

TA 1526 (1976)

USDA TECHNICAL BULLETINS

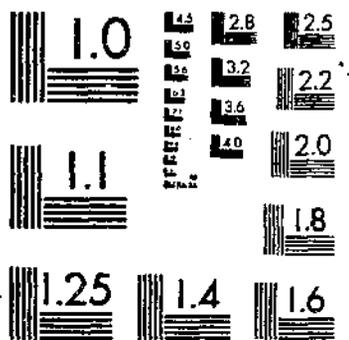
UPDATA

THE RATIONALE OF SEQUENTIAL SAMPLING, WITH EMPHASIS ON ITS USE IN

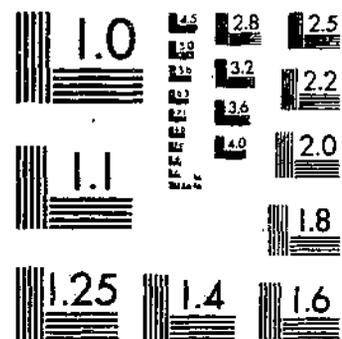
OSNAGER, J. A.

1 OF 1

START



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

R 630
0537
1000

REFERENCE
DO NOT LOAN

THE RATIONALE OF SEQUENTIAL SAMPLING, WITH EMPHASIS ON ITS USE IN PEST MANAGEMENT

U. S. DEPOSITORY

MAY 17 1976

Los Angeles Public Library

Technical Bulletin No. 1526

Agricultural Research Service
UNITED STATES DEPARTMENT OF AGRICULTURE
in cooperation with
South Carolina Agricultural Experiment Station

CONTENTS

	Page
Abstract	1
Introduction	1
Mechanics of sequential sampling	2
Sampling from a binomial distribution	8
Formulas for sequential plans	8
Binomial example	9
Sampling from a Poisson distribution	10
Formulas for sequential plans	11
Poisson example	12
Sampling from a negative binomial distribution	13
Formulas for sequential plans	13
Negative binomial example	14
General considerations	15
Literature cited	17
Appendix.—Definition of symbols	18

Illustrations

Fig.	
1. Sequential sampling chart	3
2. Operating characteristics curve for sequential sampling plan	6
3. Average sample number curve for sequential sampling plan	7

Table

1. Partial probability distribution of expected results of sampling when $p'=0.5$, $p_1=0.2$, $p_2=0.8$, and $\alpha=\beta=0.058823$	7
---	---

THE RATIONALE OF SEQUENTIAL SAMPLING, WITH EMPHASIS ON ITS USE IN PEST MANAGEMENT

By JEROME A. ONSAGER, *entomologist, Agricultural Research Service, U.S. Department of Agriculture, Charleston, S.C.*¹

ABSTRACT

The rationale of sequential sampling is described in nonmathematical terms. Formulas are given for dichotomous sampling plans based on the binomial, Poisson, and negative binomial distributions. Solved examples are given for each type of distribution, and some applications to sampling for insect pests are discussed.

INTRODUCTION

Knipling (3)² has discussed pest management programs designed to maximize the role of natural biological agents and minimize the use of insecticides in maintaining pest populations below economic threshold levels. The concept relies on critical monitoring of both destructive and beneficial insect populations and presupposes that the presence or absence of critical densities can be accurately detected. Since emphasis is essentially on proper classification of populations, rather than on estimation of population parameters, sequential sampling is eminently appropriate to present-day pest management programs (10).

According to Wald (9), the concept of sequential sampling was brought to fruition in 1943 under wartime military contract for use in quality control. The savings realized from sampling sequentially were so significant (often exceeding 50%) and the potential for other uses was so obvious that the information was released to the public in 1945. A decade later, Waters (10) observed that the application of sequential sampling to biological problems had been "somewhat limited" and gave an excellent discussion of the advantages, applications, basic requirements, and essential features of sequential sampling plans in relation to entomological problems. While the use of sequential sampling has subsequently proliferated, its maximum potential will not be realized until all those who could use it are in full command of it.

¹Present address: Rangeland Insect Laboratory, Montana State University, Bozeman 59715.

²Italic numbers in parentheses refer to items in "Literature Cited," p. 17.

Ordinary sampling methods usually require a fixed number of samples that will estimate a key density with some desired level of precision. However, the specified number of samples will invariably be inadequate for constant precision at low densities and excessive for high densities. Sequential sampling is designed simply to choose between two alternatives (i.e., is the density higher or lower than a critical level?). Sampling continues only until results dictate that one alternative is more likely true than the other at some constant acceptable level of probability. When densities are either very high or very low, the most likely alternative quickly becomes apparent. Thus, while the two methods may have similar sampling requirements within moderate population densities, the sequential method is much more efficient over the entire range of possible densities.

I believe that the apparent mathematical complexity of sequential sampling has unfortunately inhibited its use in entomology. The rationale behind sequential plans is actually quite simple, but the explanations in literature are mathematical and therefore seem complicated to the reader who is not a mathematician. Furthermore, there is a tendency for each textbook on the subject to contain a unique approach, a unique set of symbols, and a unique variation of common formulas, omitting at the same time helpful information that has been exploited elsewhere. While this lends an air of originality to each discourse and contributes to conciseness, it annoys and confuses beginning students of sequential sampling.

The intent of this bulletin is to bring together some of the information now scattered in previous treatments of the subject and present it in the simplest possible mathematical terms. Comprehension will require some facility in the use of common logarithms (log to the base 10) and familiarity with the point-slope and slope-intercept forms of the general equation of a straight line. The mechanics rather than the mathematics of sequential sampling will be emphasized, however, and methods for developing dichotomous sampling plans will be given for three distributions commonly encountered in sampling for insects: the binomial, the Poisson, and the negative binomial. Unfortunately, this requires yet another unique approach, unique set of symbols,³ and a unique variation of common formulas. It is hoped that such means can be used to make sequential sampling easier to understand.

