[8] However, in the practical case where (δa_β) is not known and $|h - g|_\beta^2$ is an estimate, choosing R_β equals some constant is a reasonable procedure. The easiest case to handle is the one where $(\delta a_\beta)^2/|h - g|_\beta^2$ is sufficiently small compared to Q_β to be set equal to zero. Of course, the same remarks hold in choosing κ_β from the desired k_β as did in the case of choosing the λ_β from the desired γ_β.

Obviously, precisely the same result as above holds true if $m_2(\beta)$ is considered instead of $m(\beta)$.

Summing up, the results of this section are primarily acceptable, when y is a single variable. Otherwise, for $p(y)$ as in (3.23), if y is a multidimensional variable, even assuming one can carry out the integration to obtain μ (which is not easy in the one variable case), it would in general be quite difficult to construct the distribution. If one can construct it then the results of this section are applicable, if one cannot, the next section is appropriate.

3.4. *The case where* $p(y) = \prod_j p_j(y_j)$. As a final concession to practical difficulties it will be required in this section that

$$p(y) = p(y_1, y_2, \cdots, y_n) = \prod_{j=1}^{n} p_j(y_j). \tag{3.30}$$

The construction of a one dimensional distribution with density $p_j(y_j)$ is a straightforward and relatively simple procedure [3], [1] pp. xxii–xxiv, and thus Objection (1), Section 3.3, is overcome. Under Condition (3.30) the expression for var $m(\beta)$ becomes

$$\text{var } m(\beta) = (1/N)\{|h - g|_\beta^2 \prod_j Q_{ji_j} - (\delta a_\beta)^2\}, \tag{3.31}$$

and for var $m_2(\beta)$

$$\text{var } m_2(\beta) = (1/N)\{\big||\Delta h|^2 - |\Delta g|^2\big|_\beta^2 \prod_j Q_{ji_j} - 4(\delta\sigma_\beta^2)^2\}, \tag{3.32}$$

where

$$\beta = \{i_1, i_2, \cdots, i_n\}, \tag{3.33}$$

and

$$Q_{jl} = \int w_j^2(l; \zeta)[p(\zeta)]^{-1}\, d\zeta \tag{3.34}$$

It is again sufficient to consider only var $m(\beta)$ because of the parallel nature

[8] See Appendix I, last paragraph.

of (3.31) and (3.32). In order to find adaptations of Criterions A and B which are consistent with the constraint on p, it is necessary to turn the problem into an essentially single variable problem. To this end, let the set of numbers $\gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jr(j)}$ where $\gamma_{ji} \geqq 0$ and $\sum_i \gamma_{ji} > 0$ be given for $j = 1, 2, \cdots, n$. Now for β as given by (3.33) define

$$\beta(j:t) = \{i_1, i_2, \cdots, i_{j-1}, t, i_{j+1}, \cdots, i_n\} \tag{3.35}$$

where t is in the jth place, $1 \leqq t \leqq r(j)$. Further, let

$$\gamma'_\beta(j:t) = \prod_l{}' \gamma_{li_l} \tag{3.36}$$

where the prime indicates $l \neq j$. Form the sum

$$\sum_{ac'} \gamma'_{\beta(j:t)} \operatorname{var} m(\beta(j:t)) \tag{3.37}$$

where ac' indicates the sum is taken over all combinations of the $\{i_1, i_2, \cdots, i_n\}$ with t in the jth place. Expression (3.37) represents the variance of $m(\beta)$ over all possible assignments in which y_j is distributed according to $w_j(t; y_j)$; it is essentially a one variable form. Then

$$\sum_t \gamma_{jt}[\sum_{ac'} \gamma'_{\beta(j:t)} \operatorname{var} m(\beta(j:t))] \tag{3.38}$$

corresponds, as far as the jth variable is concerned, to the thing which is minimized under Criterion A′. Introducing (3.31) for var $m(\beta)$ into (3.38) gives an expression which is a weighted sum of the Q_{jt}. It follows than that the proper adaptation of Criterion A is

Criterion A″. For each $j = 1, 2, \cdots, n$, given $\lambda_{j1}, \lambda_{j2}, \cdots, \lambda_{jr(j)}$, such that $\lambda_{ji} \geqq 0$ and $\sum_i \lambda_{ji} > 0$, choose $p_j(y_j)$ such that $\sum_i \lambda_{ji} Q_{ji}$ = minimum. Of course, as before,

$$p_j(y) = (1/\mu_j)[\sum_i \lambda_{ji} w_j^2(i; y)]^{\frac{1}{2}}, \tag{3.39}$$

where

$$\mu_j = \int [\sum_i \lambda_{ji}\, w_j^2(i; \zeta)]^{\frac{1}{2}}\, d\zeta. \tag{3.40}$$

One can obtain algebraic expression for the λ_{ji} in terms of the other parameters, of course. Clearly, if the $m_2(\beta)$ statistic is considered, the same $p_j(y)$ is obtained.

