BBD—Computer Software for Fitting the Beta-Binomial Distribution to Disease Incidence Data

L. V. MADDEN, Department of Plant Pathology, Ohio State University, Ohio Agricultural Research and Development Center, Wooster 44691, and G. HUGHES, Institute of Ecology and Resource Management, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, Scotland, UK

ABSTRACT

Madden, L. V., and Hughes, G. 1994. BBD—Computer software for fitting the beta-binomial distribution to disease incidence data. Plant Dis. 78:536-540.

A software program for DOS-based personal computers was developed to fit the beta-binomial distribution to the frequency of incidence of disease. The beta-binomial is a discrete distribution, which is appropriate for describing aggregated or clustered binary data such as incidence. Variance-ratio and C(α) tests are performed to determine if there is evidence that incidence is aggregated. The program then calculates distribution parameters and their standard errors using a maximum likelihood procedure, determines the expected values of the distribution, and calculates a chi-square goodness-of-fit test. For comparison purposes, the program fits the binomial distribution to the same data. The software and a detailed user's manual are available free from either author.

The analysis of spatial patterns of plant diseases is an important component of epidemiology (1,4,11). Information on disease patterns can be used to aid in the understanding of the spatio-temporal disease dynamics, transform data to meet statistical assumptions for assessing treatment effects, and develop sampling protocols that result in precise estimates of mean disease intensity (1). There are many approaches for characterizing spatial patterns, depending on the type of data collected and knowledge of the location of where observations were made. A very popular procedure is fitting discrete probability distributions to the data and quantifying aggregation by the realized values of the parameters of the distributions (4). For instance, the Poisson and negative binomial distributions can be fitted to the counts of lesions per sampling unit. These two distributions are appropriate when there is, effectively, no upper limit to the counts in the sampling units or when the counts are substantially less than the limit. If certain assumptions are met (11), a good fit (based on χ^2 goodness-of-fit test) of the negative binomial is an indication of a clustered (or aggregated) pattern, and the degree of aggregation is assessed by the estimated parameter k of the distribution. A good fit by the Poisson distribution in this scenario would indicate a random pattern, if statistical assumptions are met (1).

Accepted for publication 5 January 1993.

Unlike the situation for unlimited counts, Hughes and Madden (7,8) recently pointed out that the use of the Poisson, negative binomial, and related distributions is generally inappropriate for analyzing data on disease incidence (proportion of plants or leaves diseased). Because of the binary nature of incidence (a plant is diseased or not), it can be misleading to fit distributions that are based on counts (such as the Poission and negative binomial). For a random pattern of disease incidence (for example, diseased plants per sampling unit) the appropriate probability distribution is the binomial. For a clustered pattern of incidence, the beta-binomial distribution is an appropriate alternative to the binomial (6,8,17,22). This distribution has three terms: $n =$ the number of plants (or plant units, e.g., leaves) in a sampling unit; $p =$ the probability of a plant being diseased; and $\theta =$ the index of aggregation. The latter two parameters are estimated from data. As θ approaches 0, the beta-binomial reduces to the binomial distribution. An alternative parameterization uses $\alpha = p/\theta$ and $\beta = (1-p)/\theta$, but p and θ have better statistical estimation properties (19). The beta-binomial distribution is entirely consistent with the Taylor empirical power law (1) as modified by Hughes and Madden for incidence data (7).

A computer program by Gates and Ethridge (3) is used by many plant pathologists to fit the Poisson and negative binomial distributions to data. A new version of the program, called DISCRETE, is now available for personal computers (2). Although the binomial distribution can be fitted with this program, it is not possible to fit the beta-binomial. Smith (19) published a FORTRAN subroutine to estimate beta-binomial parameters, using maximum likelihood, which we incorporated into a mainframe program to estimate parameters for virus disease data sets (8). The mainframe program does not allow user control without revising the actual FORTRAN source code. Additionally, a program for calculating the expected values of the frequencies was not available, which is a nontrivial matter with the beta-binomial. Originally, we wrote a MINITAB (13) command file (macro) to calculate the expected values by inputting n and estimates of θ and p, as determined with the Smith algorithim. A χ^2 goodness-of-fit test was done in a separate operation with MINITAB after entering the observed frequencies.

To allow more individuals to use the beta-binomial distribution, and to combine parameter estimation, calculation of expected values, and the χ^2 goodness-of-fit test, we wrote a computer program, called BBD, for use on microcomputers. The purpose of this article is to describe the input and output of the program, present the control options available to the user, and present examples of the use of the program.

