THE 2-SPRT AS AN ALTERNATIVE TO WALD'S SPRT WHEN TESTING HYPOTHESES CONCERNING INSECT POPULATIONS

Linda J. Young
Jerry H. Young

Follow this and additional works at: https://newprairiepress.org/agstatconference

Part of the Agriculture Commons, and the Applied Statistics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.
THE 2-SPRT AS AN ALTERNATIVE TO WALD’S SPRT WHEN TESTING HYPOTHESES CONCERNING INSECT POPULATIONS

Linda J. Young and Jerry H. Young, Departments of Statistics and Entomology, Oklahoma State University, Stillwater, OK 74078

ABSTRACT: Exact and Wald’s approximations for the operating characteristic and average sample number functions are compared for a sequential probability ratio (SPRT) test commonly used in cotton integrated pest management (IPM) programs in southwestern Oklahoma. The 95th percentile of the sample size is also given. For the same set of hypotheses and stated error probabilities, the exact operating characteristic and average sample number functions and the 95th percentile of sample size are compared for the SPRT and 2-SPRT.

KEY WORDS: SPRT, 2-SPRT, negative binomial distribution

Abraham Wald (1947) developed the sequential probability ratio (SPRT) test during World War II. The purpose was to test simple versus simple hypotheses with stated probabilities of Type I and Type II errors. This is accomplished by taking observations one at a time until enough information has been gathered to make a decision at the stated levels of probability. The SPRT has been used extensively by entomologists in integrated pest management (IPM) programs. Scouts in these programs have the responsibility of informing producers when an insect population has reached the economic threshold and chemical control is needed. Two simple hypotheses are tested to determine whether the population is above the economic threshold or below a safety level. Statistically, the procedure is appealing since this may be accomplished while at the same time controlling the probabilities of making an incorrect decision; that is, exercising control when the population is actually below the safety level or failing to exercise control when the population is above the economic threshold. From a practical standpoint, large or small populations may be quickly determined to be, respectively, above the economic threshold or below the safety zone. More sampling is required when the populations are near the hypothesized levels. In this paper, we will compare Wald’s approximations of the operating characteristic and average sample number functions to their exact values. Then we will introduce an alternative to the SPRT, known as the 2-SPRT, and compare the operating characteristic and average sample number functions and the 95th percentiles of the sample size of the two tests for a particular example.

THE SPRT

As a basis for discussion, consider a commonly used sampling plan for fleahoppers in cotton. The economic threshold is 0.4 fleahoppers per terminal (the alternative hypothesis), and the safety level is 0.2 (the null hypothesis). The distribution is negative binomial with a k of 1. Consider the
error probabilities of 0.1 for each of the two cases: spraying when we should not or failing to spray when we should. An observation is made by selecting a terminal of a cotton plant at random from the field and counting the number of fleahoppers on that terminal. Using the formulae developed by Wald (1947), we would continue sampling after n observations if

\[-4.077 + 0.286n < T_n < 4.077 + 0.286n\]

where \(T_n\) represents the total number of fleahoppers on the first n cotton terminals to be included in the sample. If the lower inequality is violated first, then sampling would cease and the decision would be that the population is at or below the safety level. If the upper inequality is violated, then sampling would stop and a need for chemical control would be indicated. Graphically, we would have three areas (see Figure 1). We begin in the continuation zone. After each observation, the total fleahopper count in the first n observations is plotted against n, the sample number. If we cross this lower boundary, we immediately stop and do not exercise chemical control. If we cross the upper boundary, we immediately stop and exercise chemical control.

Wald (1947) developed approximate operating characteristic (OC) and average sample (ASN) number curves. Fowler and Lynch (1987) showed through Monte Carlo studies that the error probabilities tend to be overstated and the average sample size underestimated when these approximations are used. They noted that this is due to the fact that seldom does sampling stop exactly on the boundary. Instead there is some overshooting. The distribution of overshooting will determine the closeness of the approximation. A program has been developed for IBM compatible personal computers which will obtain exact values of the operating characteristic and average sample number functions for the binomial, negative binomial and Poisson populations (Seebeck, 1989).

For our example, we can compare the exact operating characteristic curve with the approximate one (see Figure 2). This curve represents the probability of deciding the population is below the safety level for various values of the mean. The actual error probabilities for Type I and Type II errors are 0.071 and 0.095, respectively. While these are not exactly equal to the specified levels of 0.1, the disparity is not great enough to be of practical concern in our example. In fact, we have done better than we specified since both error probabilities are smaller than their desired values. The average sample number for possible values of the mean using both the approximate and exact computations are shown in Figure 3. When the population level is between the hypothesized ones, Wald’s approximation of the ASN is substantially below the actual value. For example, if the true mean is 0.3, Wald’s approximation gives an average sample number of 44.0 while it is actually 53.7. These approximation errors can be considered
serious. Since this is a sequential procedure, the sample size is not known in advance. In addition, the sample size tends to be highly skewed (Wetherill and Glazebrook, 1986). Thus, in addition to the average sample curve, a curve of the 95th percentile of the sample size is computed exactly across a range of possible means. In Figure 4, the 95th percentile is contrasted with the approximate ASN. This shows that if we are at the economic threshold of 0.4, for example, the sample size will be 91 or less 95% of the time. This is an important consideration in the evaluation of a sampling procedure since the average sample size of 35.0, using Wald’s approximation, does not give a clear indication of the range of possible sample sizes.

