UNBIASED SEQUENTIAL ESTIMATION FOR BINOMIAL POPULATIONS[1]

BY MORRIS H. DEGROOT

Carnegie Institute of Technology

1. Introduction and summary. The subject of minimum variance unbiased estimation has received a great deal of attention in the statistical literature, e.g., in the papers of Bahadur [2], Barankin [3], and Stein [14]. The emphasis in these papers has typically been placed on the existence and construction of minimum variance unbiased estimators when the sampling plan to be used was given in advance.

In this paper, criteria are developed for the selection of an appropriate sampling plan for the family of binomial distributions. Thus, independent observations are to be taken on the random variable U so distributed that

$$\Pr\{U = 1 \mid p\} = p, \qquad \Pr\{U = 0 \mid p\} = 1 - p = q, \tag{1.1}$$

where p lies in the open interval $0 < p < 1$, and the value of a given function, $g(p)$, is to be estimated. The problem considered here is that of determining a sampling plan and an unbiased estimator of $g(p)$ that are optimal, in some sense, at a specified value, p_0, of p. Optimality will depend, not only on the variance of the estimator, but also on the average sample size of the plan. A sampling plan, S, and estimator, f, will be considered optimal at p_0 if, among all procedures with average sample size at p_0 no larger than that of S, there does not exist an unbiased estimator with smaller variance at p_0 than that of f.

The basic tool to be used is the information inequality (see Lemma 2.7 and the discussion following it) which provides a lower bound for the variance of an estimator in terms of its expected value and the average sample size of the sampling plan. If, at p_0, this lower bound is attained for a particular estimator and sampling plan, it may be immediately concluded that they are optimal at p_0. Such an estimator is said to be efficient at p_0.

In Section 2, various definitions, assumptions, and fundamental facts to be used throughout the paper are collected.

In Section 3 it is shown that the single sample plans and the inverse binomial sampling plans are the only ones that admit an estimator that is efficient at all values of p.

In Section 4 some techniques are given that are often useful in the analysis of inverse binomial sampling plans.

In Section 5 relationships between the average sample size of a sampling plan and the functions of p that are estimable optimally are explored.

Received May 26, 1958; revised September 8, 1958.

[1] This paper is based on a dissertation submitted to the Department of Statistics, University of Chicago, in partial fulfillment of the requirements for the Ph.D. degree. The research was sponsored in part by the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

In Sections 6 and 7 it is shown that there can exist two distinct sampling plans with the same average sample size for all p and some comparisons are made of such plans.

In Section 8 a new characterization of completeness is given for bounded sampling plans and it is shown that the dimension of the linear space of unbiased estimators of 0 can be determined simply by counting the number of boundary points. It is further shown that for a wide class of plans, the estimators that are efficient at a given value of p do not have uniformly minimum variance, although non-trivial uniformly minimum variance estimators do exist.

After this paper had been accepted for publication I learned that R. B. Dawson had obtained expression (4.13) and Theorems 8.2 and 8.4, as well as various other interesting results related to the material of Section 8, in his Ph.D. thesis, "Unbiased tests, unbiased estimators, and randomized similar regions," Harvard University, May, 1953.

2. Definitions, assumptions, and fundamental facts. A formal definition of the sampling plans to be considered in the paper will now be given. It will be helpful to keep in mind the interpretation of a sampling plan as a rule that specifies at each stage of a sequential sampling process whether sampling is to cease or another observation is to be taken. Furthermore, it will be helpful to visualize a sequential sample as a path in the Euclidean plane—the path starting at the origin and being extended at a given stage one unit in either the horizontal or vertical direction according as the observation at that stage is 0 or 1. Further discussion of these interpretations will be given below. A formal description of binomial sampling plans was first given by Girshick, Mosteller, and Savage in [7], and the following discussion utilizes several of the concepts presented in that paper and the paper of Lehmann and Stein [11]. In what follows, the word *point* refers only to points γ of the Euclidean plane whose coordinates $X(\gamma)$ and $Y(\gamma)$ are non-negative integers.

A *sampling plan* is a function S defined on the points γ, taking only the values 0 and 1, and such that for the point θ with coordinates $X(\theta) = 0$ and $Y(\theta) = 0$, $S(\theta) = 1$.

A *path* to γ is a sequence of points $\theta = \gamma_0, \gamma_1, \cdots, \gamma_n = \gamma$ such that $S(\gamma_k) = 1$ for $k = 0, 1, \cdots, n - 1$ and either

$$X(\gamma_{k+1}) = X(\gamma_k) + 1 \qquad (2.1)$$
$$Y(\gamma_{k+1}) = Y(\gamma_k),$$

or

$$X(\gamma_{k+1}) = X(\gamma_k) \qquad (2.2)$$
$$Y(\gamma_{k+1}) = Y(\gamma_k) + 1.$$

Thus, each point of the sequence is either one unit to the right or one unit above its predecessor.

Under a given sampling plan, γ is a *boundary point* if there exists a path to γ and $S(\gamma) = 0$. It is a *continuation point* if there exists a path to γ and $S(\gamma) = 1$, so that there is also at least one path "through" γ. It should be noted that the origin θ is always a continuation point.

A point γ is an *inaccessible point* if there does not exist any path to γ. Thus, every point can be uniquely classified as a continuation, inaccessible, or boundary point.

The *sample size* $N(\gamma)$ of any point γ is the sum of its coordinates, $X(\gamma) + Y(\gamma)$.

The *boundary* B is the set of all boundary points.

Clearly, the values of S at inaccessible points are irrelevant to the sampling process. However, some discussions to be given later will be simplified if it is assumed once and for all that $S(\gamma) = 1$ at all inaccessible points. It should be clear that under these conditions S is completely determined by the set of boundary points.

The sampling plans just discussed are somewhat restricted in two respects. The first is that for a given sequence of observations $U_1, \cdots, U_m$ the decision as to whether another observation is taken depends only on the point γ reached by the sample path prescribed by $U_1, \cdots, U_m$; i.e., the decision depends only on m and $\sum_{j=1}^{m} U_j$. The justification for considering only such plans is provided by the fact that the sequence $\sum_{j=1}^{m} U_j$, $m = 1, 2, \cdots$, is both sufficient and transitive. A thorough discussion of these concepts is given by Bahadur in [1]. Similar considerations, especially the theorem of Blackwell [4], justify the definition of an estimator, to be given below, as a statistic that depends on the observed sample sequence only through the boundary point reached by the sequence.

The second restriction on the sampling plans is that they are non-randomized. A randomized sampling plan is a function S defined on the points γ such that S may take any value in the closed unit interval. If, at a given stage in the sampling process, the point γ is reached, then another observation is taken with probability $S(\gamma)$ and sampling is terminated with probability $1 - S(\gamma)$. In this paper, only plans for which $S(\gamma)$ is 0 or 1 are considered.

The probability of reaching a particular boundary point γ is $K(\gamma)p^{Y(\gamma)}q^{X(\gamma)}$, where $K(\gamma)$ is the number of distinct paths to γ.

A sampling plan is said to be *closed* if

$$\sum_{\gamma \varepsilon B} K(\gamma)p^{Y(\gamma)}q^{X(\gamma)} = 1 \tag{2.3}$$

for all p, $0 < p < 1$. Only closed plans are considered.

An *estimator* f is a real-valued function defined on B. The only estimators to be considered are those for which

$$E(f \mid p) = \sum_{\gamma \varepsilon B} f(\gamma)K(\gamma)p^{Y(\gamma)}q^{X(\gamma)} \tag{2.4}$$

is absolutely convergent.

A sampling plan is *complete* (*boundedly complete*) if the only estimator (bounded estimator) f, such that $E(f \mid p) = 0$ for all p, is the one defined by $f(\gamma) = 0$ for all $\gamma \varepsilon B$.

A sampling plan is *simple*, if, for each positive integer m, the continuation points of sample size m form an interval on the line $X(\gamma) + Y(\gamma) = m$.

The basic facts concerning completeness follow. Lemma 2.1 was developed by Girshick, Mosteller, Savage, and Wolfowitz in a sequence of papers [7], [17], [12]. Lemma 2.2 is due to Lehmann and Stein [11].

