NASA CONTRACTOR REPORT

Report No. 61289

PROBABILITY MODELS FOR THE VARIATION IN THE NUMBER OF THUNDERSTORM HITS PER DAY

By William O. Williford and Michael C. Carter

The University of Georgia
Department of Statistics
Athens, Georgia

July 1969

Prepared for

NASA-GEORGE C. MARSHALL SPACE FLIGHT CENTER
Marshall Space Flight Center, Alabama 35812

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
NASA CR-61289		

4. TITLE AND SUBTITLE
PROBABILITY MODELS FOR THE VARIATION IN THE NUMBER OF THUNDERSTORM HITS PER DAY

5. REPORT DATE
July 1969

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
William O. Williford and Michael C. Carter

8. PERFORMING ORGANIZATION REPORT #
37

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of Georgia
Department of Statistics
Athens, Georgia

10. WORK UNIT NO.

11. CONTRACT OR GRANT NO.
NAS8-11175

13. TYPE OF REPORT & PERIOD COVERED
Contractor Report

12. SPONSORING AGENCY NAME AND ADDRESS
NASA-Marshall Space Flight Center
Aero-Astrodynamics Laboratory
Aerospace Environment Division
Marshall Space Flight Center, Alabama 35812

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES
Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

16. ABSTRACT

From among three modified discrete probability distributions investigated, a modified negative binomial distribution is recommended as the "best" model to represent the variation in the number of thunderstorms per day which move across a a launch site at Cape Kennedy, Florida.

17. KEY WORDS

18. DISTRIBUTION STATEMENT
PUBLIC RELEASE
E. D. GEISSLER
Director, Aero-Astrodynamics Lab, MSFC

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
U	U	21	

MSFC - Form 3292 (May 1969)

I. Introduction

Three probability distributions are investigated for the purpose of representing the variation in the number of thunderstorms per day which move across a particular point, for example, a launch site at Cape Kennedy. When a thunderstorm moves across the given point, we will call this a thunderstorm hit (TH). Two of the models were previously derived by Singh [13,14,15] for the distribution of the number of births to a couple during a given time interval. The third model is derived here under assumptions similar to those used by Singh and also by Neyman [11] for the distribution of the number of schools of fish caught in a fishing area.

From sample data on the number of "thunderstorm events" (see section III) per day, it was found by Falls [6] that the sample variance exceeded the sample mean. Several distributions have been devised for data which is overdispersed [1,10]. Among those devised, one which has a number of advantages in its use is the negative binomial distribution [2,3]. The new model derived here is a modification of the negative binomial distribution.

Further, since the probability distributions will involve two unknown parameters, a method developed by Neyman [12] for obtaining BAN (best asymptotically normal) estimates of the parameters will be outlined.

II. Statistical Models

We made the following assumptions.

1. A probability of $\alpha(1-\alpha)$ is assigned to the possibility of a TH occurrence (nonoccurrence) on any given day.
2. $\Pr\{\text{TH occurs in a unit of time} \mid \text{A TH not in progress}, \alpha \neq 0\} = p$.
3. T is the number of units of time in the specified time period. The positive integer h is defined by the statement

$$\Pr\{\text{TH occurs in a unit of time} \mid \text{A TH in the preceeding } h-1 \text{ units of time}\} = 0.$$

Then the maximum number of occurrences in T units of time is $n \leq [T/h] + 1$, where $[T/h]$ stands for the greatest integer not exceeding T/h.

The assumptions above and the models given below ignore a great many details; but in order to study real phenomena by statistical methods, we must begin constructing some simplified statistical model of these phenomena. See Neyman [11] section 1. We assume in the following models that the probability p (in assumption 2) remains constant throughout the day. This is a strong assumption and perhaps needs to be modified. One modification would be to consider periods of the day; e.g., afternoon hours, during which the probability of a TH may be essentially constant. Another would be to consider p as a random variable with some a priori probability distribution function.

Under the above assumptions and if X is a random variable denoting the number of TH's per time period T, we have the following models:

Modified Negative Binomial

$$\Pr\{X=0\} = (1-\alpha)+\alpha q^{T} \qquad (q = 1 - p) \tag{1}$$

$$\Pr\{X=i\} = \alpha[P^{i}q^{T-ih}\binom{T-ih+i-1}{i} + P^{i}q^{T-ih}\sum_{m=1}^{h-1}\binom{T-(i-1)h+i-m-2}{i-1}q^{h-m}] \tag{2}$$

for $0<i<n$

$$\Pr\{X=n\} = 1-\Pr\{X<n\} \tag{3}$$

Singh's Binomial

$$\Pr\{X=0\} = (1-\alpha)+\alpha q^{T} \qquad (q = 1 - p) \tag{4}$$

$$\Pr\{X=i\} = \alpha[P^{i}q^{T-ih}\binom{T-ih+i}{i} + P^{i}q^{T-ih}\sum_{m=1}^{h-1}\binom{T-(i-1)h+i-m-1}{i-1}q^{h-m}] \tag{5}$$