THE 2-SPRT

An obvious weakness in the SPRT may be seen by viewing Figure 1 again. If after a finite number of observations n, no decision has been made, there is a positive probability that no decision will be made after the next observation and that sampling must continue. Sometimes, in practice, sampling is merely terminated after a given number of observations are taken. Yet a plan which would assure the investigator that sampling would eventually end would be of value. For this purpose, we are suggesting the use of the 2-SPRT which was first developed by Lorden (1976, 1980).

When Wald developed the SPRT, his goal was to minimize the average sample size at the two hypothesized values; that is, the economic threshold of 0.4 and the safety level of 0.2 for our example. No other sampling procedure, fixed or sequential, has a smaller average sample size at the hypothesized values if the error probabilities are equal or smaller. However, if the true level of the population is between the safety level and the economic threshold, the average sample size may be substantially larger as Figure 3 illustrates. An alternative approach, known as the Kiefer-Weiss problem, is to develop a sampling plan which minimizes the maximum sample size for all possible values of the parameter, the mean in our example. Anderson (1960) showed that in the Weiner process with equal Type I and Type II error rates, high efficiencies can be attained by using a pair of converging straight lines as the stopping boundaries. Lorden (1976, 1977, 1980) studied a subclass of the tests discussed by Anderson which he called the 2-SPRT’s. For the Koopman-Darmois family of densities, he showed that the 2-SPRT could be used to asymptotically minimize the expected sample size at a third point in the parameter space as the error rates tend to zero, providing an asymptotic solution to the modified Kiefer-Weiss problem. Huffman (1983) showed how to choose the third point (a value of the mean between the economic threshold and safety level in our case) so that the 2-SPRT would give an asymptotic solution to the Kiefer-Weiss problem. Nagardeolekar (1988) studied the properties of the 2-SPRT when sampling from the negative binomial distribution.
The 2-SPRT is conducted by simultaneously performing two one-sided SPRT's. To apply the 2-SPRT to our problem, we must determine the third value of the mean \( \mu \) for which the sample size is to be minimized. To do so, we must first find an intermediate value \( \mu' \). Following Huffman (1983) and Nagardeolekar (1988), we begin the process using the Kullback-Leibler information numbers, given by

\[
I(\mu, \mu_i) = \mu \log\left(\frac{\mu}{\mu_i + k}\right) + k \log\left(\frac{\mu_i + k}{\mu + k}\right), \quad i = 0, 1
\]

where \( \mu_0 \) and \( \mu_1 \) are the values for the safety level and economic threshold, respectively. Also define the constants

\[
a_i(\mu) = I^{-1}(\mu, \mu_i) \log\left(\frac{\mu}{\mu_i + k}\right)
\]

for \( i = 0, 1 \). For desired Type I and Type II error probabilities of \( \alpha \) and \( \beta \), respectively, the constants \( A \) and \( B \) are computed as follows:

\[
A(\mu) = \frac{a_1(\mu) - a_0(\mu)}{a_1(\mu)} \alpha
\]

and

\[
B(\mu) = \frac{a_0(\mu) - a_1(\mu)}{a_0(\mu)} \beta
\]

Then \( \mu' \) is the solution of the equation

\[
\frac{\log(1/A)}{I(\mu', \mu_0)} = \frac{\log(1/B)}{I(\mu', \mu_1)}
\]

The point \( \mu \) for which the average sample number is to be minimized may now be found using

\[
\bar{\mu} = \frac{k \mu' \exp(u'/\sigma \sqrt{n})}{\mu'[1 - \exp(u'/\sigma \sqrt{n})] + k}
\]

where

\[
u' = \Phi\left(\frac{a_0(\mu^*)}{a_0(\mu^*) - a_1(\mu^*)}\right),
\]
\[ n^* = \frac{\log(1/A)}{I(\mu^*, \mu_0)} = \frac{\log(1/B)}{I(\mu^*, \mu_1)}, \]

\[ \sigma^2 = \mu^* + \mu^*/k, \]

and \( \Phi(\cdot) \) denotes the cumulative distribution function of the standard normal variate. The maximum sample size \( M \) for this procedure is

\[ M = \text{smallest integer } \geq \frac{\log(B) \log\left( \frac{\bar{\mu}}{\mu_0} \frac{\mu_0 + k}{\bar{\mu} + k} \right) + \log(A) \log\left( \frac{\mu_1}{\bar{\mu}} \frac{\bar{\mu} + k}{\mu_1 + k} \right)}{k \left( \log\left( \frac{\bar{\mu}}{\mu_0} \frac{\mu_0 + k}{\bar{\mu} + k} \right) \log\left( \frac{\mu_1}{\bar{\mu}} \frac{\bar{\mu} + k}{\mu_1 + k} \right) - \log\left( \frac{\mu_1 + k}{\bar{\mu} + k} \right) \log\left( \frac{\bar{\mu}}{\mu_0} \frac{\mu_0 + k}{\bar{\mu} + k} \right) \right)} \]

