N73-25678
NASA CR-2280

NASA CONTRACTOR REPORT

NASA CR-2280

ESTIMATION IN A MODIFIED BINOMIAL DISTRIBUTION

by Michael C. Carter

Prepared by
APPALACHIAN STATE UNIVERSITY
Boone, N.C. 28607
for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JUNE 1973

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
NASA CR-2280		

4. TITLE AND SUBTITLE	5. REPORT DATE
Estimation In A Modified Binomial Distribution	June 1973
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT #
Michael C. Carter	M110

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Appalachian State University Dept. of Mathematics Boone, North Carolina 28607	11. CONTRACT OR GRANT NO. NAS 8-29286
	13. TYPE OF REPORT & PERIOD COVERED Contractor
12. SPONSORING AGENCY NAME AND ADDRESS National Aeronautics and Space Administration Washington, D.C. 20546	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

Prepared under the technical direction of the Aerospace Environment Division, Aero-Astrodynamics Laboratory, NASA-Marshall Space Flight Center. Contract Monitor: L. W. Falls.

16. ABSTRACT

Estimation techniques in a modified binomial distribution, developed to describe thunderstorm activity over a small area at Cape Kennedy, Florida, are compared. A compound model is also developed and compared with the original model. The minimum Chi square technique is compared with the maximum likelihood and method of moments techniques. The minimum Chi square technique, although useful in complicated models, compared poorly compared to the other aforementioned techniques. The maximum likelihood and method of moments were comparable. The compound model fit better in every case but not significantly so based on a likelihood ratio test comparing the compound model with the modified binomial model using maximum likelihood estimators.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Distribution Functions Thunderstorm Activity Binomial Distribution Modified Distributions Statistical Analysis	Category 20

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	21	$3.00

* For sale by the National Technical Information Service, Springfield, Virginia 22151

SUMMARY

Estimation techniques in a modified binomial distribution, developed to describe thunderstorm activity over a small area at Cape Kennedy, Florida, are compared. A compound model is also developed and compared with the original model. The minimum Chi square technique is compared with the maximum likelihood and method of moments techniques. The minimum Chi square technique, although useful in complicated models, compared poorly compared to the other aforementioned techniques. The maximum likelihood and method of moments were comparable. The compound model fit better in every case but not significantly so based on a likelihood ratio test comparing the compound model with the modified binomial model using maximum likelihood estimators.

ACKNOWLEDGEMENTS

The author wishes to thank L. W. Falls, O. E. Smith and Dr. A. C. Cohen, Jr. for many helpful discussions.

1. Introduction

The model investigated below arose while endeavoring to determine probabilistic models for thunderstorm activity at Cape Kennedy, Florida. This model is designed to predict the frequency of thunderstorm "hits", i.e., the occurrence of thunderstorms over a small area.

According to standard United States weather observing procedure a thunderstorm is reported when thunder is heard at the station and ends 15 minutes after thunder is last heard. This standard definition of a thunderstorm may therefore include multiple occurrences and will be called a "thunderstorm event" (THE) with the individual occurrences within this THE called thunderstorms (TH's). Falls, et.al. (1971) show that THE frequencies per day are adequately described by a negative binomial model. In a related paper Carter (1972) discusses the problem of predicting multiple TH occurrences within a THE.

The occurrences of atmospheric phenomena generally form stochastic processes in continuous time. Variables such as ground temperature and wind speeds are usually analyzed in such a framework. Other phenomena such as lightning, hurricane, thunderstorm and hail occurrences are recorded in a discrete fashion and statistically analyzed using discrete models. Panofsky and Brier (1958, pp. 32-40) discuss the applications of discrete distributions in meteorology (specifically discussing the role of the binomial model in describing TH frequencies) and Thom (1957) employs the negative binomial distribution in describing the frequency of hail occurrence.

