SAMPLING FOR TARNISHED PLANT BUGS IN COTTON

J. L. Willers and D. C. Akins
Genetics and Precision Agriculture Research Unit
P. O. Box 5367
Mississippi State, MS, 39762

ABSTRACT

A line-intercept sampling (LIS) design was adapted for use with a drop cloth to create an efficient sampling technique for tarnished plant bug (*Lygus lineolaris* (Palisot de Beauvois)) (TPB) nymphs and tenereal adults in cotton. Essentially, the method creates a belt transect by placing the drop cloth in the furrow between two rows and repeating the procedure at least four more times. TPB are dislodged onto the drop cloth, counted and recorded onto a data sheet one row at a time. The method is applicable in cotton between the time of first square and canopy closure. The data collected by this sampling method can be manipulated one of two ways to estimate numbers per acre. The first way is to use the classical LIS estimator to determine the average number of TPB per acre. The second way incorporates Bayesian concepts based upon the hypergeometric probability function to estimate numbers of TPB per acre. This article describes the technical details that are necessary for adapting LIS with scouting cotton for TPB. Several illustrative examples based on samples from large commercial cotton fields in the Mississippi Delta are presented.

INTRODUCTION

Arthropod pests of cotton are one of several causes for loss of yield. Today, much is known about these pests and the various tactics and strategies implemented for their control. Despite this knowledge, losses in yield continue to occur despite numerous applications (Williams 1998) of pesticides. For example, total yearly losses to cotton in 1997 for the United States alone are estimated to be 9.42%. This percentage, when converted to dollars, represents a loss to production of more than $785 million (Williams 1998). When the cost of control (i.e., money spent to prevent loss) is considered, the combined US total is estimated at more than 1.4 billion dollars (Williams 1998). These combined losses suggest that knowledge gaps remain in our understanding about how to best control crop pests. Another interesting statistic, is that cotton insect monitoring costs across the US cotton belt during 1997 averaged $6.52 per scouted acre, for an estimated total of $56,873,735. It is clear that significant dollars are spent to gather information about the status of crop pests. This information is the foundation for management decisions made to control insect and mite pests.