Next, looking at Criterion B, it is apparent that the same, essentially single variable approach used to adapt Criterion A to Criterion A″ will work again. The result is

Criterion B″. For each $j = 1, 2, \cdots, n$, given $\kappa_{ji} > 0$, choose $p_j(y_j)$ such that $\max_l[\kappa_{jl}(Q_{jl} - R_{jl})]$ = minimum, where R_{jl} is some constant and $0 \leqq R_{jl} \leqq 1$. Obviously, remarks analogous to the remarks following Criterion B′ of Section 3.3 are again applicable. In particular, it follows that the optimal $p_j(y_j)$ again has the functional expression (3.39).

3.5. *Practical aspects.* The most important result of the section is that the expression for $p(y)$ has been obtained for reasonable meanings of optimality under constraints imposed by practical considerations. The application of this result, and others, is the subject of another paper [6]; however, let us look at the problem briefly: One starts with the premise that p be given by (3.30), (3.39) and (3.40) and hence that the variances of $m(\beta)$ and $m_2(\beta)$ are given by (3.31) and (3.32). For the moment, looking at the case where only one β is considered and the random sample is drawn from a distribution with $p(y) = w(\beta; y)$—this technique is called usual sampling in the Introduction—it is seen from (2.2), (2.3), and (2.5) that our estimators reduce to the ordinary estimators and, hence, that the ordinary estimators have variances given by (3.31) and (3.32) with, however, $Q_{ji_j} = 1$, all j. Clearly, then, the problem of achieving preset levels of precision for the estimates is, in principle, no different for multiplex sampling than for usual sampling. The chief difference is that the λ_{jl}—which determine the Q_{jm}—must be chosen to achieve the desired relative levels of precision. The approximating function g can be helpful in this respect, for, in addition to the fact that it is a variance reducer, it tends simultaneously to induce a uniformity condition, viz, a well chosen g is such that

$$(3.41)\qquad \begin{aligned} |h - g|_\beta^2 &\approx K_1, && \text{all } \beta,\\ \big||\Delta h|^2 - |\Delta g|^2\big|_\beta^2 &\approx K_2, && \text{all } \beta. \end{aligned}$$

A well chosen g simplifies the practical procedure enormously.

4. Evaluation of multiplex sampling.

4.1. *Some preliminary definitions.* In the following, to show that a quantity depends on β and that usual sampling is implied—but only where the latter fact may make a difference—the compound notation βu will be used. For example, $M_{\beta u}$ is the sample size in usual sampling for y distributed according to $w(\beta; y)$; on the other hand, $a(\beta)$ is always used for the mean since $a(\beta) = a(\beta u)$.

To compare multiplex sampling and usual sampling, at first only for $m(\beta)$, let k_β^{-1} be the preset requirement on the precision of the estimate. Thus, from Section 2.3 since $Q_{\beta u} = 1$ (see Section 3.5) for usual sampling, the requirement is

$$(4.1)\qquad k_\beta^{-1} = \operatorname{var} m(\beta u) = (1/M_{\beta u})[|h - g|_{\beta u}^2 - (\delta a_\beta)^2], \qquad \text{all } \beta,$$

where $M_{\beta u}$ has been adjusted to achieve equality. For multiplex sampling, since equality is not always possible for any specific β, the requirement is

$$(4.2)\qquad k_\beta^{-1} \geqq \operatorname{var} m(\beta) = (1/N)[|h - g|_\beta^2 Q_\beta - (\delta a_\beta)^2], \qquad \text{all } \beta,$$

where Q_β is given by (2.11). The ratio of the total number of observations required in each case is

$$(4.3)\qquad C = \frac{N}{\sum_\beta M_{\beta u}} = \frac{\max_\beta k_\beta[|h - g|_\beta^2 Q_\beta - (\delta a_\beta)^2]}{\sum_\beta k_\beta[|h - g|_{\beta u}^2 - (\delta a_\beta)^2]}.$$

This is a general formula; to make it more tractable we take the case where g is well chosen, Section 3.5. Under the circumstances we make the quite reasonable additional simplifying assumption that

$$|h - g|^2_{\beta u} \approx |h - g|^2_{\beta} \approx K, \qquad \text{all } \beta. \tag{4.4}$$

Also, only the case where $(\delta a_\beta)^2$ is negligible compared to K will be considered. Hence, (4.3) becomes

$$C = \{\max_\beta [k_\beta Q_\beta]\} / \sum_\beta k_\beta . \tag{4.5}$$

The identical result is obtained for $m_2(\beta)$ under analogous assumptions and what follows will hold for both.

The multivariable case $y = (y_1, \cdots y_n)$ of Section 3.4 is of primary interest. For $p = \prod_j p_j$, (3.30),

$$k_\beta = \prod_j k_{ji_j}, \qquad Q_\beta = \prod_j Q_{ji_j}, \qquad \beta = (i_1, \cdots, i_n). \tag{4.6}$$

Thus (4.5) becomes

$$C = \prod_j \max_{1 \leqq t \leqq r(j)} [k_{jt} Q_{jt}] \Big/ \sum_{i=1}^{r(j)} k_{ji} . \tag{4.7}$$

This result has been arrived at under a set of reasonable assumptions; indeed, in any practical application of multiplex sampling these are the assumptions one would ordinarily strive to make possible. Here, we take (4.7) as the definition of the measure of the *over-all efficiency* of multiplex sampling as compared to usual sampling in the multivariable case. For the investigation in this section we will be content to compute only a single term in the product (4.7); the fact that (4.7) is a product should be remembered, however. Also, we consider only the special case in which the k_{jt}, $t = 1, 2, \cdots, r(j)$, are all equal. Thus, the definition reduces to

$$C = \max_t [Q_{jt}]/r(j); \tag{4.8}$$

the smaller C, the more efficient the process.

