NASA CONTRACTOR
REPORT

NASA CR-129008

AN EMPIRICAL ANALYSIS OF THE DISTRIBUTION OF OVERSHOOTS IN A STATIONARY GAUSSIAN STOCHASTIC PROCESS

By Michael C. Carter and Michael W. Madison
Department of Mathematical Sciences
Appalachian State University
Boone, North Carolina 28607

July 1973

Prepared for

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

CHAPTER 1

INTRODUCTION

The purpose of this thesis is to report the results of an empirical analysis of the frequency of overshoots above an arbitrary level in a stationary gaussian stochastic process. The problem is of interest to the Terrestrial Environment Branch, Aerospace Environment Division, Aero-Astrodynamics Laboratory, George C. Marshall Space Flight Center, Alabama, and the financial support for the project was under NASA contract no. NAS8-29286. The results obtained in this analysis are applicable in the prediction of extreme properties of processes such as wind speed, ambient temperature and sea state. The methodology may also be used in other fields, i.e., electrical engineering and aerospace vehicle responses to forcing functions having known or assumed exponential autocorrelation functions.

The remainder of this chapter presents a general discussion of the scope of the work, and the organization of this analysis.

1.1 Statement of the Problem

The problem dealt with herein concerns the frequency distribution of overshoots in a stationary gaussian stochastic process with an exponential autocorrelation function. Briefly, a stationary gaussian process may be described as a stochastic process which, at any point in time, has a gaussian distribution. To be stationary the process must have a mean independent of time and an autocorrelation function dependent only on the distance between successive time points.

The problem has been of general theoretical interest for some time while little has been done to obtain numerical results. Previous work in this general area is excellently summarized in two recent texts, Cramer' and Leadbetter (1967) and Kuznetsov (1965), and both contain extensive bibliographies. The general density function for the number of crossings in a $(0,\tau)$ time interval was given by Kuznetsov and Stratonevich (1956). For a stationary gaussian process with $R(\tau) = \exp(-\beta\tau^2)$ Tikhonev (1956) approximated the probability of zero crossings in $(0,\tau)$ by expanding the proof given by Kuznetsov and Stratonevich (1956) and neglecting terms in the series of order greater than 2. Other authors have various expressions for this density function and have investigated its asymptotic behavior. A general result states that as the level increases the number of overshoots in $(0,\tau)$ is Poisson distributed. A more extensive summary of previous work in this area is presented in Appendix II. To the author's knowledge this is the first investigation conducted by extensive simulation of such a process.

1.2 Organization of the Analysis

Chapter 2 is a discussion of the simulation model and assumptions concerning the model. Using the methods developed in Chapter 2, several simulations were run on an IBM-1130 computer. The results and analysis of the simulations are presented in Chapter 3 along with the resultant distribution equations.

The modus operandi for NASA to apply this solution to their specific problems concerning atmospheric variables is presented in the concluding chapter of this analysis. Appendix I contains a computer program to utilize the algorithm obtained in this investigation.

CHAPTER 2

Model and Simulation

The first step in this development of a solution to the overshoot problem was to define a mathematical model of a stationary gaussian stochastic process with an exponential autocorrelation function. In developing the model, the following conditions were assumed:

1) The sample process had a multivariate normal distribution.
2) The process was strictly stationary, i.e., the autocorrelation function $R(t_i,t_j) = R(\tau)$ where $\tau = |t_j-t_i|$.
3) The expected value of a random variable X at time t was 0, i.e., $E(X(t)) = 0$ where E denotes the expectation operator.
4) The covariance matrix, denoted Σ, was symmetrical and positive definite.
5) The autocorrelation function, denoted $R(\tau)$, was exponential in nature, i.e., $R(\tau) = EXP(-\beta|\tau|)$.

The notation $X(t)$ will denote a stochastic process satisfying the above conditions.

The process was considered over a time interval [0,99] and a sample realization consisted of 100 equally spaced sample points in the interval. This permitted some generality in the analysis whereas for a specific application the range of interest would be some [0,T] interval. In this case $X(t)$

would be sampled at $t_0, t_1, \ldots, t_{99}$ where $t_i = (\frac{i}{100})T$, with a corresponding modification of the autocorrelation parameter β. The method of simulation was given by Odell (1971) and a summary of that technique is presented in the following discussion.

Let $\underline{X} = (X(t_0), X(t_1), \ldots, X(t_{99}))'$ ($'$ denotes matrix transposition), then the covariance matrix is given by $\Sigma = (\sigma_{ij}) = E(\underline{X} \cdot \underline{X}')$ so that for $0 \leq i, j \leq 99$

$$\sigma_{ij} = E(X(t_i)X(t_j)) = R(t_i, t_j) = R(\tau) \quad \text{where} \quad \tau = |t_i - t_j|.$$

Thus it follows that Σ is formed by evaluating $R(\tau)$ for $0 \leq \tau \leq 99$ giving

$$\Sigma = \begin{pmatrix} R(0) & R(1) & R(2) & \cdots & R(99) \\ R(1) & R(0) & R(1) & \cdots & R(98) \\ R(2) & R(1) & R(0) & \cdots & R(97) \\ \vdots & \vdots & \vdots & & \vdots \\ R(99) & R(98) & & & R(0) \end{pmatrix}.$$

By assumption $\underline{X}$ satisfies a multivariate normal distribution with mean $\underline{\mu} = \underline{0}$ and covariance matrix Σ, denoted $\underline{X} \sim N(\underline{\mu}, \Sigma)$. The following result, given by Odell (1971, pg. 37), provides the modus operandi of generating realization of $X(t)$.

Theorem: If the 100×1 vector $\underline{Y} \sim N(\underline{\mu}, \Sigma)$, and $\underline{\gamma}$ is a fixed 100×1 vector, then $\underline{V} = A\underline{Y} + \underline{\gamma}$ is distributed $N(A\underline{\mu} + \underline{\gamma}, A \Sigma A')$.

We generated a vector $\underline{Y} \sim N(\underline{0},I)$, I denoting the identity matrix, and obtained a factorization AA' of Σ, therefore, by the above theorem $\underline{V} = A\underline{Y}$ was distributed $N(\underline{0},\Sigma)$. The resultant vector $\underline{V}$ constituted a realization of $X(t)$. The generation of the vector $(y_0,y_1,\ldots,y_{99})' = \underline{Y} \sim N(\underline{0},I)$ was accomplished by generating a sequence of 100 independent standard normal variates. The Crout method was used to factor Σ into AA'.

The technique of generating $\underline{Y}$ was given by Hamming (1962). He notes that an approximation to normally distributed random numbers can be produced from a sequence of uniformly distributed random numbers by the formula

$$y_i = \frac{\sum_{k=1}^{K} x_k - \frac{K}{2}}{\sqrt{K/12}}$$

where x_k is a uniformly distributed random number in $(0,1)$, and K is the number of values of x_k used. According to the Central Limit Theorem, as K tends to infinity the value of y_i approaches a standard normal distribution. To implement this procedure on a computer we fixed the value of K at 12. The formula for y_i could then be expressed as $y_i = \sum_{k=1}^{12} x_k - 6$. This construction of y_i for $0 \leq i \leq 99$ produced a sequence $y_0,y_1,\ldots,y_{99}$ of standard normal variates with mean 0 and unit variance. This sequence is the vector $\underline{Y} \sim N(\underline{0},I)$.

We generated 250 realizations $\underline{V}_i$, $i = 1,2,\ldots,250$, for each of the autocorrelation functions simulated. This required 250 random vectors $\underline{Y}_i$, $i = 1,2,\ldots,250$ which in turn, required a sequence of $250 * 100 = 25{,}000$ standard normal variates, or $250 * (100 * 12) = 300{,}000$ uniformly distributed random numbers. The algorithm used to generate uniformly distributed random variates was:

$$r_n = \text{normalized } (S_n) \quad \text{where} \quad S_n = \lambda r_{n-1}$$

and the normalization is a reduction to (0,1).

This algorithm is the well established power residual method of generating pseudo-random sequences. The period of the sequence generated in this fashion is a function of the integer capacity of the computer being used for the generation. In the case of the IBM-1130, the largest integer, and hence the period of the sequence, was 32,767 which falls far short of the necessary 300,000.

Since the period of one number generator is too short to produce 250 realizations, we used a separate random number generator G_j, $j = 0, 1, \ldots, 99$, for each of the 100 elements of $Y_i = (y_{i,0}, y_{i,1}, \ldots, y_{i,99})$ $1 \leq i \leq 250$. Thus the generator G_j, $0 \leq j \leq 99$, produced the sequence $y_{1,j}, y_{2,j}, \ldots, y_{250,j}$ of independent standard normal variates. In this fashion, each generator G_j was required to produce $250 * 12 = 3{,}000$ uniformly distributed numbers, which is easily possible on the computer used in this analysis.

To transform each vector $\underline{Y}_i$ into a realization $\underline{V}_i$ of the process $X(t)$ via the linear transformation $\underline{V}_i = A\underline{Y}_i$, it was necessary to factor the variance covariance matrix Σ.

As noted previously, Σ is a symmetrical, positive definite matrix. A well known theorem in matrix theory states that such a matrix can be factored into the product of a lower triangular matrix and its transpose. This factorization, $\Sigma = \Lambda\Lambda'$, where Λ is lower triangular, was accomplished using the Crout factorization technique as presented by Odell (1971, pg. 38). The method is summarized in the following discussion.

The elements of $\Lambda = (a_{ij})$ will be computed in the following sequence: $a_{11}, a_{21}, a_{31}, \ldots, a_{100,1}, a_{22}, a_{32}, \ldots, a_{100,2}, \ldots, a_{99,99}, a_{100,99}, a_{100,100}$. Note that Λ is lower triangular so $a_{ij} = 0$ whenever $j > i$. Using this fact we have

$$\sigma_{ij} = \sum_{k=1}^{j} a_{ik}a_{jk} \tag{2.2}$$

from which the following algorithms were derived. For $i = j = 1$ we have $\sigma_{11} = a_{11}^2$ so it follows that

$$a_{11} = (\sigma_{11})^{1/2}. \tag{2.3}$$

For $i > j = 1$ we have $\sigma_{ij} = a_{i1}a_{11}$ so the remaining elements of the first column of Λ are given by

$$a_{i1} = \sigma_{i1}/a_{11}. \tag{2.4}$$

After $j-1$ columns of Λ have been generated we have $\sigma_{jj} = \sum_{k=1}^{j} a_{jk}a_{jk} = \sum_{k=1}^{j-1} a_{jk}^2 + a_{jj}^2$ so for the remaining diagonal elements we have

$$a_{jj} = (\sigma_{jj} - \sum_{k=1}^{j-1} a_{jk}^2)^{1/2}. \tag{2.5}$$

For the remaining elements we have $\sigma_{ij} = \sum_{k=1}^{j} a_{ik}a_{jk} = \sum_{k=1}^{j-1} a_{ik}a_{jk} + a_{ij}a_{jj}$ so we can conclude

$$a_{ij} = (\sigma_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk})/a_{jj} \quad \text{for} \tag{2.6}$$

$i = j + 1, j + 2, \ldots, 100.$

The autocorrelation function $R(\tau) = \exp(-\beta|\tau|)$ determines the degree of association between successive values of $X(t)$. The process $X(t)$ was simulated for a range of β values yielding processes where the correlation was above .98 throughout the process, to processes where $X(t)$ values could be considered independent after two time intervals. The minimum β value used was .002 which yielded $R(99) = .9802$, and the maximum β value was 5.0 which yielded $R(2) = .00004539$. The primary β values utilized were .002, .005, .0075, .01, .025, .05, .075, .1, .25, .5, .75, 1.0, 1.5, 2.0, 3.0, and 5.0. We did, however simulate processes which were outside our primary range of interest, namely 7.5 and 10.0. At each of these β values 250 realizations were generated. The selection of 250 as the number of realizations for each β value was based on available computer storage capabilities but, from a statistical viewpoint, was deemed adequate for subsequent estimation and inference activities.

