Probability Distributions for Thunderstorm Activity at Cape Kennedy, Florida[1]

LEE W. FALLS

Marshall Space Flight Center, NASA, Alabama

AND WILLIAM O. WILLIFORD AND MICHAEL C. CARTER

University of Georgia, Athens

(Manuscript received 11 August 1969, in revised form 7 September 1970)

ABSTRACT

Several statistical distribution functions are investigated as prospective models to represent the variation of thunderstorm activity at Cape Kennedy, Fla. Statistical methods are presented using the latest and most comprehensive thunderstorm data available. The conclusion is reached that the negative binomial distribution and a modification of the negative binomial distribution may be used as provisional statistical models to represent "thunderstorm events" and "thunderstorm hits," respectively, at Cape Kennedy.

1. Introduction

Statistical methods of analysis may be divided into two general categories, descriptive and analytical, both of which depend on the basic laws of probability. Descriptive methods reduce large amounts of data to a few meaningful "statistics" such as means and standard deviations. A theoretical statistical model (distribution function) is assumed for the observations, and analytical methods are used to determine how well the empirical data fit this model.

The purpose of this paper is to determine underlying, or basic, theoretical distributions for making probability inferences in regard to two types of thunderstorm activity at Cape Kennedy, Fla. A thunderstorm event will relate to a thunderstorm at Cape Kennedy and immediate surroundings. A thunderstorm hit, denoted by TH, relates to a thunderstorm passing over a point, e.g., a launch site.

Thunderstorms are of primary concern in the design of launch vehicles, in the planning of space missions, and in launch operations at Cape Kennedy because of the high winds, lightning, and extreme turbulence associated with this atmospheric phenomenon. The combinations of environmental conditions, including unstable air with a relatively high moisture content, and some type of lifting action present during the summer months make Florida one of the major thunderstorm genesis areas over the entire earth. Two distributions are presented. The first is the negative binomial distribution to represent the variation in the number of thunderstorm events per day and the second a modified negative binomial distribution to represent the variation in the number of thunderstorms per day which pass over a given point.

[1] Research partially supported under NASA Contract NAS8-11175 with Aerospace Environment Office, Aero-Astrodynamics Laboratory, Marshall Space Flight Center.

2. Statistical models

In practical statistics, a discrete probability law is required to describe events which seem to occur at random; for example, the arrivals of customers at a service point or the number of accidents and breakdowns in a factory. It is common practice to assume that the frequencies of such events fit a Poisson distribution. However, the Poisson series requires the assumption that the probability of the event remains constant. In reality, it is rarely true that the probability of the event remains constant. Any variation in the probability of the event, in particular the tendency for one event to increase the probability of another, will increase the variance of the distribution in relation to the mean. A report by the weather observer of a thunderstorm is proof that the atmosphere is in a state of instability and conditions are present for the formation of further thunderstorm cells, i.e., the probability of the event is increasing. For the Poisson distribution, the variance of the distribution equals its mean, whereas in the binomial distribution, the variance is less than the mean. A distribution having its variance greater than its mean is the negative binomial.

Let us consider the first application of the negative binomial probability distribution by Yule (1910). We will make an analogy between this application and the distribution of thunderstorms at Cape Kennedy. Suppose we have a population of people subjected to recurring exposures to a disease and that during an exposure each member of the population has an equal

TABLE 1. Frequencies of the observed number of days that experienced x thunderstorm events at Cape Kennedy, Fla., for the 11-year period of record January 1957 through December 1967.

	Month														
x	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Spring	Summer	Fall
0	335	295	308	299	266	187	177	185	228	311	321	334	873	549	860
1	4	9	20	18	43	77	80	89	54	17	6	3	81	246	77
2	2	4	9	10	25	40	47	30	33	9	3	2	44	117	45
3		2	3	3	3	17	26	24	12	4		2	9	67	16
4			1		3	6	9	10	3				4	25	3
5					0	2	2	3					0	7	
6					1	1							1	1	
n	341	310	341	330	341	330	341	341	330	341	330	341	1012	1012	1001

probability p of contracting the disease. After x exposures, the proportions who have contracted the disease 0, 1, 2, $\cdots$ times will be given by

