copy 2

DISSERTATION

INVESTIGATION OF WET AND DRY YEARS BY RUNS

Submitted by

Jaime Saldarriaga

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

December 1969

CED69-70JS-12

ABSTRACT OF DISSERTATION

INVESTIGATION OF WET AND DRY YEARS BY RUNS

A technique is advanced for testing the structure of time series, with the basic statistical parameter being the mean run-length. This technique is shown to be distribution-free, as opposed to other techniques of investigation of time series like autocorrelation and variance spectrum analyses which are based on the assumption that the underlying variable is normally distributed. Application of technique to selected precipitation and river gaging stations is presented.

Analytical expressions are developed by which the probabilities of sequences of wet and dry years of specified lengths can be calculated when the basic hydrologic time series are either independent or stationary dependent Gaussian processes, and the truncation level specified.

Numerical values of these probabilities are obtained by means of digital computer for a first order autoregressive process. A set of tables and a set of graphs are presented in order to make the numerical values useable. The probabilities of runs of dependent variables with a common distribution do not depend on the underlying univariate distribution of the variable for a given specified truncation level.

The significance of the investigation is based on the need of accurate stochastic models of hydrologic time series for the generation of samples of annual precipitation and annual river flow, for the planning design and operation of water resource systems. It is also based on the concept that runs, as statistical properties of runs, represent the best basic concept for an objective definition of drought.

Jaime Saldarriaga
Department of Civil Engineering
Colorado State University
Fort Collins, Colorado
December 1969

ACKNOWLEDGMENTS

The author appreciates sincerely the counsel and valuable suggestions given by his major professor and adviser, Dr. Vujica Yevjevich, Professor of Civil Engineering. Special thanks are extended to Dr. Mohammed M. Siddiqui, Professor in the Department of Mathematics and Statistics, whose guidance and encouragement have been definitive in the author's work.

The author is also thankful to the other members of his graduate committee, Dr. Petar Todorovic, Associate Professor of Civil Engineering, Dr. Melvin E. Holland, Assistant Professor of Civil Engineering and Dr. Stanley A. Schumm, Associate Professor of Geology for their helpful suggestions.

The sponsorship of this research is gratefully acknowledged. This research was supported by the National Science Foundation (grants GK-1661 and GK-11444). The graduate research assistanship that made possible the author's studies was obtained with the assistance of the Fulbright Commission of the U. S. Department of State Educational and Cultural Exchange Program and the Institute of International Education. To these organizations the author is also thankful.

Sincere appreciation is extended to Mrs. Betty Hutcheson for typing the text and to Miss Hanae Akari for drafting the figures of this dissertation.

Finally, to his wife for her encouragement the author gives a special vote of thanks.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF FIGURES

NOTATION

r_k	k-lag serial correlation coefficient
N	Sample size
ρ_k	k-lag autocorrelation coefficient (population value)
μ	Population mean
σ	Population standard deviation
$O(\rho^3)$	Of the order of ρ^3
S_n^+	Surplus
S_n^-	Deficit
R_n	Range
$\{e_i\}$	Sequence of independent variables with a common distribution
K_j^i	Number of runs of kind i of length j
K^i	Number of runs of kind i
K	Number of total runs. Total number of runs.
N_o, N_1	Number of elements of kinds 0 and 1 respectively in a binomial population
N	$N_o + N_1$
α	N_o/N_1. Also level of significance
u	Truncation level
q	Probability of drawing an element of kind 1 in a binomial population. Also $F(u)$
p	Probability of drawing an element of kind 0 in a binomial population. Also $1-F(u) = 1-q$
N_j^+	j-th positive run in discrete time
N_j^-	j-th negative run in discrete time
N_j	j-th total run in discrete time

$\Gamma(r)$	Gamma function of r
ρ_{ij}	Correlation coefficient between X_i and X_j
$P_m(j^+)$	Probability that $X_1, X_2, \ldots, X_j$ are simultaneously positive for a probability truncation level $q = .5$
$P(j^+)$	Same as previous one for any q
T_j^+	j-th positive run in continuous time
T_j^-	j-th negative run in continuous time
$P_m(k^-, j^+)$	Probability that $X_1, X_2, \ldots, X_k$ are simultaneously negative and $X_{k+1}, X_{k+2}, \ldots, X_{k+j}$ are simultaneously positive for a probability truncation level q.
$P(k^-, j^+)$	Same as previous one for any q.
E	Expectation
var	Variance
$\gamma(K)$	autocorrelation function
$f(x)$	Probability density function of x
$F(x)$	Probability distribution function of x
R	Matrix of correlation coefficients
X^*	Modified variables so that the limits of the multinormal integral are zero and infinity.
ρ^*_{ij}	Correlation coefficient between X_i^* and X_j^*
R^*	Matrix of modified correlation coefficients
$H_i(x)$	Hermite polynomial of order i.
δ	Parameter used in the relationship between mean run-length and truncation probability level.