Aside from the fact that thunderstorm occurrences have historically been treated as discrete events the ultimate use of the model prompted a discrete treatment. In the design and assembly schedules for launch vehicles the occurrences of thunderstorms, especially TH hits, are of primary concern. While a continuous time model should prove adequate, questions of the type "How many days in June can we expect X TH hits?" or "What is the expected number of THE's per day in June?" would require answers for scheduling purposes. A discrete model would be mathematically simpler and readily provide adequate answers.

2. Models and Data

When a TH is overhead, another TH cannot then occur for some time interval h, otherwise they would be considered a single TH hit. In general, distributions with this property are called "interrupted" distributions (See Johnson and Kotz 1969, pp. 269-273). The Geiger counter problem with finite resolving time equal to h (See Feller 1968, pg. 306) is a related problem. Singh (1963, 1968) has applied the same concept in fertility studies as has Neyman (1949a) in estimating the number of schools of fish. Our development, in an entirely different application, will parallel the model developed by Singh (1963) and Neyman (1949a).

We make the following assumptions:

(1) A probability of α is assigned to the possibility of one or more hits occurring in T time units.

(2) The probability p is the probability of a TH hit occurring in a unit of time. We assume p to be

constant during the T time units.

(3) The constant h denotes the "waiting time", i.e., given a TH overhead in a specific time interval another cannot occur within the next h-1 units of time. The maximum number of occurrences in T is $n \leq [T/h] + 1$ where $[T/h]$ denotes the largest integer in T/h.

Letting X be the random variable denoting the number of hits in T time units we have immediately

$$\Pr\{X = 0\} = (1-\alpha) + \alpha q^{T}, \quad (q = 1 - p) \tag{1.1}$$

as the sum of the mutually exclusive probabilities $1-\alpha$ for no hits possible in T time periods and αq^{T} when hits are possible but none occur. For the case $0 < x < n$ hits per T time periods there are two distinct cases. Either the hits and resultant waiting times are wholly contained in the interval T or the hits occur in such a manner that the last waiting time extends into the next time period. In the former case we have the probability

$$\binom{T-(h-1)x}{x} p^{x} q^{T-xh}$$

and in the latter case the probability that the rest period will extend k units into the next period is

$$\binom{T-k-(x-1)h+x-1}{x-1} p^{x} q^{T-(x-1)h-k}.$$

Noting that the different values of k comprise mutually exclusive events and the events of all rest periods contained in T and the

extension of the xth rest period into the next time period are also mutually exclusive, we have

$$\Pr\{X=x\} = \alpha p^{x} q^{T-xh} \left\{ \binom{T-(h-1)x}{x} + \sum_{k=1}^{h-1} \binom{T-k-(x-1)h+x-1}{x-1} q^{h-k} \right\}. \qquad (1.2)$$

Finally for $X=n$ we have

$$\Pr\{X=n\} = 1 - \sum_{x=0}^{n-1} \Pr\{X=x\}. \qquad (1.3)$$

This model (1.1, 1.2, 1.3) is of the general form

$$\Pr\{X=x\} = \begin{cases} 1-\alpha + \alpha P_0 & x = 0 \\ \alpha P_x & x > 0 \end{cases}$$

where $\Pr\{X=x\} = P_x$ $(x=0,1,2,\ldots)$ for the original distribution. Distributions of this form are called "modified" distributions (See Johnson and Kotz 1969, pp. 204-209) and are usually employed when an excess of zeros is present.

In the investigations leading to the selection of a negative binomial model to describe THE activity Falls, et.al. (1971) initially considered the Poisson distribution, a natural choice to describe the variation in THE frequencies for a specified time interval. The negative binomial gave better fits possibly because the synoptic conditions that prompted one THE occurrence increased the possibility of additional occurrences - hence creating a possible dependency between successive events. When successive events are possibly dependent the negative binomial distribution

is suggested as an alternative to the Poisson model (See Johnson and Kotz 1965, pg. 135 or Jeffreys 1961, pp. 79, 319).