$$q^x, xpq^{x-1}, \frac{x(x-1)}{2!}p^2q^{x-2}, \cdots, \tag{1}$$

where $q=1-p$. The terms given by (1) are terms of the binomial series $(q+p)^x$. If k unfavorable exposures to the disease are fatal to the individual, the proportion surviving after x exposures will be given by the first k terms of the binomial $(q+p)^x$. The proportion dying during the xth exposure will be those who contracted the disease $(k-1)$ times in the first $(x-1)$ exposures and who contract it again during the xth exposure; this proportion is

$$\binom{x-1}{k-1}p^{k-1}q^{x-k}p=\binom{x-1}{k-1}p^kq^{x-k}; \tag{2}$$

and since deaths do not begin until the kth exposure, the proportion of deaths at the kth, $(k+1)$th, $\cdots$ exposure will be

$$p^k\left[1, kq, k(k+1)\frac{q^2}{2!}, \cdots\right], \tag{3}$$

which are successive terms in the expansion of $p^k(1-q)^{-k}$, a binomial with a negative index. Thus, the proportions of the original population dying during successive exposures are given by successive terms of the negative binomial distribution with the first deaths occurring at the kth exposure.

Now, the probability of exactly x events (density function) is given by

$$\Pr(x=i)=\binom{x+k-1}{k-1}p^kq^x, \quad \text{for } i\geqslant 0. \tag{4}$$

Suppose in Yule's classic example we let the people exposed to the disease be analogous to the days in some month, say June, being exposed to the synoptic condition favorable for the formation of thunderstorms at Cape Kennedy. Now, the number of deaths that result from exposure to the disease will be analogous to the number of thunderstorms that actually develop in June. Now, we have all the days in June subjected to recurrent exposures of synoptic conditions favorable for the formation of thunderstorms. We must assume that each day in June that is exposed to the favorable synoptic conditions has an equal probability p of having a thunderstorm develop. This is a reasonable assumption. Continuing our analogy, the proportion of thunderstorms that develop at the kth, $(k+1)$th,$\cdots$ exposure will be given by (3), successive terms in the expansion of $p^k(1-q)^{-k}$, a negative binomial whose density function is given by (4). Similar applications of the negative binomial distribution have been performed by Thom (1957, 1958). Thom considers the Polya distribution (a particular case of the negative binomial) in regard to hail frequency series.

Thunderstorms over a point (TH's) require additional assumptions. Only one TH is possible at a given time and a specified time interval must elapse before new activity constitutes a new TH as opposed to a continuation of the previous TH. To incorporate these requirements into the model the negative binomial distribution was modified. The assumptions and modification are presented below.

a. Assumptions

1) A probability of α is assigned to the possibility of a TH occurrence on any given day. The complementary situation, non-occurrence of a TH, has the complementary probability, $(1-\alpha)$.

2) Pr (TH occurs in a unit of time | 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

TABLE 1a. Relative frequency of days that experienced at least one thunderstorm event at Cape Kennedy, Fla.

Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Spring	Summer	Fall
0.018	0.048	0.097	0.094	0.220	0.433	0.481	0.457	0.309	0.088	0.027	0.021	0.137	0.458	0.141

TABLE 2. Frequencies of the observed number of days that experience x TH's at Cape Kennedy, Fla., for the 11-year period of record January 1957 through December 1967.

x	June	July	August	Summer
	Month			
0	293	305	300	898
1	27	24	30	81
2	5	6	7	18
3	3	3	2	8
4 or more	2	3	2	7
Total	330	341	341	1012

Pr (TH occurs in a unit of time | a TH in the preceding $h-1$ units of time) = 0.

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

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 model:

$$\left.\begin{aligned} &\Pr(x=0)=(1-\alpha)+\alpha q^T, \quad q=1-p \\ &\Pr(x=i)=\alpha\left[p^i q^{T-ih}\binom{T-ih+i-1}{i}+p^i q^{T-ih}\times\sum_{m=1}^{h-1}\binom{T-(i-1)h+i-m-2}{i-1}q^{h-m}\right] \quad \text{for } 0<i<n \\ &\Pr(x=n)=1-\Pr(x<n) \end{aligned}\right\}. \quad (5)$$

Similar modifications have previously been made on the binomial and Poisson models by Singh (1963, 1968) and Neyman (1949a). The parallel in reasoning is best demonstrated by comparing our assumptions with those of Singh (1968) in his Poisson birth model. Singh assumes cohabitations are Poisson distributed and proceeds to show that resulting conceptions are modified-Poisson, where the value of h was 10 months. Both models have an underlying parent distribution and require special and identical modifications to realize a specific outcome resulting from the activity of the basic population.