Computer program. The program is written in Microsoft FORTRAN and compiled by version 5.1 of Microsoft's Professional Development System (One Microsoft Way, Redmond, WA 98052-6399). The program requires DOS 3.2 or higher. Extended memory and a math coprocessor are not required, but a coprocessor is used if present. The program will run with 8088/80286 and higher microprocessors.

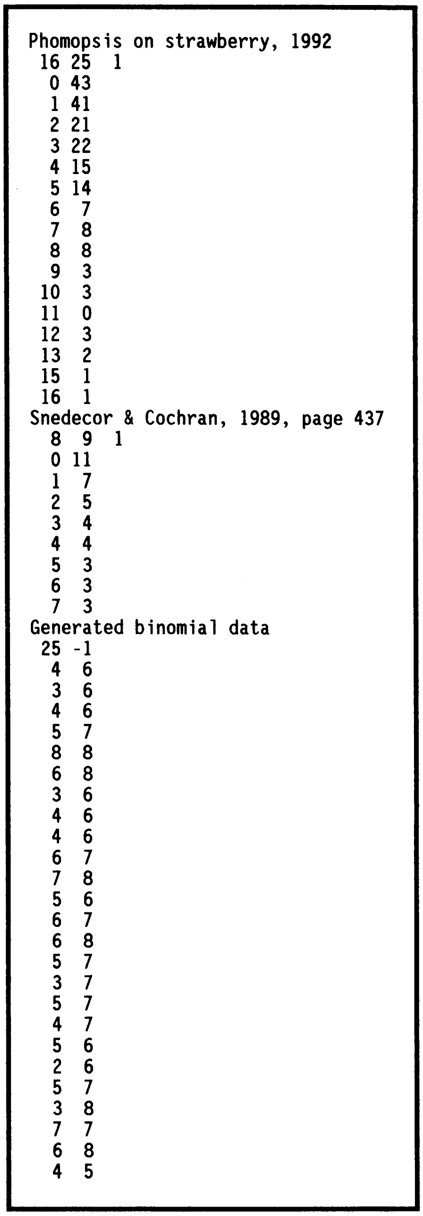
Moment estimates of p and θ, calculated using the procedure of Kleinman (9), are used as initial values in the iterative Smith (19) algorithm. Maximum likelihood estimates (MLEs) of the parameters and standard errors of the estimates are calculated using a damped Newton-Raphson technique. Expected values for the beta-binomial are determined by following the method given by Skellam (18). When possible (see below), a χ^2 statistic is calculated for goodness-of-fit between the observed and expected frequencies, and the significance level of the χ^2 value is calculated using the method in the Gates program (2).

For comparison purposes, BBD calculates the expected values of the binomial distribution (with the same estimated p) using a procedure from Press et al (15). A χ^2 goodness-of-fit test for the binomial distribution also is done. This allows the user to compare distributions without running two programs (BBD and DISCRETE [2]).

Two variance-ratio tests are calculated prior to calculating MLEs of the parameters. These tests allow one to determine if there is sufficient evidence to reject the null hypothesis of a binomial distribution, that is, that there is a random pattern. The calculated (observed) variance of diseased plants per sampling unit (see, for example, equation 21.5.1 in Snedecor and Cochran [20]) is divided by the theoretical variance for a binomial distribution [$np(1 - p)$] to produce an index of dispersion D. This is directly analogous to the standard variance-to-mean test for count data because, with counts, the theoretical variance for the Poisson distribution is the mean. Multiplying D by the number of sampling units minus one ($N - 1$) produces the first test, which is a chi-square statistic [$\chi^2(V)$] with $N - 1$ degrees of freedom (21). The significance of the statistic is calculated by BBD using the algorithm in DISCRETE (2). A large test statistic, $\chi^2(V)$, or small significance probability (P) indicates that one rejects the null hypothesis of a binomial distribution in favor of the alternative hypothesis of a distribution with a larger variance.