CHAPTER 3

Simulation Results and Analysis

Once the "data sets" had been generated the basic problem of counting overshoots came into focus. Letting A denote some arbitrary level, we counted the number of overshoots above values $A = .5, .75, 1.0. 1.25, 1.5, 1.75$, and 2.0. Since each realization has mean $\underline{0}$ and unit variance, this was equivalent to counting the number of overshoots over .5 standard deviations above the mean, .75 standard deviations above the mean, etc. In future applications the overshoots above a value of, say $A = .75$, would be equivalent to overshoots above a value of $.75\sigma + \mu$, where the process has mean μ and variance σ^2.

The value of 2.0 was selected as the upper limit of the major range of interest since, in the completely independent case, only 2.27% of the values would be above 2.0 and in the more correlated cases, the number of points, and hence the number of overshoots, would likely decrease. The value of .5 was selected as the lower limit of the range of A values. In the completely independent case 30.85% of the values lie above .5, but the more memory the system has the longer the duration of each overshoot, and hence the fewer the number of overshoots. We did, however, count overshoots above higher levels for the purpose of determining the integrity of the estimation model outside the primary range of interest. Specifically, overshoots were counted for A levels of 2.25, 2.5, 2.75, 3.0, 3.5 and 4.0 for β values of .005, .02, .05, .1, .5, 1.5 and 3.0.

To count the number of overshoots above level A, we counted the number of times $V(t_{i-1}) \leq A$ while $V(t_i) > A$ where $0 \leq i \leq 99$ and $V(t_0) = 0$.

After the number of overshoots for a particular level A and autocorrelation parameter β was determined, the sample mean, $\bar{X}$, and variance, S^2, were computed in the traditional fashion. This provided the data to complete the table of means for A and β, Table 1, and the table of variances for A and β, Table 2.

TABLE 1

MEANS FOR LEVELS OF A AND β

β Value	A Level: .5	.75	1.0	1.25	1.5	1.75	2.0	2.25	2.50	2.75	3.0	3.5	4.0
.002	1.016	.863	.724	.554	.364	.264	.120						
.005	1.524	1.258	.931	.744	.524	.452	.252	.226	.152	.099	.063	.023	.007
.0075	2.032	1.712	1.274	1.056	.752	.480	.380						
.01	2.488	2.004	1.512	1.097	.692	.472	.324	.185	.109	.062	.034	.009	.002
.025	3.368	2.752	2.240	1.980	1.374	.868	.592						
.05	4.357	3.648	2.888	2.228	1.736	1.320	.824	.525	.321	.187	.105	.029	.006
.075	5.816	4.936	3.788	2.924	2.028	1.325	.831						
.1	6.368	5.368	4.304	3.420	2.404	1.540	1.060	.623	.359	.196	.102	.023	.004
.25	9.884	8.320	6.632	4.892	3.332	2.140	1.320						
.5	13.168	11.008	8.688	6.628	4.608	2.632	1.708	.908	.472	.228	.103	.017	.002
.75	15.160	12.768	10.024	7.468	5.204	3.304	1.900						
1.0	16.9[illegible]	14.216	11.172	8.204	5.660	3.540	2.076						
1.5	18.708	15.828	12.260	8.956	6.052	3.768	2.176		.571		.108	.015	.002
2.0	20.048	16.676	12.972	9.360	6.188	3.744	2.140						
3.0	21.260	17.672	13.572	9.588	6.368	4.144	2.388		.682		.148	.025	.003
5.0	21.924	18.172	13.824	9.644	6.452	3.972	2.200						

TABLE 2

VARIANCES FOR LEVELS OF A AND β

β Value	A Level .5	.75	1.0	1.25	1.5	1.75	2.0	2.25	2.5	2.75	3.0	3.5	4.0
.002	2.875	2.669	2.168	1.716	1.076	1.030	.323						
.005	5.335	3.893	3.068	2.890	2.001	1.799	1.217	.557	.377	.268	.196	.109	.056
.0075	6.047	5.202	3.665	3.916	2.412	1.600	1.642						
.01	6.299	5.136	3.862	3.189	2.535	1.905	1.167	.741	.516	.379	.289	.182	.122
.025	6.209	6.147	6.207	5.629	4.096	2.444	1.769						
.05	6.239	5.627	5.184	4.442	4.091	3.576	1.832	1.047	.742	.555	.433	.286	.203
.075	6.713	6.228	5.670	5.412	4.116	2.849	1.647						
.1	6.531	6.458	6.734	6.020	4.314	3.093	1.880	1.154	.820	.616	.482	.320	.22[illegible]
.25	7.019	7.801	7.141	5.888	4.279	3.020	1.744						
.5	7.602	7.783	7.115	5.921	4.207	2.839	1.950	1.369	.976	.737	.579	.388	.281
.75	6.721	6.275	7.100	5.808	4.886	3.875	2.307						
1.0	6.495	8.339	7.838	6.918	5.125	3.775	2.135						
1.5	6.320	8.151	7.896	7.191	5.013	2.878	1.423		1.070		.636	.428	.311
2.0	7.074	7.143	7.047	6.609	4.450	3.733	2.169						
3.0	7.856	8.783	8.262	7.352	4.788	3.489	2.246		1.125		.670	.452	.329
5.0	7.829	8.577	8.330	7.025	5.646	3.923	2.096						

As noted in the introductory chapter, the Poisson distribution is the limiting distribution as the crossing level becomes large and it seemed reasonable to first try the Poisson as a model for lower crossing levels. An estimation model for the multivariate Poisson (multivariate in the sense that the parameter λ was assumed to be a function of A and β) was implemented and tried for various functions of A and β. The results were discouraging. We first attributed the failure to our inability to find the proper function of A and β, but later it was determined that the Poisson model was, in general, inadequate.

The next and most fruitful step was the careful examination of the means and variances for various levels of A and β. This led immediately to the following conclusions:

1) There was a strong empirical relationship between the sample means and A and β, and to a lesser extent, between the sample variances and A and β.

2) The binomial and negative binomial distributions, with parameters calculated from the sample means and variances, were more appropriate for the levels of A we investigated.

For values of $A \leq 1.5$ and $\beta \leq 1.0$ the means exceeded the variances with the discrepancy increasing as A and β decreased. As A and β increased above 1.5 and 1.0 respectively the values became approximately equal or the variances exceeded the mean. Once this trend was noticed, the reasons for observations 1 and 2 above became clear. If we assume one of three models, binomial, Poisson, or negative binomial, is appropriate, then an accepted selection criterion is the relationship between the mean and variance. These observations led us to seek those functional relationships that could best predict a process mean and variance.

In the search for the relationship between A, β and the mean μ, we first graphed the sample mean $\overline{X}$ as a function of A for each β. This graph, Figure 1, strengthened the conclusion that such a relationship existed but, due to our inability to find an appropriate approximating function of that relationship, this method of viewing the data was abandoned. However, we did note from this plot that the relationship behaved in what appeared to be an exponential fashion.

Suspecting the exponential characteristic, the next step was to graph $\ln(\overline{X})$ as a function of A for each β on semi log graph paper. This plot, Figure 2, was not a straight line as we had anticipated, but rather it seemed parabolic with the parabolas opening about the $\ln(\overline{X})$ axis. We selected the general parabolic model

$$\ln(\overline{X}) = \lambda_0(\beta) + \lambda_1(\beta)A + \lambda_2(\beta)A^2 \qquad (3.1)$$

to try as an approximating relationship. The least squares technique summarized below was used to estimate $\lambda_0(\beta)$, $\lambda_1(\beta)$, and $\lambda_2(\beta)$ for each β. Using these results, equation (3.1) was then rewritten to produce the estimate of the mean as

$$\mathrm{EST}(\mu) = \exp(\lambda_0(\beta) + \lambda_1(\beta)A + \lambda_2(\beta)A^2). \qquad (3.2)$$

For each β the estimate of μ was a good approximation of $\overline{X}$ so we concluded that if the dependency of the λ's on β could be found, then (3.2) would provide a good estimate of the mean.

The least squares technique was given by Jorgenson (1961) and was used to estimate $\Lambda = (\lambda_0, \lambda_1, \ldots, \lambda_k)$ for the general model $n_i = \lambda_0 t_0 + \lambda_1 t_1 + \ldots + \lambda_k t_k$ for $i=1,\ldots,m$ (m being the number of observations). Let $T = (T_{ij})$ where $T_{ij} = t_j$ for $j=1,\ldots,k$, and $i=1,\ldots,m$. Then $\Lambda = (T'T)^{-1}T'\underline{n}$ where $\underline{n} = (n_1, n_2, \ldots, n_m)'$ and Λ is a $k \times 1$ vector of the estimates.

Using the results of the least squares method, the first step toward determining the dependence of λ_i on β was to plot $\ln(\beta)$ vs. λ_i on semi-log graph paper for each of λ_0, λ_1, and λ_2. For all three coefficients three distinct trends were observed. For $\beta \leq .01$ the relationship was linear, for $.01 < \beta \leq 1.5$ the relationship appeared quadratic, and for $\beta > 1.5$ the relationship was again linear and essentially horizontal. Accordingly, the following models were fit using the method of least squares:

$$
\begin{array}{lll}
\beta \leq .01 & \ln(\beta) = a_0 + a_1\lambda_i, & i=1,2,3 \\
.01 < \beta \leq 1.5 & \ln(\beta) = a_0 + a_1\lambda_i + a_2\lambda_i^2, & i=1,2,3 \\
\beta > 1.5 & \beta = a_0 + a_1\lambda_i, & i=1,2,3
\end{array}
\tag{3.3}
$$

The results of the least squares estimates of the a_i's are presented in Table 3.

TABLE 3

LEAST SQUARES ESTIMATES OF COEFFICIENTS OF (3.3)

		a_0	a_1	a_2
λ_0	$\beta \le .01$	-6.0717	1.1105	-
λ_0	$.01 < \beta \le 1.5$	-6.423	1.57686	.218[illegible]1
λ_0	$\beta > 1.5$	-236.0714	71.4286	-
λ_1	$\beta \le .01$	-6.65546	-3.23315	-
λ_1	$.01 < \beta \le 1.5$	-1.0582	10.31154	10.23657
λ_1	$\beta > 1.5$	-10.2162	-54.0541	-
λ_2	$\beta \le .01$	-7.81708	-10.2928	-
λ_2	$.01 < \beta \le 1.5$	-3.9142	3.382882	16.3656
λ_2	$\beta > 1.5$	23.5	50.0	-

Using the results given in Table 3 we then solved (3.3) for λ_0, λ_1, and λ_2 yielding estimation equations for λ_0, λ_1, λ_2 as

$$\begin{array}{llll}
\beta \le .01 & \lambda_i = (\ln(\beta) - a_0)/a_1 & i=1,2,3 & \\
.01 < \beta \le 1.5 & \lambda_i = (-a_1 + [a_1^2 - 4a_2(a_0 - \ln(\beta))]^{1/2})/2a_2 & i=1,2 & \\
 & \lambda_i = (-a_1 - [a_1^2 - 4a_2(a_0 - \ln(\beta))]^{1/2})/2a_2 & i=3 & \\
\beta > 1.5 & \lambda_i = (\beta - a_0)/a_1 & i=1,2,3 & (3.4)
\end{array}$$

We then estimated all coefficients and used them in the mean prediction equation (3.2). The estimated means are given in Figure 1 along with the sample means. From this graph it is clear that, in almost all cases, the deviations are very slight and, as will be subsequently noted, the means estimation was deemed adequate.