for $0<i<n$

$$\Pr\{X=n\} = 1-\Pr\{X<n\} \tag{6}$$

Singh's Poisson

$$\Pr\{X=0\} = (1-\alpha) + \alpha e^{-\lambda T} \tag{7}$$

$$Pr\{X=i\} = \alpha[\sum_{m=0}^{i} e^{-\lambda[T-ih]}\frac{[\lambda(T-ih)]^m}{m!} - \sum_{m=0}^{i-1} e^{-\lambda[T-ih+h]}\frac{[\lambda(T-ih+h)]^m}{m!}] \qquad (8)$$

for $0<i<n$

$$Pr\{X=n\} = 1 - Pr\{X<n\}. \qquad (9)$$

These **three** models are investigated due to the nature of the data (presented in the next section). When dealing with discrete data the usual practice is to compute the mean and variance and then fit the Binomial, Poisson or Negative Binomial according to whether the mean exceeds equals, or is less than the variance. However, we are dealing with extremely "J" shaped curves and this criterion loses its usefulness just as do the mean and variance. Clearly the mode is a much more meaningful measure of "central tendency" and with few cells the usefulness of the variance (or any other measure of "dispersion") is questionable. As all three models have "J" shaped frequency functions for small p (or λ) it would seem there is little to distinguish between their usage.

Aside from any justifications of a physical nature one cannot ascertain from the available data on TH's which model does the best job statistically. The necessity of estimating small "tail" probabilities jeapordizes the use of the X^2 goodness of fit test. This problem is discussed more fully in the results section. This is particularly critical as all other well-known goodness of fit tests require continuous underlying distributions and completely specified hypotheses-neither of which is present here.

Any tests of hypotheses involving the estimated parameter values is a difficult problem for two reasons. Firstly, no information is available to formulate suitable hypotheses. To demonstrate this problem consider the following situation. In one case the p value is estimated to be .007. A logical test would be H_o: $p = 0$ vs H_i" $p\neq 0$. As the calculated value of p is quite small it is possible that H_o could be accepted. The result $Pr\{TH \text{ occurance} \mid H_o \text{ true}\} = 0$, which then says $\alpha=0$, given H_o true, is of little practical benefit. Secondly, any effective tests of hypotheses would necessarily be a multiple decision problem on α and p (or λ). The two parameters are functionally related (a complication of some importance in determining critical regions) and undoubtedly the estimators of α and p (or λ) have a very complicated joint distribution.

The crux of this discussion boils down to one fact. We can solve the problem in three different and equally acceptable ways. Statistically it is impossible to distinguish between the results by presently known techniques. This fact suggests some indeterminacy in the available data (or in the process itself). In such a situation the statistician recommends a plausible solution and makes the recipient aware that this is only one of several (apparently) equivalent alternatives.

III. Data

The thunderstorm data sample presently available for Cape Kennedy contains all the information that can objectively be extracted from the Standard Weather Observers' Form WBAN-10. (Due to the type of data taken, the number of thunderstorms per day was not available but only the number of times thunder was heard (and not heard during the previous 15 minutes) per day. These we call "Thunderstorm events.") Each "thunderstorm event" for the period January 1957 through December 1966 is identified. The information for each "thunderstorm event" includes: identification--year, month, day; beginning and ending time of the event; area by quadrant where thunderstorms were first and last observed; direction of movement; maximum and minimum intensity; frequency of thunder; whether or not one or more than one thunderstorm was observed during each event; lightning type and intensity; and other information on wind, weather, clouds, and visibility. This data is coded and available on computer cards. This is the first data sample of its kind for Cape Kennedy and is the best climatic record on thunderstorms available. The card deck was produced by ESSA, National Weather Records Center, Asheville, N. C., under a government cross-service order for the NASA, MSFC, Aerospace Enviornment Division, Huntsville, Alabama.

Those occurrences which were classified as TH's from the data sample were of the following two types.

1. A thunderstorm was actually reported overhead.
2. A thunderstorm was first reported in a sector and last reported in the opposite sector. This is assuming thunderstorms move in a straight line (over small areas, at least).

Some additional situations not accounted for in the models comes to light here. No provision for the number of thunderstorms in the immediate area is made, but thunderstorm density about the "point" undoubtedly affects the likelihood of a TH. A model in the form of Poisson occurrance of thunderstorms in an area and a binomial process over the point, given thunderstorms in the area, might be useful. However, the data to "calibrate" such a model is not available, so again we must simplify.