If there is a dependency between successive THE's the influence is both small and difficult to measure in the model for TH hits. Of the number of THE's occurring, the resulting TH hits form a small percentage. A possible dependency between THE's increases the probability of additional THE occurrences and as a result possibly increases the value of p in (1.1, 1.2, 1.3) during those time periods. This possible variability in p leads to the examination of a more general model obtained by assuming that p varied from period to period according to some probability distribution and obtaining the resultant compound distribution. A reasonable choice is to assume that p varies according to the beta law, i.e., $f(p) \propto p^{\delta-1}(1-p)^{\gamma-1}$, $0 < p < 1$, $\delta, \Gamma > 0$. Based on the richness of the beta family and the range $(0 < p < 1)$ it is reasonable to assume that a member of the beta family describes or closely approximates the variation in p. A study presented in section 4 gives additional experimental verification. With this assumption on p the model becomes

$$\Pr\{X=x\} = \begin{cases} 1-\alpha + \alpha\, \beta(\delta,\gamma+T)/\beta(\delta,\gamma) & x=0 \\ \dfrac{\alpha}{\beta(\delta,\gamma)} \left\{ \binom{T-xh+x}{x} \beta(\delta+x,\gamma+T-xh) \right. & \\ \left. + \sum_{k=1}^{h-1} \binom{T-(x-1)h+x-k-1}{x-1} \beta(\delta+x,\gamma+T-xh+x-k) \right\} & 0<x<n \\ 1 - \sum_{i=0}^{n-1} \Pr(x=i) & x=n \end{cases} \tag{2}$$

where $\beta(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$.

2.2 Data

The sample data was compiled by ESSA, National Weather Records Center, Asheville, North Carolina and was made available to the author by the Aero-Astrodynamics Laboratory, Aerospace Environment Division, Marshall Space Flight Center, Alabama. Comprehensive thunderstorm data for Cape Kennedy, Florida is provided for the years 1957-67 inclusive. In the sequel we shall discuss the peak thunderstorm activity months of June, July and August. The observation area is used as the "point" and a TH hit is recorded if

(1) A thunderstorm was actually reported overhead or

(2) A thunderstorm was first reported in a sector and last reported in an opposite sector. It is reasonable to assume thunderstorms move in a straight line (over small areas at least).

These TH hit frequencies are presented in Table 1 below.

Table 1

Frequencies of the Observed Number of Days that experienced X TH hits at Cape Kennedy, Florida for the 11-Year Period 1957-67.

X	June	July	August
0	293	305	300
1	27	24	30
2	5	6	7
3	3	3	2
4 or more	2	3	2
Total	330	341	341

The zero class data can be partitioned into the days when no thunderstorms occurred in the general area (denoted by X_{oo}) and the days when thunderstorms occurred but no hits were recorded (denoted by X_{o1}). These frequencies are presented in Table 2 below.

Table 2

Frequency of Days having no TH hits (X_{oo} and X_{o1}) and X_i (i=1), 4 (or more)) hits.

	X_{oo}	X_{o1}	X_1	X_2	X_3	X_4 (or more)	Total
June	187	106	27	5	3	2	330
July	178	127	24	6	3	3	341
August	185	115	30	7	2	2	341

The time period T is a day with individual units of time defined as 30 minutes, making T=48. The data in Tables 1 and 2 were determined using the value h=2. This meant successive hits were required to be at least 30 minutes apart, otherwise the thunderstorm activity was considered to be a continuation of the previously reported hit.

3. Estimation

This section presents estimation results for the models discussed in section 2.1.

3.1 Minimum Chi Square Estimation

Following Singh (1968), who used the modified minimum chi-square technique (MCS) in estimating the parameters in the modified Poisson distribution, the procedure was applied to the model given by equations (1.1, 1.2, 1.3). The MCS technique was introduced by Neyman (1949b) who proved the MCS technique produced consistent, efficient, BAN estimators. Letting $P_i(\alpha,p) = Pr\{X=i\}$ from equations (1.1, 1.2, 1.3) the MCS estimates of α and p were obtained by minimizing the expression

$$\chi^2 = \sum_{i=0}^{n} \frac{(X_i - p_i(\alpha,p)\Sigma X)^2}{X_i}$$

with respect to α and p. This procedure leads to complicated estimating equations difficult to evaluate without the aid of a computer and as the equations are not required in the rest of the paper they will not be presented (they are given in Falls, et.al. (1971)). The results of fitting the distribution (1.1, 1.2, 1.3) to the data in Table 1 are presented in Table 3 and the parameter estimates are presented in Table 4.