In the search for the functional relationship between A, β and the variance, σ^2, we graphed the sample variance, S^2, as a function of $\ln(\beta)$ for each value of A. On careful examination of that graph, Figure 3, the ensuing observations were made;

1) The sample variances were much more erratic than the sample means.
2) For levels of A below 1.0, the graphs of the relationships of the variances and β are, for all practical purposes, coincident (for this

reason only $A = .75$ was graphed as a representation of all $A \leq 1.0$).

3) The graphs are parabolic in appearance, opening about the $\ln(\beta)$ axis.

These observations led to the following model to estimate the relationship:

$$\ln(\beta) = \lambda_0(A) + \lambda_1(A)S^2 + \lambda_2(A)(S^2)^2, \text{ where}$$

S^2 is the sample variance. (3.7)

We ran least squares fits for each A and found that the model was acceptable provided the dependency of λ_0, λ_1, and λ_2 on A could be determined. Toward that end, we graphed each coefficient of (3.7) as a function of A. The plots of λ_0 and λ_1 appeared linear. The graph of λ_2 at first appeared to be quadratic, but was later found to be better approximated by a cubic equation. Therefore the following models for λ_0, λ_1, and λ_2 were fitted using least squares techniques:

$$\lambda_0 = a_0 + a_1 A$$

$$\lambda_1 = a_0 + a_1 A$$

$$\lambda_2 = a_0 + a_1 A + a_2 A^2 + a_3 A^3. \qquad (3.8)$$

The results of the least squares fits provided the following estimation equations for λ_0, λ_1, and λ_2:

$$\lambda_0 = -7.013 + .3871A$$
$$\lambda_1 = .2192 - .1759A$$
$$\lambda_2 = -6.1371 + 14.813A - 11.633A^2 + 3.05A^3. \qquad (3.9)$$

Tests of (3.9) yielded good approximations to the coefficients for each level of A.

The original variance model (3.7) was then re-written as

$$0 = (\lambda_0 - \ln(\beta)) + \lambda_1 S^2 + \lambda_2 (S^2)^2 \qquad (3.10)$$

and solving (3.10) for the estimated variance σ^2, we have

$$\text{EST. } \sigma^2 = (-\lambda_1 + [\lambda_1^2 - 4\lambda_2(\lambda_0 - \ln(\beta))]^{1/2})/2\lambda_2. \qquad (3.11)$$

Using the coefficient model to estimate λ_0, λ_1, λ_2, we then estimated the variances from (3.11). As noted previously, the sample variances are more erratic than the sample means and the estimated variances were not, in general, as accurate as the estimates of the means. The estimated variances are shown in Figure 3 along with the sample variances.

SAMPLE AND ESTIMATED MEANS VS. CROSSING LEVEL

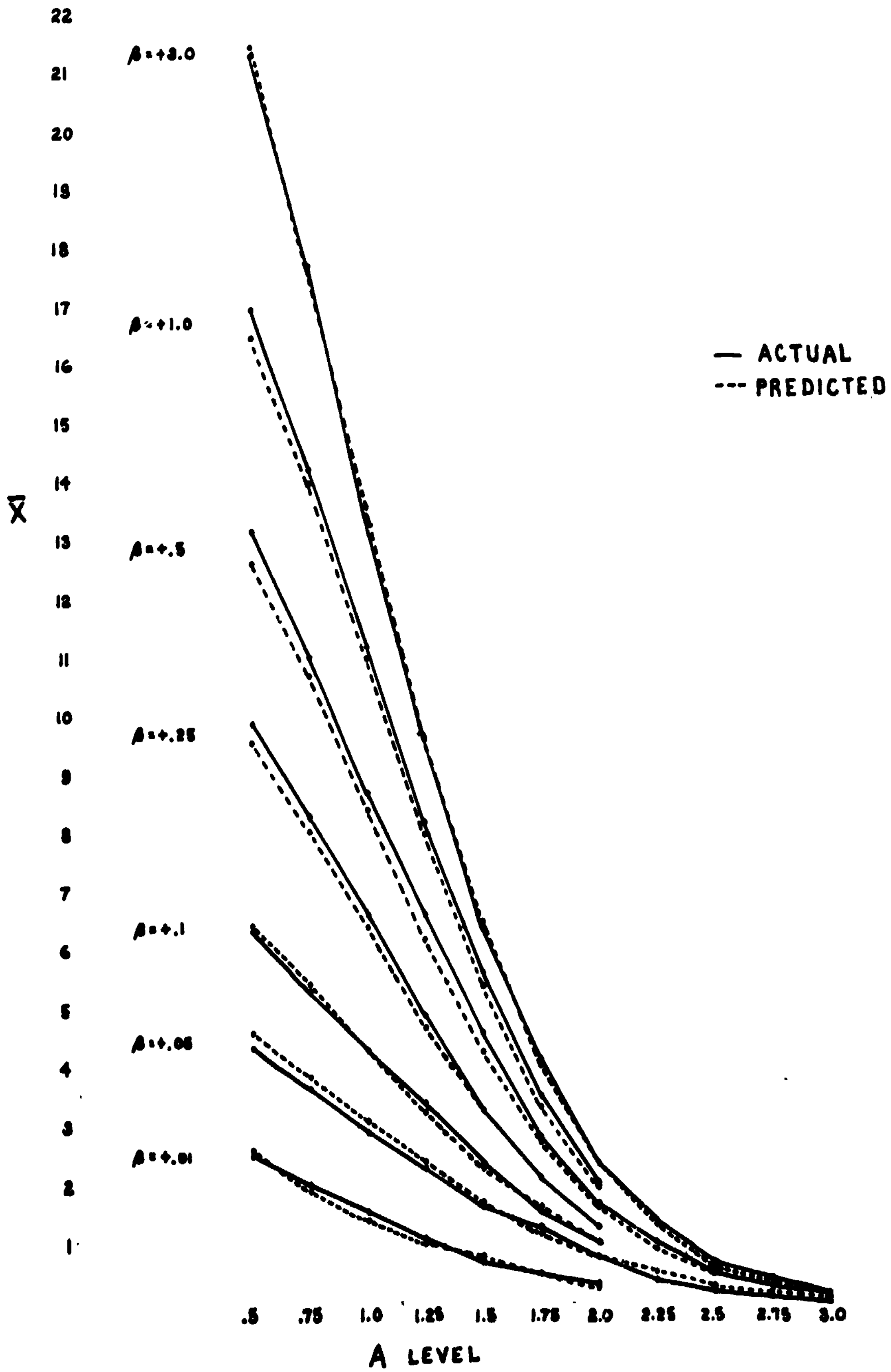


FIGURE 2

SAMPLE MEAN VS. CROSSING LEVELS

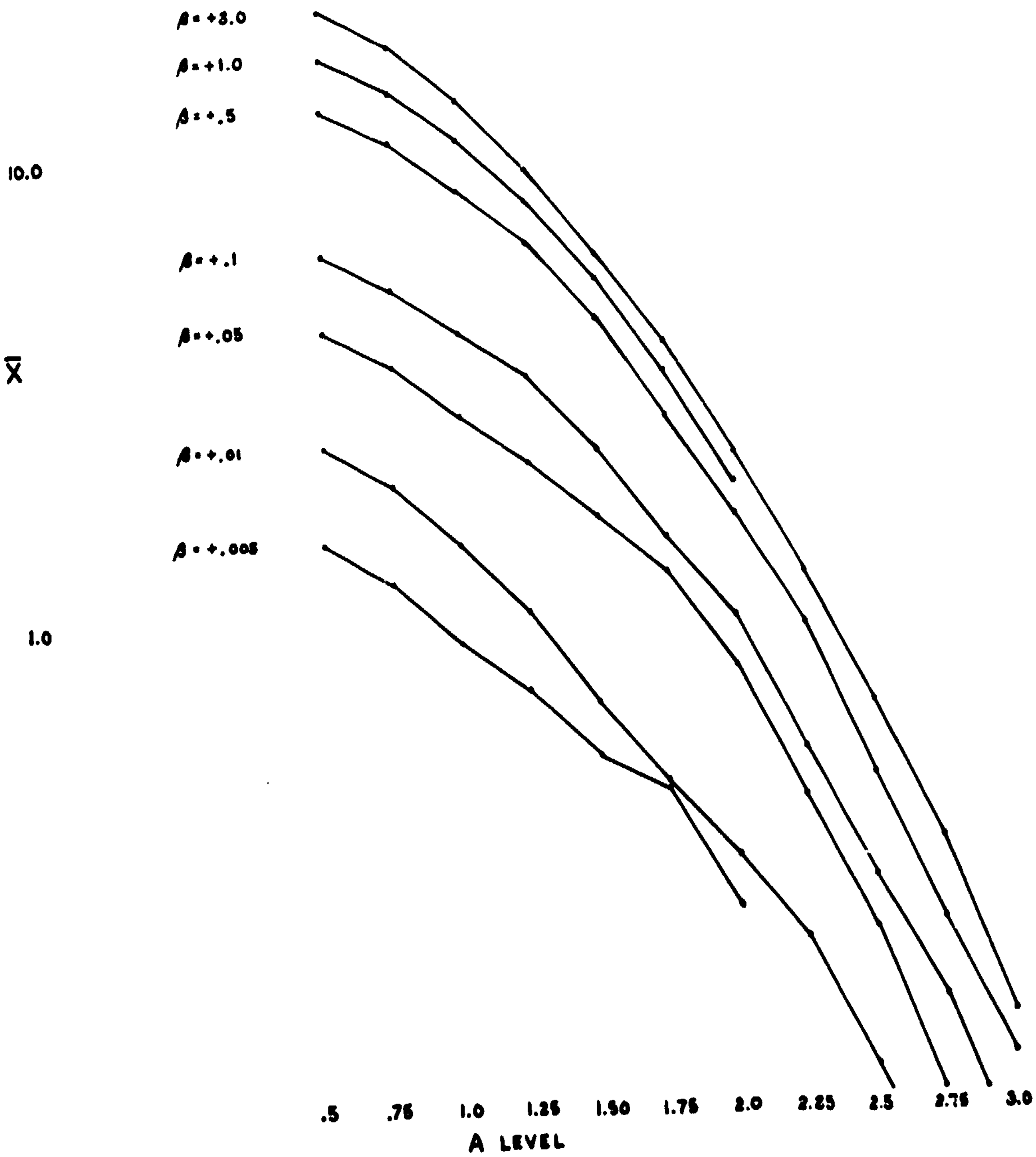
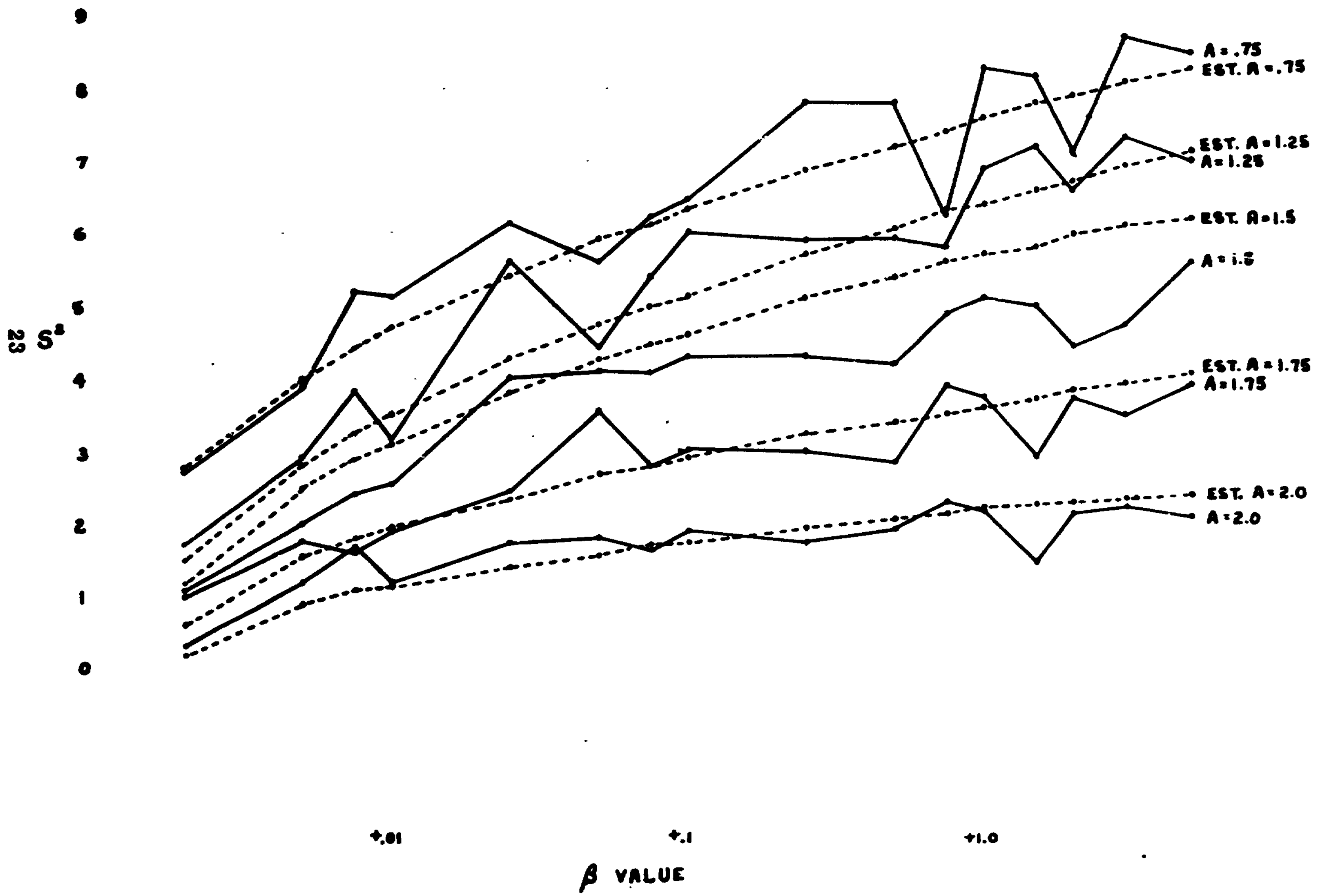