The summer months of June, July and August were selected for examination. The thunderstorm activity is more intense during this period of the year and these months demonstrate a "J" shaped curve. A preliminary review of the other months indicates the generality of such a curve and the other months differ only by possessing a larger"0" class.

The period is 24 hours and T is taken to be 48 units. The value of h is taken as 2 which means that, given a TH occuring, another cannot occur for 30 minutes. Further note that only 907 of 920 days are accounted for in the data sample.

Event (TH's per day)	June	July	August	Combined
0	263	274	269	806
1	23	22	29	74
2	5	3	7	15
3	4	1	1	6
4 or more	4	1	1	6
Total	299	301	307	907

IV. Estimation

Each of the models proposed involves two unknown parameters α and p (or q). A statistic is called a Minimum Chi-Square (MCS) estimator of α if it is obtained by minimizing, with respect to α, the expression

$$\chi^2 = \sum_{i=0}^{n} \frac{[N_i - NP_i(\alpha, p)]^2}{N_i} . \tag{10}$$

See Neyman [12], Kendall and Stuart Vol. II, 91-93 [8] and Singh [14] for a fuller explanation of BAN estimators. Neyman [12] has shown that the class of MCS estimators are also best asymptotically normal (BAN) estimators. These estimators are consistent, asymptotically normal, and asymptotically efficient. Let $P_i(\alpha, p)$ be the probability for i(i=0, 1, ..., n) TH's per day, and satisfying the regularity conditions given in Neyman [11].

Since

$$\chi^2 = \sum_{i=0}^{n} \frac{[N_i - NP_i(\alpha, p)]^2}{N_i} = \sum_{i=0}^{n} \left(\frac{[NP_i(\alpha, p)]^2}{N_i}\right) - N \tag{11}$$

to minimize equation 10 (with respect to α) set

$$\frac{\partial \chi^2}{\partial \alpha} = 2 \sum_{i=0}^{n} \left(\frac{NP_i(\alpha, p)}{N_i}\right) \frac{\partial NP_i(\alpha, p)}{\partial \alpha} = 0 \tag{12}$$

and solve for the estimator of α. The same procedure can be repeated to obtain the estimator of p.

If $P_i(\alpha, p)$ is linear in α and p, the estimates can easily be found; otherwise we can linearize them at a properly chose point $(\bar{\alpha}, \bar{p})$ and use the linearized $P_i(\alpha, p)$'s instead of the original $P_i(\alpha, p)$'s to find the estimates. The estimates obtained in this fashion are also BAN, if the point estimates $(\bar{\alpha}, \bar{p})$ are consistent.

The linearization about $(\bar{\alpha}, \bar{p})$ is accomplished by solving

$$N_0/N = P_0(\alpha, p), \quad N_1/N = P_1(\alpha, p) \tag{13}$$

and naming the solution $(\bar{\alpha}, \bar{p})$. The solutions $\bar{\alpha}$ and $\bar{p}$ can be shown to be consistent estimates of α and p. Letting $P_i'(\alpha, p)$ be the new linearized probabilities we have

$$P_i'(\alpha, p) = P_i(\bar{\alpha}, \bar{p}) + (\alpha - \bar{\alpha}) \left.\frac{\partial P_i(\alpha, p)}{\partial \alpha}\right|_{(\bar{\alpha}, \bar{p})} + (p - \bar{p}) \left.\frac{\partial P_i(\alpha, p)}{\partial p}\right|_{(\bar{\alpha}, \bar{p})} \tag{14}$$

as the general equation. In particular, we have

<u>Modified Negative Binomial</u>

$$P_0'(\alpha, p) = 1 - \alpha(1 - \bar{q}^T) + (p - \bar{p})\, \bar{\alpha}\, T\, \bar{q}^{(T-1)} \tag{15}$$

$$P_i'(\alpha, p) = \frac{\alpha}{\bar{\alpha}} P_i(\bar{\alpha}, \bar{p}) + (p - \bar{p})\bar{\alpha}\{\bar{p}^{(i-1)}\bar{q}^{(T-ih-1)}[(T-ih)\bar{p} - i\bar{q}]Q(i)$$

$$+ \sum_{m=1}^{h-1} \bar{p}^{(i-1)}\bar{q}^{(T-ih+h-m)}[i\bar{q} - (T-ih+h-m)\bar{p}]QS(i,m)\} \qquad (16)$$

where $0<i<n$, $Q(i) = \binom{T-ih+i-1}{i}$, $QS(i,m) = \binom{T-(i-1)h+i-m-2}{i-1}$.

$$P_n'(\alpha, p) = 1 - \sum_{i=0}^{n-1} P_i'(\alpha, p) \qquad (17)$$

Singh's Binomial

These formulas are of exactly the same type as those of the Negative Binomial when we redefine $Q(i) = \binom{T-ih+i}{i}$ and $QS(i,m) = \binom{T-(i-1)h+i-m-1}{i-1}$.