LEMMA 2.1. *A necessary and sufficient condition for a closed sampling plan to be boundedly complete is that it be simple.*

LEMMA 2.2. *A necessary and sufficient condition for a closed plan to be complete is that it be simple and that the conversion of any boundary point to a continuation point destroy closure.*

The functions defined on B and taking the values $X(\gamma)$, $Y(\gamma)$, and $N(\gamma)$ are denoted by X, Y, and N, respectively.

The following conditions are assumed to hold throughout the paper:

(i) For every sampling plan to be considered,

$$E(N^2 \mid p) = \sum_{\gamma \varepsilon B} N^2(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)} \tag{2.5}$$

is uniformly convergent on every closed interval of values of p;

(ii) For every estimator f to be considered,

$$g(p) = E(f \mid p) = \sum_{\gamma \varepsilon B} f(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)} \tag{2.6}$$

is differentiable termwise in the open interval, $0 < p < 1$, and the derived series is absolutely convergent.

A well-known, elementary, and useful sufficient condition for the termwise differentiability of the series in (2.6) is that the formal termwise derivative be absolutely uniformly convergent on every closed subinterval. This condition is in turn very often verified by a dominance argument (see, e.g., [5], pp. 392, 396).

Some easy but important consequences of the above assumptions will now be given. Some of the results are well-known and have been given elsewhere.

LEMMA 2.3. *$E(Y^2 \mid p)$, $E(X^2 \mid p)$, and $E(XY \mid p)$ all exist and are at most $E(N^2 \mid p)$.*

PROOF. Since $0 \leqq X \leqq N$ and $0 \leqq Y \leqq N$, the results follow from (i).

LEMMA 2.4. *N, X, and Y, considered as estimators, satisfy* (ii).

PROOF. $E(N \mid p) = \sum_{\gamma \varepsilon B} N(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)}$. The formal termwise derivative of this series is less, in absolute value, than

$$\frac{1}{pq} \sum_{\gamma \varepsilon B} N^2(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)}.$$

It follows from (i) that this series converges uniformly on every closed interval and, hence, according to the remarks following (ii), $E(N \mid p)$ is termwise differentiable. The proofs for X and Y are similar.

LEMMA 2.5. *If f satisfies* (ii) *and* $E(f \mid p) = g(p)$, *then* $E[(qY - pX)f \mid p] = pqg'(p)$.

PROOF. Termwise differentiation of the series for $g(p)$ yields the result.

LEMMA 2.6. $qE(Y \mid p) = pE(X \mid p) = E[(qY - pX)^2 \mid p] = pqE(N \mid p)$.

PROOF. These results can be derived directly from Lemma 2.5. However, they are well-known and are basic to various aspects of sequential analysis. As a result, they have been derived under various conditions; e.g., by Wald in the Appendix of [16] as consequences of the fundamental identity of sequential analysis, and by Wolfowitz in [18].

LEMMA 2.7. *For any estimator f,*

$$\text{Var } (f|p) \geq \frac{pq[g'(p)]^2}{E(N|p)}. \tag{2.7}$$

Equality holds at a particular value of p, say p_0, if and only if there exist constants a and b such that $f(\gamma) = a[q_0Y(\gamma) - p_0X(\gamma)] + b$ for all boundary points γ, where $q_0 = 1 - p_0$.

This lemma is also well-known, especially for single sample plans. A brief history of this inequality with references is given by Savage in [13], p. 238. It was first proved for sequential plans by Wolfowitz [18]. Following Savage, (2.7) will be called the *information inequality.*

A non-constant estimator f is said to be *efficient at p_0* if equality holds in the information inequality when $p = p_0$. The expected value of f is then *estimable efficiently at p_0*.

If, for a given sampling plan S, a non-constant estimator f is efficient at p for all p, then both f and S are called *efficient.* From Lemma 2.7, it is seen that f is efficient if and only if there exist two functions of p, $a(p)$ and $b(p)$, with $a(p) \neq 0$, such that

$$f(\gamma) = a(p)[qY(\gamma) - pX(\gamma)] + b(p) \tag{2.8}$$

for all p and all boundary points γ.

Two types of sampling plans of prime importance are the single sample plans and the inverse binomial sampling plans. In a single sample plan,

$$B = \{\gamma : N(\gamma) = n\}$$

for some positive integer n. It is clear that such a plan satisfies (i) since the series involved contains only a finite number of non-zero terms.

In an inverse binomial sampling plan, either $B = \{\gamma : Y(\gamma) = c\}$ for some positive integer c, or $B = \{\gamma : X(\gamma) = c\}$. This type of plan was first treated formally by Haldane in [8] and [9]; the name "inverse binomial sampling" was suggested by Tweedie [15]. It is easily seen that these plans are closed. The next lemma shows that assumption (i) is also satisfied.

LEMMA 2.8. *For an inverse binomial sampling plan $E(N^2 \mid p)$ converges uniformly on every closed interval.*

PROOF. Consider the plan for which $B = \{\gamma : Y(\gamma) = c\}$ for a given positive integer c. Then

$$\Pr\{N = c + j \mid p\} = \binom{c+j-1}{j} p^c q^j, \qquad j = 0, 1, \cdots,$$

and

$$E(N^2 \mid p) = \sum_{j=0}^{\infty} (c+j)^2 \binom{c+j-1}{j} p^c q^j.$$

On the closed interval $0 < \delta \leqq p \leqq \epsilon < 1$,

$$E(N^2 \mid p) \leqq \epsilon^c \sum_{j=0}^{\infty} (c+j)^2 \binom{c+j-1}{j} (1-\delta)^j,$$

and the series on the right is convergent, as is readily checked by the ratio test. Hence, the given series converges uniformly on the closed interval. The proof is entirely analogous when $B = \{\gamma : X(\gamma) = c\}$.

Because of the duality between the two types of inverse binomial sampling plans, any facts stated in the paper concerning these plans will be demonstrated only for the plan where $B = \{\gamma : Y(\gamma) = c\}$. The analogous proofs for the plan in which $B = \{\gamma : X(\gamma) = c\}$ can always be obtained simply by interchanging the roles of X and Y and of p and q.

3. Efficient sampling plans. In this section it is shown that the only efficient sampling plans are the single sample plans and the inverse binomial sampling plans. For a single sample plan, the efficient estimators are the non-constant linear functions of Y, and hence, the functions of p that are estimable efficiently are linear in p. For an inverse binomial sampling plan the efficient estimators are linear functions of N, and their expectations are linear in either $1/p$ or $1/q$.

LEMMA 3.1. *Let S be a given sampling plan for which there exists a non-constant estimator f that is efficient at two values of p. Then there exist constants μ, ν, and ξ, not all 0, such that $\mu X(\gamma) + \nu Y(\gamma) = \xi$ for all boundary points γ.*

PROOF. Suppose f is efficient at p_0 and p_1. Then, from Lemma 2.7, there exist constants a_0, b_0, a_1, and b_1 such that

$$\begin{aligned} f(\gamma) &= a_0[q_0 Y(\gamma) - p_0 X(\gamma)] + b_0 \\ &= a_1[q_1 Y(\gamma) - p_1 X(\gamma)] + b_1 \end{aligned} \tag{3.1}$$

for all boundary points γ. Hence

$$(a_0 q_0 - a_1 q_1) Y(\gamma) - (a_0 p_0 - a_1 p_1) X(\gamma) = b_1 - b_0, \tag{3.2}$$

and since f is not constant, neither a_0 nor a_1 is 0. Thus, not both the coefficients of $Y(\gamma)$ and $X(\gamma)$ are 0, and (3.2) is an equation of the required form.

Clearly, the boundary points of the single sample plans and the inverse binomial sampling plans lie on straight lines as the lemma demands. The following theorem shows that they are the only plans for which this is true.

THEOREM 3.1. *Let S be a given closed sampling plan for which there exist con-*

stants μ, ν, and ξ, not all 0, such that $\mu X(\gamma) + \nu Y(\gamma) = \xi$ for all boundary points γ. Then S is either a single sample plan or an inverse binomial sampling plan.

PROOF. There is no loss of generality in assuming $\mu \geqq 0$. The proof is divided into cases depending on the magnitudes of μ, ν, and ξ.