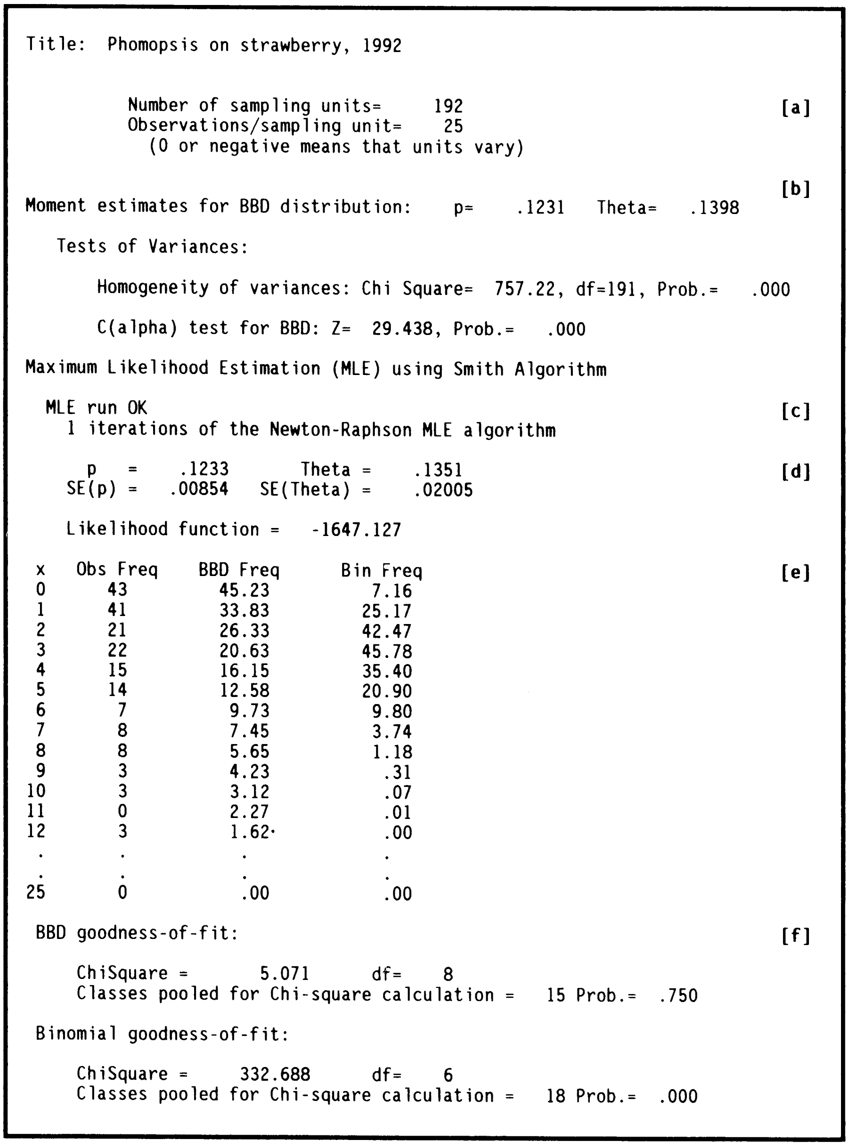
The second test, termed a $C(\alpha)$ test, also is based on D. The formula is given in Tarone (21). The calculated test statistic is the standard normal deviate (Z). A large Z or small P in this test indicates that one can reject the null hypothesis of (specifically) a binomial distribution in favor of the alternative hypothesis of a beta-binomial distribution. The significance level of Z is calculated by BBD, and the test of Z is one-sided. One could use $\chi^2(V)$ and Z to decide which distribution was more appropriate.

Input. To run BBD, data need to be in an ASCII (i.e., text) file. Input data are read by the program in either of two possible formats: 1) data from each sampling unit (diseased plants or leaves [X] and n), or 2) data representing precalculated frequency distributions (i.e., number of sampling units with 0, 1,...n



```
Phomopsis on strawberry, 1992
 16 25  1
  0 43
  1 41
  2 21
  3 22
  4 15
  5 14
  6  7
  7  8
  8  8
  9  3
 10  3
 11  0
 12  3
 13  2
 15  1
 16  1
Snedecor & Cochran, 1989, page 437
  8  9  1
  0 11
  1  7
  2  5
  3  4
  4  4
  5  3
  6  3
  7  3
Generated binomial data
 25 -1
  4  6
  3  6
  4  6
  5  7
  8  8
  6  8
  3  6
  4  6
  4  6
  6  7
  7  8
  5  6
  6  7
  6  8
  5  7
  3  7
  5  7
  4  7
  5  6
  2  6
  5  7
  3  8
  7  7
  6  8
  4  5
```