3. Estimation

Numerous estimators for the parameters of the negative binomial distribution have been proposed. We have chosen to use the first two-moment method proposed by Cohen (1965). The negative binomial density function given by (4) may be written in terms of the gamma function as

$$\Pr(x=i)=\frac{\Gamma(x+k)}{x!\Gamma(k)}p^k q^x, \quad i\geqslant 0, \quad k>0, \quad 0\leqslant p\leqslant 1. \quad (6)$$

The distribution function is given by

$$F(x)=\sum_{x=0}^{n}\frac{\Gamma(x+k)}{x!\Gamma(k)}p^k q^x, \quad (7)$$

which gives the probability of obtaining a value of x less than or equal to some particular value of x, say x_0.

Now, after some algebraic manipulation of Cohen's estimators, we have for the moment estimators of the parameters k and p

$$k^*=\frac{\bar{x}^2}{s^2-\bar{x}}, \quad p^*=\frac{k}{k+\bar{x}}, \quad (8)$$

where $\bar{x}$ is the sample mean and s^2 the sample variance.

The mean M of the negative binomial distribution is given by $M=kq/p$, and the variance V by $V=kq/p^2$.

The efficiency of estimating p and k by the method of moments is derived by Fisher (1950). In terms of the parameters used here, the reciprocal of the efficiency is given by

$$\frac{1}{E}=1+2\left[\frac{1}{3}q\frac{2}{(k+2)}+\frac{1}{4}q^2\frac{2\cdot 3}{(k+2)(k+3)}+\frac{1}{5}q^3\frac{2\cdot 3\cdot 4}{(k+2)(k+3)(k+4)}+\cdots\right].$$

TABLE 2a. Relative frequency of days that experienced at least one TH at Cape Kennedy, Fla.

June	July	August	Summer
0.112	0.106	0.121	0.113

TABLE 3. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for March.

x	f_0	$r.f.$	F_0	f_e	$F(x)$	Conditional probability i 1	2	3	4
0	308	0.902	0.902	305.4	0.896				
1	20	0.059	0.961	25.5	0.970	1			
2	9	0.026	0.987	6.7	0.990	0.277	1		
3	3	0.009	0.996	2.2	0.996	0.078	0.281	1	
4	1	0.003	1.000	0.8	0.999	0.016	0.059	0.211	1

$\bar{x}=0.150$, $s^2=0.268$, $k^*=0.189$, $p^*=0.558$, $n=341$

TABLE 4. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for April.

x	f_0	$r.f.$	F_0	f_e	$F(x)$	Conditional probability i 1	2	3
0	299	0.906	0.906	295.9	0.897			
1	18	0.055	0.961	25.3	0.973	1		
2	10	0.030	0.991	6.1	0.992	0.226	1	
3	3	0.009	1.000	1.8	0.997	0.042	0.186	1

$\bar{x}=0.142$, $s^2=0.237$, $k^*=0.214$, $p^*=0.600$, $n=330$

TABLE 5. Negative binomial distribution for thunderstorm events at Cape Kenedy, Fla., for May.

						Conditional probability i					
x	f_0	r.f.	F_0	f_e	$F(x)$	1	2	3	4	5	6
0	266	0.779	0.780	262.6	0.770						
1	43	0.126	0.906	52.4	0.924	1					
2	25	0.073	0.979	16.6	0.972	0.339	1				
3	3	0.009	0.988	5.9	0.989	0.120	0.354	1			
4	3	0.009	0.997	2.2	0.996	0.041	0.122	0.345	1		
5	0	0.000	0.997	0.9	0.998	0.013	0.037	0.106	0.306	1	
6	1	0.003	1.000	0.3	0.999	0.003	0.007	0.021	0.061	0.200	1