(i) $\mu = \nu$: If ξ/μ is not a positive integer, then there are no boundary points, and S is not closed. If $\xi/\mu = n$, a positive integer, then B is a subset of the single sample plan with boundary $B^* = \{\gamma : X(\gamma) + Y(\gamma) = n\}$. Clearly, if S is closed, then $B = B^*$.

(ii) $\mu = 0$: If ξ/ν is not a positive integer, then again B is empty, and S is not closed. If $\xi/\nu = c$, a positive integer, then B is a subset of the inverse binomial sampling plan with boundary $B^* = \{\gamma : Y(\gamma) = c\}$. Again, if S is closed, then $B = B^*$.

(iii) $\nu = 0$: As in (ii), B must be of the form $B = \{\gamma : X(\gamma) = c\}$.

(iv) $\mu > 0$, $\nu > 0$, $\mu \neq \nu$: If $\xi \leqq 0$, B is empty. If $\xi > 0$, there exists a point γ such that $\mu X(\gamma) + \nu Y(\gamma) < \xi$ and either $\mu[X(\gamma) + 1] + \nu Y(\gamma) > \xi$ or $\mu X(\gamma) + \nu[Y(\gamma) + 1] > \xi$. Thus, with positive probability, the boundary is passed and S is not closed.

(v) $\mu > 0$, $\nu < 0$: Suppose $\xi > 0$. By the strong law of large numbers ([6], p. 243), there is positive probability that, for $\epsilon > 0$,

$$qY(\gamma) - pX(\gamma) > -\epsilon N(\gamma) \tag{3.3}$$

for every point γ reached by the sample path. But for p sufficiently large and ϵ sufficiently small the line $qY(\gamma) - pX(\gamma) = -\epsilon N(\gamma)$ lies entirely above the line $\mu X(\gamma) + \nu Y(\gamma) = \xi$. Hence, no point satisfying the inequality (3.3) can be a boundary point, and there is positive probability that the boundary will not be reached. Thus, S is not closed. Analogous arguments hold if $\xi < 0$ or $\xi = 0$.

The main result of this section is

THEOREM 3.2. *The only efficient sampling plans are the single sample plans and the inverse binomial sampling plans. When $B = \{\gamma : N(\gamma) = n\}$, any non-constant function of the form $a + bY$ is an efficient estimator of $a + bnp$, and these are the only efficient estimators. When $B = \{\gamma : Y(\gamma) = c\}$, any non-constant function of the form $a + bN$ is an efficient estimator of $a + bc(1/p)$, and these are the only efficient estimators. When $B = \{\gamma : X(\gamma) = c\}$, any non-constant function of the form $a + bN$ is an efficient estimator of $a + bc(1/q)$, and these are the only efficient estimators.*

PROOF. Lemma 3.1 and Theorem 3.1 show that the only procedures satisfying certain conditions necessary for a sampling plan to be efficient are the single sample plans and the inverse binomial sampling plans. It remains to show that these procedures are indeed efficient.

Let $B = \{\gamma : N(\gamma) = n\}$. Then, for every $\gamma \varepsilon B$, $qY(\gamma) - pX(\gamma) = Y(\gamma) - np$, and hence, $Y(\gamma) = [qY(\gamma) - pX(\gamma)] + np$. This demonstrates the efficiency of Y. Since $E(qY - pX \mid p) = 0$, $E(Y \mid p) = np$.

Let $B = \{\gamma : Y(\gamma) = c\}$. Then, for every $\gamma \varepsilon B$, $qY(\gamma) - pX(\gamma) = c - pN(\gamma)$, and hence, $N(\gamma) = -(1/p)[qY(\gamma) - pX(\gamma)] + (c/p)$. Thus, N is efficient and $E(N \mid p) = c/p$.

The proof is completed by noting that if f is an efficient estimator, then so is every non-constant linear function of f. Furthermore, if two non-constant estimators are both efficient at a given value of p, then they are linearly related.

COROLLARY 3.1. *A non-constant estimator is efficient if and only if it is efficient at two distinct values of p.*

PROOF. From Lemma 3.1 and Theorems 3.1 and 3.2, it is seen that if f is not constant and is efficient at two values of p then the sampling plan admits an efficient estimator. From the comments at the end of the proof of Theorem 3.2, it follows that this estimator is linearly related to f, and hence, that f itself is efficient. The proof in the other direction is trivial.

4. Inverse binomial sampling plans. Because of Theorem 3.2, it seems worthwhile to investigate the properties of single sample and inverse binomial sampling plans in some detail.

For the single sample plan with boundary $B = \{\gamma : N(\gamma) = n\}$, Y/n is an efficient estimator of p with $\text{Var}(Y/n \mid p) = pq/n$. For any estimator f,

$$E(f \mid p) = \sum_{\gamma \in B} f(\gamma) \binom{n}{Y(\gamma)} p^{Y(\gamma)} q^{X(\gamma)}, \tag{4.1}$$

which is a polynomial in p of degree at most n. Thus, only polynomials of degree at most n are estimable unbiasedly, and since

$$E\left[\frac{Y(Y-1)\cdots(Y-k)}{n(n-1)\cdots(n-k)} \,\middle|\, p\right] = p^{k+1} \tag{4.2}$$

for $k = 0, 1, \cdots, n-1$, then every such polynomial is estimable.

The analogous properties of an inverse binomial sampling plan are less familiar and will be discussed here.

Consider the plan with boundary $B = \{\gamma : Y(\gamma) = c\}$. For each non-negative integer, k, there exists a unique point, γ_k, of B such that $N(\gamma_k) = c + k$.

Since

$$\Pr\{N = c + k \mid p\} = \binom{k+c-1}{k} p^c q^k, \qquad k = 0, 1, 2, \cdots, \tag{4.3}$$

then for any estimator f,

$$E(f \mid p) = p^c \sum_{k=0}^{\infty} f(\gamma_k) \binom{k+c-1}{k} q^k. \tag{4.4}$$

The class of functions that are estimable unbiasedly and their estimators will now be determined. For convenience, the functions are written as functions of q rather than of p.

THEOREM 4.1. *A function $h(q)$ is estimable unbiasedly if and only if it can be expanded in Taylor's series in the interval $|q| < 1$. If $h(q)$ is estimable, then its unique estimator is given by*

$$f(\gamma_k) = \frac{(c-1)!}{(k+c-1)!} \frac{d^k}{dq^k} \left[\frac{h(q)}{(1-q)^c}\right]_{q=0}, \qquad k = 0, 1, 2, \cdots. \tag{4.5}$$

PROOF. If $h(q)$ can be expanded in Taylor's series in the given interval, then so also can $h(q)/(1 - q)^c$, and conversely. Thus, suppose

$$\frac{h(q)}{(1 - q)^c} = \sum_{k=0}^{\infty} b_k q^k.$$

Then

$$h(q) = p^c \sum_{k=0}^{\infty} b_k q^k,$$

and taking

$$f(\gamma_k) = b_k \bigg/ \binom{k + c - 1}{k}$$

yields an estimator f with $E(f \mid p) = h(q)$.

Suppose now that $h(q)$ is estimable unbiasedly. Then there exists an estimator f such that

$$h(q) = p^c \sum_{k=0}^{\infty} f(\gamma_k) \binom{k + c - 1}{k} q^k.$$

Replacing p by $1 - q$ gives the required expansion. Thus,

$$\frac{h(q)}{(1 - q)^c} = \sum_{k=0}^{\infty} f(\gamma_k) \binom{k + c - 1}{k} q^k,$$

and hence

$$\binom{k + c - 1}{k} f(\gamma_k) = \frac{1}{k!} \frac{d^k}{dq^k} \left[\frac{h(q)}{(1 - q)^c} \right]_{q=0},$$

$$f(\gamma_k) = \frac{(c - 1)!}{(k + c - 1)!} \frac{d^k}{dq^k} \left[\frac{h(q)}{(1 - q)^c} \right]_{q=0}.$$

The uniqueness of f follows from the uniqueness of the Taylor's series expansion, which is of course the basis of the completeness of this sampling plan.