Fig. 1. Example input incidence data sets for use with the BBD program. Top two sets: frequency of diseased plants or leaves is recorded (e.g., 43 sampling units with 0 diseased leaves in the first set). There are 16 records of data (n = 25) in the first set, and eight records (n = 9) in the second. The "1" on the control records indicates that frequency format is used. Bottom set: data on individual sampling units are recorded. There are 25 units, and n varied among them (negative value for n given on the control record).



```
Title:  Phomopsis on strawberry, 1992


             Number of sampling units=     192                           [a]
             Observations/sampling unit=    25
               (0 or negative means that units vary)

                                                                         [b]
Moment estimates for BBD distribution:    p=    .1231   Theta=   .1398

   Tests of Variances:

      Homogeneity of variances: Chi Square=  757.22, df=191, Prob.=    .000

      C(alpha) test for BBD: Z=  29.438, Prob.=    .000

Maximum Likelihood Estimation (MLE) using Smith Algorithm

 MLE run OK                                                              [c]
   1 iterations of the Newton-Raphson MLE algorithm

      p   =    .1233       Theta =    .1351                              [d]
    SE(p) =   .00854   SE(Theta) =   .02005

   Likelihood function =   -1647.127

x     Obs Freq     BBD Freq        Bin Freq                              [e]
0        43         45.23            7.16
1        41         33.83           25.17
2        21         26.33           42.47
3        22         20.63           45.78
4        15         16.15           35.40
5        14         12.58           20.90
6         7          9.73            9.80
7         8          7.45            3.74
8         8          5.65            1.18
9         3          4.23             .31
10        3          3.12             .07
11        0          2.27             .01
12        3          1.62·            .00
.         .           .                .
.         .           .                .
25        0           .00             .00

 BBD goodness-of-fit:                                                    [f]

     ChiSquare =        5.071       df=    8
     Classes pooled for Chi-square calculation =    15 Prob.=  .750

 Binomial goodness-of-fit:

     ChiSquare =      332.688       df=    6
     Classes pooled for Chi-square calculation =    18 Prob.=  .000
```

Fig. 2. Example output from the BBD program. Data are for the incidence of strawberry leaves with Phomopsis leaf blight. Bracketed letters on the right were added to the output for annotation in the text. To save space, observed and expected frequencies between 13 and 24 were omitted.

diseased plants). If n varies among sampling units, only the first method can be used. In a typical case of variable n, there are missing plants in some sampling units. Instructions for the program on data format and output options are included in the data file (see below). Input format was designed to be similar to that used for DISCRETE (2).

Two records must precede each set of data. The first record contains the title, consisting of any arbitrary description of the data set. The second record (control record) specifics: 1) the number of records (lines) of data, 2) the number of plants (or leaves) per sampling unit (n), and 3) a code for the data style (individual sampling units or frequency data). If n varies among sampling units, a negative or zero value is given on the control record for n; the program then reads both X and n for each sampling unit. If n does not vary but the control record specifies that data from individual sampling units are given (i.e., a positive value for n), then the program reads only X for each unit. A given data set can consist of up to 500 sampling units, and the maximum n is 199. Other options can be specified which are described in the user's manual, available from the authors.

The data records follow the control record, with one X (or X and n) or one frequency class per record. Additional data sets can follow the last data record; these data sets are analyzed sequentially. An example input data file is shown in Figure 1; both the frequency (first two sets) and the individual sampling unit (last set) forms are shown.

Program execution and output. When BBD is executed, the program prompts the user for the name of the input file and the name for an output file that is created by the program. All output is printed to the screen and is stored in the output file. The output file can later be printed or manipulated with a text editor.