$\bar{x}=0.352,\ s^2=0.621,\ k^*=0.460,\ p^*=0.567,\ n=341$

The modified negative binomial distribution 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

$$\chi^2=\sum_{i=0}^{n}\frac{[N_i-NP_i(\alpha,p)]^2}{N_i}. \tag{9}$$

See Neyman (1949b), Kendall and Stuart, Vol. II, 91–93 (1963) and Singh (1963) for a fuller explanation of BAN (best asymptotically normal) estimators. Neyman (1949b) has shown that the class of MCS estimators are also BAN estimators. These estimators are consistent, asymptotically normal, and asymptotically efficient. Let $P_i(\alpha,p)$ be the probability for $i(i=0,\ 1,\cdots,\ n)$ TH's per day, and satisfying the regularity conditions in Neyman (1949b); and N_i is the observed frequency in the ith class, i.e.,

$$N=\sum_{i=1}^{n}N_i.$$

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 chosen point $(\bar{\alpha},\bar{p})$ and use the linearized $P_i(\alpha,p)$'s instead of the orginal $P_i(\alpha,p)$'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 solutions $(\bar{\alpha},\bar{p})$ to the equations

$$N_0/N=P_0(\alpha,p),\quad N_1/N=P_1(\alpha,p), \tag{10}$$

provide consistent estimates. Letting $P_i'(\alpha,p)$ be the new linearized probabilities, we have

$$P_i'(\alpha,p)=P_i(\bar{\alpha},\bar{p})+(\alpha-\bar{\alpha})\frac{\partial P_i(\alpha,p)}{\partial\alpha}\bigg]_{(\bar{\alpha},\bar{p})}+(p-\bar{p})\frac{\partial P_i(\alpha,p)}{\partial p}\bigg]_{(\bar{\alpha},\bar{p})} \tag{11}$$

as the general equation. In particular, we have

$$P_0'(\alpha,p)=1-\alpha(1-\bar{q}^T)+(p-\bar{p})\bar{\alpha}T\bar{q}^{(T-1)}, \tag{12}$$

$$P_i'(\alpha,p)=\frac{\alpha}{\bar{\alpha}}P_i(\bar{\alpha},\bar{p})+(p-\bar{p})\bar{\alpha}\times\{\bar{p}^{(i-1)}\bar{q}^{(T-ih-1)}[i\bar{q}-(T-ih)\bar{p}]Q(i)+\sum_{m=1}^{h-1}\bar{p}^{(i-1)}\bar{q}^{(T-ih+h-m)}\times[i\bar{q}-(T-ih+h-m)\bar{p}]QS(i,m)\}, \tag{13}$$

where

$$Q(i)=\binom{T-ih+i-1}{i},\quad 0<i<n,$$

$$QS(i,m)=\binom{T-(i-1)h+i-m-2}{i-1},$$

and

$$P_n'(\alpha,p)=1-\sum_{i=0}^{n-1}P_i'(\alpha,p). \tag{14}$$

TABLE 6. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for June.

						Conditional probability i					
x	f_0	r.f.	F_0	f_e	$F(x)$	1	2	3	4	5	6
0	187	0.567	0.567	181.5	0.550						
1	77	0.233	0.800	87.7	0.816	1					
2	40	0.121	0.921	36.9	0.928	0.394	1				
3	17	0.052	0.973	14.7	0.972	0.147	0.373	1			
4	6	0.018	0.991	5.7	0.989	0.051	0.130	0.348	1		
5	2	0.006	0.997	2.2	0.996	0.016	0.040	0.106	0.304	1	
6	1	0.003	1.000	0.8	0.999	0.003	0.008	0.022	0.064	0.211	1

$\bar{x}=0.752,\ s^2=1.169,\ k^*=1.354,\ p^*=0.643,\ n=330$

TABLE 7. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for July.

x	f_0	$r.f.$	F_0	f_e	$F(x)$	Conditional probability i: 1	2	3	4	5
0	177	0.519	0.519	166.2	0.487					
1	80	0.234	0.753	99.4	0.779	1				
2	47	0.138	0.891	45.4	0.912	0.399	1			
3	26	0.076	0.967	18.6	0.967	0.143	0.357	1		
4	9	0.026	0.993	7.2	0.988	0.044	0.110	0.307	1	
5	2	0.006	1.000	2.7	0.996	0.009	0.023	0.066	0.214	1

$\bar{x}=0.874$, $s^2=1.277$, $k^*=1.893$, $p^*=0.684$, $n=341$

TABLE 9. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for September.