FIGURE 3

SAMPLE VARIANCE VS. AUTOCORRELATION PARAMETER (solid line)

ESTIMATED VARIANCE VS. AUTOCORRELATION PARAMETER (dashed line)



To check the accuracy of our results, we ran estimations of overshoots for the values of A and β that were used in the estimation process. The estimated mean was calculated from equation (3.2) while the estimate of the variance was given by (3.11). To determine which distribution was appropriate, we formed the ratio

$$r = \frac{\text{estimate of mean}}{\text{estimate of variance}}. \tag{3.12}$$

If $r \leq .95$ then the variance clearly exceeded the mean so we used the negative binomial distribution

$$P\{X=i\} = \frac{\Gamma(k+i)}{\Gamma(k)\,i!}\, p^k q^i;\ 0 \leq p \leq 1,\ p+q=1,\ k>0,\ i=0,1,2\ldots. \tag{3.13}$$

If $.95 < r < 1.05$ then the mean and variance were approximately equal so we used the Poisson distribution

$$P\{X=i\} = \frac{\lambda^i e^{-\lambda}}{i!};\ \lambda>0,\ i=0,1,2,\cdots. \tag{3.14}$$

If $r \geq 1.05$ then the mean clearly exceeded the variance so we used the binomial distribution

$$P\{X=i\} = \binom{n}{i} p^i q^{n-i};\ 0 \leq p \leq 1,\ p+q=1,\ i=0,1,\cdots,n. \tag{3.15}$$

The test used to check the goodness of fit for the predicted models was the Kolmogorov goodness of fit test. Briefly, the test compares the theoretical and sample distribution functions and one concludes there is no significant difference between these distributions if the maximum absolute difference between them is less than a predetermined quantity based on the significance level and sample size.

The test is designed to compare continuous distribution functions and, as such, is not directly applicable to discrete cases. When it is applied the significance level used is conservative to an indeterminate degree. For our purposes this is quite acceptable. In general a conservative test conducted at α level of significance is, in reality, being conducted at some $\alpha' < \alpha$ level of significance. Thus in Table 4 the α levels are given as $\alpha \leq .05$ or $\alpha \leq .01$. This means that a conclusion that we have a good fit using the $\alpha = .05$ significance level really says the two distribution functions are in agreement at some α' value smaller than .05. For the cases in Table 4 where the predicted models fitted poorly we can only state that the model was rejected at some α level less than .01. The justification for using the Kolmogorov test for these data is given in Noether (1967, pp. 17-18).

The sampling distribution for the Kolmogorov test is well known and for sample sizes above 35 the maximum absolute difference between the theoretical and observed distribution functions must not exceed the value $d_\alpha/\sqrt{n}$ where $d_{.05} = 1.36$ and $d_{.01} = 1.63$. These results are available in Siegel (1956, pg. 251). Consequently using $n = 250$ the critical values are 21.50 and 25.77 respectively.

A brief glance at Table 4 shows excellent results through most of the β and A values with no rejections in the $.1 \leq \beta \leq 1.0$ range which will be the primary β values used in wind speed calculations. It is unlikely that many applications will require those β values giving poor results, namely .002 and 2.0. The .002 data set is the "end point" in our predictive process and 2.0 is the data set just above the "transitional" value where the means behavior changed drastically (previously discussed in this chapter).

Appendix III presents a spectrum of data sets in the computer format used to evaluate the goodness of fit. All pertinent information, i.e., A and β values, observed and predicted means and variances, model utilized, cumulative distribution functions and predicted probabilities, are presented.

TABLE 4

Condensed Summary of Fitted Models

Using Predictive Equations

*Good Fit at $\alpha \leq .01$

**Good Fit at $\alpha \leq .05$

$\mu\uparrow$ Bad Fit due to mean overestimation

$\mu\downarrow$ Bad Fit due to mean underestimation

$\sigma^2\uparrow$ $\sigma^2\downarrow$ Analogous to definitions on μ above

$\beta \backslash A$	.5	.75	1.0	1.25	1.5	1.75	2.0	2.5	3.0
.002	**	**	**	**	$\mu\uparrow$	$\mu\uparrow$	$\mu\uparrow$		
.005	**	**	**	*	**	**	*	$\mu\uparrow$	**
.0075	$\sigma^2\downarrow$	**	**	**	**	**	**		
.01	$\sigma^2\downarrow$	*	**	**	**	**	**	**	**
.025	**	*	**	*	*	**	*		
.05	**	**	**	**	**	**	**	$\mu\downarrow$	**
.075	**	**	**	**	**	*	*		
.1	**	**	**	**	**	**	**	**	**
.25	*	**	**	**	**	**	**		
.50	**	**	**	*	*	**	**	**	**
.75	**	**	**	**	**	**	**		
1.0	**	**	**	**	*	**	**		
1.5	**	**	**	**	**	**	**	**	**
2.0	$\mu\uparrow$	$\mu\uparrow$	*	*	*	$\mu\uparrow$	$\mu\uparrow$		
3.0	**	**	**	**	**	**	**	**	**
5.0	$\mu\downarrow$	$\mu\downarrow$	$\mu\downarrow$	**	**	**	**		

The fact that the binomial and negative binomial models fitted the data was, upon reflection, not surprising for the following reasons:

1) A well known statistical fact given by Johnson and Kotz (1969, pg. 43) states that, if one of the models is applicable, the criterion for selection depends on the relationship between the mean and variance, i.e., if $\mu > \sigma^2$ select the binomial, if μ is approximately equal σ^2 select the Poisson model, if $\mu < \sigma^2$ select the negative binomial.

2) As pointed out by Johnson and Kotz (1969, pg. 135) for β values that are quite low the negative binomial is the appropriate model since, with a small β value, the successive time points and, therefore, successive overshoots are dependent. In applications where the Poisson model seems appropriate but successive events are not independent the negative binomial model is an excellent alternative.

3) For larger β values the binomial model is required since successive time points and overshoots are, for all practical purposes, independent.

The criterion for a successful model was adequate fits on the majority of the data sets. As pointed out in Table 4, the experimental results for means and variances were, in general, approximated adequately by the prediction model.

Therefore we concluded that the model presented in this analysis is a good predictor of overshoots in a stationary gaussian stochastic process with an exponential autocorrelation function.

CHAPTER 4

Applications

The purpose of this chapter is to explain how the analyses discussed previously could be applied to problems of a general nature where the assumption of a stationary Gaussian process with an exponential autocorrelation function is plausible.

As noted in chapter 3 the A levels of .5, .75, 1.0, 1.25, 1.5, 1.75 and 2.0 were used in obtaining the basic prediction models. Figure 1 gives the predicted means for levels above 2.0 for representation values of β. It is apparent that the predictive equations are adequate for A levels above 2.0.

While the study involved counting overshoots above specified A levels it is valid to assume the model is applicable to predicting the number of "undershoots" below negative A levels - for no reason other than the symmetry of the normal distribution.

The estimating equations for the mean and variance were derived based on realizations of 100 points. Some applications, based on time periods yielding appreciably different numbers of points in a realization, require a modification in these estimating equations. In the computer program this is done directly using the standard formulas. Letting M be the number of time points desired we have

$$\text{New EST}(\mu) = \text{EST}(\mu)*M/100$$

and

$$\text{New EST}(\sigma^2) = \text{EST}(\sigma^2)*(M/100)^2$$

with the model being selected based on the new estimated mean and variance. It should be noted that these new estimates are exactly correct only for the independent case (high A and/or β values). When the correlations go to zero rapidly (β values ≥ 1 and/or moderate A values) the new estimates will be higher by an unimportant, and likely indeterminable, amount. For small β values the new estimates could be inflated if A is also low. As our primary range of interest has been in β values $\geq .1$ and A values distant from the mean the modification in the mean and variance will be satisfactory.

The computer program given in Appendix I has been developed to support applications of this study. We will, in the ensuing discussions, relate applications that can be performed using this program.

To utilize this program the following data must be provided:

1) the average, μ, of the process,
2) the standard deviation, σ, of the process,
3) the coefficient, β, of the autocorrelation function $R(\tau) = \exp(-\beta|\tau|)$,
4) the number of time points and
5) the crossing level L and maximum frequency N.

Assume that the probability distribution of the number of overshoots above some level $L > \mu$ is desired. The program will use L, μ, and σ to calculate the A level, i.e., $A = (L - \mu)/\sigma$, which will be used in the prediction. It is important to note that the model expects A to be positive and should L be less than μ the A value would be $|L - \mu|/\sigma$.

The β value used will, in most cases, correspond directly to the β values used in the analysis since empirically β values are calculated or estimated using serial correlations of lag 1, lag 2, etc. which are independent of the interval between successive time points. Should a β value be calculated using the actual time intervals it will be necessary for the user to modify the β value prior to utilizing the program. Recall that the β used in the program assumed "time" units of length 1. If a β value has been calculated using intervals of, say, .5, i.e., $\tau = .5, 1.0, 1.5,$ etc. the autocorrelation function will be $R(\tau) = \exp(-\beta|\tau|)$, $\tau = .5, 1.0, \ldots$ and this corresponds directly to $R(\tau') = \exp(-.5\beta|\tau'|)$, $\tau' - 1, 2, \ldots$. In this case the value $.5\beta$ would be the value the user supplies to the program.