The ratio (4.7) can be computed for any $p = \prod_j p_j$, (3.30), but, in particular, Section 3 has shown that the p_j of interest are given by (3.39). There is an arbitrariness in the λ_{jt}, $t = 1, 2, \cdots, r(j)$, (see the discussion following (3.26) and also [6]) to be investigated, which leads us to another kind of efficiency — an *internal efficiency*. The measure of the internal efficiency should compare the sample size required by the set of λ_{jt} under investigation to that required by an optimum set which, from (4.7), could be accomplished for each j by comparing $\max_t [k_{jt} Q_{jt}]$ evaluated for the set of λ_{jt} to its minimum value. But in general it is very difficult to find the minimum value, so this definition is not useful. It has been shown in Section 3 that the minimum is realized if $k_{j1}Q_{j1} = k_{j2}Q_{j2} = \cdots = k_{jr(j)}Q_{jr(j)}$—when equality is possible—and that any departure from the minimum is reflected in the disparity of the sizes

of the $k_{jt}Q_{jt}$. Here, we will be interested only in the case $r(j) = 2$ and $k_{j1} = k_{j2}$, and, hence, can take the measure of the internal efficiency as

$$D = Q_{j1}/Q_{j2}. \tag{4.9}$$

The choice of the λ_{jt} is optimum when this ratio is unity; any departure from unity, up or down, indicates inefficiency.

In order to carry out the investigation into these efficiencies it is necessary to go to special distributions and to develop some machinery.

4.2. *The uniform distribution.* The only set of densities $w_j(t; x_j)$, $t = 1, 2, \cdots, r(j)$, that I know for which the required integrations can be performed without undue difficulty is the set of uniform distributions. Let us simplify the notation in a manner consistent with the single variable approach of Section 3.4 and the present: drop the j from y_j, Q_{jt}, p_j, μ_j, $r(j)$, $w_j(t; y_j)$, λ_{jt}, and k_{jt} from the appropriate definitions and formulas in Section 3.4. The equations to be used here become

$$Q_t = \int [w^2(t; \xi)/p(\xi)]\, d\xi, \qquad t = 1, 2, \cdots, r, \tag{4.10}$$

$$\mu = \int [\sum_t \lambda_t w^2(t; \xi)]^{\frac{1}{2}}\, d\xi, \tag{4.11}$$

$$p(x) = (1/\mu)[\sum_t \lambda_t w^2(t; x)]^{\frac{1}{2}}. \tag{4.12}$$

The set of uniform distributions to be investigated are those with mean zero and are defined by

$$w(t; x) = \begin{cases} 1/2a_t, & |x| < a_t, \\ 0, & |x| > a_t, \end{cases} \qquad t = 1, 2, \cdots, r, \tag{4.13}$$

where

$$0 = a_0 < a_1 < a_2 < \cdots < a_r. \tag{4.14}$$

One finds that

$$\begin{aligned} \mu = {} & (a_1 - a_0)\left[\sum_{i=1}^{r} \lambda_i/a_i^2\right]^{\frac{1}{2}} + (a_2 - a_1)\left[\sum_{i=2}^{r} \lambda_i/a_i^2\right]^{\frac{1}{2}} \\ & + \cdots + (a_j - a_{j-1})\left[\sum_{i=j}^{r} \lambda_i/a_i^2\right]^{\frac{1}{2}} + \cdots + (a_r - a_{r-1})[\lambda_r/a_r^2]^{\frac{1}{2}}, \end{aligned} \tag{4.15}$$

and

$$a_t^2 Q_t = a_{t-1}^2 Q_{t-1} + \left\{\mu(a_t - a_{t-1}) \Big/ \left[\sum_{i=t}^{r} \lambda_i/a_i^2\right]^{\frac{1}{2}}\right\}, \tag{4.16}$$

$$t = 1, 2, \cdots, r; Q_0 = 0.$$

Thus, given the λ's the Q's can be found.

The converse problem of being given the desired Q_t, $t = 1, 2, \cdots, r$—or, more precisely, being given the desired ratios among the Q_t's—is the more important problem. This is embodied in Criterion B″ of Section 3, which is recopied here, for convenience.

Criterion B″. Given k_1, $k_2 \cdots k_r$ such that $k_i > 0$, choose p such that $\max_t[k_t(Q_t - R_t)] =$ minimum where the R_t are constants and $0 \leqq R_t \leqq 1$. For the case now being considered $R_t = 0$, for all t.