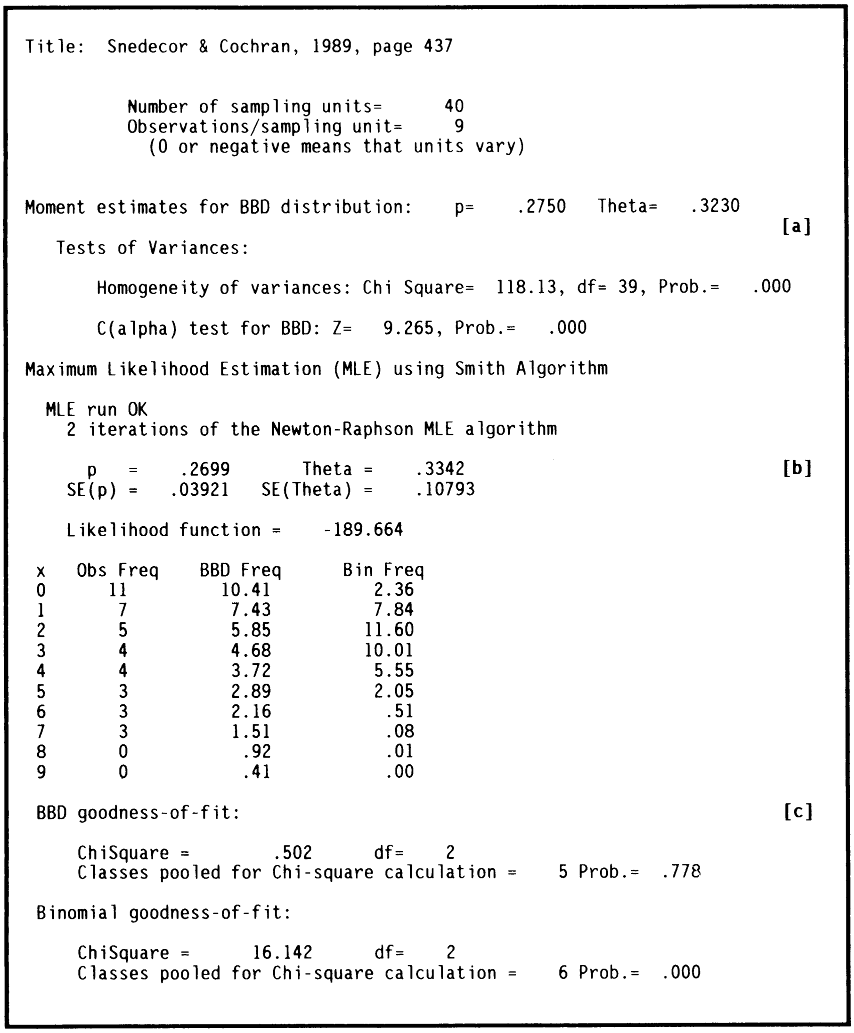
Example output for the three data sets of Figure 1 are given in Figures 2–4. Other input and output data sets are given with the user's manual. In Figure 2, data are for the incidence of Phomopsis leaf blight of strawberry, caused by *Phomopsis obscurans* (10). Observations were made by visually assessing leaves in a commercial strawberry planting (cultivar Jewell) near Wooster, Ohio, on 25 June 1992. In each of 192 locations in an area of about 20 × 30 m, the number of leaves out of 25 with symptoms of leaf blight was determined (L. V. Madden, *unpublished*). Note that the input data are in frequency form (Fig. 1) and that there are 16 records with data. A record for frequency class 14 is omitted because no sampling units with 14 diseased leaves were found. A data record could have been included for this class, and the code for number of data records on the control record would then be 17.

After printing some header information, the number of sampling units (quadrats) ($N = 192$) and observations (that is, plants or leaves) per sampling unit ($n = 25$) are printed by the program (Fig. 2[a]). A negative (or zero) value for n here would mean that n varies among units (n_i). Moment estimates of p and θ are given (Fig. 2[b]), followed by the $\chi^2(V)$ and Z statistics. Both tests in the example indicate that the binomial distribution was not appropriate (i.e., $P << 0.05$). BBD then prints results from the maximum likelihood procedure. In this case, it is shown that the MLE was successful ("run OK", Fig. 2[c]) and that it took one iteration for convergence. Convergence indicates that the change in estimated parameters in the iterative procedure fell below a threshold (default of 0.01). When the method is not successful, an error message number is printed. The user's manual gives the meaning of the error numbers.

BBD then prints the MLE values for p and θ, together with their standard errors (Fig. 2[d]). As usual, the MLEs in the example are close to the moment estimates. The natural logarithm of the likelihood (likelihood function) is printed immediately below the parameters and their standard errors. The maximum value of the likelihood function is found in MLE. Likelihood functions can also be used to test whether two or more distributions can be regarded as having common p and θ values (see reference 22 for an example).

The observed frequency distribution is printed (Fig. 2[e]) unless it is suppressed by an optional code on the control record (see the user's manual). The expected frequencies are printed next to the observed values unless n varies among sampling units (where it is impossible to define the expected values), or the program could not successfully estimate the parameters (e.g., a non-negative θ could not be obtained).