One major hemipterous pest (King et. al. 1996) for a large portion of the cotton belt is the tarnished plant bug (*Lygus lineolaris* (Palisot de Beauvois))(TPB). The potential exists to improve the management of this pest while reducing the cost. Creating new knowledge about
TABLE 1. Listing of References on Sampling for Tarnished Plant Bug for Diverse Commodities Including Cotton.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black (1976)</td>
<td></td>
</tr>
<tr>
<td>Byerly, Gutierrez, Jones, and Luck (1978)</td>
<td></td>
</tr>
<tr>
<td>Callahan, Holbrook, and Shaw (1966)</td>
<td></td>
</tr>
<tr>
<td>Curtis (1942)</td>
<td></td>
</tr>
<tr>
<td>Dempster (1961)</td>
<td></td>
</tr>
<tr>
<td>Ellington, Kiser, Cardenas, Duttle, and Lopez (1984a)</td>
<td></td>
</tr>
<tr>
<td>Ellington, Kiser, Ferguson, Cardenas (1984b)</td>
<td></td>
</tr>
<tr>
<td>Ellington, Southward, and Carrillo (1997)</td>
<td></td>
</tr>
<tr>
<td>Fenton and Howell (1957)</td>
<td></td>
</tr>
<tr>
<td>Fleischer, Gaylord, and Edelson (1985)</td>
<td></td>
</tr>
<tr>
<td>Fleischer and Gaylord (1988)</td>
<td></td>
</tr>
<tr>
<td>Fleischer, Gaylord, Hue, and Graham (1986)</td>
<td></td>
</tr>
<tr>
<td>Fye (1969)</td>
<td></td>
</tr>
<tr>
<td>Gaylord, Buchannan, Gilliland, and Davis (1983)</td>
<td></td>
</tr>
<tr>
<td>Glick and Lattimore (1954)</td>
<td></td>
</tr>
<tr>
<td>Gonzalez, Ramsey, Leigh, Ekboth, and van den Bosch (1977)</td>
<td></td>
</tr>
<tr>
<td>Goodell, Kerby, Young, and Plant (1992)</td>
<td></td>
</tr>
<tr>
<td>Graham, Jackson, and Lakin (1984)</td>
<td></td>
</tr>
<tr>
<td>Hagler, Cohen, Bradley-Dunlop, and Enriquez (1992)</td>
<td></td>
</tr>
<tr>
<td>Harding and Dupnik (1976)</td>
<td></td>
</tr>
<tr>
<td>Holman, Tugwell, Oosterhuis, and Bourland (1994)</td>
<td></td>
</tr>
<tr>
<td>King, Phillips, and Coleman (1996)</td>
<td></td>
</tr>
<tr>
<td>Kuehl, and Fye (1972)</td>
<td></td>
</tr>
<tr>
<td>Layton (1995)</td>
<td></td>
</tr>
<tr>
<td>Leigh, Grimes, Dickens, and Jackson (1974)</td>
<td></td>
</tr>
<tr>
<td>Leigh, Gonzalez, and van den Bosch (1970)</td>
<td></td>
</tr>
<tr>
<td>Leigh, Kerby, and Wynholds (1988)</td>
<td></td>
</tr>
<tr>
<td>Lewis, Haney, and Phatek (1996)</td>
<td></td>
</tr>
<tr>
<td>Lincoln, Boyer, Dowell, Barnes, and Dean (1970)</td>
<td></td>
</tr>
<tr>
<td>Lyle (1926)</td>
<td></td>
</tr>
<tr>
<td>Menhinick (1963)</td>
<td></td>
</tr>
<tr>
<td>Mukerji (1974)</td>
<td></td>
</tr>
<tr>
<td>Pieters (1977)</td>
<td></td>
</tr>
<tr>
<td>Pieters (1978)</td>
<td></td>
</tr>
<tr>
<td>Prokopy, Hubbell, Adams, and Hauschild (1982)</td>
<td></td>
</tr>
<tr>
<td>Pyke, Sterling, and Hartstack (1980)</td>
<td></td>
</tr>
<tr>
<td>Race (1960)</td>
<td></td>
</tr>
<tr>
<td>Ruppel (1974)</td>
<td></td>
</tr>
<tr>
<td>Scott, Snodgrass, and Smith (1988)</td>
<td></td>
</tr>
<tr>
<td>Schotzko and O'Keeffe (1986a)</td>
<td></td>
</tr>
<tr>
<td>Schotzko and O'Keeffe (1986b)</td>
<td></td>
</tr>
<tr>
<td>Schotzko and O'Keeffe (1989)</td>
<td></td>
</tr>
<tr>
<td>Schotzko and O'Keeffe (1990)</td>
<td></td>
</tr>
<tr>
<td>Sevacherian (1976)</td>
<td></td>
</tr>
<tr>
<td>Sevacherian and Stern (1972a)</td>
<td></td>
</tr>
<tr>
<td>Severacherian and Stern (1972b)</td>
<td></td>
</tr>
<tr>
<td>Slaymaker and Tugwell (1984)</td>
<td></td>
</tr>
<tr>
<td>Snodgrass (1993a)</td>
<td></td>
</tr>
<tr>
<td>Snodgrass (1993b)</td>
<td></td>
</tr>
<tr>
<td>Snodgrass (1996)</td>
<td></td>
</tr>
<tr>
<td>Snodgrass and Scott (1997)</td>
<td></td>
</tr>
<tr>
<td>Stern and Mueller (1968)</td>
<td></td>
</tr>
<tr>
<td>Stewart and Gaylor (1990)</td>
<td></td>
</tr>
<tr>
<td>Stewart and Gaylor (1991)</td>
<td></td>
</tr>
<tr>
<td>Stewart and Gaylor (1993)</td>
<td></td>
</tr>
<tr>
<td>Thomas, Neyman, Polya, and Sevacherian (1976)</td>
<td></td>
</tr>
<tr>
<td>Tugwell, Young, Dumas, and Phillips (1976)</td>
<td></td>
</tr>
<tr>
<td>Williams, Phillips, and Tugwell (1987)</td>
<td></td>
</tr>
<tr>
<td>Williams, Wagner, Willers (1995)</td>
<td></td>
</tr>
<tr>
<td>Wipfli, Peterson, Hogg, and Wedberg (1992)</td>
<td></td>
</tr>
<tr>
<td>Young (1989)</td>
<td></td>
</tr>
<tr>
<td>Young and Tugwell (1975)</td>
<td></td>
</tr>
<tr>
<td>Zhang and Chen (1993)</td>
<td></td>
</tr>
</tbody>
</table>

TPB ecology requires improving the ability of field scouts to estimate TPB population abundance. Knowledge of how abiotic factors influence these trends is also essential. Collectively, these data can be graphed and analyzed to discover new management policy. Therefore, employing good sampling methods to garner these data is crucial to developing better insight.
The literature on sampling for the TPB alone is quite large, more than 60 references (Table 1) if crops other than cotton are considered. The chief aim of this paper is to describe the technical foundation of a novel sampling design for TPB. Several diverse elements are integrated to develop this unique sampling approach. One element is the adaptation of a line-intercept (Kaiser 1983; Lucas and Seber 1977; McDonald 1980; Thompson 1992) sampling design, another is the drop cloth which allows the rapid examination of numerous segments of row, and the final one is the employment of Bayesian concepts (Gazey and Staley 1986; Johnson 1977; Schmitt 1969) to best describe the sample data. It is anticipated that this combined sample design can collect those sets of data that will lead to the development of improved management practices that control this important cotton pest.