In Appendix II the realization of Criterion B″ for the uniform distributions with mean zero is accomplished for general k_i, and, while the realization is not difficult, it is tedious. For use in this Section, for $k_1 = k_2 \cdots = k_r = 1$—for which it is shown in Appendix II that

$$Q_1 = Q_2 = \cdots = Q_r \equiv Q \tag{4.17}$$

is always realizable—the λ_i are given by

$$\begin{aligned} &\lambda_i = a_i^2/(a_i + a_{i-1})^2 - a_i^2/(a_{i+1} + a_i)^2, \quad i = i, 2, \cdots, r-1, \\ &\lambda_r = a_r^2/(a_r + a_{r-1})^2. \end{aligned} \tag{4.18}$$

4.3. *A suggested approximation to Criterion* B″. For the uniform distribution (4.13), σ_t is proportional to a_t where σ_t^2 is the variance of y distributed according to $w(t; y)$; thus, one is led to expect that for families of distributions which resemble the set of centered uniform distributions, a_t may be replaced by σ_t to obtain approximate formulas for the λ's. It is understood that the σ's are ordered such that

$$\sigma_1 < \sigma_2 < \cdots < \sigma_r . \tag{4.19}$$

This is only a suggested approximation. No relations have been found which put a quantitative measure on the word "resemble". A partial check on this approximation will be given when we consider two triangular distributions. It will be seen that the approximation is excellent for this specific example.

4.4. *Comparison for two distributions.* The remainder of this section will be devoted to the purpose of justifying multiplex sampling from the standpoint of showing a significant reduction in the number of observations required when compared to the usual method applied over and over again. In this section we will examine the effect of the disparity of the ranges or "widths" of two distributions on the efficiencies; in Section 4.5 we will examine the effect of interpolating distributions on the over-all efficiency, only.

Here we will examine the efficiencies for two uniform distributions—(4.13) with $r = 2$—and the efficiencies for two triangular distributions

$$w(t; x) = \begin{cases} (1 - |x|/a_t)/a_t , & 0 \leqq |x| < a_t , \\ 0, & |x| > a_t , \end{cases} \qquad t = 1, 2. \tag{4.20}$$

Define

$$\rho = a_2/a_1 , \qquad \rho > 1; \tag{4.21}$$

all the results will be plotted against the parameter ρ. Since only the relative sizes of λ_1 and λ_2 are important, it is convenient to normalize to

$$\lambda_1 = 1, \qquad \lambda_2 = \lambda. \tag{4.22}$$

What values of λ do we wish to use? Certainly we would like the result for the optimum λ, which is

$$\lambda = \rho/(\rho + 2) \tag{4.23}$$

for the uniform distributions by (4.18), i.e., for this choice $Q_1 = Q_2$. For the triangular distributions the optimum λ is not known; the suggested approximation from Section 4.3 gives the same value for λ as (4.23). Finally, to get some idea of the effect of the choice of λ, we will also use the value $\lambda = 1$; it is a natural choice to examine since this choice gives equal weights to $w(1; x)$ and $w(2; x)$. Q_1 and Q_2 are easily calculated from (4.15) and (4.16) for the uniform distributions, and they are calculated from the basic formulas (4.10) thru (4.12) for the triangular distributions; the results are shown in Figure 1.

There are a number of important observations about the behavior of the curves. The measure of over-all efficiency is obtained by dividing the value given on the appropriate curve, for any fixed ρ, by 2, (4.8). The over-all efficiency is

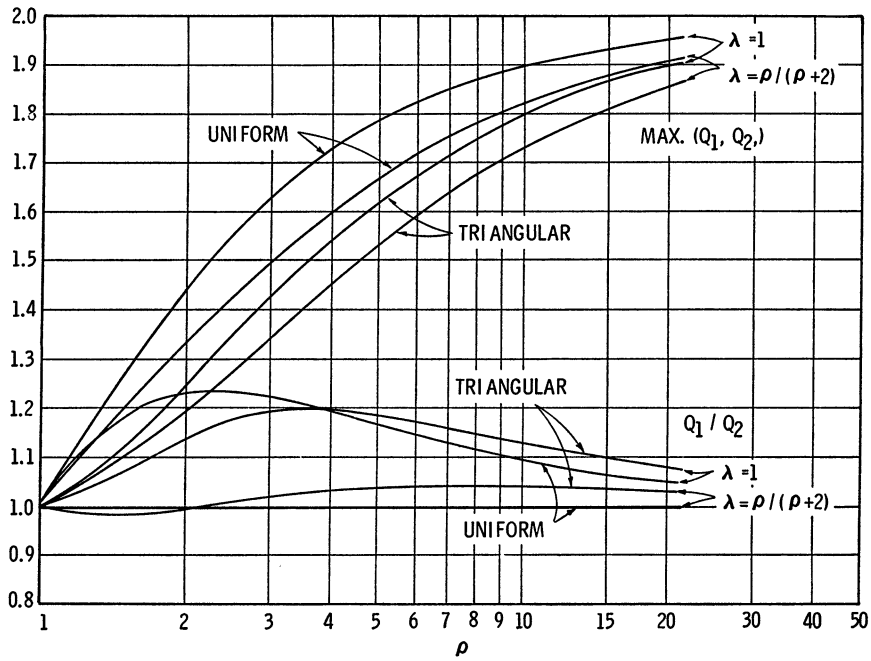

FIG. 1. Measures of efficiency as functions of ρ for the four cases of Section 4.4. For the upper curves the ordinate is max (Q_1, Q_2) and the measure of over-all efficiency is $C = \max (Q_1, Q_2)/2$. For the lower curves the ordinate is the ratio $D = Q_1/Q_2$, the measure of internal efficiency.