3.2 "Exact Zero Class" Estimation Procedure

This procedure is suggested by Johnson and Kotz (1969, pp. 205-206). For the modified binomial distribution the technique is:

(1) Ignoring the zero class, estimate p using a "truncated" modified binomial distribution and

(2) Estimate the value of α by equating the observed expected zero class frequencies.

The method of moments was used to estimate the parameter p via an iterative process utilizing a computer. Table 3 presents the results of this procedure and Table 4 lists the parameter estimates.

3.3 Maximum Likelihood Estimation

The likelihood function for (1.1, 1.2, 1.3) can be written

$$L \propto \left\{1-\alpha\left(1-q^{T}\right)\right\}^{X_o} \prod_{i=1}^{4} \left\{\alpha p^{i} q^{T-ih}\right\}^{X_i}. \qquad (3)$$

The term in equation (1.2) corresponding to the case where the last rest period extends into the next time period is not present as such a situation does not occur in the data. Had it occurred the particular frequency X_i would have been partitioned into classifications according to the number of time units $(k=0,1,2,\ldots,h-1)$ the rest period extended into the next period with a likelihood term proportional to $\alpha p^{i} q^{T-(i-1)h+k}$ corresponding to each subclassification.

Taking the logarithm of L, differentiating with respect to α and q and setting the derivatives equal to zero gives the estimating equations. Letting $m = \sum_{i=0}^{n} X_i$ and $\bar{x} = \sum_{i=0}^{n} iX_i/m$ we have

$$\frac{X_o(1-\hat{q}^{T})}{1-\hat{\alpha}(1-\hat{q}^{T})} \quad \frac{m-X_o}{\hat{\alpha}} = 0 \qquad (4.1)$$

and

$$\frac{X_o \hat{\alpha} T \hat{q}^{T-1}}{1-\hat{\alpha}(1-\hat{q}^T)} - \frac{m\bar{x}}{1-\hat{q}} + \frac{(m-X_o)T-hm\bar{x}}{\hat{q}} = 0. \tag{4.2}$$

Estimation is actually accomplished by solving (4.1) for $\hat{\alpha}$, $\hat{p}$ substituting the expression into (4.2) and finding an iterative solution via a computer. The asymptotic covariance matrix for the estimates from (4.1) and (4.2) are obtained by obtaining the second partials of log L and evaluating the expression

$$V(\hat{\alpha},\hat{q}) = \begin{bmatrix} E\left(\frac{-\partial^2 \log L}{\partial \alpha^2}\right) & E\left(\frac{-\partial^2 \log L}{\partial \alpha \partial q}\right) \\ \vdots & \\ E\left(\frac{-\partial^2 \log L}{\partial \alpha \partial q}\right) & E\left(\frac{-\partial^2 \log L}{\partial q^2}\right) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{(1-q^T)^2 E(X_o)}{\{1-\alpha(1-q^T)\}^2} - \frac{E(m-X_o)}{\alpha^2} & -\frac{Tq^{T-1}E(X_o)}{\{1-\alpha(1-q^T)\}^2} \\ \frac{-Tq^{T-1}E(X_o)}{\{1-\alpha(1-q^T)\}^2} & \frac{^2Tq^{2(T-1)}E(X_o)}{\{1-\alpha(1-q^T)\}^2} \\ \vdots & \\ & + \frac{mE(\bar{x})}{(1-q)^2} + \frac{TE(m-X_o)-hmE(\bar{x})}{q^2} \end{bmatrix}^{-1} \tag{5}$$

Using the estimates $\hat{\alpha}$ and $\hat{q}$ numerical values for $E(X_o)$, $E(m-X_o)$, $E(\bar{x})$ and, subsequently, approximate values for $V(\hat{\alpha},\hat{q})$ are calculated.