MATERIALS AND METHODS

Sampling Procedure. Briefly, the main concepts of this sample design can be described as follows. In a cotton field, a transect line (in most cases, at least one planter-width and no longer than four planter-widths in length) is established perpendicular to the row direction (Fig. 1). The sample line is placed in an area of the field where the crop can be judged to be homogenous for such parameters as growth rate, height, canopy characteristics, and fruit load. This determination is now assumed to be based upon the judgement and experience of the sampler. Eventually, however, high resolution, multi-spectral imagery (Schrader and Pouncy 1997) will be utilized to best make these determinations. In fact, the data employed in this paper is based upon sample sites selected with the use of imagery (Willers, et al. 1999; Willers unpublished).

Concepts and Terminology. Classical LIS sampling involves the use of a transect of known length that encounters, or crosses two dimensional 'objects or particles' in a study area (Lucas and Seber 1977; McDonald 1980; 1991). This well established sample design can be adapted for use in scouting cotton for populations of TPB. In the discussion that follows we describe in detail the numerous changes that are necessary to use the original LIS sample plan to estimate insect abundance in row crops such as cotton. For example, the objects in this study are the drop cloth samples along the transect line. Each drop cloth sample is called a 'quadrat' instead of an 'object or particle'.

The modified LIS method is derived from the row spacing of the crop and relationships between two imaginary lines (McDonald 1991) placed in the field (Fig. 1). One of these lines is a reference baseline (W) that runs down the center of a cotton furrow (forming the long side of a rectangle). The other is a transect line (L) that originates from a random point on the baseline and runs perpendicular across the rows (forming the short side of the rectangle, or \( W \times L = A \)). The larger rectangle thus formed is scaled to one acre in size (43,560 square feet) and forms the basic reference area (A) where a density estimate of TPB is made. The length of the transect line (L) baseline (W) must be uniquely determined for different row spacings (and planting patterns) in order to maintain a reference area one acre in size. This is convenient because many management decisions against this pest are made based on how many insects are present per acre.

In row crops it is more convenient to describe L in terms of a number of rows rather than a linear distance. The chief reason the transect line lies perpendicular to the row direction is that a convenient feature is created that insures that the transect line length is kept constant. This avoids the need to use a rope, tape measure, or other measuring device to determine the transect line's length. In all cases, the transect line must be perpendicular to the row direction; it should
FIG. 1. Diagrammatic, bird’s eye view of an 8 row, drop cloth transect line randomly placed within a 1-acre reference area. Present in dashline form are three additional, 8 row sample lines that correspond to different scales of study as noted to the left. The baseline length \(W\) and transect line length \(L\) define 1 land acre and are based on 38 in row spacing planted with an 8 row planter.

be kept straight by ‘line of sight’ and not excessively curved or zig-zagged (Anderson et al. 1979). The process of moving across rows gives the LIS method an additional advantage by capturing the variability in the crop due to planting irregularities or other causes (Willers et al. 1992). A feature of particular importance is the fact that these two lines (defined by \(W\) and \(L\)) create an inherent, but simple, randomization scheme for collecting sample data without bias (Legg and Moon 1994).

Let the variable, \(n\), be the maximum number of rows (at a particular furrow spacing) used to establish a transect line of length \(L\); thus, for row crops the lengths of these two reference lines are selected so that the product of \(L \times W = 1\) acre, or 43,560 ft\(^2\). However, if fewer rows (i.e., \(k < n\) rows) are selected to establish the transect line length, then the reference area can be subset into smaller areas; for example, \(\frac{1}{4}, \frac{1}{2},\) or \(\frac{3}{4}\) acre sample tracts (Fig. 1). In cotton, these elemental transect line lengths \((k)\) are best selected if they correspond to the planter configuration that planted the crop. The maximum length \((L)\) of a transect line is therefore, a multiple of the number of planter passes sampled and still satisfies the constraint (along with the baseline length \((W)\)) that the area of the reference rectangle be 1 acre. For example, the transect line’s maximum length is 101.33 ft (32 rows = \(L = n\)) if an 8-row planter is set to a 38 in row spacing and four planter passes are sampled (8 rows x 4 passes x 38 in/furrow ÷ 12 in/ft). If it is convenient to let four planter passes define the maximum transect line length that spans one acre, then by definition the baseline length \((W)\) is 429.9 ft \((43,560 \text{ ft}^2 ÷ 101.33 \text{ ft})\). Similar calculations would apply for other row spacings that are different than 38 in.

In scouting cotton for TPB, the drop cloth represents the LIS ‘object or particle’ crossed by a transect (which have been renamed as ‘quadrats’ to stress a distinction from traditional LIS
terminology). The drop cloth length (often, 3 ft) comprises the length dimension, \( l \), of the
quadrat. Similarly, the furrow spacing comprises the width dimension, \( w \), which is a constant
in solid stands (i.e., all rows are planted at equal spacings; skip row planting patterns require
additional modifications that are not described here). In row crops, the line segment that
represents the intersection of the transect line and the \( k \)th quadrat is (by definition) the span
between furrows, \( w \). The length, \( l \), is the dimension (i.e., the drop cloth width) of the \( k \)th quadrat
perpendicular to the transect line. The subscript \( i \) (\( i = 1, 2, 3, ..., k \leq n \)) counts the number of
individual quadrats intercepted (i.e., sampled or encountered) by the transect line. In row crops,
both \( l \) and \( w \) are constants.