MECHANICS OF SEQUENTIAL SAMPLING

All dichotomous sequential plans begin with the establishment of two alternative hypotheses, here designated H_1 and H_2 . Consecutive samples are examined and evaluated until cumulative results dictate that

³All symbols used in this bulletin are defined in the appendix.

one hypothesis is more likely to be correct than the other at some preestablished degree of reliability.

With the help of figure 1, let us develop a trivial example based on the binomial distribution. Suppose we are examining ears of sweet corn for the presence of corn earworm larvae. We expect to find no larvae in noninfested ears, and exactly one larva per infested ear, because the larvae are cannibalistic. Let us assume that (1) lots having 20 percent or fewer infested ears are "good" lots, and it is very important that they be so classified; (2) lots having 80 percent or more infested ears are "bad" lots, and it is very important that they be so classified; and (3) lots having more than 20 percent but less than 80 percent infestation can be classified as either "good" or "bad" without serious consequence.⁴ Then, for any given lot to be sampled, we state, " H_1 is that the true proportion (p') of infested ears is equal to or less than 0.2"; or, more concisely, " $H_1: p' \leq 0.2$." (H_1 always contains the lowest level of whatever we are sampling for.) Likewise, $H_2: p' \geq 0.8$.

Intuition tells us that when p' is extremely low or extremely high, our samples will yield a preponderance of noninfested or infested ears, respectively, and will quickly lead to acceptance of H_1 or H_2 , respectively. However, if we are examining a lot for which p' is exactly 0.5 (halfway between the specified limits H_1 and H_2), it is possible for

⁴The reader should not feel disturbed about the wide range between H_1 and H_2 in this simple example. It will later become apparent that H_1 and H_2 can be set with any practical difference between their specified limits. The given levels are simply convenient for subsequent discussion.

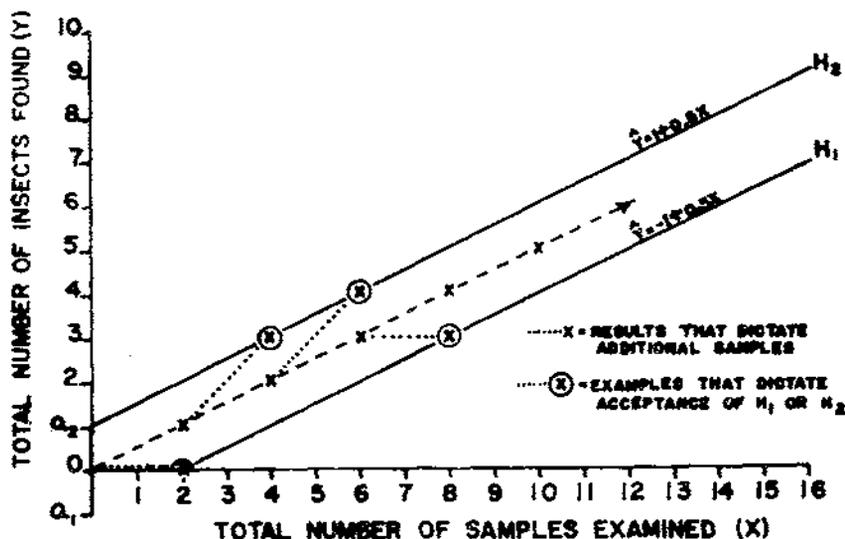


FIGURE 1.—Sequential sampling chart.

several consecutive pairs of samples to each yield one infested and one noninfested ear. This could theoretically continue indefinitely (fig. 1), so the only way our results could approach H_1 or H_2 would be through obtaining by chance an overabundance of noninfested or infested ears, respectively.⁵ Let us therefore agree to sample by examining pairs of ears chosen at random from a lot and to pause after each pair to evaluate results. Let us further agree that (1) if neither ear is infested, we shall accept H_1 (that is, we will stop sampling and classify the lot as "good"); (2) if both ears are infested, we shall accept H_2 (that is, we will stop sampling and classify the lot as "bad"); and (3) if only one of the ears is infested, we shall withhold judgment and examine another pair of ears chosen at random.

Figure 1 shows four of the possible sample points that dictate acceptance of a hypothesis. Given these points, we can calculate the slope-intercept equation, $Y = a + bX$, for two parallel lines which will intersect all possible sample results that will yield a decision to stop sampling. The point-slope form of the equation defines slope (b) as the change in Y per unit change in X . For H_1 , $b = (4 - 3)/(6 - 4) = 0.5$. Since the lines are parallel, b for H_2 is also 0.5. Given b and any point on a line, the Y -intercept (a) is calculated as $a = Y - bX$. For H_1 , $a_1 = -1$; for H_2 , $a_2 = +1$. With the solid lines in figure 1 for guidance, it is no longer essential that we examine a pair of ears before we pause to assess results. However, we may note that it is impossible to reach a decision on an odd-numbered ear in this example, so nothing would be lost by continuing to examine two at a time.

Clearly, our sampling plan is not devoid of error. It is entirely possible to draw a pair of infested ears and accept H_2 when H_1 is true. This is often called a type I error, and the maximum allowable risk of its occurrence is referred to as alpha (α). Likewise, we could accept H_1 and H_2 is true. This is often called a type II error, and its maximum allowable risk is referred to as beta (β).