Maximum likelihood estimation in the compound model given by equation (2) presents a formidable task. The likelihood equation is

$$L \propto \{1-\alpha+\alpha\ \beta(\delta,\gamma+T/\beta(\delta,\gamma)\}^{X_o} \prod_{i=1}^{4} \{\alpha\ \beta(\delta+i,\gamma+T-ih)/\beta(\delta,\gamma)\}^{X_i}. \qquad (6)$$

Estimation by the standard method, i.e., differentiating log L and solving the three nonlinear equations simultaneously was not feasible. The approach adopted was the maximization of log L using a "direct search" computer routine. Initial estimates were obtained using the $\hat{\alpha}$ obtained in (4.1, 4.2) while initial values of δ and γ were obtained by using the estimated $\hat{p}$ and var $(\hat{p})$ from (4.1, 4.2) and $V(\hat{\alpha},\hat{q})$.

The fitted models are again presented in Table 3 with parameter estimates presented in Table 4.

3.4 ML Estimation with Zero Class Partitioned

Usually the reason for applying a modified distribution is to compensate for an inflated zero class and a direct interpretation of the associated parameter is not possible. If we assume a specific interpretation for the parameter α, namely, no THE's per day implies no TH hits are possible, the result if a partition of the zero class into days with no THE's (X_{oo}) and days with THE's but no hits (X_{o1}). This partition leads to sufficient estimators for α and p. It should be noted that in usual applications of modified distributions the data cannot be partitioned in this manner. Consequently, the feasibility of such an interpretation of the parameter α cannot often be investigated.

The likelihood function for a specified month becomes

$$L \propto (1-\alpha)^{X_{oo}} (\alpha q^{T})^{X_{o1}} \prod_{i=1}^{4} (\alpha p^{i} q^{T-ih})^{X_i}$$

$$\propto (1-\alpha)^{X_{oo}} \alpha^{m-X_{oo}} p^{m\bar{x}} q^{T(m-X_{oo})-hm\bar{x}} . \qquad (7)$$

Taking the logarithm of L, differentiating with respect to α and p and setting the derivatives equal to zero gives the maximum likelihood estimators

$$\hat{\alpha} = 1 - X_{oo}/m$$

$$\hat{p} = \bar{x}/\{\bar{x} + T(1-X_{oo}/m) - h\bar{x}\}. \qquad (8)$$

The likelihood (7) can be written in the form

$$L \propto (1-\alpha)^{m(1-\hat{\alpha})} \alpha^{m\hat{\alpha}} p^{Tm\hat{\alpha}\hat{p}/\{1-(1-h)\hat{p}\}}$$

$$\times q^{Tm\hat{\alpha}-hTm\hat{\alpha}\hat{p}/\{1-(1-h)\hat{p}\}} \qquad (9)$$

which shows the estimates $\hat{\alpha}$ and $\hat{p}$ in (8) are sufficient statistics for α and p.

Likelihood equations (3), (6) and (7) are approximate in the sense that the term $p^{4} q^{T-4h}$ instead of $1 - \sum_{x=0}^{3} \Pr(X=x)$ is used for X_4 (or more). Frequencies greater than 4 are quite infrequent and the mathematical ease gained is considerable.

Table 3

Observed and Expected Frequencies for the Models and Estimation Procedures in Sections 2 and 3.