We emphasize that these quadrats are artificial sampling units (Ludwig and Reynolds
1988) improvised for convenience. However, when quadrats are arranged along a transect line
(Fig. 1) a unique property results that provides both precision and utility. The chance that the
sample line (or belt transect) will encounter at least one attribute of interest (i.e., a TPB)
approaches a certainty as the number of sample units strung together along the transect line
increases. This process creates a unique type of sample unit that dynamically changes in size as
more sub-units are examined (Fig. 2). For a series of transect lines, the average length for the
occurrence of the first encounter with a pest is related to the TPB density for that area of the field
(Willers et al. 1990). In practical terms, for a sparse TPB population transect lines need to be
longer to encounter the first insect. On the other hand, transect line lengths to the point of first
encounter are shorter when the insects are more common.

When collecting a sample, the observer ideally locates a baseline's origin and then moves
to the point on the baseline where the transect line begins. In practice, however, only the transect
line origin needs to be selected. Next, the transect line is traversed using the drop cloth on each
row. User's can elect to shake each row onto the cloth one at a time and move to a new row each
time, or shake two rows onto the same side of the cloth, one row at a time (Avoidance of
excessively disturbing the plants on the adjacent row is an important consideration). If the latter
style is done, it may be necessary at times to reverse the cloth to remove debris from the previous
sample. The key is to collect a data value for each unique row by drop cloth combination.
Recall these combinations define the fundamental sample unit, or quadrat, as mentioned earlier.

To estimate the attribute totals per acre, \( \hat{Y} \), for all \( y \) in \( A \), a simple equation can be used.
Sample data \( (y_i) \) collected from the \( i \)-th quadrat along a transect line of known length is
substituted into the following equation:

\[
\hat{Y} = \frac{W}{l} \sum_{i=1}^{n} (y_i/l) = \frac{W}{l} \sum_{i=1}^{n} (y_i)
\]  
(1)

If the transect line length is \( L \), then the result \( \hat{Y} \) needs to be multiplied by a correction
factor to correctly scale the estimate to numbers per acre. The correction factor is the ratio, \( L/k \), where
\( k \) is the transect line length actually used. Convenient values to use for both \( L \) and \( k \) is
length of these two lines expressed as the number rows, rather than their lengths in feet (e.g., \( L/k = 32 \) rows/8 rows; See Fig. 1). In this discussion, eq. (1) estimates the TPB density from counts
of TPB on drop cloth samples (i.e., quadrats) placed along a transect line. A few examples using
eq. 1 follow later.

Under field conditions, the estimate of the mean number of TPB per acre can be obtained
if a table of constants appropriate for the transect line length and width is available (An example
of the same a table (e.g., Table 2) can be found in Williams et al. 1995). Alternately, use of eq. (1)
can be used to derive these constants for different types of crop row spacings (whether solid or
FIG. 2. Diagram showing how sample units of different sizes along a transect line are dynamically created by merging together smaller sized units. This behavior is quite useful for the analysis of spatial pattern.

skip row) and transect line and quadrat lengths (Willers et al. 1999). The sum (e.g., the $\sum y$) is multiplied by the appropriate constant to convert sample data from a transect line to per acre estimates. These constants are convenient tools for immediate use in the field.

In commercial fields, transect lines should be associated with unique 1 acre cells (i.e., a reference area, $A$; Fig. 1) within a strata (or management unit) and begin no less than 100 feet from the perimeter (Williams et al. 1995). The use of this latter restriction helps guard against bias that may occur due to 'edge effects' (e.g., damage due to herbicide drift from an adjacent field, dust from field traffic, etc.) that are not representative of the interior portion of the management unit. If edge effects are common and extensive for a field, then such areas should be considered as another management unit and sampled as another strata. We anticipate that the use of remote sensing (Riley 1989) imagery with GPS (Global Positioning Systems) and GIS (Geographic Information Systems) will lead to delineation of rules that best govern the placement of transect lines. At present, the relationships among characteristics such as plant vigor, invertebrate populations and spectral reflectance are too poorly understood to be able to state many formal rules or guidelines that assist in the selection of sample sites. Additional research is planned to address these issues; however, at present a few rules seem practical. These rules are to (1) select homogenous patches (regions) of the crop, (2) centrally place a sample line within a region, and (3) not have a transect line longer than the minimum breadth of that stratum at the sample site. The latter rule provides a reason why few transect lines would ever be longer than 8 rows when estimating TPB density for a strata. A line too long would simply cross over into the next stratum. Also, if more than one transect line is used within a strata (an issue ultimately left up to the user), quadrats from one line should not overlap quadrats associated with another line. In either instance, if overlap occurs, the transect lines cannot provide independent
samples (Anderson et al. 1979; McDonald 1991). However, more than one 1 acre reference area (i.e., multiple samples) can be placed within a stratum if the above rules are followed.