high for ρ small while up around 10 or 20 the saving is insignificant, as is to be expected. In the neighborhood of $\rho = 2$—which is a common value in practice in statistical tolerancing, e.g., see [6]—the savings are significant for all distributions, varying from a low of about 55% up to a high of about 80%. I believe that this shows that multiplex sampling is worthwhile. Next, looking at the internal efficiency, the approximation $\lambda = \rho/(\rho + 2)$ is excellent for the triangular distributions. One sees that Q_1 differs from Q_2 by just a little over 4% at the worst, which for practical purposes would make the precision of the estimates for the two triangular distributions identical. However, even if equal mixtures of the two distributions are used, i.e., $\lambda = 1$, the precision of the estimates are not completely out of line; unfortunately, the difference is most serious in the neighborhood of $\rho = 2$. Finally, comparing the uniform and triangular distributions for like values of λ, for both $\lambda = \rho/(\rho + 2)$ and for $\lambda = 1$ the curve $\max[Q_1, Q_2]$ for the triangular distributions lies well below the curve for the uniform distribution.

4.5. *Interpolating distributions.* It is heuristically evident that the maximum benefit of multiplex sampling is realized in interpolation. For a quantitative indication, consider the special case of r uniform distributions with mean zero, where the extreme distributions have half ranges a_1 and a_r and each intermediate distribution has a range which is the geometric mean of its nearest neighbor distributions, i.e., set

$$a_j = \theta a_{j-1} = \theta^{j-1} a_1, \qquad j = 1, 2, \cdots, r, \tag{4.24}$$

and, as before,

$$\rho = a_r/a_1, \qquad \rho > 1, \tag{4.25}$$

and, thus,

$$\theta = \rho^{1/(r-1)}. \tag{4.26}$$

Taking the case $Q_j = Q(r), j = 1, 2, \cdots, r \geqq 2$, from (4.18) it is found, after a suitable normalization, that

$$\begin{aligned} \lambda_1 &= \theta(\theta + 2)/(\theta^2 - 1), \\ \lambda_t &= 1, \qquad t = 2, 3, \cdots, r - 1, \\ \lambda_r &= \theta^2/(\theta^2 - 1). \end{aligned} \tag{4.27}$$

From (4.15) and (4.16), after some algebra, one finds that

$$Q(r) = \frac{r(\theta - 1) + 2}{\theta + 1} = \frac{r[\rho^{1/(r-1)} - 1] + 2}{\rho^{1/(r-1)} + 1}. \tag{4.28}$$

The expression for Q holds for all $r \geqq 2$ and for $\rho > 1$. The limiting form for Q as r increases without bound is

$$\lim_{t \to \infty} Q(r) = Q(\infty) = 1 + \tfrac{1}{2} \log \rho. \tag{4.29}$$

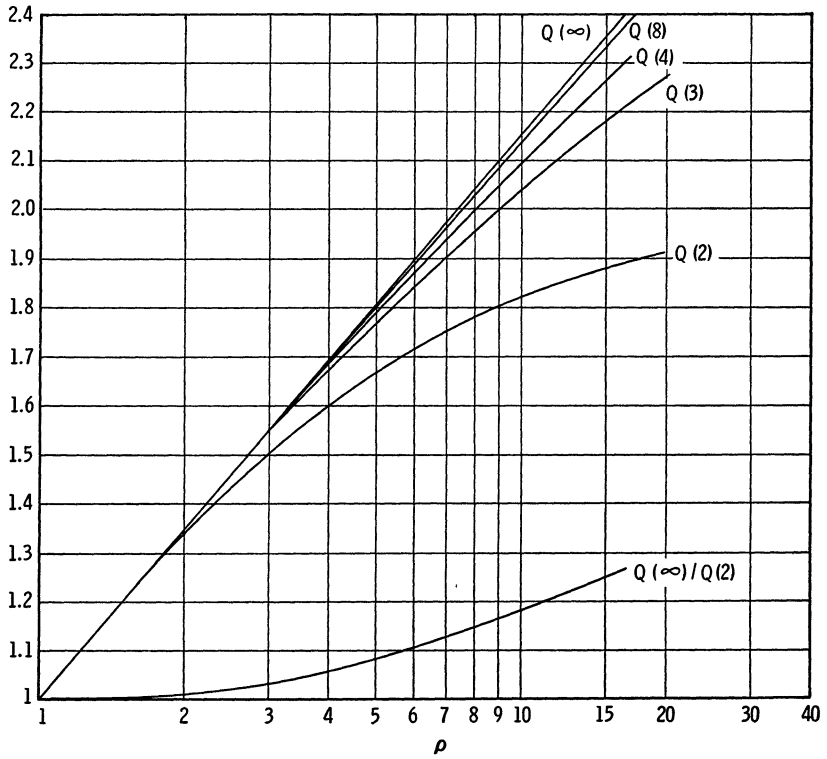

FIG. 2. Measures of over-all efficiency as function of ρ for cases of Section 4.5. For the upper curves the ordinate is Q and the measure of over-all efficiency is $Q(r)/r$ where r is the number of interpolating distributions. For the lower curve the ordinate is the ratio $Q(\infty)/Q(2)$.