The significance of α and β can be illustrated by calculating these values for our example. To make such calculations, some additional symbols will be helpful. Let us designate the specified limits of H_1 and H_2 as p_1 and p_2 , respectively. The quantities $(1 - p_1)$ and $(1 - p_2)$ are designated q_1 and q_2 , respectively. Thus, in our example, $p_1 = 0.2$, $q_1 = 0.8$, $p_2 = 0.8$, and $q_2 = 0.2$.

⁵The perceptive reader will here declare that we are relying on sampling errors to steer us to one of two classifications, neither of which is technically correct. This observation is absolutely true. We plan to capitalize on the fact that sampling errors are governed by the laws of probability. We are completely agreeable to making incorrect classifications in borderline cases where there is no serious consequence. Such error is inevitable in any decisionmaking process, and the concession protects us from making serious errors in classification.

When the true proportion of infestation is exactly as the specified limits of H_1 or H_2 (that is, when $p'=p_1$ or $p'=p_2$), the associated p is the binomial probability that any given single ear drawn at random will contain a larva; the associated q is the probability that the same ear will not contain a larva.

If $p'=p_1$, the probability of finding a larva in each of two random ears—which would lead to incorrect acceptance of H_2 —is p_1^2 or 0.04 (the logic is akin to calculating that the chances of observing consecutive "heads" in two flips of a coin is $(\frac{1}{2})^2 = \frac{1}{4}$). Likewise, the probability of no larva in two ears is q_1^2 or 0.64, which would lead to correct acceptance of H_1 . Since α may be estimated as the ratio of incorrect classifications to total classifications, $\alpha = 0.04 / (0.04 + 0.64) = 0.058823$. Obviously, as p' becomes progressively less than p_1 , the probability of accepting H_2 becomes progressively less than α ; the converse is also true.

If $p'=p_2$, the probability of a larva in each of two ears and correct classification is $p_2^2 = 0.64$. The probability of incorrect classification is $q_2^2 = 0.04$, so $\beta = 0.04 / (0.04 + 0.64) = 0.058823$. (It is not essential that $\alpha = \beta$; however, $\alpha + \beta$ must be less than 1.)

It can also be demonstrated that α and β are constant, regardless of the number of samples required to yield acceptance of H_1 or H_2 . For example, if $p'=p_2$, the probability of finding three larvae in four ears and correctly accepting H_2 is $p_2^3 q_2 = (p_2 q_2)(p_2^2) = (0.16)(0.64)$. The probability of one larva in four ears and incorrect classification is $p_2 q_2^3 = (p_2 q_2)(q_2^2) = (0.16)(0.04)$. The quantity $p_2 q_2$, or any equivalent value derived from a different number of samples, will always cancel out in the division process, leaving a ratio of 0.04/0.68 as before.

Figure 1 has now become an empirical sequential sampling plan with $H_1: p_1 \leq 0.2$, $H_2: p_2 \geq 0.8$, and $\alpha = \beta = 0.058823$. At this point we are ready to examine the plan to assure ourselves that it is feasible. This requires calculation of an operating characteristic (OC) curve and an average sample number (ASN) curve.

The typical OC curve (solid line in fig. 2) gives the level of probability (LP) of accepting H_1 for any given p' . Four of the points on the OC curve are always known. (1) When $p'=0$, $LP=1$; that is, in our example, accepting H_1 is 100 percent certain when all ears in a lot are free of earworms. Likewise, (2) when $p'=1$, $LP=0$. (3) When $p'=p_2$, LP has by definition been set at β . Likewise, (4) when $p'=p_1$, $(1-LP)$ has been set at α so $LP=(1-\alpha)$. Completion of the OC curve essentially involves proper interpolation between these known values. The inverse of the typical OC curve (the broken line in fig. 2) gives the LP of accepting H_2 for any given p' . This inverse curve is seldom plotted but always understood to exist.

The ASN curve (fig. 3) gives the number of samples that will, on the average, dictate acceptance of either H_1 or H_2 for any given p' . We must

now develop an ASN curve to determine whether or not our plan is feasible. If the ASN dictates examination of several bushels of ears per decision, we may be forced to either relax our standards or abandon the idea of sampling altogether. Conversely, if the sampling of only a few ears per decision is required, we may either accept the plan in its present form or decide that we can afford greater accuracy. The ASN is reduced at the expense of accuracy by increasing α , β , or the interval between H_1 and H_2 . Conversely, accuracy is increased at the expense of additional sampling requirements by reducing α , β , or the interval between H_1 and H_2 .

Two of the points on our ASN curve may be determined by inspection of figure 1. When $p'=0$, the only possible number of samples is 2. Likewise, when $p'=1$, the ASN must also be 2. Other points are not found so easily, but intuition tells us that the maximum ASN will occur when $p'=0.5$, because the example has to this point been symmetrical in all aspects. Table 1 gives the expected distribution of results possible when $p'=0.5$. Notice that the probability of accepting either H_1 or H_2 after examining any pair of ears equals the probability that we will have to examine at least one additional pair of ears. Thus, 50 percent of the time we can stop after two samples, 25 percent of the time after four

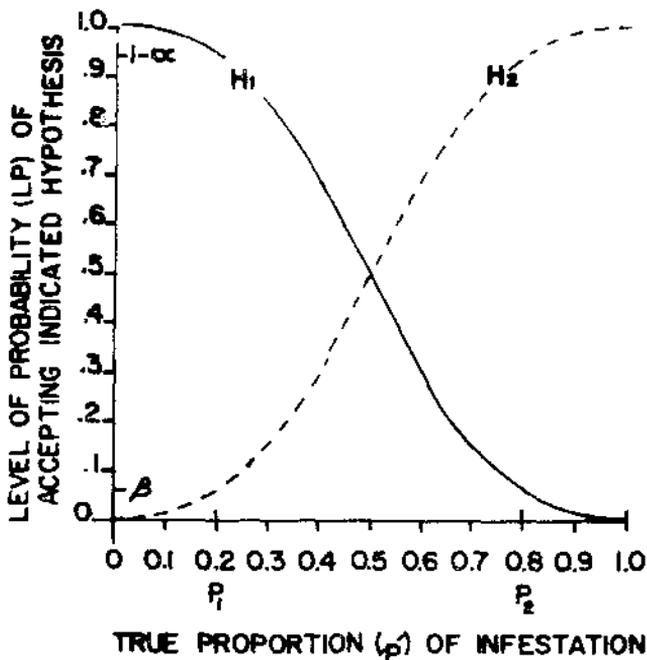


FIGURE 2.—Operating characteristics curve for sequential sampling plan.