An alternate approach based on Bayesian methods can also be applied to these data. In fact, the LIS estimator (eq.1) for TPB totals per acre in cotton is equivalent to the mode of a discrete event probability distribution used to model these sample data. Presented next are major concepts of the Bayesian approach and algorithm developed to describe the results obtained from a belt transect.

Essential Features of Bayesian Statistical Methods. Several excellent references of Bayesian methods have been used to develop the algorithm, and include Gazey and Staley (1986), Johnson (1977) and Schmitt (1969). Several points from these informative works can be summarized as follows. First, Bayesian methods combine information from two sources: (a) the sample data, and (b) prior knowledge or information that is known before sampling begins. Using a rule known as Bayes' theorem these two sources can be combined to represent (as a probability distribution) the revised state of knowledge after sampling is finished. The sample data is described by a distribution known as the 'likelihood function', the prior knowledge is represented by a probability distribution known as the 'prior', and the resultant, or final, probability distribution is called the 'posterior'. The following expression summarizes these relationships as follows:

\[ \text{Posterior Distribution} \propto \text{Likelihood} \times \text{Prior Distribution} \]

The prior distribution can assume several forms. The distribution can be 'informative' or 'uninformative'. An uninformative prior essentially states that we have no prior knowledge. An informed prior describes what is known before a sample is collected. The prior distribution can also be 'proper' or 'improper'. If it is a proper prior, then the area 'under the curve' is equal to one. An improper prior violates this common property of probability distributions. The algorithm we describe below assumes an proper, uninformative prior (Gazey and Staley 1986) as a starting point. The final posterior distribution that results after all sample data have been iteratively processed represents, as a probability distribution, the definitive statement of how many TPB are present per acre.

Application of Hypergeometric Distribution. Assume that the sample data acquired from quadrats along a transect line can be modeled by the hypergeometric distribution (eq. 2). From these elements, a Bayesian approach can be derived to estimate the probability associated with different discrete values of the number of TPB per acre. The algorithm is an adaptation of the sequential Bayes algorithms found in Gazey and Staley (1986) and Zucchini and Channing (1986). Schmitt's (1969) discussion on using the hypergeometric distribution for a polling issue in a small community has been helpful in developing the algorithm. These references contain numerous details not presented in this discussion.

The algorithm can be explained as follows. We know some factors \textit{a priori}, those being the set, \(m\), of plants sampled and the number \(r\) of TPB found from each quadrat along a transect line of \(k \leq n\) quadrats. Analysis of the set \(m\) gives an estimate of the population of plants \(N\) in the reference area, \(A\), and by extension for the particular strata from which the data were collected. From this information, we can compute the probability of encountering \(r\) infested plants by assuming at least one TPB per plant for quadrats where \(r > 0\) and that the true population of infested plants in the field is some value \(R \leq N\).

The sample size for plants sampled is simply the sum \(\sum m\) of plants collected from all rows (i.e., quadrats or drop cloths) 'crossed' by the transect line. The category of interest, infested plants (i.e., TPB), is the sum \(\sum r\) of the number of TPB collected from quadrats along
the transect line that had one or more insects. Given the current logic of the algorithm, it is assumed that the maximum possible number of TPB per acre is by definition equivalent to the maximum number of plants per acre (N). This assumption provides convenience and does not imply a one-to-one relationship between the plants and the insects since it takes too much time to examine each plant in every quadrat along a sample line. At present, the algorithm uses stand count data collected from each quadrat on the same transect line as was used to compute the density estimate of TPB. The minimum number of plants infested will be the minimum possible population of insects in the field, which is of course zero.

For computational purposes, the minimum (m₀) and maximum (mₖ) plant populations in the reference area, A, are divided into 100 equally spaced intervals, giving 101 (K) discrete infestation levels. These 101 discrete values are used when computing the TPB count posterior distributions (R₁, R₂, R₃, . . . Rₖ). The resolution, d, of each of the population increments is given by d = [(mₖ - m₀) + 1]/K. Note that choice of 100 for the number of equally spaced intervals is an arbitrary value and is not intended to have any implied significance; it does not imply, for example, zero to one hundred percent. In this case, 100 is simply an artifact of the original procedure adapted from Schmitt (1969). At present, this resolution is adequate for sampling in row crops such as cotton. For example, if the density estimate is 4,000 insects per acre, the bias introduced by this rigid selection of the number of possible class intervals is at most only 39.6 insects per acre (e.g., 4,000/101 classes).