Given $h(q)$, Theorem 4.1 provides a rule for finding its unbiased estimator. It is often possible to find the expectation of a given estimator in closed form by using the fact that if the series

$$\phi(x) = \sum_{k=0}^{\infty} b_k x^k \tag{4.6}$$

is differentiated m times within its interval of convergence, then

$$\frac{\phi^{(m)}(x)}{m!} = \sum_{k=0}^{\infty} b_{m+k} \binom{k + m}{k} x^k. \tag{4.7}$$

As illustrations of the technique involved, the variance of the unbiased estimator of p and the moment generating function of N will be determined.

It is well-known (and can be checked from Theorem 4.1) that if f is given by

$$f(\gamma_k) = (c-1)/(k+c-1), \tag{4.8}$$

then, for $c \geqq 2$, f is an unbiased estimator of p. It should be emphasized that this is *not* an efficient estimator. Using the result given above,

$$\begin{aligned} E(f^2 \mid p) &= p^c(c-1)^2 \sum_{k=0}^{\infty} \frac{1}{(k+c-1)^2} \binom{k+c-1}{k} q^k \\ &= p^c(c-1)^2 \frac{1}{(c-1)!} \frac{d^{c-1}}{dq^{c-1}} \left[\sum_{k=1}^{\infty} \frac{1}{k^2} q^k \right]. \end{aligned} \tag{4.9}$$

Note that the constant term in the last series of (4.9) is taken to be 0. Its value can be assigned arbitrarily since it does not appear in the derived series. But

$$\frac{d}{dq}\left[\sum_{k=1}^{\infty} \frac{1}{k^2} q^k\right] = \sum_{k=1}^{\infty} \frac{1}{k} q^{k-1} = \frac{1}{q} \sum_{k=1}^{\infty} \frac{1}{k} q^k = \frac{1}{q}\,[-\log(1-q)]. \tag{4.10}$$

Hence,

$$\frac{d^{c-1}}{dq^{c-1}}\left[\sum_{k=1}^{\infty} \frac{1}{k^2} q^k\right] = -\frac{d^{c-2}}{dq^{c-2}}\left[\frac{\log(1-q)}{q}\right], \tag{4.11}$$

and

$$E(f^2 \mid p) = -\frac{p^c(c-1)}{(c-2)!} \frac{d^{c-2}}{dq^{c-2}}\left[\frac{\log(1-q)}{q}\right]. \tag{4.12}$$

Using the easily verified fact that

$$\frac{d^m}{dq^m}\left[\frac{\log(1-q)}{q}\right] = \sum_{i=0}^{m} \binom{m}{i} \left[\frac{d^i}{dq^i}(\log(1-q))\right]\left[\frac{d^{m-i}}{dq^{m-i}}\left(\frac{1}{q}\right)\right],$$

one obtains

$$\frac{d^m}{dq^m}\left[\frac{\log(1-q)}{q}\right] = \frac{(-1)^m m! \log(1-q)}{q^{m+1}} + m! \sum_{i=1}^{m} \frac{(-1)^{m-i+1}}{i(1-q)^i q^{m-i+1}}.$$

Thus, referring to (4.12),

$$\begin{aligned} E(f^2 \mid p) &= \frac{(-1)^{c-1}(c-1)p^c \log(1-q)}{q^{c-1}} \\ &\quad + (c-1)p^c \sum_{i=1}^{c-2} \frac{(-1)^{c-i}}{i(1-q)^i q^{c-i-1}} \\ &= \frac{(c-1)p^c}{q^{c-1}}\left[(-1)^{c-1}\log p + \sum_{i=1}^{c-2} \frac{(-1)^{c-i}}{i}\left(\frac{q}{p}\right)^i\right]. \end{aligned} \tag{4.13}$$

It is interesting to note that Haldane [9] gives $E(f^2 \mid p)$ in the form of an integral that, in order to be evaluated, requires repeated integration by parts.

In Section 3 it was shown that N is an efficient estimator of its expected value Using the technique just illustrated, its moment generating function can be found. For $t < \log(1/q)$,

$$
\begin{aligned}
E(e^{tN} \mid p) &= p^c \sum_{k=0}^{\infty} e^{t(c+k)} \binom{k+c-1}{k} q^k \\
&= (pe^t)^c \sum_{k=0}^{\infty} \binom{k+c-1}{k} (qe^t)^k \\
&= (pe^t)^c \frac{1}{(c-1)!} \frac{d^{c-1}}{dx^{c-1}} \left[\sum_{k=0}^{\infty} x^k \right]_{x=qe^t} \\
&= (pe^t)^c \frac{1}{(c-1)!} \frac{d^{c-1}}{dx^{c-1}} \left(\frac{1}{1-x} \right)_{x=qe^t} \\
&= (pe^t)^c \frac{1}{(c-1)!} (c-1)!\,(1-qe^t)^{-c} \\
&= \left(\frac{pe^t}{1-qe^t} \right)^c.
\end{aligned} \tag{4.14}
$$

It should be noted that $N - c$ has a negative binomial distribution and that (4.14) could also be derived using techniques appropriate to that distribution (see, *e.g.*, Feller [6], pp. 155, 252).

Differentiating (4.14) at $t = 0$ yields

$$
\begin{aligned}
E(N \mid p) &= c/p \\
E(N^2 \mid p) &= (c^2 + cq)/p^2 \\
\mathrm{Var}(N \mid p) &= cq/p^2.
\end{aligned} \tag{4.15}
$$

Since N is an efficient estimator of c/p, its variance should attain the lower bound of the information inequality; *i.e.*,

$$
\mathrm{Var}\,(N \mid p) = \frac{pq[g'(p)]^2}{E(N \mid p)}. \tag{4.16}
$$

Setting $g(p) = E(N \mid p) = c/p$ does yield the value found in (4.15).

5. Relationships between sampling plans and estimable functions. For a given sampling plan, the only estimators that are efficient at a given value p_0 are the non-constant functions f^* of the form

$$
f^*(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + b, \tag{5.1}
$$

for some constants a and b and all boundary points γ.

LEMMA 5.1. *If f^* is given by* (5.1), *then f^* is constant if and only if $a = 0$.*

PROOF. If $a = 0$, then $f^*(\gamma) = b$ for all boundary points γ. If $a \neq 0$, then f^* is constant only if $q_0 Y(\gamma) - p_0 X(\gamma)$ is constant for all boundary points γ. But that would mean that the points of B lie on a straight line and, hence, the sampling plan must be either a single sample plan or an inverse binomial sampling plan. It is readily checked that for neither of these is $q_0 Y(\gamma) - p_0 X(\gamma)$ constant when $0 < p_0 < 1$.

Thus, a non-constant estimator f^* is efficient at p_0 if and only if it is of the form (5.1) with $a \neq 0$. The next theorem determines the class of functions that are estimable efficiently at a given point, simply by evaluating $E(f^* \mid p)$.

THEOREM 5.1. *For a given sampling plan, a non-constant function $g(p)$ is estimable efficiently at p_0 if and only if there exist constants a and b with $a \neq 0$, such that*

$$g(p) = a(p - p_0)E(N \mid p) + b; \tag{5.2}$$

or alternatively, if and only if there exists a constant k, $k \neq 0$, such that

$$\begin{aligned} E(N \mid p) &= k[g(p) - g(p_0)]/(p - p_0) \qquad \text{for } p \neq p_0, \\ E(N \mid p_0) &= kg'(p_0). \end{aligned} \tag{5.3}$$

PROOF. By Lemma 2.6,

$$\begin{aligned} E(q_0Y - p_0X \mid p) &= E(qY - pX + pN - p_0N \mid p) \\ &= E(qY - pX \mid p) + (p - p_0)E(N \mid p) \\ &= (p - p_0)E(N \mid p). \end{aligned}$$

If $g(p)$ is estimable efficiently at p_0 then it must be constant or the expectation of an estimator of the form (5.1) with $a \neq 0$. Hence,

$$g(p) = E(f^* \mid p) = a(p - p_0)E(N \mid p) + b. \tag{5.4}$$

Setting $p = p_0$ in (5.4) gives $b = g(p_0)$. Differentiating both sides of (5.4) at p_0 gives $g'(p_0) = aE(N \mid p_0)$. Thus, both (5.2) and (5.3) are satisfied (with $k = 1/a$).