In general we can summarize this procedure as follows: Assume the autocorrelation parameter β' has been calculated using equally spaced intervals $\tau' = h, 2h, 3h, \ldots,$ giving $R(\tau') = \exp(-\beta'|\tau'|)$. This corresponds directly to $R(\tau) = \exp(-\beta' h|\tau|)$, $\tau = 1, 2, \ldots$ which means β (for program input) $= \beta' h$.

The program output will consist of:

1) L, μ, σ, β and calculated A value,

2) Predicted mean and variance for the number of crossings and the model selected based on these values and

3) Predicted probabilities for 0, 1, 2, ..., n (or more), $n \leq 40$, overshoots.

As an example consider the situation below. For the month of January at 12 km the scalar wind speed at Cape Kennedy has the following properties:

1) $R(\tau') = \exp(-\beta'\tau')$, $\tau' = 12, 24, 36, \ldots$ with $\beta' = .0247$,

2) $\sigma = 8$ m/sec and

3) $\mu = 24$ m/s.

We desire to predict the probabilities of 0, 1, 2, 3, 4 and 5 (or more) overshoots above the level $L = 39$ m/s.

The β value for the program is not .0247 but rather is $12 * .0247 = .2964$. This makes $R(\tau') = \exp(-.0247\tau')$, $\tau' = 12, 24, 36, \ldots$ equal to $R(\tau) = \exp(-.2964\tau)$, $\tau = 1, 2, 3, \ldots$. The program input is $L = 39$, $\mu = 24$, $\sigma = 8$, $\beta = .2964$ $N = 5$ and $M = 62$. The program calculates the standardized crossing level A as 1.875 and utilizes these A and β values to calculate the predicted probabilities. Table 5 gives the resultant computer output.

The formats for program input parameters are given in Appendix I.

TABLE 5

LEVEL 39.0000 MEAN 24.0000 STD. DEVIATION 6.0000
ADJUSTED LEVEL 1.8750 DIST. MEAN 1.1259 DIST. VAR 0.9686
AUTOCORRELATION PARAMETER 0.2964

MEAN EXCEEDES VARIANCE, BINOMIAL MODEL SELECTED

NUMBER OF CROSSINGS	PREDICTED PROBABILITY
0	0.3001
1	0.3898
2	0.2215
3	0.0719
4	0.0146
5	0.0021

This algorithm was, of course, developed under the assumption that the process was stationary normal with an exponential autocorrelation function. This process is the most widely applied and, despite the intractability of the mathematics, in the only such continuous stochastic process that can be analyzed to any appreciable extent. The question of the degree of applicability to suspected or confirmed non-normal processes will certainly arise and while there appears to be no answers in the literature there are some statistical results of a general nature that have some bearing.

Most skewed distributions, e.g., lognormal or gamma are, for certain ranges of parameter values, almost normal and certainly the algorithm is useful in these cases. As an example of this term "almost" normal consider a gamma distribution with parameters $\alpha = 2f$ and $\beta = 1/2$ where $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$. As f increases this distribution rapidly approaches normality. With these values of α and β this is the χ^2 pdf and it is common statistical practice to use a normal approximation to determine critical values when the degrees of freedom (denoted by f) is large.

The equations to estimate the process mean and variance do not utilize the properties of any distribution. The assumptions of stationarity and exponential-like autocorrelation function are certainly required whether the distribution is or is not highly skewed. The Central Limit Theorem (CLT) permits one to determine the average number of points above a certain level in this process just as if it were normal (the CLT is certainly applicable with sample sizes of 250) and an extension to the average number of overshoots is reasonable. In skewed distributions the mean and variance are not independent - this possibly makes our independent equations for estimating the process mean and variance a bit tenuous. The relationship would certainly be difficult to determine but the proper procedure would likely be to incorporate the estimated mean (which should be adequate if the assumptions are satisfied) into the estimating equation for the variance. The three discrete distributions, i.e., binomial, negative binomial and Poisson, are applicable in any case.

One additional result is the fact that overshoots frequencies approach the Poisson distribution as the crossing level increases regardless of the process distribution. This "cutoff" value was approximately two standard deviations above the mean in our study and would likely be close to the "cutoff" value for most distributions.

REFERENCES

Cramer´, H., and Leadbetter, M. R., (1967); Stationary and Related Stochastic Processes; New York, Wiley.

Hamming, R. W., (1962); Numerical Methods for Scientists and Engineers; New York, McGraw-Hill.

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

Jorgenson, Dale W., (1961); Multiple Regression Analysis of a Poisson Process; Journal of the American Statistical Association, 56, pp. 235-245.

Kuznetsov, T. I., (1965); Non Linear Transformations of Stochastic Processes; Oxford, Pergamon.

Noether, G. E., (1967); Elements of Nonparametric Statistics; New York, Wiley.

Odell, Patrick L., and Newman, Thomas G., (1971); The Generation of Random Variates; New York, Hafner.

Siegel, S., (1956); Non Parametric Methods; New York, McGraw-Hill.

APPENDIX I

PREDICTION PROGRAM LISTING AND

INPUT FORMATS

NOTE: DO NOT USE THIS FORM WHEN ORDERING NEW OR REVISED CARD FORM COMPOSITION.

IBM

INTERNATIONAL BUSINESS MACHINES CORPORATION

INFORMATION RECORDS DIVISION
MULTIPLE-CARD LAYOUT FORM

Form X24 6599 1
Printed in U.S.A.

Company ______

Application: Program Input   by: M. Madison   Date ______   Job No ______   Sheet No ______

PROCESS MEAN (F10.3)	PROCESS STANDARD DEVIATION (F10.3)	AUTO-CORRELATION PARAMETER β (F10.3)	CROSSING LEVEL (F10.3)	MAX FREQ. (I5)	NO. OF PTS. (I5)	
1–10	11–20	21–30	31–40	41–45	46–50	51–80

```
C
C
C      PROGRAM ID.
C          MMU1.
C
C      AUTHOR.
C          MIKE MALISON.
C
C      DATE WRITTEN.
C          04/10/73.
C
C      PURPOSE.
C          CALCULATE FREQUENCY DISTRIBUTION FOR OVERSHOOTS IN
C                A STATIONARY GASUSSIAN STOCHASTIC PROCESS WITH
C                EXPONENTIAL AUTOCORRELATION FUNCTION.
C
C      SUBROUTINES REQUIRED.
C          DLGGM: CALCULATE THE LOG OF THE GAMMA DISTRIBUTION.
C
C
C**********************************************************************
C
C
C
      INTEGER RDR,PRNTR,RATIO
      REAL MEAN,LEVEL,JP1
C     SET UP IO ASSIGNMENTS
      RDR=1
      PRNTR = 3
C
C     READ PARAMETERS AND CONVERT TO STANDARD CROSSING LEVEL
C
  100 READ (RDR,1000,END=700) MEAN,STDEV,BETA,LEVEL,NUM,RATIO
      NUM = NUM + 1
      B = -BETA
      APRT =(LEVEL - MEAN) / STDEV
      A = ABS (APRT)
      ALOGB = ALOG (BETA)
C
C     COMPUTE EST OF DISTRIBUTION MEAN
C
      IF (BETA .GT. 1.5) GO TO 250
      IF (BETA .GT. .01) GO TO 200
      F1 = (ALOGB + 6.07171) / 1.1105
      F2 = (ALOGB + 6.65596) / (-3.23315)
      F3 = (ALOGB + 7.81708) / (-10.2928)
      GO TO 300
  200 D1 = 1.57686**2 - 4.0*.21851*(-6.423 - ALOGB)
      D2 = 10.31154**2 - 4.0*10.23657*(-1.0582 - ALOGB)
      D3 = 3.382882**2 - 4.0*16.3656*(-3.9142 - ALOGB)
      F1 = (-1.57686 + SQRT(D1)) / .43702
          F2 = (-10.31154 + SQRT(D2)) / 20.47314
      F3 = (-3.382882 - SQRT(D3)) / 32.7312
      GO TO 300
```

```
  250 F1 = 3.305 + .014*B
      F2 = -.169 - .0185*B
      F3 = -.47 + .02*B
  300 TEMP = F1 + F2*A + F3*A**2
      EMEAN = EXP (TEMP)
      EMEAN=EMEAN*RATIO/100.
C
C     COMPUTE EST OF DISTRIBUTION VARIANCE
C
      IF (A .GE. 1.0) GO TO 350
      G1 = -6.32749
      G2 = -.3902259
      G3 = .16156
      GO TO 400
  350 G1 = -7.012996 + .387066*A
         G2 = .219193 - .1758552*A
      G3 = -6.137068 + 14.81297*A - 11.63295*A**2 + 3.04961*A**3
  400 DISC = G2**2 -4.0*G3*(G1 - ALOGB)
      EVAR =(-G2 + SQRT(DISC)) / (2*G3)
      EVAR=EVAR*(RATIO/100.)**2
C
C     DETERMINE APPROPRIATE MODEL AND REASON AND PRT HEADINGS
C
      RLTN = EMEAN / EVAR
      IF (A .GT. 2.0) GO TO 500
      IF (.95 .LT. RLTN .AND. RLTN .LT. 1.05) GO TO 510
      IF (EVAR .GT. EMEAN) GO TO 520
C     MEAN > VAR -> BINOMIAL
      MODEL = 1
      PP = 1 - EVAR / EMEAN
      XTEMP = EMEAN**2/(EMEAN-EVAR)
      XTEMP = XTEMP + .5
      ITEMP = XTEMP
      IF (ITEMP .LT. (NUM-1)) GO TO 490
      PX = ITEMP
      GO TO 495
  490 PX = NUM - 1
  495 CONTINUE
      GO TO 550
C     A > 2.0 -> POISSON
  500 MODEL = 2
      GO TO 550
C     MEAN = VAR -> POISSON
  510 MODEL = 3
      GO TO 550
C     VAR > MEAN -> NEG BINOMIAL
  520 MODEL = 4
      PK = EMEAN**2 / (EVAR - EMEAN)
      PP = PK / (PK + EMEAN)
  550 WRITE (PRNTR,1010)LEVEL,MEAN,STDEV,APRT,EMEAN,EVAR,BETA
      GO TO (551,552,553,554),MODEL
  551 WRITE (PRNTR,1011)
      GO TO 560
  552 WRITE (PRNTR,1012)
```

```
      GO TO 560
  553 WRITE (PRNTR,1013)
      GO TO 560
  554 WRITE (PRNTR,1014)
  560 WRITE (PRNTR,1015)
C
C     CALCULATE AND PRINT ALL BUT LAST CELL
C
      PSUM = 0
      NM1 = NUM - 1
      IF (NUM .EQ. 1) GO TO 600
      DO 590 I=1,NM1
      J = I - 1
      JP1 = J + 1
      GO TO (571,572,573,574),MODEL
C     BINOMIAL MODEL
  571 PXP1 = PX + 1
      OMPP = 1 - PP
      ALPR = DLGGM(PXP1)+J*ALOG(PP)+(PX-J)*ALOG(OMPP)-DLGGM(JP1)
     1-DLGGM(PX-J+1.)
      GO TO 580
C     POISSON MODEL
  572 CONTINUE
  573 ALPR = J * ALOG(EMEAN) - EMEAN - DLGGM (JP1)
      GO TO 580
C     NEGATIVE BINOMIAL
  574 PKPJ = PK + J
      OMPP = 1 - PP
      ALPR = DLGGM(PKPJ)+PK*ALOG(PP)+J*ALOG(OMPP)-DLGGM(PK)-DLGGM(JP1)
C
C     CALC PROBABILITY AND PRT CELL
C
  580 PROB = EXP (ALPR)
      PSUM = PSUM + PROB
  590 WRITE (PRNTR,1020)J,PROB
C
C     COMPUTE LAST CELL
C
  600 PROB = 1 - PSUM
      WRITE (PRNTR, 1020)NM1,PROB
      GO TO 100
C
C     END OF JOB
C
   00 STOP
C
C     ************************ F O R M A T S **************************
C
 1000 FORMAT(4F10.3,2I5)
 1010 FORMAT ('1','LEVEL',F11.4,8X,'MEAN',F11.4,10X,'STD. DEVIATION',
     1F11.4/' ADJUSTED LEVEL',F11.4,3X,'DIST. MEAN',F11.4,5X,
     2'DIST. VAR',F11.4/' AUTOCORRELATION PARAMETER ',F11.4/)
 1011 FORMAT (  MEAN EXCEEDES VARIANCE, BINOMIAL MODEL SELECTED')
 1012 FORMAT (' LEVEL ABOVE 2.0, POISSON MODEL SELECTED')
```

```
 1013 FORMAT (' MEAN APPROX = VARIANCE, POISSON MODEL SELECTED')
 1014 FORMAT (' VARIANCE EXCEEDES MEAN, NEGATIVE BINOMIAL MODEL SELECT',
     1'ED')
 1015 FORMAT (/6X,'NUMBER OF CROSSINGS        PREDICTED PROBABILITY')
 1020 FORMAT (' ',10X,I6,21X,F7.4)
      END
```

```
      FUNCTION DLGGM(DX)
      DY=DX
      DTERM=1.
      IF(DX)1,1,2
    1 DLGGM=0.
      RETURN
    2 IF(DY-18.)3,3,4
    3 DTERM=DTERM*DY
      DY=DY+1.
      GO TO 2
    4 DLGGM=(DY-.5)*ALOG(DY)-DY+1./(12.*DY)-1./(360.*DY**3)+1./(1260.*
     1DY**5)-1./(1680.*DY**7)+.918938533204673-ALOG(DTERM)
      RETURN
      END
```