Chapter I

INTRODUCTION

1.1 Significance of this Investigation

Annual precipitation and annual river flow vary from year to year. This variation is generally referred to as the sequence of wet and dry years, represented either by the annual precipitation or by the annual river flow processes. These sequences are hydrologic time series. For all practical purposes in water resources development, they can be assumed to be stationary time series. The hydrologic stationary time series of annual river flows are dependent, which means that the successive values are linked in some persistent manner. In other words, sequences of annual river flows are stationary serially dependent time series. However, the sequences of annual precipitation are very close to stationary serially independent time series.

The water resource engineer is greatly concerned with predicting the future behavior of these sequences for the planning, design and operation of water resource systems. Overyear flow regulation, which permits storage of water during wet years to be released during dry years, is a common practice. Recent optimization techniques are being introduced into the water resource field. The optimization of both design of new systems and operation of already existing systems on an annual basis requires the use of accurate stochastic models of time series as inflows. This permits the prediction of statistical properties and the eventual generation of various samples of annual river flows by the use of the Monte Carlo method.

One of the least known properties of hydrologic variables is the occurrence of severe and prolonged droughts. Their properties are not known with sufficient accuracy to allow the prediction of their occurrence and duration with any real degree of reliability. It is believed that runs as statistical properties of time series represent the best basic concept for an objective definition of droughts (Yevjevich, 1967)*. Therefore, this investigation of runs and their application to series of wet and dry years is related to some of the most significant problems of hydrology and water resource development.

1.2 Objectives of the Study

The objective of this study is to develop a technique for investigating stationary independent and dependent hydrologic time series, whose basic statistical parameters are runs. Four phases are involved in this investigation:

(a) Mathematical formulation of the problem;

(b) Selection of suitable parameters for testing

---

*References are designated by a year or years in parenthesis associated with the author's name and may be found in the list of references.

hypotheses of stationarity and time dependence;

(c) Statistical inference for stationarity and time dependence models, and

(d) Application of technique to selected precipitation and river gaging stations.

There are basically four methods based on specific statistical parameters for the investigation of hydrologic time series:

1. <u>Autocorrelation analysis</u>. The parameters involved are autocorrelation coefficients as functions of the time-lag between correlated values which are defined as

$$\rho_k = \frac{\mathrm{cov}(x_i x_{i+k})}{\mathrm{var}\ x_i} \qquad 1.1$$

for stationary sequences.

The use of autocorrelation analysis as an investigation technique of hydrologic time series is based on the concept of analogy. One should know the correlograms of particular time series, and then by statistical inference determine whether a computed correlogram of a hydrologic time series is well approximated by the correlogram of known processes (Yevjevich, 1969). An illustrative example is presented in Chapter 5.

2. <u>Variance spectrum analysis</u>. Basically this is the Fourier series analysis in which an infinite number of small oscillations with a continuous distribution of

frequencies is fitted to an observed series. The parameters involved are the variance densities of various harmonics represented against their frequencies. The variance of a harmonic is equal to its amplitude squared. This type of analysis is a representation of the process in the frequency domain while the autocorrelation is a representation in the time domain. It might be noted that for a second order stationary process the spectral density function is the Fourier transform of the autocorrelation function in the frequency domain. The use of variance spectrum analysis as an investigation technique of hydrologic time series is also based on the concept of analogy. One should determine by statistical inference whether a computed variance spectrum of a hydrologic time series is well approximated by the variance spectrum of known processes. An example is given in Chapter 5.

3. Ranges. The parameters involved are defined in terms of differences between maximum and minimum cumulative sums of departures of values from the average or other values for given subseries sizes, with the ranges represented as random variables for each subsample size. Let $\{X_i;\ i = 1,\ldots,N\}$ be the observed sequence, and let u be a specified truncation level which in general represents a target level. Then surplus is defined by Yevjevich (1965) as

$$S_n^+ = \max \{0, \sum_{i=1}^{j} (X_i - u);\ j = 1,2,\ldots,n\} \qquad 1.2$$

and deficit is defined as

$$S_n^- = \min \{0, \sum_{i=1}^{j} (X_i - u);\ j = 1,2,\ldots,n\}\ , \qquad 1.3$$

where n is the size of a subsample taken from $\{X_i\}$. Finally range is defined as

$$R_n = S_n^+ - S_n^- \ . \qquad 1.4$$

As in the case of autocorrelation and variance spectrum analyses the use of ranges as an investigation technique of hydrologic time series may be based on the concept of analogy. The comparison of the observed sample function of R_n as a function of n with the expected function of the same parameter of known processes, and the distribution of the observed parameter allows one to make statistical inference of the goodness of fit of theoretical models.