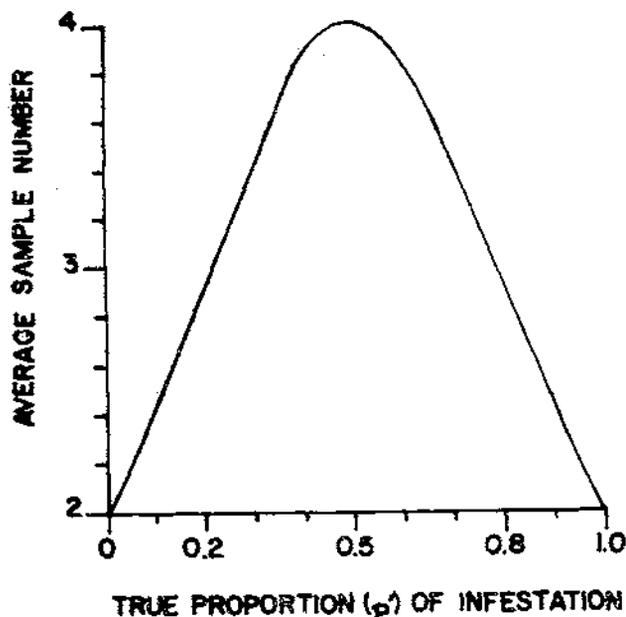


FIGURE 3.—Average sample number curve for sequential sampling plan.

samples, and so forth. Obviously, as we approach an infinite number of samples, the probability that additional samples will be required approaches zero. If we attempt to average all possible sample numbers weighted by their probabilities, i.e.,

$$\frac{(2)(0.50) + (4)(0.25) \dots + (\infty)(0)}{0.50 + 0.25 \dots + 0}$$

we will find that the numerator approaches 4 while the denominator approaches 1 as the number of samples approaches infinity. Thus, we say that the limit of the above equation is 4, which is then the ASN for $p'=0.5$. Stated otherwise, if we were to sample ears from a million different lots, each of which was exactly 50 percent infested, we would expect to examine a total of 4 million ears in the process. It can be confirmed by similar calculations for other values of p' that 4 is indeed a maximum on the ASN curve for our example.

TABLE 1.—Partial probability distribution of expected results of sampling when $p'=0.5$, $p_1=0.2$, $p_2=0.8$, and $\alpha=\beta=0.058823$

Expected result	Number of samples examined					
	0	2	4	6	8	10
Accept H_1	0	0.25	0.125	0.0625	0.03125	0.015625
Keep sampling.....	1	.50	.250	.1250	.0625	.03125
Accept H_2	0	.25	.125	.0625	.03125	.015625

Now, to relate our example to real life. In practice, the steps are as follows: (1) The parameters by which populations of an insect can be estimated are determined and (2) tolerable and intolerable population levels that can be set as limits of H_1 and H_2 are determined. These are logical tasks of the research scientist. Next, (3) α and β are set arbitrarily at acceptable levels of risk for committing a type I and type II error, respectively. Ideally, this is not the responsibility of the scientist, but should be done by parties that are affected by the decision; e.g., the farmer, the manager of a scouting program, the company that has contracted to buy the crop, etc. Then (4) appropriate general formulas (which are given later for binomial, Poisson, and negative binomial distributions) are used to calculate a_1 , a_2 , and b ; OC and ASN values are calculated and evaluated; and, if necessary or feasible, H_1 , H_2 , α , or β are adjusted to yield an adequate compromise between accuracy and practicality. Finally, (5) a sampling chart like figure 1 or a table containing equivalent information is prepared by using the formulas $y_1 = a_1 + bX$ and $y_2 = a_2 + bX$ to determine the points of acceptance for H_1 and H_2 , respectively. Now, subject to certain considerations mentioned later, we are ready to sample in the field.

SAMPLING FROM A BINOMIAL DISTRIBUTION

The binomial distribution applies when random samples from an infinite population (or at least a very large one) fall into one of two categories (e.g., units are either round or square, male or female, infested or noninfested, etc.). The parameter of interest is p , the proportion of samples that bear the attribute of interest. The proportion that lacks that attribute is q ; obviously then, $q = 1 - p$.

Formulas for Sequential Plans

Given a binomial distribution, $H_1 : p' \leq p_1$, $H_2 : p' \geq p_2$, and acceptable values for α and β , necessary calculations follow the general formulas below, all of which are variations of or derivations from Waters' formulas (10).