Conversely, if either (5.2) or (5.3) is satisfied then $g(p)$ is the expectation of an estimator of the form (5.1) with $a = 1/k$ and $b = g(p_0)$.

From Theorem 5.1, it is clear that there does not always exist a sampling plan that admits estimation of a given function efficiently at a given value of p. Since $E(N \mid p)$ can never be smaller than unity, obvious restrictions on $g(p)$ and p_0 are that $[g(p) - g(p_0)]/(p - p_0)$ cannot change sign in the open interval $0 < p < 1$ and that $g'(p_0)$ cannot vanish. However, no general result characterizing the class of functions $\phi(p)$ for which there exist sampling plans with $E(N \mid p) = \phi(p)$ has, to the author's knowledge, yet been found.

6. Some sampling plans. Theorem 5.1 reveals that the functions $g(p)$ that are estimable at a given value p_0 depend only on $E(N \mid p)$ in a very simple way. From the specific form of this relationship it is seen that if it is desired to estimate a polynomial in p efficiently at a particular value, then $E(N \mid p)$ must itself be a polynomial in p of a certain form. In this section some sampling plans for which $E(N \mid p)$ is a polynomial will be described. It should be noted that if, for a given sampling plan, there exists a positive integer n such that $\Pr\{N \leqq n \mid p\} = 1$, then $E(N \mid p)$ is a polynomial of degree at most $n - 1$. One of the interesting things about the plans given below is that they yield polynomials for $E(N \mid p)$ even though they do not satisfy this condition. In particular, one of the procedures will be an unbounded sequential plan for which $E(N \mid p)$ is constant.

Hence, considering the single sample plan with the same $E(N \mid p)$, it may be concluded that $E(N \mid p)$ does not determine the sampling plan.

Scheme I. Let n and m be two positive integers and let

$$B = \{\gamma : N(\gamma) = n,\ Y(\gamma) < n\} \cup \{\gamma : Y(\gamma) = n + m\}. \tag{6.1}$$

That is, a single sample plan of sample size n is used unless all of the observations are equal to 1. When this happens, sampling is continued under an inverse binomial sampling plan until m additional 1's are obtained.

The function $E(N \mid p)$ for the inverse binomial sampling plan was found in Section 4. Using this result, it is easy to obtain $E(N \mid p)$ for the above plan.

$$\begin{aligned} E(N \mid p) &= E(N \mid N \leqq n, p)\Pr(N \leqq n \mid p) + E(N \mid N > n, p)\Pr(N > n \mid p) \\ &= n(1 - p^n) + [n + (m/p)]p^n = n + mp^{n-1}. \end{aligned} \tag{6.2}$$

For example suppose $n = 2$, $m = 4$. Then $E(N \mid p) = 2 + 4p$, and it follows from Theorem 5.1 that p^2 is estimable efficiently at $1/2$. Indeed,

$$\begin{aligned} E(N \mid p) &= 4(p + 1/2) = 4(p^2 - 1/4)/(p - 1/2) \\ &= 4[g(p) - g(p_0)]/(p - p_0), \end{aligned} \tag{6.3}$$

when $g(p) = p^2$, $p_0 = 1/2$. In the same way, $p^2 + (1/2 - p_0)p$ is estimable efficiently at p_0.

If $n = 1$, then, from (6.2), $E(N \mid p) = m + 1$ for all p. From Theorem 5.1 it is seen that this plan admits an efficient estimator of p at any value p_0. This does not mean that the plan is efficient—indeed, it is known not to be. Thus, although for every p_0 there is an estimator of p that is efficient at p_0, there is no single estimator that is efficient at all values of p. As has been shown, such an estimator exists only under a single sample plan. From Lemmas 2.1 and 2.2 it is seen that this plan (and every plan included in Scheme I) is boundedly complete but not complete.

It is interesting to compare the various estimators of p for the plan with $n = 1$. The only bounded unbiased estimator of p is

$$f_0(\gamma) = [1/(m + 1)]Y(\gamma), \tag{6.4}$$

which equals 0 or 1 according as the first observation is 0 or 1. Hence, $\text{Var}(f_0 \mid p) = pq$. Since $E(N \mid p) = m + 1$, $N - (m + 1)$ is an unbounded, unbiased estimator of 0. Hence, any estimator of the form

$$f_c(\gamma) = \frac{1}{m + 1}\,Y(\gamma) + c\left[1 - \frac{N(\gamma)}{m + 1}\right], \tag{6.5}$$

for some constant c, is an unbiased estimator of p. It is easily checked that f_{p_0} is the efficient estimator at p_0. Thus, the efficient estimators of p are unbounded; this serves as a reminder that unbiased estimators—even efficient ones—are not necessarily desirable estimators.

Using the moments of N found in Section 4 for the inverse binomial sampling plan, $\operatorname{Var}(f_{p_0} \mid p)$ is found to be

$$\frac{pq}{m+1}\left[1 + m\left(\frac{p_0}{p} - 1\right)^2\right]. \tag{6.6}$$

Comparing this result with the variance of the bounded estimator, f_0, it is seen that $\operatorname{Var}(f_{p_0} \mid p) < \operatorname{Var}(f_0 \mid p)$ for $p > p_0/2$.

Finally, it should be noted that the functions that are estimable efficiently at each p under this plan are precisely the functions that are estimable efficiently under a single sample plan of sample size $m + 1$. The latter plan yields the same $E(N \mid p)$ as the above plan and admits an unbiased estimator of p with uniformly smaller variance than any of the estimators discussed above.

The following variant of Scheme I also yields polynomials for $E(N \mid p)$.

Scheme II. Let n and m be two positive integers with $n \geqq 2$, and let

$$B = \{\gamma : N(\gamma) = n,\ Y(\gamma) < n - 1\} \cup \{\gamma : Y(\gamma) = n + m\}. \tag{6.7}$$

Thus, the boundary is the same as in Scheme I except that the point with coordinates $X(\gamma) = 1$, $Y(\gamma) = n - 1$, is omitted. In other words, a single sample of size n is taken. If all of the observations are equal to 1, sampling is continued until m additional 1's are obtained. If all but one of the initial n observations are equal to 1, sampling is continued until $m + 1$ additional 1's are obtained. If at least two of the initial n observations are not equal to 1, sampling ceases.

Thus, denoting the number of successes in the first n observations by Y_n,

$$\begin{aligned} E(N \mid p) &= E(N \mid Y_n < n - 1, p)\Pr(Y_n < n - 1 \mid p) \\ &\quad + E(N \mid Y_n = n - 1, p)\Pr(Y_n = n - 1 \mid p) \\ &\quad + E(N \mid Y_n = n, p)\Pr(Y_n = n \mid p) \\ &= n(1 - p^n - np^{n-1}q) \\ &\quad + \left(n + \frac{m+1}{p}\right)(np^{n-1}q) + \left(n + \frac{m}{p}\right)p^n \\ &= n + n(m+1)p^{n-2} + (m - n - mn)p^{n-1}. \end{aligned} \tag{6.8}$$

The sampling plans included in Scheme II are also boundedly complete but not complete.

There are obvious variants of Schemes I and II obtained by removing other points on the line $N(\gamma) = n$ from the boundary, or by interchanging the roles of X and Y.

7. Selection of the sampling plan. Theorem 5.1 shows that in order to estimate a function $g(p)$ efficiently at p_0 a sampling plan must be selected for which

$$E(N \mid p) = k[g(p) - g(p_0)]/(p - p_0). \tag{7.1}$$

The efficient estimator at p_0 is defined by

$$f(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + g(p_0), \tag{7.2}$$

where $a = 1/k$. As discussed in Section 5, there does not always exist a sampling plan under which a given function $g(p)$ is estimable efficiently at a given value, p_0. On the other hand, it is made clear in Section 6 that for a given $g(p)$ and p_0, there may exist more than one sampling plan satisfying (7.1). Indeed, it has been shown that there may exist more than one plan with a given $E(N \mid p)$ (i.e., with a given value of k).