x	f_0	$r.f.$	F_0	f_e	$F(x)$	Conditional probability i: 1	2	3	4
0	228	0.691	0.691	219.2	0.664				
1	54	0.164	0.855	73.1	0.886	1			
2	33	0.100	0.955	24.8	0.961	0.316	1		
3	12	0.036	0.991	8.5	0.987	0.089	0.283	1	
4	3	0.009	1.000	2.9	0.995	0.018	0.057	0.203	1

$\bar{x}=0.509$, $s^2=0.777$, $k^*=0.967$, $p^*=0.655$, $n=330$

Replacing $P_i(\alpha,p)$ by $P_i'(\alpha,p)$ in (9), we obtain a modified form of χ^2:

$$(\chi^2)'=\sum_{i=0}^{n}\frac{[N_i-P_i'(\alpha,p)]^2}{N_i}, \tag{15}$$

which is minimized to give the MCS estimates α^* and p^*.

4. Data sample

According to standard United States weather observing procedure, a thunderstorm is reported whenever thunder is heard at the station. It is reported along with other atmospheric phenomena on the standard weather observer's form WBAN-10 when thunder is heard and ends 15 min after thunder is last heard. Notice that the standard definition of a thunderstorm may include multiple occurrences of thunderstorms. For this reason, we have chosen to use the term "thunderstorm event" as a more appropriate definition for our statistical analysis.

The type of statistical analysis presented is useful primarily for the planning of missions rather than for application to operations. Statistics may be useful up to a few days before a mission. However, at this time the weather forecaster's predictions should be more accurate, and the transition is made from statistical inference to weather forecasting dependent upon the synoptic situation prevailing a few days before the mission.

The data sample used was produced by the National Weather Records Center, Asheville, N. C., for the Terrestrial Environment Branch, Aerospace Environment Division, and is the latest and most comprehensive thunderstorm data available for Cape Kennedy. The period of record is January 1957 through December 1967.

Table 1 summarizes observed frequencies of days that experienced x thunderstorm events for all months, and for the spring, summer and fall seasons at Cape Kennedy. Table 1a gives the relative frequency of occurrence of days that experienced at least one thunderstorm event at Cape Kennedy for the same reference periods.

Those occurrences which were classified as thunderstorm hits (TH's) from the data sample were such that 1) a thunderstorm was actually reported over head, or 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).

Only the summer months of June, July and August were selected for TH examination since there were not enough TH data in the sample during the remainder of the year.

The period is 24 hr and T is taken to be 48 units. The value of h is taken as 2 which means that, given a TH occurring, another cannot occur for 30 min. The data and relative frequencies are given in Tables 2 and 2a.

5. Analysis

Theoretical summaries of the months and seasons that experience significant thunderstorm events at Cape Kennedy are given in Tables 3–13. In all cases, the sample variance was significantly greater than the sample mean, indicating the negative binomial distribution as the appropriate model.

TABLE 8. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for August.

x	f_0	$r.f.$	F_0	f_e	$F(x)$	Conditional probability i: 1	2	3	4	5
0	185	0.542	0.542	180.2	0.528					
1	89	0.261	0.803	92.2	0.799	1				
2	30	0.088	0.891	40.5	0.918	0.399	1			
3	24	0.070	0.961	16.9	0.967	0.146	0.366	1		
4	10	0.029	0.990	6.8	0.987	0.046	0.116	0.316	1	
5	3	0.009	1.000	2.7	0.995	0.010	0.026	0.070	0.221	1

$\bar{x}=0.809$, $s^2=1.280$, $k^*=1.391$, $p^*=0.632$, $n=341$

TABLE 10. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for October.

x	f_0	$r.f.$	F_0	f_e	$F(x)$	Conditional probability i: 1	2	3
0	311	0.911	0.911	307.7	0.902			
1	17	0.050	0.961	24.2	0.973	1		
2	9	0.026	0.987	6.1	0.991	0.235	1	
3	4	0.012	1.000	1.9	0.997	0.045	0.192	1

$\bar{x}=0.138$, $s^2=0.242$, $k^*=0.182$, $p^*=0.570$, $n=341$

TABLE 11. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for spring (March, April, May).