Singh's Poisson

$$P_0'(\alpha,\lambda) = 1-\alpha(1-e^{-\bar{\lambda}T}) - (\lambda-\bar{\lambda})\bar{\alpha}Te^{-\bar{\lambda}T} \qquad (18)$$

$$P_i'(\alpha,\lambda) = \frac{\alpha}{\bar{\alpha}} P_i(\bar{\alpha},\bar{\lambda}) + (\lambda-\bar{\lambda})\bar{\alpha} \{e^{-\lambda(T-ih+h)} \frac{(T-ih+h)^i}{i!} (\bar{\lambda})^{i-1}$$

$$-e^{-\bar{\lambda}(T-ih)} \frac{(T-ih)^{i+1}}{i!} (\bar{\lambda})^i \qquad (19)$$

for $0<i<n$

$$P_n'(\alpha,\lambda) = 1 - \sum_{i=0}^{n-1} P_i'(\alpha,\lambda). \qquad (20)$$

Replacing $P_i(\alpha,p)$ by $P_i'(\alpha,p)$ in equations (11) and (12), we obtain a modified form of χ^2

$$(\chi^2)' = \sum_{i=0}^{n} \frac{[N_i-P_i'(\alpha,p)]^2}{N_i} \qquad (21)$$

which is minimized: (with respect to α) by setting

$$\frac{\partial(\chi^2)'}{\partial\alpha} = 2\sum_{i=0}^{n}\left(\frac{NP_i'(\alpha,p)}{N_i}\right)\frac{\partial NP_i'(\alpha,p)}{\partial\alpha} = 0 \tag{22}$$

and (with respect to p) by setting

$$\frac{\partial(\chi^2)'}{\partial p} = 2\sum_{i=0}^{n}\left(\frac{NP_i'(\alpha,p)}{N_i}\right)\frac{\partial NP_i'(\alpha,p)}{\partial p} = 0. \tag{23}$$

From (22) and (23) we divide out unnecessary constants and solve the resulting linear equations simultaneously for $\hat{\alpha}$ and $\hat{p}$. Then the solutions $\hat{\alpha}$ and $\hat{p}$ of equations (22) and (23) are those values which minimize $(\chi^2)'$ and are BAN estimates of α and p.

To facilitate the considerable computational task involved in estimation, a FORTRAN program was written. The program fits all three models and the output consists of a simple probability, frequency table along with the calculated χ^2 value. Maximum generality has been allowed in the determination of T, h, and number of cells. A listing of the FORTRAN program and variable definitions is appended.

V. Results

This section discusses the results obtained and possible areas of future investigations. One table (I) is presented and it contains all pertinent data from the analysis.

As mentioned in Section 2 the χ^2 goodness of fit test does not work well with these data. The June or Combined data furnish the best results. With the exception of the Negative Binomial on Combined data every set of expected frequencies would need to be "lumped" into two cells in order to get the usual expected frequency value of 5 deemed necessary to efficiently apply the test. Considering the number of parameters estimated this is an unacceptable procedure.

However, it should be noted that all models fit particularly well in the 0, 1, 2 frequencies and generally very poorly in the 3,4 classes. The notable exception is the Negative Binomial on the August data.

The other interesting result is the relationship between the three sets of parameter estimates. To make comparisons easier the data is reproducted below:

	June		July		August		Combined	
	α	p(or λ)	α	p(or λ)	α	p(or λ)	α	p(or λ)
N. Binomial	.2007	.0155	.2979	.0072	.3102	.0105	.2526	.0114
Binomial	.1887	.0159	.2822	.0073	.2993	.0107	.2402	.0116
Poisson	.1871	.0162	.2783	.0075	.2948	.0110	.2375	.0118

The Negative Binomial consistently estimates α higher and p lower with the Poisson at the other end of the spectrum. On combined data the difference in α-estimates is 1.5% which might be deemed important for some purposes. Assuming α is of considerable interest the possible range of this difference seems to merit some investigation.

As the Negative Binomial and Poission are the extremes and the Binomial is sort of an "average", two possibilities seem worthy of future interest. Firstly, a mixture of the two extreme distributions and possibly the Binomial might better describe the data better in some situations and would be as good in any case. The mixing weights would likely be functions of the parameter estimates. Secondly, as the distributions are extremely close together they are essentially independent of the parameter estimates (Independence meaning any one of the three sets of estimates would give good results in one of the other models). This suggests they are "distribution-free" in a rough sense. Using this result some investigations into a "general" way to estimate α and p and a general distribution function based on conditional probabilities could be of benefit.