In this section it is assumed that for a given $g(p)$, p_0, and k, there does exist more than one sampling plan satisfying (7.1). Let $\mathcal{S}$ denote the class of such plans. Since every plan of $\mathcal{S}$ yields the same $E(N \mid p)$, and since, under every plan of $\mathcal{S}$, the estimator f given by (7.2) is efficient at p_0, it follows from the information inequality that $\text{Var}(f \mid p_0)$ is the same under every plan of $\mathcal{S}$. In general, however, for values of p other than p_0, $\text{Var}(f \mid p)$ will be different under the various plans of $\mathcal{S}$. The problem considered here is that of determining the plan of $\mathcal{S}$ for which $\text{Var}(f \mid p)$ is smallest at some value of p other than p_0, or the plan that minimizes $\text{Var}(f \mid p)$ in the neighborhood of p_0. It will be shown that this is equivalent to determining the plan for which $\text{Var}(N \mid p)$ is minimized at the relevant values of p.

In the following derivation of $\text{Var}(f \mid p)$ it is assumed for simplicity that, in (7.2), $a = 1$ and $g(p_0) = 0$. Thus,

$$\begin{aligned} f &= q_0 Y - p_0 X = (qY - pX) + (p - p_0)N, \\ f^2 &= (qY - pX)^2 + (p - p_0)^2 N^2 + 2(p - p_0)N(qY - pX). \end{aligned} \tag{7.3}$$

This yields, upon application of Lemmas 2.5 and 2.6,

$$\begin{aligned} E(f \mid p) &= (p - p_0)E(N \mid p), \\ E(f^2 \mid p) &= pqE(N \mid p) + (p - p_0)^2 E(N^2 \mid p) + 2(p - p_0)pqE'(N \mid p), \end{aligned} \tag{7.4}$$

and hence,

$$\text{Var}(f \mid p) = (p - p_0)^2 \text{Var}(N \mid p) + pqE(N \mid p) + 2(p - p_0)pqE'(N \mid p). \tag{7.5}$$

This expression yields the following theorem, where for any estimator T and sampling plan S, $E(T \mid p, S)$ is the expectation of T at p under the plan S.

Theorem 7.1. *Let $f = a(q_0 Y - p_0 X) + b$, Let p^* be a value of p other than p_0. Let S_1 and S_2 be two sampling plans such that $E(N \mid p, S_1) = E(N \mid p, S_2)$ for all p. Then $\text{Var}(f \mid p^*, S_1) \lesseqgtr \text{Var}(f \mid p^*, S_2)$ if and only if $\text{Var}(N \mid p^*, S_1) \lesseqgtr \text{Var}(N \mid p^*, S_2)$.*

It follows from assumption (i) of Section 2 that $\text{Var}(N \mid p)$ is a continuous function of p. Hence, again referring to (7.5), the following result is immediate.

Theorem 7.2. *Let f, S_1, and S_2 satisfy the hypotheses of Theorem 7.1, and suppose $\text{Var}(N \mid p_0, S_1) < \text{Var}(N \mid p_0, S_2)$. Then there exists an interval I of positive length containing p_0 such that*

$$\begin{aligned} \text{Var}(f \mid p_0, S_1) &= \text{Var}(f \mid p_0, S_2), \\ \text{Var}(f \mid p, S_1) &< \text{Var}(f \mid p, S_2), \qquad \text{for } p \in I, p \neq p_0. \end{aligned} \tag{7.6}$$

One method that might at first seem useful in determining the plan of $\mathcal{S}$ for which $\text{Var}(N \mid p_0)$ is minimized would be to show that, for a particular plan, this variance attained the lower bound provided by the information inequality; i.e., that

$$\text{Var}\,(N \mid p_0) = \frac{p_0 q_0 [E'(N \mid p_0)]^2}{E(N \mid p_0)}. \tag{7.7}$$

The next lemma and theorem show that there exists such a plan only in the trivial situations where $\mathcal{S}$ contains an efficient sampling plan.

LEMMA 7.1. *For a given sampling plan,*

$$\text{Var}\,(N \mid p) = \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \tag{7.8}$$

for all p if and only if the sampling plan is efficient.

PROOF. Suppose the plan is efficient. In a single sample plan, N is constant. Hence, $\text{Var}(N \mid p) = E'(N \mid p) = 0$ for all p, and (7.8) holds. In an inverse binomial sampling plan, N is efficient and hence again (7.8) holds.

Suppose the plan is not efficient. Since N is constant only under a single sample plan, it is not constant under the given plan. Thus, (7.8) cannot hold for all p, for if it did, N would be an efficient estimator.

THEOREM 7.3. *For any sampling plan, either*

$$\text{Var}\,(N \mid p) = \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \qquad \text{for all } p, \tag{7.9}$$

or

$$\text{Var}\,(N \mid p) > \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \qquad \text{for all } p. \tag{7.10}$$

PROOF. From the information inequality,

$$\text{Var}\,(N \mid p) \geqq \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \qquad \text{for all } p. \tag{7.11}$$

Suppose equality holds in (7.11) for $p = p_0$. Then N must be of the form

$$N(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + b$$

for some constants a and b. Hence,

$$(1 - aq_0)Y(\gamma) + (1 + ap_0)X(\gamma) = b$$

for all boundary points γ. Since the coefficients of X and Y cannot both vanish, the boundary points lie on a straight line, and, by Theorem 3.1, the plan is efficient. The conclusion follows from Lemma 7.1.

It would be reassuring to know that, despite Theorem 7.3, there always exists a plan in $\mathcal{S}$ for which $\text{Var}(N \mid p_0)$ is smallest. This would be true if $\mathcal{S}$ contained only a finite number of sampling plans. No general results have as yet been obtained in this direction, but the following theorem concerning the specific case where $\mathcal{S}$ contains all plans with $E(N \mid p) = 2$ indicates the type of result that might be expected.

THEOREM 7.4. *There exist exactly three sampling plans for which* $E(N \mid p) = 2$ *for all* p.

PROOF. Three plans have already been given for which $E(N \mid p) = 2$; namely, the single sample plan, the plan S^* given under Scheme I of Section 6, with $n = 1$, $m = 1$, and the symmetric image of this plan obtained by interchanging the roles of X and Y. It will now be shown that these are the only three plans. Throughout the proof, points will be specified by their coordinates $(X(\gamma), Y(\gamma))$.

Consider a sampling plan S with boundary B for which $E(N \mid p) = 2$ for all p. The points (0, 1) and (1, 0) cannot both be in B, for then $E(N \mid p) = 1$. If neither (0, 1) nor (1, 0) is in B, then $N \geqq 2$. Thus, if $E(N \mid p) = 2$, then $\Pr(N = 2 \mid p) = 1$ and S is the single sample plan. Suppose then that exactly one of the points (0, 1) and (1, 0) is in B and, for the moment, assume it to be (1, 0). It will be shown that S is the plan S^* described above with boundary

$$B^* = \{(1, 0), (0, 2), (1, 2), (2, 2), (3, 2), \cdots\}.$$

Suppose $(0, 2) \not\in B$. Then $N \geqq 3$ whenever the sample path goes through (0, 2). Hence, $E(N \mid p) \geqq 3p^2$, which is greater than 2 for values of p arbitrarily close to 1. Thus, $(0, 2) \in B$. Also, $(1, 1) \not\in B$, for, otherwise, $\Pr\{N \leqq 2 \mid p\} = 1$, $\Pr\{N < 2 \mid p\} > 0$, and $E(N \mid p) < 2$. The proof is now completed by an inductive argument.

Assume, for a given integer n, $n \geqq 0$, the points (1, 0) and (0, 2), (1, 2), $\cdots$, $(n, 2)$ are in B and the points (0, 1), (1, 1), $\cdots$, $(n + 1, 1)$ are not in B. It must be shown that $(n + 1, 2) \in B$ and $(n + 2, 1) \not\in B$.