						Conditional probability i					
x	f_0	$r.f.$	F_0	f_e	$F(x)$	1	2	3	4	5	6
0	873	0.863	0.863	863.6	0.853						
1	81	0.080	0.943	103.7	0.956	1					
2	44	0.043	0.986	29.2	0.985	0.312	1				
3	9	0.009	0.995	9.8	0.994	0.106	0.339	1			
4	4	0.004	0.999	3.5	0.998	0.035	0.113	0.335	1		
5	0	0.000	0.999	1.3	0.999	0.011	0.034	0.101	0.303	1	
6	1	0.001	1.000	0.5	1.000	0.002	0.007	0.022	0.066	0.217	1

$\bar{x}=0.215$, $s^2=0.386$, $k^*=0.271$, $p^*=0.557$, $n=1012$

The notation used in Tables 3–13 is as follows:

- x number of thunderstorm events per day
- f_0 observed number of days during the 11-year period of record that experienced x thunderstorm events
- $r.f.$ relative frequency of occurrence of x thunderstorm events
- F_0 observed distribution function
- f_e expected frequencies using the negative binomial distribution
- $F(x)$ negative binomial distribution function
- $\bar{x}$ sample mean
- s^2 sample variance
- k^*,p^* parameter estimators of the negative binomial distribution
- n sample size.

Conditional probabilities are also included in the tables.

Consider the month of June (Table 6) as an example. There were 40 out of 330 days (11 years of Junes) that had exactly two thunderstorm events. This gives a relative frequency (probability) of occurrence of 0.121 of having exactly two thunderstorm events during any day in June. The observed distribution function F_0 gives a probability of 0.921 of having two or less thunderstorm events during any day in June, or a probability of (1−0.921) or 0.079 of having more than two thunderstorm events during any day in June. The negative binomial distribution predicts that 36.9 days in June will experience exactly two thunderstorm events with the probability $F(x)=0.928$ of having two or less thunderstorm events during any day in June, or a probability of (1−0.928) or 0.072 of having more than two thunderstorm events during any day in June. The agreement between theory and observation is very good for this example.

The conditional probabilities are computed from the theoretical frequencies (f_e) by using a double summation technique. The following tabulation is an example of this technique using the month of June (see Table 6).

				Conditional probabilities i					
x 0	181.5	Σ	$\Sigma\Sigma$	1	2	3	4	5	6
1	87.7	148.0	244.2	$\frac{244.2}{244.2}=1$					
2	36.9	60.3	96.2	$\frac{96.2}{244.2}=0.394$	$\frac{96.2}{96.2}=1$				
3	14.7	23.4	35.9	$\frac{35.9}{244.2}=0.147$	$\frac{35.9}{96.2}=0.373$	$\frac{35.9}{35.9}=1$			
4	5.7	8.7	12.5	$\frac{12.5}{244.2}=0.051$	$\frac{12.5}{96.2}=0.130$	$\frac{12.5}{35.9}=0.348$	$\frac{12.5}{12.5}=1$		
5	2.2	3.0	3.8	$\frac{3.8}{244.2}=0.016$	$\frac{3.8}{96.2}=0.040$	$\frac{3.8}{35.9}=0.106$	$\frac{3.8}{12.5}=0.304$	$\frac{3.8}{3.8}=1$	
6	0.8	0.8	0.8	$\frac{0.8}{244.2}=0.003$	$\frac{0.8}{96.2}=0.008$	$\frac{0.8}{35.9}=0.022$	$\frac{0.8}{12.5}=0.064$	$\frac{0.8}{3.8}=0.211$	$\frac{0.8}{0.8}=1$

Each element in the second summation ($\Sigma\Sigma$) is divided by the appropriate top element in each column as indicated in order to obtain the conditional probabilities; i.e., in each column under conditional probabilities, given i thunderstorm events ($i=1, 2, 3, \cdots$), the probability of having j additional thunderstorm events ($j=0, 1, 2, \cdots$) is given by

$$P(i+j|i)=\frac{(i+j)\text{th element}}{(i)\text{th element}}, \tag{16}$$

where (16) refers only to the previous table. For example

TABLE 12. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for summer (June, July, August).