Intercepts:

$$a_1 = - \frac{\log \left(\frac{1 - \alpha}{\beta} \right)}{\log \left(\frac{p_2 q_1}{p_1 q_2} \right)}$$

$$a_2 = \frac{\log \left(\frac{1-\beta}{\alpha} \right)}{\log \left(\frac{p_2 q_1}{p_1 q_2} \right)}$$

Slope:

$$b = \frac{\log \left(\frac{q_1}{q_2} \right)}{\log \left(\frac{p_2 q_1}{p_1 q_2} \right)}$$

Binomial OC curve.—It has already been given that when $p'=0$, $LP=1$; when $p'=p_1$, $LP=1-\alpha$; when $p'=p_2$, $LP=\beta$; and when $p'=1$, $LP=0$. Also, when $p'=b$, $LP=a_2/(a_2-a_1)$. These are the only values essential for crude evaluations of tentative sampling plans, since the rest of the OC curve can be estimated by interpolation. For any p' between 0 and 1, the exact LP and the associated ASN can be calculated by an implicit method through use of a "dummy" variable (10). This is not a simple task unless it is done by a programmable calculator or a computer, however.

Binomial ASN curve.—Wald (9) reported that the weighted-average method we used earlier to calculate ASN's is usable only under certain rigid conditions. To circumvent these limitations, he developed an alternate method that yields what he termed "arbitrarily fine approximations" of values that cannot be calculated exactly. The general formula is

$$ASN = \frac{LP(a_1 - a_2) + a_2}{p' - b}$$

By substitution, we observe that when $p'=0$ and $LP=1$, $ASN=a_1/(-b)$. Likewise, when $p'=1$ and $LP=0$, $ASN=a_2/(1-b)$. We also observe that the general formula is not valid when $p'=b$ because we cannot divide by zero. However, Wald (9) has shown that the ASN when $p'=b$ is the product of the ASN's at $p'=0$ and $p'=1$; that is, when $p'=b$,

$$ASN = \frac{(a_1)(a_2)}{(-b)(1-b)}$$

The ASN at $p'=b$ will be maximum ASN when $\alpha=\beta$; it will also suffice as an estimate of maximum ASN when α and β are similar.

Binomial Example

While all constants in the trivial example could be verified here, let us rather develop a more significant hypothetical example. Assume that an infestation of aphids on alfalfa has been classified as "threatening" and that we want to assess the potential of a population of chalcid parasites.

Previous experience has established that (1) aphid populations are apt to increase, and that some remedial action is required if parasitism is 8 percent or less, (2) aphid populations are likely to be reduced if 13 percent or more of the aphids exhibit symptoms of parasitism, and (3) the balance is precarious when parasitism is between 8 and 13 percent. Thus, $H_1: p' \leq 0.08$ and $H_2: p' \geq 0.13$. If H_1 is true we definitely want to be alerted to that fact, so let us set α at 0.05. If H_1 is not true, it will not be an extremely serious mistake to incorrectly assume it is true because we retain the option of correcting the error in subsequent monitoring. Let us therefore set β at 0.15. Calculations may proceed as follows:

$$\frac{p_1 q_1}{p_1 q_2} = \frac{(0.13)(0.92)}{(0.08)(0.87)} = \frac{0.1196}{0.0696} = 1.718391$$

$$a_1 = - \frac{\log \left(\frac{0.95}{0.15} \right)}{\log 1.718391} = - \frac{0.801632}{0.235122} = -3.409$$

$$a_2 = \frac{\log \left(\frac{0.85}{0.05} \right)}{\log 1.718391} = \frac{1.230449}{0.235122} = 5.233$$

$$b = \frac{\log \left(\frac{0.92}{0.87} \right)}{\log 1.718391} = \frac{0.0242686}{0.235122} = 0.1032$$

p'	LP		ASN	
	Formula	Value	Formula	Value
0	Definition	1	$a_1 / -b$	33.0
p_1	$1 - \alpha$.95	General	128.2
b	$\frac{a_2}{(a_2 - a_1)}$.6055	$\frac{(a_1)(a_2)}{(-b)(1-b)}$	192.8
p_2	β	.15	General	147.0
1	Definition	0	$a_2 / (1 - b)$	5.8

A true maximum ASN of 193.7 was determined by several iterations of the implicit method. The above value for $p' = b$ is obviously an excellent approximation of the maximum ASN.

SAMPLING FROM A POISSON DISTRIBUTION

The Poisson is a distribution in its own right, but is most easily visualized as an approximation to the binomial when p is very small and

the number of samples is large. We therefore assume a random distribution of insects and expect a preponderance of samples to be devoid of insects. Unlike the binomial, however, the Poisson distribution allows the presence of two or more insects per sample. The required parameter of the Poisson is the mean, which is usually designated as lambda (λ).

Formulas for Sequential Plans

The mechanics of the Poisson sequential sampling plan is similar to that of the binomial plan. However, the mathematics differs, because Poisson probabilities are more complex. Given a Poisson distribution, $H_1: \lambda' \leq \lambda_1, H_2: \lambda' \geq \lambda_2$, and acceptable values for α and β , necessary calculations follow the general formulas below, which are variations of formulas by Waters (10).

Intercepts:

$$a_1 = - \frac{\log \left(\frac{1-\alpha}{\beta} \right)}{\log \left(\frac{\lambda_2}{\lambda_1} \right)}$$

$$a_2 = \frac{\log \left(\frac{1-\beta}{\alpha} \right)}{\log \left(\frac{\lambda_2}{\lambda_1} \right)}$$

Slope:⁶

$$b = \frac{0.4343(\lambda_2 - \lambda_1)}{\log \left(\frac{\lambda_2}{\lambda_1} \right)}$$

Poisson OC curve.—Obviously, when $\lambda' = 0$, $LP = 1$; when $\lambda' = \lambda_1$, $LP = 1 - \alpha$; and when $\lambda' = \lambda_2$, $LP = \beta$. When $\lambda' = b$, $LP = a_2 / (a_2 - a_1)$, as for the binomial distribution. These four LP values will suffice for crude evaluation of sampling plans. Exact LP's and ASN's for any λ' again may be calculated by an implicit method (10).