The value not known a priori is the actual number of infested plants (i.e., TPB) per acre. For each value (Rₖ), the probability of being the ‘true’ value, given the sample, is computed iteratively using the sample data from each successive quadrat along the transect line. Since we are assuming sampling without replacement, we use the hypergeometric distribution function (eq. 2) (Gazey and Staley 1986; Schmitt 1969) as the basis to derive \( P(R_j \mid N, m_i, r_i) \). The posteriors (or alternately, informed priors) are computed for each quadrat along the transect line by calculating the probability of \( R_j \) infested plants for each of the 101 discrete population values expressed as numbers per acre:

\[
P(R_j \mid N, m_i, r_i) = \frac{\left( \frac{\sum_j^K R_j}{\sum_i^n r_i} \right) \left( \frac{N_K - \sum_j^K R_j}{\sum_i^m n_i - \sum_j^m r_i} \right)}{\left( \frac{N_K + 1}{\sum_i^m m_i} + 1 \right)}
\]

The numerical algorithms for computing the hypergeometric function (eq. 2) were adapted from the source code found in Press et al. (1989). Specifically, the subroutines BICO and FACTLN were used to compute the binomial coefficients given two general arguments, A and B. The procedure for calculating binomial coefficients for two numbers A and B was based on the statement:

\[
BICO = \exp(FACTLN(A) - FACTLN(B) - FACTLN(A-B))
\]
Making the appropriate substitutions into equation 2 yields an equation of the form:

$$P_r = \frac{e^\alpha * e^\beta}{e^\gamma}$$  \hspace{1cm} (4)

where:

$$\alpha = \ln(A!) - \ln(B!) - \ln((A-B)!),$$

$$\beta = \ln(C!) - \ln(D!) - \ln((C-D)!),$$

$$\gamma = \ln(E!) - \ln(F!) - \ln((E-F)!).$$

and the arguments $A-F$ correspond to the elements in each of the three binomial coefficients of eq. (2).

However, solving these binomial coefficients for certain values of $A-F$ resulted in an arithmetic overflow, even for extra precision representations. Zucchini and Canning (1986) noted this problem as well and limited the size of populations being considered in their algorithm to 1,000. Since populations of TPB infestations in cotton fields may exceed 30,000 or more, this limitation due to overflow was unacceptable. Given that $e^{a}e^{b} = e^{a+b}$ and $e^{y}/e^{b} = e^{y-b}$, eq. 4 can be re-arranged into a more manageable form:

$$P_r = e^{(\alpha + \beta - \gamma)}$$  \hspace{1cm} (5)

Returning to the original code developed by Press et al. (1989), the calling equation (eq. 3) was modified to:

$$BICO = FACTLN(A) - FACTLN(B) - FACTLN(A-B)$$  \hspace{1cm} (6)

Next, the probabilities ($P_r$) for the 101 discrete population density values, $R_i$, were calculated using:

$$P_r = \exp(BICO(A,B) + BICO(C,D) - BICO(E,F))$$  \hspace{1cm} (7)

Using these concepts, the problem of machine underflow or overflow was avoided. These modifications to the computer code were computationally cleaner and provided faster performance.

In addition, calculating the intermediate posteriors for each $R_i$ for each quadrat along the transect line is not necessary before computing the final posterior; the final posterior can be derived immediately. However, the results from intermediate posteriors can be a useful visual diagnostic tool (Gazey and Staley 1986). For example, intermediate posterior distributions should tend to mass about a modal value if the population is randomly distributed and exhibits a homogenous variance among quadrats along a transect within a management unit. A continuous trend for intermediate posterior distributions that increases or decreases to a final posterior indicates that these initial assumptions (just mentioned) about the dispersion pattern of the population in the sample area has been violated (Gazey and Staley 1986). This result indicates that the sampled area, in fact, should be subdivided (or stratified) into separate study areas. Exploring the potential of this feature is an area of current research.
RESULTS

Sampling. The LIS data can be arranged into a table that can be viewed as a matrix of \( r \) rows and \( c \) columns. Each row in the matrix corresponds to a sample transect and each column corresponds to each quadrat (i.e., crop row) along the line where the number of TPB discovered on the drop cloth were counted one row at a time. The results of an analysis of sample data of this nature by two different approaches are described next.

Modified LIS Examples. These examples assume a 38 in row spacing, planted in a solid stand, with an 8 row planter.

Example 1. If an 8 row sample is zero for all 3 ft row lengths sampled along the line, the estimate by eq. (1) of TPB density is zero. Actually, for this instance, the result is best thought of as an interval bounded by the limits 0 ≤ \( \hat{Y} \leq 0.573 \) per acre. (The upper limit is found by setting \( y \leq 1 \) for only one quadrat of the 8 possible quadrats and solving for \( Y \).) For an answer of exactly zero, no line of any length can be provided to actually conclude that the field is completely free of TPB.