VI. Summary

All three models fit the data well and would likely be equivalent were a good statistical criterion available to judge them by. The Negative Binomial Modification fits best in most cases and is always best in the 0, 1, 2 classes. Based on this fact and considering the possible usage, The Modified Negative Binomial is recommended as the "best" model.

TABLE I

MONTH	DATA		NEG. BINOMIAL		BINOMIAL		POISSON	
	Class # TH's	Frequency	Expected Frequency	Probability	Expected Frequency	Probability	Expected Frequency	Probability
JUNE	0	263	267.25	.8938	268.64	.8984	268.75	.8988
	1	23	21.33	.0713	20.62	.0689	20.64	.0690
	2	5	7.54	.0252	7.65	.0255	7.61	.0254
	3	4	1.66	.0055	1.77	.0059	1.71	.0057
	4 or more	4	1.22	.0041	.32	.0011	.29	.0009
			α = .2007 P = .0155	χ^2 = 10.69	α = .1887 P = .0159	χ^2 = 45.83	α = .1871 λ = .0162	χ^2 = 51.37
JULY	0	274	274.68	.9125	275.78	.9162	275.82	.9163
	1	22	21.78	.0723	21.27	.0706	21.25	.0706
	2	3	3.51	.0116	3.55	.0118	3.54	.0117
	3	1	.35	.0011	.37	.0012	.36	.0011
	4 or more	1	.67	.0022	.03	.00009	.03	.00009
			α = .2979 P = .0072	χ^2 = 1.43	α = .2822 P = .0073	χ^2 = 34.32	α = .2783 λ = .0075	χ^2 = 37.67
AUGUST	0	269	269.01	.8762	269.74	.8786	269.75	.8786
	1	29	28.98	.0944	28.87	.0940	28.86	.0940
	2	7	6.87	.0223	7.16	.0233	7.18	.0233
	3	1	1.01	.0033	1.11	.0036	1.09	.0035
	4 or more	1	1.12	.0036	.13	.0004	.12	.0003
			α = .3102 P = .0105	χ^2 = .014	α = .2993 P = .0107	χ^2 = 5.84	α = .2948 λ = .0110	χ^2 = 6.40
COMBINED	0	806	809.65	.8926	813.32	.8967	813.50	.8969
	1	74	72.60	.0800	71.03	.0783	71.03	.0783
	2	15	18.74	.0206	19.05	.0210	19.00	.0209
	3	6	3.01	.0033	3.19	.0035	3.09	.0034
	4 or more	6	2.99	.0033	.41	.0004	.37	.0004
			α = .2526 P = .0114	χ^2 = 6.77	α = .2402 P = .0116	χ^2 = 80.32	α =.2375 λ = .0118	χ^2 = 88.88

REFERENCES

1. Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. Biometrika 37, 358-382.

2. Bliss, C. I. and Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data, with a note on the efficient fitting of the negative binomial, Biometrics 9, 176-200.

3. Bliss, C. I. and Owen, A. R. G. (1958). Negative binomial distributions with a common k, Biometrika 55, 37-58.

4. Cohen, A. C. (1966). A note on certain discrete mixed distributions, Biometrics 22, 566-572.

5. Dandekar, V. M. (1955). Certain modified forms of binomial and Poisson distributions, Sankhya 15, 237-251.

6. Falls, Lee W. (1969). A probability distribution for the number of thunderstorms events at Cape Kennedy, Florida," NASA TM X-53816, Marshall Space Flight Center, Alabama.

7. Feller, W. (1960). An Introduction to Probability Theory and its Applications, Vol. 1, (Second Edition, Fourth Printing), John Wiley and Sons, Inc., New York.

8. Greenwood, M. and Yule, G. U. (1920). An enquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, J. R. Statist. Soc. 83, 255-279.

9. Kendall, M. G. and Stuart, A. (1963). The Advanced Theory of Statistics, Vols. I and II, Charles Griffin and Company, London.

10. Neyman, J. (1939). On a new class of "contagious" distributions, applicable in entomology and bacteriology, Ann. Math. Stat. 10, 35-57.

11. Neyman, J. (1949). On the problems of estimating the numbers of schools of fish, University of California Publications in Statistics 1, 21-36.

12. Neyman, J. (1949). Contribution to the theory of χ^2 test, Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 239-273.

13. Singh, S. N. (1963). Probability models for the variation in the number of births per couple, Journal of the American Statistical Association 58, 721-727.

14. Singh, S. N. (1963). A probability model for couple fertility, Sankhya 26, 89-94.

15. Singh, S. N. (1968). A chance mechanism of the variation in the number of births per couple, Journal of the American Statistical Association 63, 209-213.