APPENDIX II

COMMENTS ON THE THEORETICAL APPROACH

COMMENTS ON THE THEORETICAL APPROACH

Two interesting problems in the theory of stochastic processes are first to find the probability density of the duration of a crossing of a given level by a random process $X(t)$ and second to find the probability density of the number of crossings of a level by the process. The problem of obtaining the average number of crossings of a level has received much attention in the literature. In fact, if $X(t)$ is a stationary gaussian process, the complete solution has been given by Ito (1964) and Ylvisaker (1965). For non-stationary gaussian processes, Leadbetter and Cryer (1965) have given a similar result. And finally, Leadbetter (1966) has considered the average number of crossings for a wide class of non-gaussian processes. However, solutions in closed form for the original two problems have not been obtained even in the more desirable case when $X(t)$ is gaussian. Several approximations to these probabilities have been obtained and we shall give some with references.

Let $X(t)$ be a random process with correlation function $R(\tau)$. Following Rice (1945), the probability density function of the interval between the ith and the $(i + m + 1)$th crossing of a level A by $X(t)$ is denoted by $P_m(\tau)$; and the probability of exactly n crossings of the level A in the interval $(t, t + \tau)$ is denoted by $p(n,\tau)$. For a basic relationship between $P_m(\tau)$ and $p(n,\tau)$ see Appendix I of McFadden (1958). Let $f_0 = X(t_0) = A$, $f_i = X(t_i)$, $g_i = f'(t_i)$

for $i = 0, 1, \ldots, n$ and $d_k(t_1, \ldots, t_k) =$

$$(1/N(\Lambda))\int_0^\infty dg_0\int_{-\infty}^0\cdots\int_{-\infty}^0 |g_0 g_1 \cdots g_k| W_{k+1}(\Lambda,\ldots,\Lambda; g_0,\ldots g_k)\, dg_1\cdots dg_k$$

where $N(\Lambda) = \int_0^\infty g\, W_1(\Lambda;g)dg$ is the expected number of crossings of the level Λ by $X(t)$ in the interval $(t, t+\tau)$ and $W_{k+1}(f_0,\ldots,f_k;g_0,\ldots,g_k)$ is the joint probability density of $f_0,\ldots,f_k$ and $g_0,\ldots,g_k$. Then according to Kuznetsov and Stratonovich (1956)

$$p(n,\tau) = 1/n!\sum_{k=0}^{\infty}(-1)^k/k!\int_0^\tau\cdots\int_0^\tau d_{n+k}(t_1,\ldots t_{n+k})dt_1\cdots dt_{n+k}. \tag{1}$$

And by Kuznetsov, Stratonovich and Tikhonov (1954) the probability density for the duration τ is

$$p(\tau) = \frac{-d}{d\tau}\, p(0,\tau). \tag{2}$$

It is apparent that the desired probabilities (1) and (2) are very complicated and consequently only approximations have been given. To illustrate this point let us consider the case where $X(t)$ is gaussian with correlation function $R(\tau)$. It is known that the normal property is retained for any linear transformation of a normal random function. Consequently, the joint probability density for the values of the random function and its derivatives will also be normal. Thus $W_{k+1}(f_0,\ldots,f_k;g_0,\ldots,g_k)$

$$= (1/(2\Pi)^{k+1}\Delta^{1/2})\exp[-1/2\sum_{i,j=0}^{2k+1}L_{ij}f_i f_j]$$

where $f_{k+i} = g_{i-1}$, $i=1,\ldots,k+1$, $\Delta = |(r_{ij})|$, $L_{ij} = (r_{ij})^{-1}$,

$$(r_{ij}) = \begin{pmatrix} R_{00} & \cdots & R_{0k} & R'_{00} & \cdots & R'_{0k} \\ \vdots & & \vdots & \vdots & & \vdots \\ R_{k0} & & R_{kk} & R'_{k0} & \cdots & R'_{kk} \\ -R'_{00} & & -R'_{0k} & -R''_{00} & & -R''_{0k} \\ \vdots & & \vdots & \vdots & & \vdots \\ -R'_{k0} & \cdots & -R'_{kk} & -R''_{k0} & \cdots & -R''_{kk} \end{pmatrix}_{2k+2,2k+2}$$

and $R_{ij} = R(t_i - t_j)$.

Thus we have $W_{k+1}(A,\ldots,A;g_0,\ldots,g_k)$

$$= (1/(2\Pi)^{k+1}\Delta^{1/2}) \exp[-1/2(\sum_{i,j=0}^{k} A^2 L_{ij} + \sum_{i,j=0}^{k} L_{k+1+i,k+1+j} g_i g_j)]$$

$$= (1/(2\Pi)^{k+1}\Delta^{1/2}) \exp[-A^2/2 \sum_{i,j=0}^{k} L_{ij}] \exp[-1/2 \sum_{i,j=0}^{k} L_{k+1+i,k+1+j} g_i g_j].$$

Now let us denote by $p(g_0,\ldots,g_k|f_0,\ldots,f_k)$ the conditional probability density of $(g_0,\ldots,g_k)$ given $(f_0,\ldots,f_k)$ and let $p(f_0,\ldots,f_k)$ be the probability density of $(f_0,\ldots,f_k)$ then

$$p(f_0,\ldots,f_k) = (1/(2\Pi)^{k+1/2}D^{1/2}) \exp(-1/2 \sum_{i,j=0}^{k} M_{ij} f_i f_j)$$

where (M_{ij}) is the inverse of (R_{ij}) and

$$D = \begin{vmatrix} R_{00} & \cdots & R_{0k} \\ \vdots & & \vdots \\ R_{k0} & \cdots & R_{kk} \end{vmatrix} .$$

The last summation of (3) involves only the last $k+1$ rows and columns of (L_{ij}). Denote the inverse of this matrix by (m_{ij}); that is,

$$(m_{ij}) = \begin{pmatrix} L_{k+1,k+1} & \cdots & L_{k+1,2k+1} \\ \vdots & & \vdots \\ L_{2k+1,k+1} & \cdots & L_{2k+1,2k+1} \end{pmatrix}^{-1}$$

It is clear that (m_{ij}) is the covariance matrix of $(g_0,\ldots,g_k)$ given that $f_0 = f_1 = \ldots = f_k = 0$ and by Jacobi's Theorem the (i,j)th element of this matrix is the bordered determinant

$$m_{ij} = \begin{vmatrix} R_{00} & \cdots & R_{0k} & R'_{0j} \\ \vdots & & \vdots & \vdots \\ R_{k0} & & R_{kk} & R'_{kj} \\ -R'_{i0} & & -R'_{ik} & -R''_{ij} \end{vmatrix} \div D.$$

The determinant of (m_{ij}) is given by $|(m_{ij})| = \Delta/D$. So now we have $p(g_0,\ldots,g_k|0,\ldots,0)$

$$= \frac{(1/(2\Pi)^{k+1}\Delta^{1/2}) \exp[-1/2 \sum_{i,j=0}^{k} L_{k+1+i,k+1+j}g_i g_j]}{(1/(2\Pi)^{k+1/2}D^{1/2})}$$

$$= (1/(2\Pi)^{k+1/2}|(m_{ij})|^{1/2}) \exp[-1/2 \sum_{i,j=0}^{k} L_{k+1+i,k+1+j}g_i g_j]$$

$$= Z(\vec{g}, \vec{m}).$$

Thus $W_{k+1}\ (A,\ldots,A;g_0,\ldots,g_k)$

$$= (1/(2\Pi)^{k+1}\Delta^{1/2})\ \exp[-A^2/2 \sum_{i,j=0}^{k} L_{ij}]\ \exp[-1/2 \sum_{i,j=0}^{k} L_{k+1+i,k+1+j}g_ig_j]$$

$$= (1/(2\Pi)^{k+1/2}D^{1/2})\ \exp[-A^2/2 \sum_{i,j=0}^{k} L_{ij}]\ .\ Z(\vec{g},\ \vec{m}).$$

Therefore, $d_k(t_1,\ldots,t_k)$

$$= (1/N(A)(2\Pi)^{k+1/2}D^{1/2})\ \exp[-A^2/2 \sum_{i,j=0}^{k} L_{ij}]\cdot\int_0^{\infty}dg_0\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}|g_0\cdots g_k|$$

$$\cdot Z(\vec{g},\ \vec{m})dg_1\ldots dg_k.$$

Let

$$n_{ij} = \begin{cases} -m_{ij}/(m_{ii}m_{jj})^{1/2} & i \neq 0 \\ m_{ij}/(m_{ii}m_{jj})^{1/2} & i = 0 \end{cases}$$

and $h_i = g_i/(m_{ii})^{1/2}$ then $d_k(t_1,\ldots,t_k)$

$$= ((m_{00}\ldots m_{kk})^{1/2}/N(A)(2\Pi)^{k+1/2}D^{1/2})\ \exp[-A^2/2 \sum_{i,j=0}^{k} L_{ij}]$$

$$\cdot\int_0^{\infty}\cdots\int_0^{\infty}h_0\ldots h_k Z(\vec{h},\ \vec{n})dh_0\ldots dh_k$$

where $Z(\vec{h},\ \vec{n})$ is the ordinary normal probability density function in $k+1$ variables $h_0,\ldots,h_k$ with covariance matrix (n_{ij}).

In an attempt to find $p(\tau)$ when $X(t)$ is gaussian and $R(\tau) = \exp(-\alpha\tau^2)$, Tikhonov (1956) has approximated $p(0,\tau)$ in (1) by neglecting all terms of the series greater than 2. He claims that his results give satisfactory agreement with the experimental results of Rice (1953). It is clear that the smaller τ is, the better the approximation. However, if τ is very large or if we wish to find $p(n,\tau)$ for large n we must find some other means of approximation.