Month	X		Observed Frequency	MCS (Sec. 3.1)	Exact Zero Frequency (Sec. 3.2)	ML (Sec. 3.3)	Compound Model (Sec. 3.3)	ML - Zero Class Partitioned (Sec. 3.4)
June	oo		187					168.52
	o1		106					127.57
		0	293	296.47	293.00	293.00	291.46	
		1	27	25.58	23.76	23.80	25.03	29.68
		2	5	6.71	10.03	10.02	9.36	[3.89
		3	3	[1.10	[2.64	[2.63	3.00	.32
		4 or more	2	.14]	.56]	.56]	1.16	.02]
		χ^2		12.077*	3.926	3.978	2.808	13.853*
July	oo		178					162.98
	o1		127					147.70
		0	305	308.31	305.00	305.00	304.07	
		1	24	22.45	20.56	20.64	21.07	27.39
		2	6	8.09	10.87	10.84	9.44	[2.74
		3	3	[1.82	[3.58	[3.55	3.76	.17
		4 or more	3	.32]	.99]	.97]	2.05	.01]
		χ^2		8.601*	3.204	3.192	2.255	33.989*
August	oo		185					164.68
			115					137.49
		0	300	301.81	300.00	300.00	299.03	
		1	30	29.20	27.47	27.50	27.64	33.75
		2	7	8.31	10.58	10.52	10.15	[4.66
		3	2	[1.47	[2.52	[2.51	3.09	.40
		4 or more	2	.20]	.48]	.48]	1.08	.03]
		χ^2		3.450	1.754	1.750	2.357	13.479*

Brackets indicate classes have been combined in χ^2 calculations.
*χ^2 with 1 d.f at $\alpha = .01$ is 6.635

Table 4

Estimated Parameter Values for the Models Presented in Table 3.

		June	July	August
MCS	α	.2403	.1816	.2562
(Sec. 3.1)	p	.0114	.0155	.0123
Exact Zero Frequency	α	.1921	.1591	.2188
(Sec. 3.2)	p	.0181	.0224	.0165
ML*	α	.1925	.1597	.2192
(Sec. 3.3)	p	.0180	.0223	.0164
Compound Model	α	.2662	.2158	.2753
(Sec. 3.3)	δ	1.844	1.537	2.225
	γ	130.096	85.875	156.750
ML-Zero Partitioned	α	.4248	.4656	.4502
	p	.00575	.00440	.00605

*The variance-covariance matrices are

June		July		August	
1.1523×10^{-3}	5.2372×10^{-5}	7.6465×10^{-4}	4.2235×10^{-5}	1.3753×10^{-3}	5.4399×10^{-5}
5.2372×10^{-5}	8.3707×10^{-6}	4.2235×10^{-5}	1.1043×10^{-5}	5.4399×10^{-5}	6.7839×10^{-6}

4. Discussion

This section compares the several models and estimation techniques investigated and should be prefaced by the following observations. The MCS technique was initially adopted from Singh (1968). The presence of low tail frequencies prompted very poor fits and as the model appeared satisfactory from a physical standpoint other forms of estimation were investigated. The modified Poisson used by Singh (1968) was also investigated and the model (1.1, 1.2, 1.3) always yielded smaller χ^2 values.

The compound model gave a better fit in every case. This is undoubtedly due to the presence of an extra parameter. The assumption that p follows a beta distribution can, to an extent, be verified by examining the actual data. The randomness alone is a very plausible assumption. One simple method is to take small groups of days, estimate p for each and examine the resultant data. Maximum likelihood estimates for p were calculated for successive five-day frequencies for August. There were 20 such periods where no TH hits occurred giving $\hat{\alpha} = 0$ and any $0 \leq \hat{p} \leq 1$ as estimates. These were not included in the calculations. A crude sketch indicated a beta model was plausible and the calculated mean and variance, .01265 and .01413 respectively, compared favorably with the mean and variance values, .01399 and .00928 respectively, obtained using $\hat{\delta}$ and $\hat{\gamma}$ from the compound model fit for August. The smaller variance obtained from $\hat{\delta}$ and $\hat{\gamma}$ can be explained by the much larger sample size.