APPENDIX

This program is written in the IBM 7094 version of FORTRAN IV. It is presently set up to handle values of T up to 200 and number of cells up to 50. It is not anticipated that usage would exceed these values but changing the dimensioned values of FACT, FACE, FADE and QS would be the only alteration required to use a larger T, and those with present dimension values of 50 would require changing to accomodate larger cell numbers.

Variables required are:

N : number of cells

T : number of units in T interval

H : value of h.

FN(I) : number of observations for class I-1 (computer doesn't recognize 0 subscripts, hence the correction)

Using this program requires the following cards after program deck

DATA : (or Monitoring System Counterpart)

CARD I : N, T, H in (I2, 2F, 5.0) format

CARD 2 : FN(I), I = 1, N in (10F 8.0) Format for as many cards
⋮ as required

CARD N

CARD N + 1 : N, T, H in (I2, 2F 5.0) format, etc.

As many data sets as desired can be processed at one time. The final data card should be followed by the "end of file" card (a 7-8 card or its counterpart).

As stated in the report, the output is very simple with all numbers being adequately identified for immediate interpretation.

```
      DIMENSION FACE(200),FN(50),PA(50),PD(50),PL(50),P(50),EP(50),QS(50
     1,200),FACT(200),Q(50),FADE(200)
 1000 READ(5,400)N,T,H
 400  FORMAT(I2,2F5.0)
      READ(5,401)(FN(I),I=1,N)
  401 FORMAT(10F8.0)
      FNT=0.
      DO 216 I=1,N
 216  FNT=FNT+FN(I)
      NDF=N-3
      IH=H
      IHMI1=IH-1
      IT=T
      FADE(1)=0.
      DO 402 I=2,IT
      FI=I
 402  FADE(I)=FADE(I-1)+ALOG(FI)
      NMIN1=N-1
      NMIN2=N-2
      Q(1)=T-H
      DO 416 I=2,NMIN2
      J=IT-I*IH+I-1
      JJ=J-I
      Q(I)=EXP(FADE(J)-FADE(I)-FADE(JJ))
      DO 417 M=1,IHMI1
      J=IT-M-(I-1)*IH+I-2
      JJ=J-(I-1)
 417  QS(I,M)=EXP(FADE(J)-FADE(I-1)-FADE(JJ))
 416  CONTINUE
      IJK=1
```

```
      14
      GO TO 418
419   Q(1)=T-H+1.
      DO 500 I=2,NMIN1
      JI=IT-I*IH+I
      JII=JI-I
      Q(I)=EXP(FADE(JI)-FADE(I)-FADE(JII))
      DO 500 M=1,IHMI1
      JI=IT-(I-1)*IH-M+I-1
      JII=JI-(I-1)
500   QS(I,M)=EXP(FADE(JI)-FADE(I-1)-FADE(JII))
      IJK=2
418   DO 501 M=1,IHMI1
 501 QS(1,M)=1.
      NN=101
      B=1.
      A=100.
 502 DO 503 IL=1,NN
      AI=IL-1
      PI=B-AI/A
      IF((1.-PI)**IT-.000000001)503,503,505
 505 SUM=0.
      DO 507 M=1,IHMI1
 507 SUM=SUM+(1.-PI)**(IH-M)
      F=(1.-FN(1)/FNT)*PI*(1.-PI)**(IT-IH)*(Q(1)+SUM)/(1.-(1.-PI)**IT)-F
     1N(2)/FNT
      IF(F)503,508,509
 503 CONTINUE
 509 IF(ABS(F)-.000001)508,508,510
 510 B=B-(AI-1.)/A
      A=10.*A
```

```
      NN=(NN-1)*10+1
  511 GO TO 502
  508 PB=PI
      AB=(1.-FN(1)/FNT)/(1.-(1.-PI)**IT)
      PB1= 1.-PI
      DO 514 I=1,NMIN2
      TI=I