The χ^2 goodness-of-fit test is presented for each distribution if the expected values can be calculated (Fig. 2[f]). For the test, frequency classes are pooled (starting at the largest X) so that all expected values are greater than 5. This is a conservative method to insure a proper χ^2 test. The pooled numbers are



```
Title:  Snedecor & Cochran, 1989, page 437


          Number of sampling units=      40
          Observations/sampling unit=     9
            (0 or negative means that units vary)


Moment estimates for BBD distribution:    p=     .2750    Theta=    .3230
                                                                           [a]
   Tests of Variances:

       Homogeneity of variances: Chi Square=  118.13, df= 39, Prob.=    .000

       C(alpha) test for BBD: Z=   9.265, Prob.=    .000

Maximum Likelihood Estimation (MLE) using Smith Algorithm

  MLE run OK
    2 iterations of the Newton-Raphson MLE algorithm

      p   =     .2699        Theta =     .3342                             [b]
    SE(p) =   .03921    SE(Theta) =    .10793

    Likelihood function =     -189.664

x    Obs Freq     BBD Freq       Bin Freq
0       11          10.41           2.36
1        7           7.43           7.84
2        5           5.85          11.60
3        4           4.68          10.01
4        4           3.72           5.55
5        3           2.89           2.05
6        3           2.16            .51
7        3           1.51            .08
8        0            .92            .01
9        0            .41            .00

 BBD goodness-of-fit:                                                      [c]

     ChiSquare =         .502       df=   2
     Classes pooled for Chi-square calculation =     5 Prob.=  .778

 Binomial goodness-of-fit:

     ChiSquare =      16.142       df=   2
     Classes pooled for Chi-square calculation =     6 Prob.=  .000
```

Fig. 3. Example output from the BBD program. Data are from an example in Snedecor and Cochran (20, page 437) on incidence of infected plants. Bracketed letters on the right were added to the output for annotation in the text.

not printed. Degrees of freedom (*df*) are equal to the number of classes after pooling minus the number of estimated parameters and minus one. In general, the *df* will differ for the two distributions because of differences in the number of parameters estimated and the expected frequencies calculated. The null hypothesis here is that the specified distribution is appropriate. Small χ^2 value or large P ("Prob.") indicate that one cannot reject the null hypothesis. In this example, BBD was found to provide a good fit to the data ($\chi^2 = 5.07$, $df = 8$, $P = 0.75$), but the binomial did not ($P = 0.00$), which is consistent with the earlier variance tests (Fig. 2[b]).

It is possible for calculated degrees of freedom to be 0 or negative when there is a small number of frequency classes (after pooling). This often happens at low p, when most of the X's are 0 or very low, or when N is small. In these cases, the program arbitrarily lists P as 0.000 because the test cannot be performed.

A second example is shown in Figure 3 by analyzing a plant disease data set given on page 437 of Statistical Methods, 8th ed. (20). Data on the number of infected plants (disease and host not specified) out of nine ($n = 9$) in each of 40 locations is presented. This is another case where the variance tests indicated that the binomial distribution (or a random pattern) was not appropriate (Fig. 3[a]). The estimate of θ was 0.33 and about three times its standard error (Fig. 3[b]), an indication that diseased plants were highly aggregated. Only the beta-binomial distribution provided a good fit to the data (Fig. 3[c]).

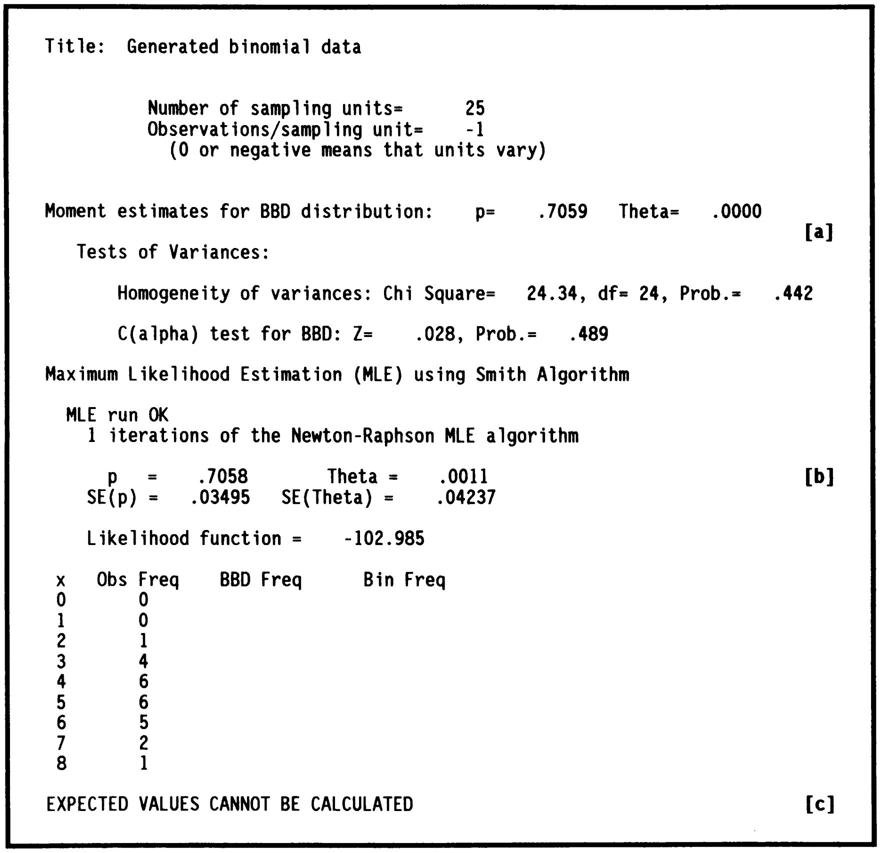
As a final example, simulated data with variable n were analyzed with BBD (Fig. 4). A random sample of 25 observations (see Fig. 1) was generated for the binomial distribution ($p = 0.7$) using a version of the RANDOM command in MINITAB (13). It was assumed here that n varied from 5 to 8 based on a uniform distibution. Generated uniform values were rounded to the nearest integer. The $\chi^2(V)$ and Z statistics both indicated, as expected, that the binomial distribution was appropriate ($P > 0.05$) (Fig. 4[a]). The estimate of θ was considerably less than its standard error (Fig. 4[b]). Because n was variable, expected values for the two distributions could not be calculated, and therefore, goodness-of-fit could not be determined. Observed frequencies were calculated (Fig. 4[c]) to help the user visualize the data.