Poisson ASN curve.—The general formula is

$$ASN = \frac{LP(a_1 - a_2) + a_2}{\lambda' - b}$$

By substitution, when $\lambda' = 0$ and $LP = 1$, $ASN = a_1 / -b$. When $\lambda' = b$, the

⁶The constant 0.4343 is $1/\ln 10$, a conversion that allows use of the more familiar logarithms to the base 10, rather than the natural logarithms that Waters (10) specified.

general formula again is not valid, but the ASN can be calculated as $(a_1)(a_2)/-b$. Thus, four points on the Poisson ASN curve may be easily calculated.

The maximum ASN will be very close to one of the three ASN's for $\lambda' = 0$, $\lambda' = \lambda_1$, or $\lambda' = b$. Exactly which of the three ASN's will be greatest varies with the levels of α and β , and with the magnitude of λ_2 relative to λ_1 . Experience has shown that when α and β are 0.05 or greater, and when λ_2 does not exceed 10 λ_1 , the highest ASN calculated by the formulas above will be at least 94 percent of the true maximum ASN.

Poisson Example

Consider a sequential plan for detecting the presence or absence of an economic infestation of Pacific Coast wireworms in soil to be planted to potatoes. The distribution of larvae in soil cores having a surface area of one-fourth square foot and a depth of 9 inches is Poisson, an average of 0.022 larva per core will cause no significant damage, and band treatments with chemical insecticide become economical when larvae average 0.030 per core (6, 7). Let us arbitrarily set $\alpha = 0.4$ and $\beta = 0.1$. Then, $H_1: \lambda \leq \lambda_1 = 0.022$, and $H_2: \lambda' \geq \lambda_2 = 0.030$.

$$a_1 = - \frac{\log\left(\frac{0.6}{0.1}\right)}{\log\left(\frac{0.030}{0.022}\right)} = - \frac{\log 6}{\log 1.3636} = - \frac{0.778151}{0.134698} = -5.78$$

$$a_2 = \frac{\log\left(\frac{0.9}{0.4}\right)}{0.134698} = \frac{\log 2.25}{0.134698} = \frac{0.352182}{0.134698} = 2.61$$

$$b = \frac{(0.4343)(0.03 - 0.022)}{0.134698} = 0.0258$$

λ'	LP		ASN	
	Formula	Value	Formula	Value
0	Definition	1	$a_1/-b$	224.0
λ_1	$1-\alpha$.60	General	637.9
b	$a_2/(a_2-a_1)$.3111	$(a_1)(a_2)/-b$	584.7
λ_2	β	.10	General	421.7

The exact maximum ASN was determined by several iterations of the implicit method to be 639. This dictates examination of up to nearly 4.5 cubic yards of soil per decision. Clearly, this plan is not feasible, and no reasonable manipulation of H_1 , H_2 , α , or β will make it so.

SAMPLING FROM A NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial has often been used to describe the distribution of insects that tend to occur in aggregates or clumps. Sampling from this distribution typically yields more zero counts and more very high counts than one expects from a random distribution like the Poisson. The required parameters are the mean (\bar{X}) and an exponent, called k , which is a function of the degree of aggregation.

Formulas for Sequential Plans

Given a negative binomial distribution, $H_1 : \bar{X}' \leq \bar{X}_1$, $H_2 : \bar{X}' \geq \bar{X}_2$, a value of k that is valid when $\bar{X}_1 \leq \bar{X}' \leq \bar{X}_2$, and acceptable values for α and β , the subsequent procedures are described by Oakland (5).

Two new parameters, P and Q , are calculated for each hypothesis as follows:

Parameter	H_1	H_2
$P = \frac{\bar{X}}{k}$	$P_1 = \frac{\bar{X}_1}{k}$	$P_2 = \frac{\bar{X}_2}{k}$
$Q = 1 + P$	$Q_1 = 1 + P_1$	$Q_2 = 1 + P_2$

The new parameters are used in all subsequent calculations.

Intercepts:

$$a_1 = - \frac{\log \left(\frac{1-\alpha}{\beta} \right)}{\log \left(\frac{P_2 Q_1}{P_1 Q_2} \right)}$$

$$a_2 = \frac{\log \left(\frac{1-\beta}{\alpha} \right)}{\log \left(\frac{P_2 Q_1}{P_1 Q_2} \right)}$$

Slope:

$$b = k \frac{\log \left(\frac{Q_2}{Q_1} \right)}{\log \left(\frac{P_2 Q_1}{P_1 Q_2} \right)}$$

OC curve.—Again, when $\bar{X}' = 0$, $LP = 1$; when $\bar{X}' = \bar{X}_1$, $LP = 1 - \alpha$; when $\bar{X}' = \bar{X}_2$, $LP = \beta$; and when $\bar{X}' = b$, $LP = a_2 / (a_2 - a_1)$. These four LP values

will suffice for crude evaluation of sampling plans. Oakland (5) and Waters (10) give formulas whereby other LP's and associated ASN's may be calculated implicitly.

ASN curve.—Oakland's (5) formula is essentially as follows:

$$ASN = \frac{a_2 + (a_1 - a_2)(LP)}{\bar{X}' - b}$$

When $\bar{X}' = 0$ and $LP = 1$, the formula again reduces to $ASN = a_1 / -b$. When $\bar{X}' = b$, the general formula again is not valid, but the ASN can be calculated as $(a_1)(a_2) / -(b^2/k + b)$.⁷ Once again, four points on the ASN curve may be calculated easily.