      DSUM=0.
      SUM=0.
      DO 515 M=1,IHMI1
      TM=M
      SUM=SUM+QS(I,M)*PB1**(IH-M)
  515 DSUM=DSUM+PB1**(IT-I*IH+IH-M-1)*(TI*PB1-(T-TI*H+H-TM)*PB)*QS(I,M)
      DSUM=DSUM*PB**(I-1)
      PA(I)=PB**I*PB1**(IT-I*IH)*(Q(I)+SUM)
      PL(I)=AB*PA(I)
      PD(I)=PB**(I-1)*PB1**(IT-I*IH-1)*(TI*PB1-(T-TI*H)*PB)*Q(I)+DSUM
  514 PD(I)=AB*PD(I)
      PAZER=PB1**IT-1.
      PBZER=-T*AB*PB1**(IT-1)
      PLZER=1.-AB*(1.-(1.-PB)**IT)
      PA(NMIN1)=-PAZER
      PD(NMIN1)=-PBZER
      PL(NMIN1)=1.-PLZER
      DO 523 I=1,NMIN2
      PA(NMIN1)=PA(NMIN1)-PA(I)
      PD(NMIN1)=PD(NMIN1)-PD(I)
  523 PL(NMIN1)=PL(NMIN1)-PL(I)
      BAA=0.
      BPP=0.
      A11=0.
```

```
      A12=0.
      A22=0.
      DO 516 I=1,NMIN1
      BAA=BAA+PL(I)*PA(I)/FN(I+1)
      A11=A11+PA(I)**2/FN(I+1)
      BPP=BPP+PL(I)*PD(I)/FN(I+1)
      A12=A12+PA(I)*PD(I)/FN(I+1)
516   A22=A22+PD(I)**2/FN(I+1)
      BP=-BPP-PLZER*PBZER/FN(1)
      BA=-BAA-PLZER*PAZER/FN(1)
      A11=A11+PAZER**2/FN(1)
      A12=A12+PAZER*PBZER/FN(1)
      A22=A22+PBZER**2/FN(1)
      E=A11*A22-A12**2
      AHAT=AB+(A22*BA-A12*BP)/E
      PHAT=PB+(A11*BP-A12*BA)/E
      PZER=1.-AHAT*(1.-(1.-PHAT)**IT)
      EPZER=PZER*FNT
      CHISQ=(EPZER-FN(1))**2/EPZER
      DSUM=0.
      DO 517 I=1,NMIN2
      SUM=0.
      DO 518 M=1,IHMI1
518   SUM=SUM+QS(I,M)*(1.-PHAT)**(IH-M)
      P(I)=AHAT*PHAT**I*(1.-PHAT)**(IT-I*IH)*(Q(I)+SUM)
      DSUM=DSUM+P(I)
      EP(I)=P(I)*FNT
517   CHISQ=CHISQ+(EP(I)-FN(I+1))**2/EP(I)
      P(NMIN1)=1.-PZER-DSUM
      EP(NMIN1)=FNT*P(NMIN1)
```

```
      CHISQ=CHISQ+(FN(N)-EP(NMIN1))**2/EP(NMIN1)
      IF(IJK-1)530,530,531
  530 WRITE(6,532)
  532 FORMAT(1H1,30H   MODIFIED NEGATIVE BINOMIAL )
      GO TO 533
  531 WRITE(6,534)
  534 FORMAT(1H1,17H SINGHS BINOMIAL )
  533 WRITE(6,17)FNT
      WRITE(6,535)T,H
  535 FORMAT(1H ,30X,12H T INTERVAL ,3X,F5.0,10X,12H H INTERVAL ,3X,F5.0
     1)
      WRITE(6,536)AHAT,PHAT
  536 FORMAT(1H0,30X,10H ALPHA IS ,3X,F10.5,10X,6H P IS ,3X,F10.5)
      WRITE(6,18)
      WRITE(6,19)FN(1),EPZER,PZER
      DO 520 I=1,NMIN1
      J=I+1
  520 WRITE(6,21)FN(J),EP(I),P(I),I
      WRITE(6,22)NDF,CHISQ
      IF(IJK-1)419,419,600
  600 CONTINUE
  200 NN=101
      B=1.
      A=100.
  205 DO 201 IK=1,NN
      AI=IK-1
      ALP=B-AI/A
      PI=-(1./T)*ALOG(((FN(1)/FNT)+ALP-1.)/ALP)
      GT=EXP(-PI)
      F=ALP*GT**IT*((1./GT)**IH*(1.+PI*(T-H))-1.)-FN(2)/FNT
```

```
      IF(F)203,202,201
  201 CONTINUE
  203 IF(ABS(F)-.00001)202,202,204
  204 B=B-(AI-1.)/A
      A=10.*A
      NN=(NN-1)*10+1
 250  GO TO 205
  202 AB=ALP
      PB=PI
      GT=EXP(-PB)
      Q(1)=GT**(IT-IH)*(1.+PB*(T-H)-GT**IH)
      DO 220 I=1,N
 220  FACE(I)=EXP(FADE(I))
      DO 207 I=2,NMIN2
      Q(I)=0.
      QS(I,1)=0.
      SUM=0.
      FI=I