Test of the program. BBD has executed successfully when analyzing published data tests, in plant pathology and other fields, and generated (simulated) data sets (8; L. V. Madden, *unpublished*). Normal completion of the program occurs even when the maximum likelihood procedure does not converge. To demonstrate this program, we analyzed the published data on the incidence of potyvirus diseases of tobacco (12). Briefly, the incidence of tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV) was assessed in six fields in 2 yr and in four fields a third year. Fields were divided into either 75 or 50 quadrats (depending on the field size and dimension) with either 40 or 60 plants per quadrat ($n = 40, 60$). Fields were assessed multiple times for virus symptoms, resulting in 188 data sets for analysis. Mean disease incidence ranged from 10^{-4} to 0.89.

Moment estimates of the parameters were obtained in all cases. The maximum likelihood procedure converged in about 80% (151) of the data sets. That is, it was possible to obtain MLEs of p and θ, as well as their standard errors, 80% of the time. Lack of convergence was related to mean number of diseased plants (estimated by p). In 75% of the times when convergence was not achieved, p was less than 0.017. In these situations, almost all quadrats had no diseased plants, and very few quadrats had more than one diseased plant. For this reason, there were insufficient degrees of freedom to test for goodness-of-fit by either distribution in 28 of the 37 cases. For the remaining nine cases where convergence was not achieved but goodness-of-fit could be tested, four were well fitted by the beta-binomial (using moment estimates of p and θ). In none of these cases did the binomial distribution fit the data.

Further assessment of the results was conducted by considering only the data sets with enough degrees of freedom to determine goodness-of-fit (i.e., 131 data sets with $df > 0$). The maximum likelihood procedure converged in 93% of the cases, and the beta-binomial distribution provided a good fit in 88%. The moment and MLEs of the parameter were similar. For instance, the average absolute difference between the moment and maximum likelihood estimates of p was 0.0004. For θ, the average absolute difference between the two types of estimates was 0.0053. The variance ratio test was significant ($P < 0.05$) for all of these cases. Of these 131 data sets, only nine were well fitted by the binomial distribution. Of these nine, eight were better fitted by the beta-binomial based on a higher P value from the χ^2 test. Therefore, only one data set was better fitted by the binomial compared to the beta-binomial distribution.