Suppose $(n + 1, 2) \not\in B$. Then $N \geqq n + 4$ whenever the sample path goes through $(n + 1, 2)$. This happens with probability p^2q^{n+1}. Thus, by considering the other points of B,

$$\begin{aligned} E(N \mid p) &\geqq q + 2p^2 + 3p^2q + \cdots + (n + 2)p^2q^n + (n + 4)p^2q^{n+1} \\ &= q + p^2(1 + 2q + \cdots + (n + 1)q^n + (n + 2)q^{n+1}) \\ &\qquad\qquad + p^2(1 + q + \cdots + q^n + q^{n+1}) + p^2q^{n+1} \\ &= q + p^2 \frac{d}{dq}\left[\frac{1 - q^{n+3}}{1 - q}\right] + p^2\left[\frac{1 - q^{n+2}}{1 - q}\right] + p^2q^{n+1} \\ &= 2 + p^2q^{n+1} - (n + 4)pq^{n+2} - q^{n+3}. \end{aligned}$$

Now let $p = 1 - \delta$, $q = \delta$. Then

$$E(N \mid p) \geqq 2 + \delta^{n+1} + o(\delta^{n+1}),$$

and for δ positive but arbitrarily close to 0, $E(N \mid p) > 2$. It follows that $(n + 1, 2) \in B$.

Now, if $(n + 2, 1)$ were in B, then S would be a bounded procedure and its continuation points would be a proper subset of the continuation points of S^*. Since S and S^* yield the same $E(N \mid p)$, this is impossible. Thus, $(n + 2, 1) \not\in B$ and the induction is completed. It follows that $S = S^*$.

If, originally, it was assumed that $(0, 1) \in B$, $(1, 0) \not\in B$, then an entirely analogous demonstration would show that S must be the symmetric image of S^* described at the beginning of the proof.

8. On the completeness of bounded sampling plans. A *bounded* sampling plan is one for which there exists a positive integer n such that

$$\Pr\{N \leqq n \mid p\} = 1. \tag{8.1}$$

The *size* of a bounded sampling plan is the smallest n for which (8.1) holds.

In the following Theorems 8.1–8.5 it is shown that if the boundary of a plan of size n contains $n + 1 + k$ points, $k \geqq 0$, then k is the dimension of the linear space of unbiased estimators of $\mathbb{Q}$.

THEOREM 8.1. *The boundary of a sampling plan of size n contains at least $n + 1$ points.*

PROOF. If $n = 1$, then (0, 1) and (1, 0) must both be boundary points. Proceeding by induction, suppose the theorem to be true for $n = m$ and consider a plan S of size $m + 1$. Since $m + 1 \geqq 2$, the points (0, 1) and (1, 0) cannot both be boundary points and there must exist a path through at least one of them that extends to a boundary point of sample size $m + 1$. Without loss of generality, suppose this is true of the point (1, 0). Then, given that the sample path has reached (1, 0), the sampling plan (or what remains of it) is now of size m and hence, by the induction hypothesis, involves at least $m + 1$ boundary points. In other words, there are at least $m + 1$ boundary points that can be reached by paths through (1, 0). Since S is closed and bounded there must also exist a boundary point of the form $(0, y)$. Since this point cannot be reached from (1, 0) it is not counted above, and the boundary contains at least $m + 2$ points.

THEOREM 8.2. *If the boundary of a sampling plan of size n contains more than $n + 1$ points, the plan is not complete.*

PROOF. The probability, $P(p; \gamma)$, of reaching a particular boundary point γ is

$$P(p; \gamma) = K(\gamma)p^{Y(\gamma)}q^{X(\gamma)},$$

a polynomial in p of degree at most n. Hence, the expectation

$$\sum_{\gamma \in B} f(\gamma)P(p; \gamma)$$

of any estimator f is also such a polynomial. Conversely, every linear combination of the $P(p; \gamma)$ can be attained as the expectation of some estimator. Thus, there exists a non-trivial estimator f such that $E(f \mid p) = 0$ for all p if and only if there exist constants $a(\gamma)$, not all 0, such that $\sum_{\gamma \in B} a(\gamma)P(p; \gamma) = 0$ for all p. But the linear space of polynomials in p of degree at most n has dimension $n + 1$ and hence, if B contains more than $n + 1$ points, the set of polynomials $\{P(p; \gamma): \gamma \in B\}$ must be linearly dependent.

Theorem 8.3. *Under a sampling plan of size n, all polynomials in p of degree at most n are estimable unbiasedly.*

Proof. The polynomial 1 is trivially estimable unbiasedly. Furthermore, since the plan is of size n, there exists a boundary or continuation point γ_j of sample size j for each j, $1 \leqq j \leqq n$. Let $\gamma_j = (x_j, y_j)$, where $x_j + y_j = j$. Girshick Mosteller, and Savage [7], have shown explicitly how to construct an unbiased estimator of $p^{y_j}q^{x_j}$, a polynomial in p of degree j. Thus, the polynomials $1, p^{y_1}q^{x_1}, \cdots, p^{y_n}q^{x_n}$ are all estimable unbiasedly and, hence, so is any linear combination of them. Since no two of these polynomials are of the same degree they are linearly independent and thus form a basis for the $n + 1$-dimensional space of all polynomials in p of degree at most n. It follows that every such polynomial is estimable unbiasedly.

Theorem 8.4. *If the boundary of a sampling plan of size n contains exactly $n + 1$ points the plan is complete.*

Proof. The expectation of every estimator is a linear combination of the $n + 1$ polynomials $P(p; \gamma)$, $\gamma \varepsilon B$, and by Theorem 8.3, every polynomial of degree at most n can be expressed as such a linear combination. Thus, the set $\{P(p; \gamma): \gamma \varepsilon B\}$ spans the $n + 1$-dimensional linear space of polynomials. It follows that the $n + 1$ polynomials $P(p; \gamma)$ must be linearly independent and, hence, there exist no non-trivial unbiased estimators of 0.

Theorem 8.5. *If the boundary of a sampling plan of size n contains $n + 1 + k$ points, $k > 0$, then there exist exactly k linearly independent unbiased estimators of* 0.

Proof. Consider the $n + 1 + k$ boundary points $\gamma_1, \cdots, \gamma_{n+1+k}$. Each estimator f can be considered as a vector $(f_1, \cdots, f_{n+1+k})$ where $f(\gamma_j) = f_j$, $j = 1, \cdots, n + 1 + k$. Thus, the space of estimators can be considered as an $n + 1 + k$-dimensional linear space, V. The expectation operator E is a linear mapping from V onto the $n + 1$-dimensional linear space of polynomials in p of degree at most n. The subspace $V^0 = \{f: E(f \mid p) = 0$ for all $p\}$ is the null space of this mapping and it follows from the standard theorems concerning rank and nullity that V^0 has dimension k.

The remainder of this section is devoted to bounded sampling plans that are complete only after the removal of some points from the boundary. It will be shown that for a plan of this type it is easy to explicitly construct a basis for the space of unbiased estimators of 0.

The following notation is used in Theorems 8.6–8.8. It is assumed that S is a bounded sampling plan with boundary B, and that $\beta_1, \cdots, \beta_t$, with $t > 0$, are t points of B such that the sampling plan with boundary $B - \{\beta_1, \cdots, \beta_t\}$ is closed and complete.

The sampling plan with boundary $B_j = B - \{\beta_j\}$ is denoted by S_j for $j = 1, \cdots, t$.

For any point $\gamma \varepsilon B$, $K(\gamma)$ is the number of paths to γ under the plan S and $K_j(\gamma)$ is the number of paths to γ under the plan S_j for $j = 1, \cdots, t$. If γ is not a boundary point for a particular plan then the number of paths to γ under that plan is taken to be 0. Thus, $K_j(\beta_j) = 0$ for $j = 1, \cdots, t$.

It should be noted that $K(\gamma)$ does not vanish for any $\gamma \in B$. Also, if $\gamma \neq \beta_j$ then $K_j(\gamma) \geqq K(\gamma)$. This follows from the fact that any path to γ under S is obviously a path to γ under any sampling plan with fewer boundary points.

Consider now, under the sampling plan S, the estimators ϕ_j defined for $j = 1, \cdots, t$ by

$$\phi_j(\gamma) = 1 - [K_j(\gamma)/K(\gamma)], \tag{8.2}$$

for $\gamma \in B$.

THEOREM 8.6. *Under the sampling plan S, the estimators $\phi_1, \cdots, \phi_t$ form a basis for the linear space of unbiased estimators of* 0.