The maximum ASN will be very close to the ASN for $\bar{X}' = b$ when $\alpha \ll \beta$. As α becomes progressively greater than β , the reliability of this approximation rapidly diminishes and use of the implicit method becomes mandatory for a realistic estimate of the maximum ASN.

Negative Binomial Example

Let us consider a sequential plan reported by Sylvester and Cox (8) for sampling sugar beet plants to detect aphid infestations warranting treatment. Given $H_1: \bar{X}' \leq \bar{X}_1 = 0.9$ aphids per plant, $H_2: \bar{X}' \geq \bar{X}_2 = 1.1$, $k = 0.81$, and α and $\beta = 0.1$, then

Parameter	H_1	H_2
P	$P_1 = \frac{0.9}{0.81} = 1.111111$	$P_2 = \frac{1.1}{0.81} = 1.358025$
Q	$Q_1 = 1 + P_1 = 2.111111$	$Q_2 = 1 + P_2 = 2.358025$

Intercepts:

$$\alpha_1 = - \frac{\log \left(\frac{0.9}{0.1} \right)}{\log \left(\frac{(1.358025)(2.111111)}{(1.111111)(2.358025)} \right)} = - \frac{\log 9}{\log 1.094241} = - \frac{0.954242}{0.039113} = -24.40$$

$$\alpha_2 = \frac{\log \left(\frac{0.9}{0.1} \right)}{\log 1.094241} = \frac{0.94242}{0.039113} = 24.40$$

⁷I am not aware of this formula having been published previously, but it has worked for every type of example that I have been able to contrive.

Slope:

$$b = \frac{0.81 \log \left(\frac{2.358025}{2.111111} \right)}{\log 1.094241} = \frac{(0.81)(0.048037)}{0.039113} = 0.9948$$

X'	LP		ASN	
	Formula	Value	Formula	Value
0	Definition	1	a_1/b	24.5
\bar{X}_1	$1-\alpha$.90	General	205.8
b	$a_2/(a_2-a_1)$.50	$\frac{(a_1)(a_2)}{-(b^2/k+b)}$	268.5
\bar{X}_2	β	.10	General	185.5

The above values differ only by rounding errors from reported values. The exact maximum ASN was determined by several iterations of the implicit method to be 268.95.

GENERAL CONSIDERATIONS

Morris (4) reported that persons engaged in sampling preferred to evaluate results with the aid of tables rather than graphs like figure 1.

Situations will arise occasionally in which a rather large number of samples have failed to yield a decision. Several authors have offered arbitrary suggestions for resolving this predicament, but none of these solutions appears to have universal merit. The necessary considerations are based largely on economics and available options, which are apt to be unique for each sampling plan. In sampling for foliage-feeding insects, for example, if the maximum ASN has been examined without a decision, we could assume that the need for corrective action is not imminent. However, in the case of soil insects that must be controlled by insecticide broadcast before the crop is planted, the decision cannot be postponed indefinitely. In such a situation, as in the case of disease vectors, we might decide to apply control measures whenever sampling fails to dictate that controls are not required. In general, arbitrary decisions of this sort force us to make a guess, and thereby defeat the basic reason for sampling. Granted that sampling should have increased the probability that the guess will be a good one, the guess nevertheless remains a guess.

One possible solution to indecisiveness lies in careful study of tentative sampling plans. The nature of the problem often dictates that most sampling will be of populations in the vicinity of H_1 and H_2 . This is especially true for scouting programs, because (1) natural trends are to populations in excess of H_1 , or the insect would not be an economic pest in

the first place, (2) infestations greater than H_2 are constantly reduced by control measures, and (3) reinfestation to the level of H_2 does not proceed in a predictable manner. (If it did, we could justifiably adapt a rigid treatment schedule and dispense with sampling.) We should therefore be prepared to cope with maximum rather than minimum ASN's, as well as the occasional examination of an unusually large number of samples. One hopes that H_1 , H_2 , α , or β can be adjusted to preclude a high incidence of indecisiveness.

An alternate solution to indecisiveness that appears feasible when units are sampled several times in succession is to consider the estimated mean from each examination in scheduling the subsequent examination. The rules of maximum likelihood tell us that the population is probably lower in field I where H_1 was accepted after examining n samples than in field II where H_1 was accepted after $2n$ samples. Likewise, field II probably has fewer insects than field III, where a decision was not reached after examining an arbitrary limit of $3n$ samples. Field III obviously needs to be resampled sooner than the others, and we can efficiently give it the attention it needs if we ignore field I for a while.

A word of caution concerning the selection of a value for k in negative binomial sequential plans appears to be in order. For some insects, such as potato beetles and leatherjackets, k is essentially constant (1). That is, k is apparently independent of density, so a common value (referred to as common k or k_c) is an adequate parameter of aggregation or clumping for any typical population density. A common k poses no problem to sequential sampling. However, with other insects, such as certain wireworms (1) and aphids (8), aggregation (and therefore k also) varies with density. It is therefore essential that the k used in a sequential sampling plan be indicative of aggregation between the critical densities that are specified as limits of H_1 and H_2 .

Sylvester and Cox (8) and Ingram and Green (2) report that when the negative binomial distribution is true and critical levels are reached before 100 percent of the sample units become infested, it may under some conditions be more expedient to determine only whether samples are infested or not infested, as opposed to the tedious chore of counting each insect in each infested sample. Given a negative binomial distribution, \bar{X} , and k , the probability of a given sample being noninfested is

$$\left(\frac{k}{k + \bar{X}} \right)^k$$

and is calculated as the antilog of k times

$$\log \left(\frac{k}{k + \bar{X}} \right)$$

This probability is equivalent to q of the binomial distribution. Since the

binomial $p=1-q$, \bar{X}_1 and \bar{X}_2 of the negative binomial can be converted to p_1 and p_2 of the binomial. A binomial sequential plan based on the converted parameters will theoretically yield the same decisions as the negative binomial plan with identical values of α and β . The binomial plan will require the higher ASN's, but compensation lies in the fact that only noninfested samples are examined in total.