Longuet-Higgins (1962) has obtained an infinite series for $p(n,\tau)$ and $P_m(\tau)$ similar to (1) where each term is an integral of the joint probability $W(+, -, -, \ldots, -)dt_1 \ldots dt_n$ that $X(t)$ has an up-crossing in the infinitesimal interval $(t_1, t_1 + dt_1)$ and a down-crossing in the remaining $(n-1)$ intervals $(t_i, t_i + dt_i)$ $(i = 2, 3, \ldots, n)$. He also gives a general relation between $P_m(\tau)$, $p(n,\tau)$ and $W(S)$ where S is a series of plus and minus signs (plus if $X(t)$ has an up-crossing and minus if $X(t)$ has a down-crossing). Using the infinite series he obtains the asymptotic behavior of $P_m(\tau)$ and $p(n,\tau)$ for small τ.

Based on their experimental results, Faureau, Low and Pfeffer (1956) hypothesised the distribution of $P_0(\tau)$ for a gaussian process $X(t)$ whose spectrum is $(1 + \sigma^2)^{-2}$ to be negative exponential. However, using his asymptotic expression Longuet-Higgins (1962) was able to disprove this conjecture.

Other experimental and analytical approximations of the desired probabilities have been given but almost all are asymptotic approximations for small τ or approximations as the level A approaches ∞. Although, we cannot obtain the exact probabilities $p(n,\tau)$ and $p(\tau)$, we desire approximations which are valid for intermediate level τ and $n \geq 1$.

References

Favreau, R. R., Low, H. and Pfeffer, I. (1956). "Evaluation of Complex Statistical Functions by an Analog Computer". IRE Convention Record. pt 4, 31-37.

Ito, K. (1964). "The Expected Number of Zeros of Continuous Stationary Gaussian Processes". J. Math. Kyoto Univ. 3, 207-216.

Kuznetsov, P. I. and Stratonovich, R. L. (1956). "A Note On the Mathematical Theory of Correlated Random Points". Izvestiya Akod. Nauk SSSR, ser. mat., 20, 167. Also, Non-Linear Transformations of Stochastic Processes; Pergamon Press, Oxford. 6, 101-115.

Kuznetsov, P. I., Stratonovich, R. L. and Tikhonov, V. I. (1954). "On the Duration of Excursions of Random Functions". Zhur. tekh. fig., 24, No. 1, 103. Also, Non-Linear Transformations of Stochastic Processes; Pergamon Press, Oxford. 28, 341-353.

Leadbetter, M. R. (1966). "On Crossings of Levels and Curves by a Wide Class of Stochastic Processes". Ann. Math. Stat. 37, 260-267.

Leadbetter, M. R. and Cryer, J. D. (1965). "On the Mean Number of Curve Crossings by Non-Stationary Normal Processes". Ann. Math. Stat. 36, 509-517.

McFadden, J. A. (1958). "The Axis-Crossing Intervals of Random Functions II". IRE Trans. on Inf. Theory IT-4,14-24.

Rice, S. O. (1945). "Mathematical Analysis of Random Noise". Bell Sys. Tech. J. 24, 46-156.

Rice, S. O. (1953). "Theory of Fluctuating Noise". Collection of papers. Foreign Literature Publishing House.

Tikhonov, V. I. (1956). "The Distribution of the Duration of Excursions of Normal Fluctuations". Radio tekh. i. Elekt., 1, No. 1, 23. Also, Non-Linear Transformations of of Stochastic Processes; Pergomon Press, Oxford. 29, 354-367.

Ylvisaker, N. D. (1965). "The Expected Number of Zeros of a Stationary Gaussian Process". Ann. Math. Stat. 36, 1043-1046.

APPENDIX III

COMPARISON OF SAMPLE DISTRIBUTIONS TO PREDICTED DISTRIBUTIONS

A LEVEL 0.75 AUTOCORRELATION PARAMETER B 0.01000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	97.0	73.7	0.2948	23.27	
1	136.0	133.6	0.2399	2.30	
2	163.0	175.5	0.1675	-12.59	
3	187.0	203.2	0.1106	-16.25	
4	212.0	220.9	0.0709	-8.98	
5	226.0	232.1	0.0446	-6.14	
6	237.0	239.0	0.0277	-2.08	
7	242.0	243.3	0.0171	-1.36	
8	249.0	245.9	0.0104	3.02	
9	250.0	249.9	0.0160	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 1.9531 ESTIMATED VARIANCE 4.6889
SAMPLE MEAN 2.0040 SAMPLE VARIANCE 5.1365

A LEVEL 1.50 AUTOCORRELATION PARAMETER B 0.01000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
C	182.0	182.6	0.7306	-0.65	
1	210.0	212.6	0.1199	-2.64	
2	230.0	226.6	0.0560	3.33	
3	236.0	234.6	0.0319	1.36	
4	240.0	239.5	0.0197	0.42	
5	242.0	242.7	0.0128	-0.78	
6	243.0	244.9	0.0085	-1.93	
7	247.0	246.4	0.0058	0.59	
8	247.0	247.4	0.0040	-0.42	
9	250.0	250.0	0.0102	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 0.7169 ESTIMATED VARIANCE 3.1306
SAMPLE MEAN 0.6920 SAMPLE VARIANCE 2.5352

A LFVEL 2.00   AUTOCORRELATION PARAMETER B 0.01000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	221.0	216.4	0.8656	4.57	
1	231.0	233.6	0.0689	-2.65	
2	236.0	240.6	0.0281	-4.69	
3	241.0	244.3	0.0145	-3.33	
4	243.0	246.4	0.0083	-3.41	
5	248.0	247.6	0.0050	0.31	
6	249.0	248.4	0.0031	0.52	
7	250.0	249.9	0.0061	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 0.3024   ESTIMATED VARIANCE 1.1484
SAMPLE MEAN 0.3240   SAMPLE VARIANCE 1.1676

A LEVEL 2.50 AUTOCORRELATION PARAMETER B 0.01000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	241.0	224.1	0.8966	16.84	
1	243.0	248.6	0.0978	-5.62	
2	246.0	249.9	0.0053	-3.95	
3	246.0	250.0	0.0001	-4.00	
4	248.0	250.0	0.0000	-2.00	
5	248.0	250.0	0.0000	-2.00	
6	250.0	250.0	-0.0000	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.1091 ESTIMATED VARIANCE 0.5158
SAMPLE MEAN 0.1120 SAMPLE VARIANCE 0.4613

A LEVEL 3.00   AUTOCORRELATION PARAMETER B 0.01000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	248.0	241.7	0.9668	6.27	
1	250.0	250.0	0.0331	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.0336   ESTIMATED VARIANCE 0.2894
SAMPLE MEAN 0.0080   SAMPLE VARIANCE 0.0079

A LEVEL 0.75   AUTOCORRELATION PARAMETER B 0.05000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	19.0	10.8	0.0435	8.11	
1	54.0	38.7	0.1115	15.23	
2	88.0	79.2	0.1619	8.74	
3	128.0	123.0	0.1751	4.96	
4	164.0	162.2	0.1570	1.70	
5	197.0	193.1	0.1233	3.86	
6	217.0	215.0	0.0877	1.92	
7	231.0	229.5	0.0578	1.47	
8	242.0	238.4	0.0357	3.53	
9	248.0	243.7	0.0210	4.27	
10	250.0	250.0	0.0250	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 3.8901   ESTIMATED VARIANCE 5.9067
SAMPLE MEAN 3.6480   SAMPLE VARIANCE 5.6266

A LEVEL 1.50 AUTOCORRELATION PARAMETER B 0.05000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	93.0	84.2	0.3368	8.79	
1	148.0	144.9	0.2430	3.03	
2	174.0	184.7	0.1591	-10.75	
3	204.0	209.9	0.1006	-5.92	
4	230.0	225.5	0.0625	4.43	
5	235.0	235.1	0.0384	-0.17	
6	241.0	241.0	0.0234	-0.04	
7	245.0	244.6	0.0142	0.39	
8	247.0	246.7	0.0086	0.23	
9	249.0	248.0	0.0051	0.94	
10	250.0	250.0	0.0077	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 1.7514 ESTIMATED VARIANCE 4.2508
SAMPLE MEAN 1.7360 SAMPLE VARIANCE 4.0906

A LEVEL 2.00 AUTOCORRELATION PARAMETER B 0.05000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	155.0	140.3	0.5613	14.66	
1	194.0	199.4	0.2366	-5.49	
2	220.0	226.3	0.1074	-6.35	
3	238.0	238.8	0.0499	-0.83	
4	242.0	244.6	0.0234	-2.69	
5	246.0	247.4	0.0111	-1.47	
6	249.0	248.7	0.0052	0.20	
7	250.0	249.9	0.0048	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 0.8206 ESTIMATED VARIANCE 1.5973
SAMPLE MEAN 0.8240 SAMPLE VARIANCE 1.8323

A LEVEL 2.50 AUTOCORRELATION PARAMETER B 0.05000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	214.0	181.3	0.7255	32.61	
1	238.0	239.5	0.2328	-1.58	
2	247.0	248.9	0.0373	-1.92	
3	248.0	249.9	0.0039	-1.92	
4	248.0	250.0	0.0003	-2.00	
5	249.0	250.0	0.0000	-1.00	
6	250.0	250.0	-0.0000	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.3208
SAMPLE MEAN 0.2240

ESTIMATED VARIANCE 0.7417
SAMPLE VARIANCE 0.4717

A LEVEL 3.00 AUTOCORRELATION PARAMETER B 0.05000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	241.0	225.1	0.9006	15.84	
1	248.0	248.7	0.0942	-0.72	
2	250.0	250.0	0.0050	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.1047 ESTIMATED VARIANCE 0.4330
SAMPLE MEAN 0.0440 SAMPLE VARIANCE 0.0582

A LEVEL 0.75 AUTOCORRELATION PARAMETER B 0.10000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	4.0	1.6	0.0067	2.30	
1	14.0	9.5	0.0313	4.47	
2	32.0	28.1	0.0745	3.62	
3	64.0	58.6	0.1221	5.30	
4	95.0	97.2	0.1543	-2.29	
5	137.0	137.4	0.1607	-0.47	
6	170.0	173.3	0.1433	-3.30	
7	193.0	201.4	0.1125	-8.44	
8	220.0	221.3	0.0794	-1.30	
9	237.0	234.0	0.0511	2.91	
10	245.0	241.6	0.0303	3.32	
11	248.0	245.8	0.0167	2.12	
12	249.0	248.0	0.0087	0.94	
13	250.0	249.9	0.0077	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 5.4116 ESTIMATED VARIANCE 6.3429
SAMPLE MEAN 5.3680 SAMPLE VARIANCE 6.4584

A LEVEL 1.50   AUTOCORRELATION PARAMETER B 0.10000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	58.0	48.9	0.1956	9.08	
1	104.0	106.6	0.2310	-2.68	
2	142.0	155.1	0.1937	-13.12	
3	177.0	190.2	0.1403	-13.20	
4	203.0	213.6	0.0936	-10.61	
5	226.0	228.4	0.0592	-2.43	
6	242.0	237.4	0.0361	4.52	
7	248.0	242.8	0.0214	5.15	
8	249.0	245.9	0.0124	3.03	
9	250.0	250.0	0.0161	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 2.3430   ESTIMATED VARIANCE 4.6488
SAMPLE MEAN 2.4040   SAMPLE VARIANCE 4.3140

A LEVEL 2.00 AUTOCORRELATION PARAMETER B 0.10000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	124.0	115.6	0.4624	8.38	
1	179.0	184.5	0.2755	-5.51	
2	208.0	219.4	0.1396	-11.41	
3	232.0	236.0	0.0665	-4.05	
4	243.0	243.7	0.0307	-0.74	
5	249.0	247.2	0.0139	1.77	
6	250.0	250.0	0.0111	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 1.0225 ESTIMATED VARIANCE 1.7547
SAMPLE MEAN 1.0600 SAMPLE VARIANCE 1.8799