The results have been plotted as a function of ρ in Figure 2. By (4.8) the numbers given in Figure 2 are to be compared with r to determine the over-all efficiency of multiplex sampling. These curves speak eloquently for interpolation as does the limiting form for large r, also.

Acknowledgment. Many people have contributed to the development of this paper by their suggestions, criticisms, and general comments. It would be impractical to list all of them, but I do want to acknowledge the contribution of J. W. Tukey, however. He gave a series of lectures in a seminar on tolerancing at Bell Telephone Laboratories and many refinements resulted directly from that series.

APPENDIX I

Criterion A′ is applied against the Q's which were obtained through the mean value theorem to get (2.9); strictly speaking this procedure fixes p and precludes varying it under the integral sign as was done in applying Criterion A′. Actually

we could argue around this objection easily, but it is instructive to sketch out a direct attack. First we define by some suitable set of restrictions a class C_a of d functions such that C_a contains p as given by (3.23). There exists bounds $\overline{|h-g|^2_{\beta a}}$ such that

$$\text{(1)} \qquad \text{var } m(\beta) \leqq \frac{1}{N}\{\overline{|h-g|^2_{\beta a}}\, Q_\beta - (\delta a_\beta)^2\},$$

where Q_β is defined by (2.11), for all p in C_a. For the λ's given in Criterion A′,

$$\text{(2)} \qquad \frac{1}{N}\sum_\beta \lambda_\beta(Q_\beta - (\delta a_\beta)^2/\overline{|h-g|^2_{\beta a}}) \geqq \sum_\beta [\lambda_\beta/\overline{|h-g|^2_{\beta a}}] \text{ var } m(\beta).$$

Choosing p such that the left hand side is a minimum is precisely Criterion A′ since the second term in the sum is a constant. This p is given by (3.23) because we required that it be a member of C_a; note that it is *independent* of the bounds. Obviously $\overline{|h-g|^2_{\beta a}} \geqq |h-g|^2_\beta$, where $|h-g|^2_\beta$ is as in (2.12) with p as in (3.23). The above is true for every suitably restricted C_a containing p given by (3.23). Considering all such C_a and taking the infimum over all bounds

$$\text{(3)} \qquad \begin{aligned} &\inf\left(\min_p\left[\frac{1}{N}\{\sum_\beta \lambda_\beta(Q_\beta - (\delta a_\beta)^2/\overline{|h-g|^2_{\beta a}})\}\right]\right) \\ &\quad = \frac{1}{N}\sum_\beta \lambda_\beta(Q_\beta - (\delta a_\beta)^2/|h-g|^2_\beta) = \sum_\beta [\lambda_\beta/|h-g|^2_\beta] \text{ var } m(\beta). \end{aligned}$$

Thus, we see that Criterion A′ corresponds to a two stage minimization process.

The analogous procedure cannot be carried through for Criterion B′ because the last step fails if R_β is not taken as a constant since then p depends on the bounds. Of course, with the approximation R_β independent of p everything goes through.

APPENDIX II[9]

Using the results of Section 3, p satisfying Criterion B is given by (4.12). First we invert (4.16); thus,

$$\text{(1)} \qquad \left[\sum_{i=t}^{r} \lambda_i/a_i^2\right]^{\frac{1}{2}} = \frac{\mu(a_t - a_{t-1})}{a_t^2 Q_t - a_{t-1}^2 Q_{t-1}}, \qquad t = 1, 2, \cdots, r; \qquad a_0 = 0.$$

Since the left hand side must be non-negative and finite the Q's must satisfy the restriction

$$\text{(2)} \qquad a_t^2 Q_t > a_{t-1}^2 Q_{t-1}, \qquad t = 1, 2, \cdots, r.$$

Squaring both sides of (1) and subtracting from the expression using t the expression with t replaced by $t+1$, $t = 1, 2, \cdots, r-1$, gives (after dropping

[9] The results here will be only for $R_t = 0$; it will be obvious how to handle the case $R_t > 0$.

the factor of proportionality)

$$\begin{aligned}\lambda_i/a_i^2 &= \frac{(a_i - a_{i-1})^2}{(a_i^2 Q_i - a_{i-1}^2 Q_{i-1})^2} - \frac{(a_{i+1} - a_i)^2}{(a_{i+1}^2 Q_{i+1} - a_i^2 Q_i)^2}, \quad i = 1, 2, \cdots, r-1, \\ \lambda_r/a_r^2 &= \frac{(a_r - a_{r-1})^2}{(a_r^2 Q_r - a_{r-1}^2 Q_{r-1})^2}.\end{aligned} \tag{3}$$

Again, since the λ's must be non-negative, the Q's must satisfy

$$\frac{(a_i - a_{i-1})^2}{(a_i^2 Q_i - a_{i-1}^2 Q_{i-1})^2} \geqq \frac{(a_{i+1} - a_i)^2}{(a_{i+1}^2 Q_{i+1} - a_i^2 Q_i)^2}. \tag{4}$$