      DO 208 M=1,I
 208  SUM=SUM+(PB*(T-FI*H))**M/FACE(M)
      DSUM=0.
      IM=I-1
      DO 209 M=1,IM
 209  DSUM=DSUM+(PB*(T-FI*H+H))**M/FACE(M)
      Q(I)=GT**(IT-I*IH)*(1.+SUM)-GT**(IT-I*IH+IH)*(1.+DSUM)
      QS(I,1)=GT**(IT-I*IH)*PB**(I-1)*(GT**IH*(T-FI*H+H)**I/FACE(I-1)-(T
     1-FI*H)**(I+1)*PB/FACE(I))
  207 CONTINUE
      QS(1,1)=GT**(IT-IH)*(GT**IH*T-(T-H)**2*PB)
      QZER=1.-GT**IT
```

```
      QSZER=T*GT**IT
      PLZER=1.-AB*(1.-GT**IT)
      PAZER=GT**IT-1.
      PDZER=-T*AB*GT**IT
      PL(NMIN1)=1.-PLZER
      PD(NMIN1)=-PDZER
      PA(NMIN1)=-PAZER
      DO 225 I=1,NMIN2
      PL(I)=Q(I)*AB
      PA(I)=Q(I)
      PD(I)=AB*QS(I,1)
      PA(NMIN1)=PA(NMIN1)-PA(I)
      PD(NMIN1)=PD(NMIN1)-PD(I)
  225 PL(NMIN1)=PL(NMIN1)-PL(I)
C     SOLVING FOR ALPHA-HAT AND LAMBDA-HAT
      BAA=0.
      BPP=0.
      A11=0.
      A12=0.
      A21=0.
      A22=0.
      DO 210 I=1,NMIN1
      BAA=BAA+PL(I)*PA(I)/FN(I+1)
      BPP=BPP+PL(I)*PD(I)/FN(I+1)
      A11=A11+PA(I)**2/FN(I+1)
      A12=A12+PA(I)*PD(I)/FN(I+1)
  210 A22=A22+PD(I)**2/FN(I+1)
      BA=-BAA-PLZER*PAZER/FN(1)
      BP=-BPP-PLZER*PDZER/FN(1)
      A11=A11+PAZER**2/FN(1)
```

```
      A12=A12+PAZER*PDZER/FN(1)
      A22=A22+PDZER**2/FN(1)
      D=A11*A22-A12**2
      AHAT=AB+(A22*BA-A12*BP)/D
      PHAT=PB+(A11*BP-A12*BA)/D
C     HAT ESTIMATES COMPUTED
      GT=EXP(-PHAT)
      PZER=1.-AHAT*(1.-GT**IT)
      P(1)=AHAT*(GT**(IT-IH)*(1.+PHAT*(T-H)-GT**IH))
      EP(1)=FNT*P(1)
      VAT=0.
      DO 211 I=2,NMIN2
      P(I)=0.
      FI=I
      IMIN1=I-1
      SUM=0.
      DSUM=0.
      DO 212 M=1,I
  212 SUM=SUM+(PHAT*(T-FI*H))**M/FACE(M)
      DO 221 M=1,IMIN1
  221 DSUM=DSUM+(PHAT*(T-FI*H+H))**M/FACE(M)
      P(I)=AHAT*(GT**(IT-I*IH)*(1.+SUM)-GT**(IT-I*IH+IH)*(1.+DSUM))
      VAT=VAT+P(I)
  211 EP(I)=FNT*P(I)
      P(NMIN1)=1.-PZER-VAT-P(1)
      EP(NMIN1)=FNT*P(NMIN1)
      CHISQ=0.
      DO 213 I=1,NMIN1
  213 CHISQ=CHISQ+(FP(I)-FN(I+1))**2/EP(I)
      CHISQ=CHISQ+(PZER*FNT-FN(1))**2/(PZER*FNT)
```

```
      EPZER=PZER*FNT
      WRITE(6,214)
  214 FORMAT(1H1,12H POISSON FIT)
      WRITE(6,17)FNT
      WRITE(6,535)T,H
      WRITE(6,222)AHAT,PHAT
222   FORMAT(1H0,30X,10H ALPHA IS ,3X,F10.5,10X,11H LAMBDA IS ,3X,F10.5)
      WRITE(6,18)
      WRITE(6,19)FN(1),EPZER,PZER
      DO 215 I=1,NMIN1
      J=I+1
  215 WRITE(6,21)FN(J),EP(I),P(I),I
      WRITE(6,22)NDF,CHISQ
   21 FORMAT(1H0,18X,F8.2,16X,F8.2,16X,F10.5,12X,I2)
22    FORMAT(1H0,16H CHI-SQUARE WITH,1X,I2,8H D.F. IS,3X,E12.5)
   17 FORMAT(1H0,30X,19H TOTAL OBSERVATIONS ,6X,F5.0)
   19 FORMAT(1H0,18X,F8.2,16X,F8.2,16X,F10.5,12X,3H 0 )
   18 FORMAT(1H0,10X,24H   OBSERVED FREQUENCY    ,24H   EXPECTED FREQUENC
     2Y    ,26H  PROBABILITY OF  I  THS  ,10H     I    )
      GO TO 1000
      END
```