A LEVEL 2.50 AUTOCORRELATION PARAMETER B 0.10000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	195.0	174.5	0.6982	20.44	
1	237.0	237.2	0.2508	-0.26	
2	246.0	248.5	0.0450	-2.52	
3	249.0	249.9	0.0053	-0.87	
4	250.0	249.9	0.0004	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.3392 ESTIMATED VARIANCE 0.8196
SAMPLE MEAN 0.2920 SAMPLE VARIANCE 0.4003

A LEVEL 3.00 AUTOCORRELATION PARAMETER B 0.10000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	242.0	225.8	0.9033	16.15	
1	249.0	248.7	0.0917	0.20	
2	250.0	249.9	0.0048	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.1016
SAMPLE MEAN 0.0360

ESTIMATED VARIANCE 0.4815
SAMPLE VARIANCE 0.0428

A LEVEL 0.75 AUTOCORRELATION PARAMETER B 0.50000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	0.0	0.0	0.0000	-0.00	
1	0.0	0.0	0.0000	-0.01	
2	0.0	0.0	0.0003	-0.09	
3	0.0	0.4	0.0015	-0.48	
4	0.0	1.8	0.0056	-1.89	
5	2.0	5.7	0.0155	-3.77	
6	9.0	14.3	0.0343	-5.35	
7	26.0	30.0	0.0628	-4.07	
8	53.0	54.3	0.0969	-1.32	
9	77.0	86.2	0.1278	-9.29	
10	111.0	122.7	0.1456	-11.71	
11	149.0	158.8	0.1445	-9.86	
12	173.0	190.3	0.1258	-17.31	
13	202.0	214.4	0.0964	-12.43	
14	222.0	230.7	0.0654	-8.79	
15	235.0	240.6	0.0393	-5.62	
16	244.0	245.8	0.0210	-1.87	
17	248.0	248.3	0.0099	-0.36	
18	249.0	249.4	0.0042	-0.42	
19	249.0	249.8	0.0015	-0.81	
20	249.0	249.9	0.0005	-0.94	
21	250.0	250.0	0.0002	0.00	

THE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 10.[illegible]697 ESTIMATED VARIANCE 7.2353
SAMPLE MEAN 11.0080 SAMPLE VARIANCE 7.7830

A LEVEL 1.50   AUTOCORRELATION PARAMETER B 0.50000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	3.0	5.8	0.0233	-2.83	
1	7.0	25.2	0.0777	-18.26	
2	35.0	59.7	0.1378	-24.72	
3	84.0	102.9	0.1730	-18.99	
4	129.0	146.1	0.1724	-17.12	
5	172.0	182.3	0.1450	-10.38	
6	205.0	209.1	0.1069	-4.13	
7	229.0	226.8	0.0709	2.12	
8	241.0	237.6	0.0431	3.33	
9	247.0	243.7	0.0243	3.24	
10	247.0	246.9	0.0129	0.01	
11	249.0	248.6	0.0064	0.39	
12	250.0	250.0	0.0055	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 4.2630   ESTIMATED VARIANCE 5.4600
SAMPLE MEAN 4.6080   SAMPLE VARIANCE 4.2071

A LEVEL 2.00   AUTOCORRELATION PARAMETER B 0.50000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	50.0	59.0	0.2362	-9.06	
1	128.0	134.8	0.3030	-6.82	
2	183.0	191.5	0.2267	-8.50	
3	226.0	223.8	0.1291	2.19	
4	242.0	239.3	0.0621	2.67	
5	247.0	245.9	0.0265	1.03	
6	248.0	248.5	0.0103	-0.56	
7	249.0	249.5	0.0038	-0.51	
8	250.0	250.0	0.0019	-0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 1.6308   ESTIMATED VARIANCE 2.0733
SAMPLE MEAN 1.7080   SAMPLE VARIANCE 1.9505

A LEVEL 2.50   AUTOCORRELATION PARAMETER B 0.50000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	155.0	155.9	0.6239	-0.97	
1	226.0	229.5	0.2943	-3.56	
2	248.0	246.9	0.0694	1.07	
3	250.0	249.9	0.0123	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.4717   ESTIMATED VARIANCE 0.9764
SAMPLE MEAN 0.4840   SAMPLE VARIANCE 0.4756

A LEVEL 3.00   AUTOCORRELATION PARAMETER B 0.50000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	222.0	225.4	0.9019	-3.49	
1	248.0	248.7	0.0930	-0.76	
2	250.0	249.9	0.0049	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.1032   ESTIMATED VARIANCE 0.5785
SAMPLE MEAN 0.1200   SAMPLE VARIANCE 0.1220

A LEVEL 0.75 AUTOCORRELATION PARAMETER B 1.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	0.0	0.0	0.0000	-0.00	
1	0.0	0.0	0.0000	-0.00	
2	0.0	0.0	0.0000	-0.00	
3	0.0	0.0	0.0000	-0.00	
4	0.0	0.0	0.0000	-0.02	
5	0.0	0.1	0.0004	-0.13	
6	1.0	0.5	0.0015	0.46	
7	2.0	1.7	0.0047	0.26	
8	5.0	4.7	0.0120	0.24	
9	13.0	11.2	0.0259	1.76	
10	24.0	23.2	0.0478	0.79	
11	41.0	42.4	0.0768	-1.40	
12	72.0	69.2	0.1075	2.70	
13	100.0	102.2	0.1320	-2.29	
14	132.0	137.9	0.1425	-5.94	
15	172.0	171.8	0.1357	0.11	
16	195.0	200.3	0.1140	-5.39	
17	225.0	221.5	0.0845	3.47	
18	234.0	235.3	0.0552	-1.32	
19	239.0	243.2	0.0317	-4.26	
20	242.0	247.2	0.0159	-5.26	
21	249.0	249.0	0.0070	-0.02	
22	250.0	250.0	0.0039	-0.00	

THE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 13.9498 ESTIMATED VARIANCE 7.5813
SAMPLE MEAN 14.2160 SAMPLE VARIANCE 8.3386

A LEVEL 1.50   AUTOCORRELATION PARAMETER B 1.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	0.0	1.3	0.0053	-1.33	
1	6.0	8.0	0.0270	-2.08	
2	21.0	25.4	0.0693	-4.41	
3	46.0	55.4	0.1199	-9.40	
4	73.0	94.8	0.1576	-21.81	
5	118.0	136.7	0.1677	-18.74	
6	169.0	174.3	0.1504	-5.35	
7	199.0	203.6	0.1170	-4.61	
8	224.0	223.7	0.0806	0.23	
9	238.0	236.2	0.0499	1.75	
10	244.0	243.2	0.0281	0.72	
11	248.0	246.9	0.0145	1.08	
12	249.0	248.6	0.0069	0.33	
13	250.0	249.9	0.0053	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 5.4084 ESTIMATED VARIANCE 5.7737
SAMPLE MEAN 5.6600 SAMPLE VARIANCE 5.1248

A LEVEL 2.00   AUTOCORRELATION PARAMETER B 1.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	33.0	38.5	0.1543	-5.59	
1	103.0	106.9	0.2734	-3.96	
2	156.0	171.0	0.2561	-15.00	
3	208.0	213.1	0.1686	-5.17	
4	238.0	235.0	0.0875	2.93	
5	244.0	244.6	0.0381	-0.60	
6	249.0	248.2	0.0145	0.76	
7	250.0	249.9	0.0070	0.00	

THE NEGATIVE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 1.9724   ESTIMATED VARIANCE 2.1960
SAMPLE MEAN 2.0760   SAMPLE VARIANCE 2.1347

A LEVEL 0.75   AUTOCORRELATION PARAMETER B 3.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	0.0	0.0	0.0000	-0.00	
1	0.0	0.0	0.0000	-0.00	
2	0.0	0.0	0.0000	-0.00	
3	0.0	0.0	0.0000	-0.00	
4	0.0	0.0	0.0000	-0.00	
5	0.0	0.0	0.0000	-0.00	
6	0.0	0.0	0.0000	-0.00	
7	0.0	0.0	0.0001	-0.03	
8	0.0	0.1	0.0003	-0.13	
9	0.0	0.4	0.0012	-0.44	
10	1.0	1.3	0.0035	-0.34	
11	2.0	3.5	0.0087	-1.51	
12	12.0	8.1	0.0186	3.82	
13	23.0	16.9	0.0351	6.03	
14	36.0	31.6	0.0586	4.38	
15	59.0	53.2	0.0866	5.71	
16	82.0	81.7	0.1138	0.24	
17	111.0	114.9	0.1329	-3.99	
18	159.0	149.4	0.1379	9.51	
19	193.0	181.2	0.1271	11.71	
20	207.0	207.2	0.1039	-0.27	
21	221.0	226.0	0.0751	-5.05	
22	237.0	238.0	0.0478	-1.01	
23	244.0	244.6	0.0267	-0.69	
24	246.0	247.9	0.0129	-1.94	
25	249.0	249.3	0.0054	-0.31	
26	250.0	250.0	0.0027	0.00	

THE BINOMIAL MODEL WAS SELECTED.

ESTIMATED MEAN 17.5446   ESTIMATED VARIANCE 8.0941
SAMPLE MEAN 17.6720   SAMPLE VARIANCE 8.7835

A LEVEL 1.50 AUTOCORRELATION PARAMETER B 3.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	0.0	0.3	0.0015	-0.37	
1	1.0	2.8	0.0098	-1.84	
2	6.0	10.8	0.0319	-4.83	
3	24.0	28.1	0.0691	-4.13	
4	49.0	56.2	0.1122	-7.20	
5	87.0	92.6	0.1457	-5.62	
6	140.0	132.0	0.1576	7.96	
7	175.0	168.5	0.1461	6.43	
8	209.0	198.1	0.1185	10.80	
9	232.0	219.5	0.0854	12.43	
10	241.0	233.4	0.0554	7.56	
11	247.0	241.6	0.0327	5.38	
12	248.0	246.0	0.0177	1.95	
13	249.0	248.2	0.0088	0.74	
14	250.0	250.0	0.0069	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 6.4899 ESTIMATED VARIANCE 6.2380
SAMPLE MEAN 6.3680 SAMPLE VARIANCE 4.7877

A LEVEL 2.00   AUTOCORRELATION PARAMETER B 3.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	23.0	22.6	0.0906	0.34	
1	79.0	77.0	0.2175	1.95	
2	137.0	142.3	0.2612	-5.34	
3	193.0	194.6	0.2090	-1.61	
4	230.0	225.9	0.1255	4.00	
5	242.0	241.0	0.0602	0.93	
6	249.0	247.0	0.0241	1.90	
7	250.0	249.9	0.0116	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 2.4012   ESTIMATED VARIANCE 2.3773
SAMPLE MEAN 2.3880   SAMPLE VARIANCE 2.2464

A LEVEL 2.50   AUTOCORRELATION PARAMETER B 3.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	144.0	126.4	0.5057	17.55	
1	226.0	212.6	0.3447	13.35	
2	246.0	242.0	0.1175	3.98	
3	250.0	249.9	0.0319	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.6816
SAMPLE MEAN 0.5360

ESTIMATED VARIANCE 1.1250
SAMPLE VARIANCE 0.5067

A LEVEL 3.00 AUTOCORRELATION PARAMETER B 3.00000

NUMBER OF OVERSHOOTS	CUMMULATIVE SAMPLE FREQUENCY	CUMMULATIVE PREDICTED FREQUENCY	PREDICTED PROBABILITY (X=I)	SAMPLE MINUS PREDICTED	COMMENTS
0	236.0	215.5	0.8620	20.48	
1	250.0	250.0	0.1379	0.00	

THE POISSON MODEL WAS SELECTED.

ESTIMATED MEAN 0.1484 ESTIMATED VARIANCE 0.6699
SAMPLE MEAN 0.0560 SAMPLE VARIANCE 0.0530