Since all the terms inside the parentheses are positive (cp., (2) and (3)), the terms may be unsquared. One can then obtain the continued inequality

$$\begin{aligned}0 < \frac{a_1^2 Q_1}{a_1} &\leqq \frac{a_2^2 Q_2 - a_1^2 Q_1}{a_2 - a_1} \leqq \frac{a_3^2 Q_3 - a_2^2 Q_2}{a_3 - a_2} \leqq \cdots \\ &\leqq \frac{a_{r-1}^2 Q_{r-1} - a_{r-2}^2 Q_{r-2}}{a_{r-1} - a_{r-2}} \leqq \frac{a_r^2 Q_r - a_{r-1}^2 Q_{r-1}}{a_r - a_{r-1}}.\end{aligned} \tag{5}$$

The zero may be appended at the left because Q_1 and/or a_1 equals zero is nonsensical; thus (5) supersedes (2). (5) is *the* test to determine whether or not, for a given set of Q_i, a set of λ_i exists; if it does exist they are given by (3).

Now let us use (3) and (5) to find the λ's under Criterion B″. First, take the simplest case; assume that the k's are such that Q's can be found satisfying

$$\min_\lambda [\max_i (k_i Q_i)] = B = k_1 Q_1 = k_2 Q_2 = \cdots = k_r Q_r. \tag{6}$$

Then (5) becomes

$$\begin{aligned}0 < \frac{a_1^2/k_1}{a_1} &\leqq \frac{a_2^2/k_2 - a_1^2/k_1}{a_2 - a_1} \leqq \frac{a_3^2/k_3 - a_2^2/k_2}{a_3 - a_2} \leqq \cdots \\ &\leqq \frac{a_i^2/k_i - a_{i-1}^2/k_{i-1}}{a_i - a_{i-1}} \leqq \frac{a_{i+1}^2/k_{i+1} - a_i^2/k_i}{a_{i+1} - a_i} \leqq \cdots \\ &\leqq \frac{a_{r-1}^2/k_{r-1} - a_{r-2}^2/k_{r-2}}{a_{r-1} - a_{r-2}} \leqq \frac{a_r^2/k_r - a_{r-1}^2/k_{r-1}}{a_r - a_{r-1}}.\end{aligned} \tag{7}$$

In order for (6) to be true (7) must also be true; if (7) is satisfied, then, from (3), the essentially unique set of λ's is given by

$$\begin{aligned}\lambda_i/a_i^2 &= \frac{(a_i - a_{i-1})^2}{(a_i^2/k_i - a_{i-1}^2/k_{i-1})^2} - \frac{(a_{i+1} - a_i)^2}{(a_{i+1}^2/k_{i+1} - a_i^2/k_i)^2}, \qquad i = 1, 2, \cdots, r-1, \\ \lambda_r/a_r^2 &= \frac{(a_r - a_{r-1})^2}{(a_r^2/k_r - a_{r-1}^2/k_{r-1})^2}.\end{aligned} \tag{8}$$

This takes care of the case when (6) holds. What about the case in which not

all the k_iQ_i are equal? Actually, as will be seen, this has already been solved; all that is needed is to put the various pieces together in proper relation. Recall that it was found in Section 3 that if for any j

$$k_jQ_j < B = \min_\lambda [\max_i(k_iQ_i)], \quad \text{then} \quad \lambda_j = 0. \tag{9}$$

The effect of this, i.e., $\lambda_j = 0$, is the same as if $w(j; x)$ had never existed—with one minor exception. That is, one deletes $w(j; x)$ and pretends that he started with the set of probability density functions $w(t; x)$, $t = 1, 2, \cdots, r$, $t \neq j$, for the purposes of computing μ, Q_t, λ_t, $t = 1, 2, \cdots, r$, $t \neq j$. The minor exception is that there still exists a positive value for Q_j, this may be computed at the end if it is desired.

In particular, the effect of the above on (7)—assuming for now that (9) is true for only one j—is to rewrite it, in part, as

$$\cdots \leqq \frac{a_{j-1}^2/k_{j-1} - a_{j-2}^2/k_{j-2}}{a_{j-1} - a_{j-2}} \leqq \frac{a_{j+1}^2/k_{j+1} - a_{j-1}^2/k_{j-1}}{a_{j+1} - a_{j-1}} \leqq \frac{a_{j+2}^2/k_{j+2} - a_{j+1}^2/k_{j+1}}{a_{j+2} - a_{j+1}} \leqq \cdots \tag{10}$$

instead of

$$\cdots \leqq \frac{a_{j-1}^2/k_{j-1} - a_{j-2}^2/k_{j-2}}{a_{j-1} - a_{j-2}} \underset{(1)}{\leqq} \frac{a_j^2/k_j - a_{j-1}^2/k_{j-1}}{a_j - a_{j-1}} \underset{(2)}{\leqq} \frac{a_{j+1}^2/k_{j+1} - a_j^2/k_j}{a_{j+1} - a_j} \underset{(3)}{\leqq} \frac{a_{j+2}^2/k_{j+2} - a_{j+1}^2/k_{j+1}}{a_{j+2} - a_{j+1}} \leqq \cdots . \tag{11}$$