A common lamentation these days is that accurate economic thresholds are not available for several important pests that, like the proverbial mountain, have often been assaulted simply because they were there. However, sequential sampling techniques can help prevent this lack of information from inhibiting the inception of a pest management program that otherwise would proceed. In the first place, economic thresholds that can be accurately defined as a point probably do not exist, and I would question the wisdom of efforts to refine them to that degree. Given that most thresholds vary within a range from season to season and field to field, the limits of that range can be set as the specified limits of H_1 and H_2 of the sequential sampling plan. Consequently, as infestations increase from the upper limit of H_1 to the lower limit of H_2 , it becomes increasingly probable that they will indeed cause economic injury, and corrective measures accordingly will be dictated with increasing probability. Secondly, in the absence of precise data, a minimum of research combined with experience and astute judgment can often produce an estimated range that will subtend most of the true range of given economic thresholds. Serious reservations about the accuracy of initial estimates for H_1 and H_2 can be compensated for by setting α and β at low levels (about 0.05 or lower). The OC curve will then be relatively flat in the neighborhood of the specified limits of H_1 and H_2 (see fig. 2), indicating that relatively large errors in economic thresholds will have relatively little effect on the proportion of erroneous classifications rendered by use of the plan. Wood (11) has aptly discussed the justification for proceeding on the basis of the best available information in response to urgency.

LITERATURE CITED

- (1) Bliss, C. I., and Owen, A. R. G. 1958. Negative binomial distributions with a common k . *Biometrika* 45: 37-58.
- (2) Ingram, W. R., and Green, S. M. 1972. Sequential sampling for bollworms on raingrown cotton in Botswana. *Cotton Grow. Rev.* 49: 265-275.
- (3) Knippling, E. F. 1972. Entomology and the management of man's environment. *J. Austr. Entomol. Soc.* 11: 153-167.
- (4) Morris, R. F. 1945. A sequential sampling technique for spruce budworm egg surveys. *Can. J. Zool.* 32: 302-313.
- (5) Oakland, G. B. 1950. An application of sequential analysis to whitefish sampling. *Biometrics* 6: 59-67.
- (6) Onsager, J. A. 1969. Sampling to detect economic infestations of *Limoniuss* spp. *J. Econ. Entomol.* 62: 183-189.

- (7) ———, and Foiles, L. L. 1970. Relative efficiency of three methods for chemical control of Pacific Coast wireworms in potatoes. *Am. Potato J.* 47: 379-385.
- (8) Sylvester, E. S., and Cox, E. L. 1961. Sequential plans for sampling aphids on sugar beets in Kern County, California. *J. Econ. Entomol.* 54: 1080-1085.
- (9) Wald, A. 1945. Sequential tests of statistical hypotheses. *Ann. Math. Stat.* 16: 117-186.
- (10) Waters, W. E. 1955. Sequential sampling in forest insect surveys. *For. Sci.* 1: 68-79.
- (11) Wood, B. J. 1971. Development of integrated control programs for pests of tropical perennial crops in Malaysia. In C. B. Huffaker (ed.), *Biological Control*, pp. 422-455. Plenum Press, New York.

APPENDIX.—DEFINITION OF SYMBOLS

General Discussion

- H_1 = the hypothesis that an attribute of interest in an unknown population equals or is less than a predescribed level which is less than a level predescribed for H_2 .
- H_2 = the hypothesis that an attribute of interest in an unknown population equals or is greater than a predescribed level which is greater than a level predescribed for H_1 .
- α = (alpha) the maximum allowable probability of committing a type I error; i.e., the probability of accepting H_2 when H_1 is true.
- β = (beta) the maximum allowable probability of committing a type II error; i.e., the probability of accepting H_1 when H_2 is true.
- a = the Y -intercept in an equation of a straight line.
- b = the measure of slope in an equation of a straight line; i.e., the rate of change in Y per unit change in X .
- LP = level of probability.
- OC = operating characteristic.
- ASN = average sample number.

Binomial Distribution

- p = a general abbreviation for the proportion of samples in an unknown population that bear the attribute of interest.
- q = $1-p$ = a general abbreviation for the proportion of samples in an unknown population that lack the attribute of interest.
- p_1 = the proportion specified as the upper limit of H_1 .
- $q_1 = 1-p_1$.
- p_2 = the proportion specified as the lower limit of H_2 .
- $q_2 = 1-p_2$.
- p' = any proportion that is known or assumed to be true.

Poisson Distribution

- λ = (lambda) a general symbol for a Poisson mean.

- λ_1 = the mean that is specified as the upper limit of H_1 .
 λ_2 = the mean that is specified as the lower limit of H_2 .
 λ' = any mean that is known or assumed to be true.

Negative Binomial Distribution

- \bar{X} = a general symbol for the mean.
 \bar{X}_1 = the mean that is specified as the upper limit of H_1 .
 \bar{X}_2 = the mean that is specified as the lower limit of H_2 .
 \bar{X}' = any mean that is known or assumed to be true.
 k = an exponent which is a parameter of the degree of aggregation.
 P = a contrived parameter calculated as $P = \bar{X}/k$.
 Q = a contrived parameter calculated as $Q = 1 + P$.
 $P_1 = \bar{X}_1/k$.
 $Q_1 = 1 + P_1$.
 $P_2 = \bar{X}_2/k$.
 $Q_2 = 1 + P_2$.

END