PROOF. It was shown by Girshick, Mosteller, and Savage [7], and it is very easy to verify, that, for each j, ϕ_j is not identically 0 and $E(\phi_j \mid p) = 0$ for all p. Since the boundary of S contains t more points than that of a complete plan of the same size as S it follows from Theorem 8.5 that the dimension of the space of unbiased estimators of 0 is t. Thus, the theorem will be proven if it is shown that the estimators $\phi_1, \cdots, \phi_t$ are linearly independent.

Let the sample size of β_j be n_j (i.e., $X(\beta_j) + Y(\beta_j) = n_j$) and assume that $\beta_1, \cdots, \beta_t$ have been ordered so that $n_1 \leqq n_2 \leqq \cdots \leqq n_t$. Consider the matrix A with elements $a_{ij} = \phi_j(\beta_i)$; thus,

$$A = \begin{pmatrix} \phi_1(\beta_1) & \cdots & \phi_t(\beta_1) \\ \vdots & & \vdots \\ \phi_1(\beta_t) & \cdots & \phi_t(\beta_t) \end{pmatrix}. \tag{8.3}$$

For $j > i$, $K_j(\beta_i) = K(\beta_i)$. This follows by noting that a path to β_i under S_j goes through points of sample size less than n_i and then through β_i. It does not go through β_j since $n_j \geqq n_i$, and hence, it is also a path to β_i under S. Thus, $K_j(\beta_i) \leqq K(\beta_i)$. The reverse inequality has been stated in the comments preceding the theorem. It follows from (8.2) that $\phi_j(\beta_i) = 0$ for $j > i$. Since $\phi_j(\beta_j) = 1$ for each j, the matrix A is seen to be triangular with each of its diagonal elements equal to 1. Hence, $|A| = 1$. It can now be concluded that $\phi_1, \cdots, \phi_t$ are linearly independent, for otherwise $|A|$ must vanish. This completes the proof.

Under a given sampling plan, an estimator f^* is said to be a *uniformly minimum variance estimator* of its expected value, if, for any other estimator f with the same expectation as f^*, $\mathrm{Var}(f^* \mid p) \leqq \mathrm{Var}(f \mid p)$ for all p. Consider the estimator $f^* = q_0Y - p_0X$ that is efficient at p_0. The following theorem shows that under a sampling plan of the type now being considered f^* is not a uniformly minimum variance estimator. On the other hand, if f is any other estimator with the same expectation as f^*, then $\mathrm{Var}(f^* \mid p_0) < \mathrm{Var}(f \mid p_0)$ and f is not a uniformly minimum variance estimator. Thus, under the sampling plan S, if a non-constant function $g(p)$ is estimable efficiently at p_0 then there is no uniformly minimum variance estimator of $g(p)$.

THEOREM 8.7. *Under the sampling plan S, the estimator $q_0Y - p_0X$ is not a uniformly minimum variance estimator.*

PROOF. Lehmann and Scheffé [10] have shown that for bounded sampling plans a necessary and sufficient condition for an estimator T to be a uniformly minimum variance estimator is that $E(T\phi \mid p) = 0$ for every ϕ such that $E(\phi \mid p) = 0$. Suppose $q_0Y - p_0X$ was a uniformly minimum variance estimator. Then, for $j = 1, \cdots, t$,

$$(8.4)\qquad \begin{aligned} 0 &= E[\phi_j(q_0Y - p_0X) \mid p] \\ &= E[\phi_j(qY - pX) \mid p] + (p - p_0)E(\phi_jN \mid p). \end{aligned}$$

By Lemma 2.5, $E[\phi_j(qY - pX) \mid p] = 0$ for all p, and hence, $E(\phi_jN \mid p) = 0$ for all p and each j. (Actually, equation (8.4) yields $E(\phi_jN \mid p) = 0$ only for $p \neq p_0$. That $E(\phi_jN \mid p_0) = 0$ follows from continuity.) In particular, $E(\phi_tN \mid p) = 0$ where again it is assumed that the points $\beta_1, \cdots, \beta_t$ have been ordered so that $n_1 \leqq \cdots \leqq n_t$.

It follows from Theorem 8.6 that there exist constants $r_1, \cdots, r_t$ such that

$$(8.5)\qquad N(\gamma)\phi_t(\gamma) = \sum_{j=1}^{t} r_j\phi_j(\gamma)$$

for all $\gamma \varepsilon B$. In particular,

$$(8.6)\qquad n_i\phi_t(\beta_i) = \sum_{j=1}^{t} r_j\phi_j(\beta_i)$$

for $i = 1, \cdots, t$. Recalling that $\phi_j(\beta_j) = 1$ and $\phi_j(\beta_i) = 0$ for $j > i$, the values of $r_1, \cdots, r_t$ are readily found from (8.6) to be $r_1 = \cdots = r_{t-1} = 0$, $r_t = n_t$. Thus, from (8.5),

$$(8.7)\qquad N(\gamma)\phi_t(\gamma) = n_t\phi_t(\gamma)$$

for all $\gamma \varepsilon B$. It follows that $\phi_t(\gamma) = 0$ for all γ such that $N(\gamma) \neq n_t$. But, as argued in the proof of Theorem 8.6, $\phi_t(\gamma) = 0$ for all boundary points of sample size n_t with the exception of β_t. Hence,

$$(8.8)\qquad E(\phi_t \mid p) = \phi_t(\beta_t)K(\beta_t)p^{Y(\beta_t)}q^{X(\beta_t)} \neq 0,$$

a contradiction. It follows that $q_0Y - p_0X$ is not a uniformly minimum variance estimator.

The next theorem shows that, despite Theorem 8.7, there do exist non-constant uniformly minimum variance estimators.

THEOREM 8.8. *Under the sampling plan S, there exist non-constant uniformly minimum variance estimators.*

PROOF. Since S is bounded, there exists a boundary point $\alpha_0 = (x, 0)$ and a boundary point $\alpha_1 = (0, y)$. Neither α_0 nor α_1 is one of the points $\beta_1, \cdots, \beta_t$ since if either is removed from B the resulting plan is not closed. Furthermore, $\phi_j(\alpha_0) = \phi_j(\alpha_1) = 0$ for $j = 1, \cdots, t$ since, under any of the plans S or S_j, there is only one path to either α_0 or α_1. Now consider an estimator f of the form

$$f(\alpha_0) = f_0,$$
$$f(\alpha_1) = f_1, \tag{8.9}$$
$$f(\gamma) = c \qquad \text{for } \gamma \,\varepsilon\, B - \{\alpha_0, \alpha_1\}.$$

Then $E(f\phi_j \mid p) = cE(\phi_j \mid p) = 0$ for $j = 1, \cdots, t$, and it follows from Theorem 8.6 and the condition of Lehmann and Scheffé given at the beginning of the proof of Theorem 8.7 that f is a uniformly minimum variance estimator.

9. Acknowledgments. I am deeply indebted to L. J. Savage, under whose careful and considerate guidance this research was done, and to S. G. Ghurye and W. H. Kruskal for several helpful comments.

REFERENCES

[1] R. R. Bahadur, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[2] R. R. Bahadur, "On unbiased estimates of uniformly minimum variance," *Sankhya*, Vol. 18 (1957), pp. 211–224.

[3] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477–501.

[4] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[5] R. Courant, *Differential and Integral Calculus*, Vol. 1, Interscience Publishers, Inc., New York, 1937.

[6] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd Edition, John Wiley and Sons, New York, 1957.

[7] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[8] J. B. S. Haldane, "A labour-saving method of sampling," *Nature*, Vol. 155 (1945), pp. 49–50.

[9] J. B. S. Haldane, "On a method of estimating frequencies," *Biometrika*, Vol. 33 (1945), pp. 222–225.

[10] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation—Part I," *Sankhya*, Vol. 10 (1950), pp. 305–340.

[11] E. L. Lehmann and C. Stein, "Completeness in the sequential case," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 376–385.

[12] L. J. Savage, "A uniqueness theorem for unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 295–297.

[13] L. J. Savage, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954.

[14] C. Stein, "Unbiased estimates with minimum variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 406–415.

[15] M. C. K. Tweedie, "Inverse statistical variates," *Nature*, Vol. 155 (1945), p. 453.

[16] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1947.

[17] J. Wolfowitz, "On sequential binomial estimation," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 489–493.

[18] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.