Example 2. If an 8 row sample provides the sample data of \( y = (0,0,0,1,0,1,0,0) \) then by eq. (1) and adjustment by the correction factor, \( L/k = 32 \) rows/8 rows, the estimate is \( \hat{Y} = (429.9/3)*2*32/8 = 1,146.4 \) bugs per acre. Alternately, using the scaling constants found in Table 2 of Williams et al. (1995), the estimate is \( \hat{Y} = 2 * 573.2 = 1,090 \) TPB per acre. The occurrence of numerous zeros in this sample line suggest a coarse-grained spatial pattern (Pielou 1977) of TPB throughout the management unit. Specifically, there are localized areas where the TPB density is beginning to become common and other areas of the field have little or no TPB. More samples among various management units within the field are necessary to confirm this supposition.

Example 3. This example suggests both a fine-grained spatial pattern (Pielou 1977) and a severe TPB problem; i.e., any area selected for sampling will more than likely be found infested with TPB. Here, drop cloth samples of an 8 row line provide the following data: \( y = 2, 2, 2, 0, 2, 0, 2, 3 \). Using the same scaling factor used above, \( \hat{Y} = 573.2 * 13 = 7,451.6 \) bugs per acre. It is left to the reader as a didactic exercise to obtain the same result using eq. (1) and the correct correction factor (i.e., \( L/k = 32 \) rows/8 rows).

Bayesian Examples. The examples just discussed require no knowledge about the number of plants present for each row along the transect. On the other hand, the Bayesian approach requires information about the stand of plants on each row along the sample line. With Bayesian methods, the median, mode and selected percentiles of interest can be readily determined from the posterior distribution of population estimates. The mode obtained by Bayesian methods is equivalent to the LIS estimate of the mean.

The following examples apply to similar data as given above but are solved using the Bayesian algorithm. The graphs (Figs. 3-5) represent samples from higher to lower numbers, and show each of the iterative sequences until the final posterior is achieved. The process stops when the data from all quadrats in a transect line have been accounted for by the algorithm. The number of plants on each row, as well as the actual counts of insects, are omitted for brevity.

The Bayesian approach also provides a point estimate of population density. Note, however, that with this approach, the point estimate is a mode and not an average. (See Gazey and Staley (1986) for a discussion of how the average and the mode of a discrete event probability distribution differ). In addition, to an estimate of the mode, but the probabilities for other outcomes (i.e., \( R \), numbers of TPB per acre) are also derived.

We can also examine how these probabilities vary as the information for each quadrat along the sample line is successively incorporated into the estimates using Bayes' rule. These intermediate probability distributions clearly describe the variation for a sample line. Figures
3 - 5 illustrate examples that are representative of how TPB density can be expected to vary as populations range from several thousand to as few as a one-hundred and fifty. Note that each figure is a combined representation of the three 8-row transect lines by assuming that only one, 24-row sample line was used in that stratum. This has been done to illustrate how the algorithm can be used to easily summarize data from a series of smaller transect lines. This 'pooling' will work for any series of smaller transect lengths whose combined length is \( \leq L \).

The density of TPB for a population of several thousand is illustrated in Fig. 3. The intermediate estimates are represented by the numerous, fine lines shown on the graph. The final estimate, known as the final posterior distribution, is represented in the figure as the thicker, dark black line having a mode near 3,210 TPB per acre.

In Figs. 4 and 5, plots are drawn for typical, non-zero estimates of sample sites from where the population is small or smaller.

The graphs (especially, Figs. 4 and 5) obtained by Bayesian methods clearly show the phenomenon of skewness, particularly for low densities. This type of graphical behavior increases our confidence that the algorithm is performing correctly. Further, while the ordering of the data values for each quadrat does cause the pattern of the intermediate posteriors to vary slightly from sample to sample, the estimate of population size is the same the distribution for the final posterior.

**DISCUSSION**

Sampling insect populations is a traditional topic among entomologists (e.g., Pedigo and Buntin 1994; Wilson et al. 1989) and with each issue of any one of the many journals the number of citations continues to grow. It is easy to see why. If the data obtained for use in making a spray or no spray decision against an insect pest of a crop are flawed, errors in management will result. The sample design described here, as a comparison among the illustrative results shows, will never err in failing to detect a severe population, especially if TPB nymphs or callow adults are present (Willers, unpublished). The only instance where the sample technique may not work well is with immigrant adults moving into cotton from wild hosts found along field margins (Willers, unpublished).

Previously, Ludwig and Reynolds (1988) discussed how the size of the sample unit influences the assessment of spatial pattern for arbitrarily defined sample units. Sample unit size has a tremendous impact upon the precision of population estimates. Consideration of the effect of sample unit size (i.e., size of a unit searched for insects) is as important as sample number (i.e., how many units were searched). The two influences work together. The result of a sample design that uses both considerations simultaneously yields a better estimation of TPB abundance and with less effort. Only a few minutes are needed to complete an 8 row transect sample if only stand and TPB are recorded.