Discussion. We have found the FORTRAN program BBD to be useful in fitting the beta-binomial and binomial distributions to disease-incidence data. Because the program can run on virtually any DOS-based personal computer, it should provide a relatively easy means for others to estimate parameters, com-



```
Title:  Generated binomial data


          Number of sampling units=      25
          Observations/sampling unit=    -1
            (0 or negative means that units vary)


Moment estimates for BBD distribution:    p=    .7059   Theta=   .0000
                                                                          [a]
   Tests of Variances:

      Homogeneity of variances: Chi Square=   24.34, df= 24, Prob.=   .442

      C(alpha) test for BBD: Z=    .028, Prob.=   .489

Maximum Likelihood Estimation (MLE) using Smith Algorithm

  MLE run OK
   1 iterations of the Newton-Raphson MLE algorithm

      p   =    .7058      Theta =    .0011                                 [b]
    SE(p) =   .03495   SE(Theta) =   .04237

    Likelihood function =    -102.985

x    Obs Freq     BBD Freq       Bin Freq
0      0
1      0
2      1
3      4
4      6
5      6
6      5
7      2
8      1

EXPECTED VALUES CANNOT BE CALCULATED                                       [c]
```

Fig. 4. Example output from the BBD program. Data were generated (using MINITAB [13]) as a random sample from the binomial distribution with a variable n. Bracketed letters on the right were added to the output for annotation in the text.

pare the goodness-of-fit of both discrete distributions, and test for aggregation or clustering of disease incidence. Depending on the information available on the position of diseased plants, BBD can be used in conjunction with other programs to analyze aggregation based on distance between diseased plants (14), or on autocorrelation of incidence values among locations throughout fields (5,16).

ACKNOWLEDGMENT

Salaries and research support provided by state and federal funds appropriated to the Ohio Agricultural Research and Development Center, the Ohio State University. Manuscript No. 124-93.

LITERATURE CITED

1. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.
2. Gates, C. E. 1988. Discrete, a computer program for fitting discrete frequency distributions. Pages 458-466 in: Lecture Notes in Statistics. Vol. 55, Estimation and Analysis of Insect Populations. L. McDonald, B. Manly, J. Lockwood, and J. Logan, eds. Springer-Verlag, Berlin.
3. Gates, C. E., and Ethridge, F. G. 1970. A generalized set of discrete frequency distributions with FORTRAN program. Int. Assoc. Math. Geol. 4:1-24.
4. Gilligan, C. A. 1988. Analysis of the spatial pattern of soilborne pathogens. Pages 86-98 in: Experimental Techniques in Plant Disease Epidemiology. J. Kranz and J. Rotem, eds. Springer-Verlag, Berlin.
5. Gottwald, T. R., Richie, S. M., and Campbell, C. L. 1992. LCOR2—Spatial correlation analysis software for the personal computer. Plant Dis. 76:213-215.
6. Griffiths, D. A. 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of disease. Biometrics 29:637-648.
7. Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. Plant Pathol. 41:657-660.
8. Hughes, G., and Madden, L. V. 1993. Using the beta-binomial distribution to describe aggregated patterns of disease incidence. Phytopathology 83:759-763.
9. Kleinman, J. C. 1973. Proportions with extraneous variance: Single and independent samples. J. Am. Stat. Assoc. 68:46-54.
10. Maas, J. L., ed. 1984. Compendium of Strawberry Diseases. American Phytopathological Society, St. Paul, MN.
11. Madden, L. V. 1989. Dynamic nature of within-field disease and pathogen distributions. Pages 96-126 in: Spatial Components of Plant Disease Epidemics. M. J. Jeger, ed. Prentice Hall, New Jersey.
12. Madden, L. V., Pirone, T. P., and Raccah, B. 1987. Analysis of spatial patterns of virus-diseased tobacco plants. Phytopathology 77:1409-1417.
13. Minitab, Inc. 1991. MINITAB Reference Manual, Release 8. Minitab Incorporated, State College, PA.
14. Nelson, S. C., Marsh, P. L., and Campbell, C. L. 1992. 2DCLASS, a two-dimensional distance class analysis software for the personal computer. Plant Dis. 76:427-432.
15. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. 1986. Numerical Recipes. Cambridge University Press, Cambridge.
16. Reynolds, K. M., and Madden, L. V. 1988. Analysis of epidemics using spatio-temporal autocorrelation. Phytopathology 78:240-246.
17. Shiyomi, M., and Takai, A. 1979. The spatial pattern of infected or infested plants and negative hypergeometric series. (In Japanese.) Jpn. J. Appl. Entomol. Zool. 23:224-229.
18. Skellam, J. G. 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. J. R. Stat. Soc. B 10:257-261.
19. Smith, D. M. 1983. Maximum likelihood estimation of the parameters of the beta binomial distribution. Appl. Stat. 32:192-204.
20. Snedecor, G. W., and Cochran, W. C. 1989. Statistical Methods. 8th ed. Iowa State University, Ames.
21. Tarone, R. E. 1979. Testing the goodness of fit of the binomial distribution. Biometrika 66:585-590.
22. Williams, D. A. 1975. Analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31:949-952.