While this compound model is plausible and fits the data well for all three months, the small variance for p (indicated by either

the variance calculated using $\hat{\delta}$ and $\hat{\gamma}$ or the asymptotic variance for $\hat{p}$ given in Table 4) suggests that the treatment of p as a constant is not a serious simplification. A likelihood ratio (modified binomial/compound) was calculated for June, July and August yielding values of .3549, .3315 and .4834 respectively. We can immediately conclude that the compound model is better supported by the data (See Hacking 1965, pp. 70-71). Since the question is whether or not an additional parameter is required, the testing procedure given by Jeffreys (1961, pp. 433-434) can be used. If the null hypothesis (additional parameter not necessary) and the alternative hypothesis (additional parameter is necessary) are equiprobable then, to paraphrase Jeffreys, a value between $1/\sqrt{10} = .3162$ and 1 gives evidence against the null hypothesis but is not worth more than a bare mention. The conclusion is that the compound model with its additional parameter is not significantly better than the modified binomial model, i.e., p can reasonably be treated as a constant.

An examination of Tables 3 and 4 shows that the parameter estimates and fitted models obtained using the method prepared by Johnson and Kotz (1969, pp. 205-206) and maximum likelihood differ a negligible amount. This was an expected result since the comparison is essentially one of ML vs the method of moments for larger samples. As the asymptotic properties are similar and the numerical complexities seem equivalent there is little to choose between the estimators in this application. Perhaps a deciding factor could be the availability of estimator variances through the M.L. technique.

From Table 3 it is evident that, while data to partition the zero class is available, a strict physical interpretation cannot be given the parameter α. The estimated values for α agree with the negative binomial probability of one or more THE's obtained by Falls, et.al. (1971). As noted earlier, the low tail frequencies apparently prompted the poor showing made by the MCS estimators which, in turn, prompted a rejection of the proposed model. It is worthwhile to note that in each case the MCS estimators yielded the best estimates of the X_1 and X_2 frequencies and closely approximated the zero class frequencies.

For these data we may draw the following conclusions. The MCS estimation technique, although useful in complicated distributions of this type, should be used with care when a portion of the sample frequencies are small. Here the results were quite misleading. The technique proposed by Johnson and Kotz (1965, pp. 205-206) compares favorably with the maximum likelihood technique. The compound model obtained by assuming the modified binomial parameter p has a beta distribution gives a better fit in all cases but likelihood ratio tests show the extra parameter does not yield significantly better results. The exact interpretation of the parameter α yields sufficient estimators for α and p but the fitted models were unacceptable. One possible explanation is that α and p likely are not independent and this seemingly reasonable partition of the zero class yields "independent" estimators (the asymptotic variance matrix is diagonal).

References

Carter, M. C. (1972). A model for thunderstorm activity: Use of the compound negative binomial-positive binomial distribution. J. Roy. Statist. Soc., Series C, 21, part 3.

Falls, L. W., Williford, W. O. and Carter, M. C. (1971). Probability distributions for thunderstorm activity at Cape Kennedy, Florida. J. Appl. Meteorology 10, pp. 97-104.

Feller, W. (1968). An Introduction to Probability Theory and its Applications, Vol. I, 3rd. Edition. New York: Wiley.

Hacking, I. (1965). Logic of Statistical Inference. Cambridge: University Press.

Jeffreys, H. (1961). Theory of Probability, 3rd Edition. Oxford: Clarendon Press.

Johnson, N. and Kotz, S. (1969). Distribution in Statistics: Discrete Distributions. Boston: Houghton-Mifflin.

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

Neyman, J. (1949b). On the problem of estimating the number of schools of fish. University of California Publications in Statistics 1, pp. 21-36.

Panofsky, H. A. and Brier, G. W. (1958). Some Applications of Statistics to Meteorology. Pennsylvania State University Press.

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

Singh, S. N. (1968). A chance mechanism of the variation in the number of births per couple. J. Amer. Statist. Assoc. 58, pp. 721-727.

Thom, H. C. S. (1957). The frequency of hail occurrence. Arch. Meteor., Geophys. Bioklim B8, No. 2, pp. 185-194.

NASA-Langley, 1973 —— 20