Since the assumption for the present case is that it is not possible to find a set of λ's such that the k_iQ_i are all equal and since the complete expression of which (11) is a part tests whether or not a set of λ's can be found such that a specific relation among the Q's is possible, it follows that one or more of the inequalities (1), (2), (3) must be violated. From (9) it follows that the given k_j is too small; it is patently impossible that the given k_j be too large. Therefore, the only inequality which is violated is (2); (1) and (3) are not violated. As a corollary, note that

$$k_rQ_r = B \tag{12}$$

always. ((12) is obvious for another reason; $\lambda_r = 0$ implies $Q_r = \infty$, from the definition of Q_r(4.10) and p (4.12).) The process of going from (11) to (10) will be called *collapsing* the continued inequality. The terms which are lost in the collapsing process will be called *deleted* terms.

The above has supposed only one $k_iQ_i < B$. For two *non-consecutive* $kQ < B$, e.g., in (11) $j - 1$ and $j + 1$ and with $j - 2$, j, $j + 2$ all satisfying equality, the results can immediately be seen to be no different, i.e., only the inequality with the $j - 1$ terms appearing on either side and the inequality with the $j + 1$

terms appearing on either side are violated. The same result holds for any number of non-consecutive terms, regardless of the spacing.

For *consecutive* terms such that $kQ < B$, e.g., in (11) $j - 1, j, j + 1$ with $j - 2, j + 2$ satisfying equality, the results are straightforward but quite possibly more laborious. In the example cited, (11), any combination of the five inequalities shown may or may not be violated (e.g., k_{j-1} may be too small, but k_j may be much too small, etc.)—subject to the proviso that at least one must be violated (otherwise $kQ = B$ is possible). What then must happen is that after (11) is collapsed about the offended inequalities one or more of the remaining inequalities must be violated unless all of the terms in question have been deleted. The resultant continued inequality is collapsed again and so on until all of the consecutive terms satisfying $kQ < B$ have been deleted. The above obviously holds for any allowable number of consecutive terms.

After all of the terms corresponding to $k_jQ_j < B$ have been deleted by any or all of the above processes and the ultimately collapsed version of the continued inequality (7) is obtained, then the λ_i corresponding to the terms remaining (i.e., the terms not deleted) in the collapsed continued inequality may be calculated from the collapsed version of (8). The collapsed version of (8) is obtained as follows: Suppose the remaining terms are numbered, in ascending order,

$$0, e, \cdots, g, h, i, j, l, \cdots, q, r. \tag{13}$$

For λ_i, replace a_{i-1} and k_{i-1} by a_h and k_h; replace a_{i+1} and k_{i+1} by a_j and k_j. For λ_e, $a_{e-1} = 0$. For λ_r replace a_{r-1} and k_{r-1} by a_q and k_q. For the deleted terms the corresponding λ's are, of course, zero.

An important special case of the set of centered uniform distributions under Criterion B″ is

$$k_1 = k_2 = \cdots = k_r = 1. \tag{14}$$

The continued inequality (7) is satisfied throughout because of the strict monotonicity of the a_i; hence, no terms are deleted and (6) is satisfied. Thus (4.17) and (4.18).

REFERENCES

[1] *A Million Random Digits with 100,000 Normal Deviates* (1955). The RAND Corporation, Free Press, Glencoe, Ill.

[2] MEYER, H. A. (ed.) (1956). *Symposium on Monte Carlo Techniques*. Wiley, New York.

[3] BUTLER, J. W. (1956). Machine sampling from given probability distributions. In [2], pp. 249–264.

[4] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[5] EVANS, D. H. (1957). Optimum tolerance assignment to yield minimum manufacturing cost. *Bell System Tech. J.* **37** 461–484.

[6] EVANS, D. H. Applied multiplex sampling. *Technometrics*. **5** 341–359.

[7] HARDY, G. H., LITTLEWOOD, J. E., and POLYÁ, G. (1952). *Inequalities* (2nd ed). Cambridge Univ. Press.

[8] KAHN, H. (1956). Use of different Monte Carlo sampling techniques. In [2], pp. 146–190.

[9] TROTTER, H. F. and TUKEY, J. W. (1956). Conditional Monte Carlo for normal samples. In [2], pp. 64–79.
[10] TUCKER, A. W. (1957). Linear and nonlinear programming. *Operations Res.* **5** 244–257.
[11] TUKEY, J. W. (1957). Propagation of errors, fluctuations and tolerances, No. 1: basic generalized formulas. Tech. Report No. 10, Statist. Tech. Res. Group, Princeton Univ.
[12] TUKEY, J. W. (1957). Propagation of errors, fluctuations and tolerances, No. 2: supplementary formulas. Tech. Report No. 11, Statist. Tech. Res. Group, Princeton Univ.
[13] TUKEY, J. W. (1957). Propagation of errors, fluctuations and tolerances, No. 3: exercise in differentiation. Tech. Report No. 12, Statist. Tech. Res. Group, Princeton Univ.