						Conditional probability i					
x	f_0	$r.f.$	F_0	f_e	$F(x)$	1	2	3	4	5	6
0	549	0.542	0.542	527.8	0.522						
1	242	0.243	0.785	279.6	0.798	1					
2	117	0.116	0.901	122.7	0.919	0.404	1				
3	67	0.066	0.967	50.1	0.969	0.153	0.379	1			
4	25	0.025	0.992	19.7	0.988	0.054	0.133	0.352	1		
5	7	0.007	0.999	7.6	0.995	0.017	0.041	0.108	0.307	1	
6	1	0.001	1.000	2.9	0.998	0.004	0.009	0.023	0.067	0.216	1

$\bar{x}=0.812$, $s^2=1.245$, $k^*=1.523$, $p^*=0.652$, $n=1012$

for $i=2$, given two thunderstorm events on any day in June, what is the probability of having two additional thunderstorm events ($j=2$) on that day in June? From (16) we have

$$P(4|2)=\frac{\text{4th element}}{\text{2nd element}}=\frac{12.5}{96.2}=0.130.$$

Also, given four thunderstorm events on any day in June ($i=4$), the probability of having one additional thunderstorm event ($j=1$) on that same day in June is 0.304.

Using the modified negative binomial distribution as a model for TH activity, we see from Table 14 that the model fits particularly well in the 0, 1 and 2 classes and not so well in the 3 and 4 classes. The notable exception is the August data.

Notations used in Table 14 are as follows:

- x number of TH's per day
- f_0 observed number of days during the 11-year period of record that experienced x TH's
- $r.f.$ relative frequency of occurrence of x TH's
- F_0 observed distribution function
- f_e expected frequencies using the modified negative binomial distribution
- $F(x)$ modified negative binomial distribution function
- α^*, p^* parameter estimators of the modified negative binomial distribution
- n sample size.

6. Conclusions

There are many advantages in the use of a theoretical statistical model for predicting variables such as thunderstorm events or thunderstorm hits. Once sufficient representative samples have been collected and analyzed and the validity of the theory is established, the theoretical model becomes "deterministic" and may be applied universally to the variable under consideration. Another advantage of theory over empirical statistics is the use of the acceptable theoretical function for making probability inferences concerning values of the variable outside of the range of observation. It is often desired to make predictions relating to these "never observed" values, and the theoretical approach permits one to do so.

The physical properties necessary for the application of the negative binomial distribution have been shown to be present in our experiment concerning the number of thunderstorm events at Cape Kennedy. In all the samples considered, the sample variance exceeded the sample mean, suggesting that the negative binomial might yield a satisfactory representation. Our comparison with Yule's classic application of the negative binomial substantiates its validity to represent the number of thunderstorm events at Cape Kennedy. Both the negative binomial and Poisson distributions were tried as prospective models for thunderstorm events. The negative binomial gave a "better" fit in all cases than the Poisson distribution.

Comparisons were also made using the same modification on the binomial and Poisson distributions as was used here on the negative binomial distribution to represent the variation in the number of thunderstorm hits per day at a launch site. The modified negative binomial distribution gave a "better" fit in all cases.

The χ^2 "goodness of fit" test was applied to all cases where data were sufficiently large to admit its use. For both the negative binomial and the modified negative binomial distributions, the computed χ^2 values were a minimum relative to those of the other distributions mentioned above.

In summary, the negative binomial distribution and a modification of it have been studied as provisional representations of the frequency distributions by

TABLE 13. Negative binomial distribution for thunderstorm events at Cape Kennedy, Fla., for fall (September, October, November).

						Conditional probability i			
x	f_0	$r.f.$	F_0	f_e	$F(x)$	1	2	3	4
0	860	0.859	0.859	845.2	0.844				
1	77	0.077	0.936	109.5	0.954	1			
2	45	0.045	0.981	30.6	0.984	0.286	1		
3	16	0.016	0.997	10.1	0.994	0.080	0.281	1	
4	3	0.003	1.000	3.6	0.998	0.017	0.058	0.208	1

$\bar{x}=0.227$, $s^2=0.397$, $k^*=0.302$, $p^*=0.571$, $n=1001$

TABLE 14. June, July, August and summer-modified negative binomial distribution for TH's at Cape Kennedy, Fla.