For \( n < M \), sampling continues as long as

\[ \frac{\log(A) + nk \log\left( \frac{\bar{\mu} + k}{\mu_0 + k} \right)}{\log\left( \frac{\bar{\mu}}{\mu_0} \frac{\mu_0 + k}{\bar{\mu} + k} \right)} < T_n < \frac{\log(1/B) + nk \log\left( \frac{\mu_1 + k}{\bar{\mu} + k} \right)}{\log\left( \frac{\mu_1}{\bar{\mu}} \frac{\bar{\mu} + k}{\mu_1 + k} \right)} \]  \hspace{1cm} (1)

If the lower inequality is not satisfied for an \( n \), sampling stops and the hypothesis that the population is below the safety level is accepted. If the upper inequality is violated for an \( n < M \), sampling is discontinued and the hypothesis that the population is above the economic threshold is accepted. At \( n = M \), a decision is made as follows:

a) accept that the population is below the safety level if \( T_n < M \).
b) accept that the population is above the economic threshold if \( T_n > M \).
c) randomize with equal probability if \( T_n = M \).

For our set of hypotheses and error probabilities, \( \mu^* = 0.288 \) which gives \( \bar{\mu} = 0.284 \). The maximum sample size is \( M = 121 \). Thus (1) states to continue sampling as long as

\[ -5.909 + 0.338n < T_n < 6.000 + 0.239n \]

The stopping boundaries for both the SPRT and 2-SPRT are shown in Figure 5. Notice that the continuation zone is now a triangular region. Initially the continuation zone is broader, forcing more samples to be taken. Yet the two boundaries converge to a point at \( M \).

The computations required to obtain the stopping boundaries may be quite extensive. However, a program, again for IBM
compatible PC's, has been written which allows the user to quickly develop a plan for stated economic threshold, safety level, and error probabilities when sampling from the binomial, negative binomial, or Poisson distributions (Lim, 1989). Exact properties of this sampling process may be computed in a manner similar to that used for the SPRT. We compare the exact properties of the 2-SPRT and SPRT. The operating characteristic curve, the probability of deciding the population is below the safety level for varying values of the mean, is shown in Figure 6. Note that while differences do exist, they are not large ones. For example, at the economic threshold of 0.4, we have specified an error rate of 0.1. For the SPRT the actual error rate is 0.095, and it is 0.121 for the 2-SPRT. Figure 7 compares the average sample number function numbers for the two tests. In some regions more observations are required to stop, but the increases seem to be more than offset by the decrease between the economic threshold and safety levels. The savings are even more dramatic when we look at the 95th percentile of the sample size in Figure 8. As seen in the figure, the 95th percentile is substantially below the maximum sample size in all regions of the parameter space. Extensive simulation studies have yet to produce a sample for which a decision occurs at n = M.

SUMMARY

Wald's approximation of the operating characteristic function works well in many practical situations. His approximation of the average sample number function may seriously underestimate the true value of the function, especially if the population parameter value is between the two hypothesized ones. Since the distribution of the sample size tends to be highly skewed, the 95th percentile of the sample size provides valuable insight into a proposed sampling plan.

The 2-SPRT has many of the features of the SPRT with the advantage that the continuation zone is a closed region. When using the 2-SPRT, a user encounters slightly larger expected regions of the parameter space, but a substantial reduction in the average sample size occurs at most values intermediate to the hypothesized ones. When considering the 95th percentile of sample size, the savings using the 2-SPRT are even more dramatic. Extensive field testing is now being done on the 2-SPRT.

REFERENCES


Lim, Siew Joo. 1989. A computer program to develop and evaluate a 2-SPRT for the negative binomial, binomial, and Poisson distributions. Masters report. Oklahoma State University, Stillwater, OK.


Seebeck, Katherine. 1989. A computer program to develop and evaluate a Wald's sequential probability ratio test for the parameters of three discrete distributions. Masters report. Oklahoma State University, Stillwater, OK.


Figure 1.
SPRT Boundaries

Treatment Zone
Continue
To Sample
Zone
Do Not Treat Zone

Figure 2
Approximate and Exact OC of SPRT

Figure 3
Approximate and Exact ASN of SPRT

Figure 4
ASN and 95th Percentile for SPRT
Figure 5.
SPRT and 2-SPRT Boundaries

Figure 6
Exact OC for SPRT and 2-SPRT

Figure 7
ASN for SPRT and 2-SPRT

Figure 8
95th Per. for SPRT and 2-SPRT

alpha and beta 0.10