To better understand the results of this study, consider several small technical illustrations. First, the resolution (apart from pure sample error) of the desired sample determines the length of the transect sample line. The resolution of a sample can be determined by letting \( \sum y = 1 \) for a line of length \( x \leq a \) (or \( L \)) and solving for \( \hat{Y} \) (eq. 1). Most drop cloths are 3 ft wide; thus, with a 38 in row spacing and no skip rows, an 8 row sample can measure at least 573 TPB per acre whereas a 16 row sample (with the same crop conditions) can measure at least 287 TPB per acre. By comparison, a combination of a 5 ft drop cloth and a 32 row sample could measure at least 86 TPB per acre. Therefore, the choice of drop cloth width and length of successive samples across adjacent rows determines the sensitivity of the method to detect low densities. Typically, an 8 row sample line has nearly a 100% chance to detect randomly dispersed populations of TPB.
FIG. 3. Distribution of sample density estimates for plant bugs from three eight row samples assuming they comprise a single 24-row sample. The thicker line is the final posterior distribution, whereas the more numerous, thinner lines represent intermediate posterior distributions calculated sequentially from each quadrat along the line.

in homogeneous habitats as rare as 1.4% (e.g., 573 TPB per acre found in stands of 40,000 plants per acre and 38 in row spacings). Processing the sample data with Bayesian methods also leads to a direct determination of sample reliability as illustrated in Figs. 3 - 5.

Whenever the TPB densities dramatically increase above a treatment threshold, the line length can be shortened and the estimate similarly scaled by the number of rows (f) actually sampled with reference to an ideal 8- or 16-row sample length (k). For example, if 5 TPB are found from a 2 row drop cloth sample, the estimate (using the scaling factor for an 8 row sample in a 38 in row spacing and an alternate, revised correction factor, k/f = 8/2) is 5 * 573.2 * 8/2 = 11, 464 TPB per acre. The same problem is solved by application of eq. (1) and use of the primary correction factor (L/k) results in \( \hat{Y} = 429.9/3 * 5 * 32/2 = 11, 464 \) bugs per acre. With experience and judgment, the sampler will learn how to balance the length of the sample line against the value of the information needed to make a management decision.
FIG. 4. Distribution of sample density estimates for plant bugs from three eight row samples assuming these smaller lines comprise a single 24-row sample. Note how the mode is smaller than the mode from samples (Fig. 3) where the population is larger. Note the smaller scale of densities along the x-axis. The algorithm can be scaled for smaller or larger populations as necessary.

Also, if plant stand counts are collected along with the insect counts, then the number of insects per acre is automatically scaled to the plant density with the Bayesian algorithm (eq. 2). The posterior distribution of the estimated density offers several advantages. This distribution can improve the robustness of simulation models on TPB ecology because it explicitly describes the variance. Additionally, it can be used to describe an interval about the estimate of population abundance (θ) with a specified probability, e.g. $P(\theta_1 \leq \theta \leq \theta_2) = 0.95$. Here, $\theta_1$ and $\theta_2$ correspond to the 0.05 and 0.95 percentiles of the cumulative distribution function of the final posterior. This interval is not equivalent to a 95% confidence interval (Johnson 1977), but serves a similar purpose (Software is available by request from the authors to calculate the posterior distribution and the upper and lower probability limits).
FIG. 5. A typical density estimate of plant bugs at a very low level of population abundance. These data are comprised of three eight row samples pooled into one 24-row sample.

Both the LIS and Bayesian approach for estimating density solves several problems that other sampling designs cannot avoid. Namely, the frequent occurrence of 'zeros' in the data is not a limitation and the need to estimate the dispersion parameter (k) of the negative binomial distribution is completely avoided. Estimating the dispersion parameter is essential in the application of sequential sampling designs (Poole 1977) and obtaining correct estimates of that parameter is difficult, especially over time.

In conclusion, the estimation of density by the classical LIS function (eq. 1) can be employed to develop a 'lookup table' of constants that are applicable to any row spacing and skip pattern (e.g., Williams et al. 1995). These constants, once defined can easily be used for commercial scouting purposes. For training or research purposes the estimation of TPB density by Bayesian methods (eq. 2) is more descriptive. However, for commercial purposes the Bayesian method is too computer intensive and not likely to be used under field conditions. The next step of the research presented here is to demonstrate how the behavior of intermediate posteriors (also known as 'informed priors (Gazey and Staley 1986)) are (1) diagnostic for the
granularity of the dispersion pattern and (2) how the posterior distribution can compare populations to a treatment threshold and be used to directly obtain the chance of an incorrect decision (Legg and Moon 1994; Mendenhall et al. 1986). Other important issues such as optimizing sample number size for different fixed sample unit sizes, minimizing the cost of sampling, and additional details on the spatial dispersion or dispersal capacities of TPB need to be explored. These issues are the focus of current research efforts and should be addressed in future papers.

ACKNOWLEDGMENT

Portions of this study supported by funds from USDA-CSREES Special Grants 97-34374-3990. Appreciation is expressed to Kenneth Hood of Perthshire Farms. Thanks are expressed to Elgenaid Hamadain, Shane Minor and Robert Turner for assistance in collection of the sample data. The comments of the editors and several anonymous reviewers for portions of this manuscript are also appreciated.

LITERATURE CITED


56