June						July					
x	f_0	$r.f.$	F_0	f_e	$F(x)$	x	f_0	$r.f.$	F_0	f_e	$F(x)$
0	293	0.888	0.888	295.18	0.894	0	305	0.894	0.894	306.99	0.900
1	27	0.082	0.970	26.21	0.973	1	24	0.070	0.964	23.05	0.968
2	5	0.015	0.985	6.54	0.993	2	6	0.018	0.982	7.97	0.991
3	3	0.009	0.994	1.02	0.996	3	3	0.009	0.991	1.71	0.996
4 or more	2	0.006	1.000	1.05	0.999	4 or more	3	0.009	1.000	1.28	1.000
$\alpha^*=0.25467$, $p^*=0.01108$, $n=330$						$\alpha^*=0.19135$, $p^*=0.01523$, $n=341$					

August						Summer					
x	f_0	$r.f.$	F_0	f_e	$F(x)$	x	f_0	$r.f.$	F_0	f_e	$F(x)$
0	300	0.879	0.879	300.55	0.881	0	898	0.887	0.887	902.38	0.892
1	30	0.088	0.967	29.66	0.968	1	81	0.080	0.967	79.00	0.970
2	7	0.021	0.988	8.12	0.992	2	18	0.018	0.985	22.88	0.993
3	2	0.006	0.994	1.38	0.996	3	8	0.008	0.993	4.12	0.997
4 or more	2	0.006	1.000	1.29	1.000	4 or more	7	0.007	1.000	3.61	1.000
$\alpha^*=0.26740$, $p^*=0.01214$, $n=341$						$\alpha^*=0.23461$, $p^*=0.01282$, $n=1012$					

months, of thunderstorms and thunderstorm hits, respectively, at Cape Kennedy, Fla. The method of moments and minimum chi-square were used for parameter estimation.

Examination of the results (Tables 3–14) shows deviations of the data from the fitted negative binomial distributions with systematic (statistically significant) patterns of signs. For example, considering the eight separate monthly distributions (Tables 3–10) as independent, and looking only at the signs of departures in the first class, we see that all eight signs are positive. In the second class the signs are all negative except one (August). In the remaining classes there are patterns of signs of departure although not as pronounced as in the first two categories.

These patterns of deviations of the data from the fitted negative binomial distributions indicate that this distribution may not be satisfactory as a theoretical model, although the general correspondence of shape might be good enough for some practical purposes. The modified negative binomial improved the fit, but the signs of the departures still showed some regularity of pattern.

Using statistical methods, it has been demonstrated that the negative binomial and the modified negative binomial distributions may be used as provisional statistical models to represent the variation in thunderstorm events and thunderstorm hits at Cape Kennedy, Fla.

Acknowledgments. The authors wish to acknowledge the assistance of Mr. O. E. Smith, Chief, Terrestrial Environment Branch, Aerospace Environment Division, Aero-Astrodynamics Laboratory, Marshall Space Flight Center, for his contribution of the double summation technique for computing conditional probabilities, and for proposing the method used by Singh for the modification of the negative binomial distribution to develop the "thunderstorm hit" model.

REFERENCES

Cohen, A. C., 1965: Estimation in the negative binomial distribution. NASA Tech. Memo. TM X-53372, Marshall Space Flight Center, Alabama.

——, 1966: A note on certain discrete mixed distributions. *Biometrics*, **22**, 566–572.

Fisher, R. A., 1950: *Contributions to Mathematical Statistics*. New York, Wiley, 38.182–38.187.

Kendall, M. G., and A. Stuart, 1963: *The Advanced Theory of Statistics*, Vols. I and II. New York, Hafner Publ. Co., 1109 pp.

Neyman, J., 1939: On a new class of "contagious" distributions, applicable in entomology and bacteriology. *Ann. Math. Statist.*, **10**, 35–57.

——, 1949a: On problems of estimating the number of schools of fish. *Univ. Calif. Publ. Statist.*, **1**, 21–36.

——, 1949b: Contribution to the theory of χ^2 test. *Proceedings Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 239–273.

Singh, S. N., 1963: Probability models for the variation in the number of births per couple. *J. Amer. Statist. Assoc.*, **58**, 721–727.

——, 1968: A chance mechanism of the variation in the number of births per couple *J. Amer. Statist. Assoc.*, **63**, 209–213.

Thom, H. C. S., 1957: The frequency of hail occurrence. *Arch. Meteor., Geophys. Bioklim.*, **B8**, No. 2, 185–194.

——, 1958: A method for the evaluation of hail suppression. *Z. Angew. Math. Phys.*, **9**, 37–64.

Yule, G. U., 1910: On the distribution of deaths with age when the causes of death act cumulatively, and similar frequency distributions. *J. Roy. Statist. Soc.*, **73**, 26–35.