4. Runs. For the purposes of this study the concept of run is identical to the concept of run-length. Basically they are the number of consecutive positive or negative departures from a specified constant value called

the truncation level. Positive runs are associated with positive departures and negative runs with negative departures. The structure of a time series is analyzed by studying the properties of runs at different truncation levels. On the other hand, these statistics have a very practical meaning in hydrology because a positive run can be immediately associated with the duration of a wet period or a surplus time interval, while a negative run can be associated with the duration of a drought or a deficit time interval.

1.3 Runs as Compared to other Techniques of Time Series Investigation

The two classical techniques for investigation of stationary time series are the autocorrelation analysis and the variance spectrum analysis. The use which is made of either of them in exploring the internal structure of a time series depends to some extent on the purpose of the inquiry and prior knowledge of the generating system. The correlogram tells us something about the relationship between consecutive values of the series which are separated in time. The spectrum exhibits the extent to which the series is in step with certain fundamental rhythms (Kendall and Stuart 1966).

Ranges and runs are two techniques, which can be used advantageously in the hydrologic decision making process because they can be immediately associated with the concepts of storage and droughts, respectively, and with the concepts of surplus and deficit, which are of particular interest to the hydrologist.

If a target level is specified, the time associated with a negative run represents the duration of a deficit relative to the target level. The frequency of specified deficit periods is relevant for the planning, design and operation of hydrologic systems.

Furthermore, the structure of a time series is reflected in the properties of runs that it generates at specified truncation levels. For example, independent variables with a common distribution are characterized by a mean run-length equal to two for a truncation level equal to the median of the distribution of the variables. Identically distributed variables with a highly positive first order serial correlation coefficient are characterized by a mean run-length greater than two at the same level. On the other hand, identically distributed variables with a highly negative first order serial correlation coefficient are characterized by a mean run-length smaller than two at the same level. These properties, which will be described in detail in later chapters, justify the use of runs not only as a technique to use in the decision making process but also as a technique for the investigation of time series and more specifically for testing stationarity and dependence models of hydrologic time series.

1.4 Whitening and Use of Original Series

Two different approaches may be taken for the analysis of hydrologic time series. The first approach consists in analyzing the original sequence by one or several of the techniques outlined in section 1.2. It is here referred to as the use of original series. The second approach assumes a model for the process that is composed of a systematic component and an independent component. A residual series is obtained by subtracting the systematic component from the original series. Under the hypothesis that the assumed model is an adequate representation of the process, the residual series must be a sequence of independent variables. The independence of the residual series is tested and the assumed model is accepted or rejected depending on whether the independence of the residual series was accepted or rejected. This procedure is here referred to as "whitening", meaning that the residual series is expected to be "white noise".

It is perhaps interesting to emphasize a basic difference between these two approaches. The first approach does not assume a model for the process, but rather it leads to the discovery of the structure of the process so that eventually a model can be fitted to it. The second approach starts by assuming a model for the process so that a previous knowledge of the process is required as a basis for its assumption.

1.5 Inference on Time Series Mathematical Models and their Properties

Inference on time series mathematical models is an important phase in the statistical analysis of hydrologic time series because it permits the testing of hypotheses about mathematical models ascribed to the observed sequences. The statistical inference is made about the significance of the parameters involved in the type of analysis that is being performed, so as to test whether or not the expected parameters, under the hypothetical model, are significantly different from the observed parameters at a specified confidence level. If they are significantly different, the hypothetical model is rejected. Otherwise, it is accepted.

In order to perform this test it is necessary to know the distribution of the parameter that is going to be tested under the hypothesis that the assumed model is valid. Finally the $1 - \alpha$ confidence limits of the parameter are set up and the observed value is tested at this level, where α is the level of significance.

In order to illustrate the problem of statistical inference, consider the autocorrelation analysis of a hydrologic time series. Let $\{X_i; i = 1,\ldots N\}$ be the observed time series and r_k the lag k serial correlation coefficient of $\{X_i\}$. The shape of the correlogram gives condensed information about the process. For an independent sequence all population autocorrelation coefficients ρ_k are zero and therefore all sample serial correlation coefficients r_k must be not significantly different from zero. Under the hypothesis that $\{X_i\}$ is a sequence of normal independent variables, r_1 is approximately normally distributed with mean and variance given by Siddiqui (1957) as

$$Er_1 = -\frac{1}{N} \qquad \text{var}(r_1) = \frac{N^3-3N^2+4}{N^2(N^2-1)} \tag{1.5}$$

The 95% confidence limits of r_1 are easily obtained by the writer as

$$-\frac{1}{N} \pm 1.96\sqrt{\frac{N^3-3N^2+4}{N^2(N^2-1)}} \tag{1.6}$$

If r_1 is inside these confidence limits the sequence $\{X_i\}$ is accepted as a linearly independent sequence. Otherwise it is accepted as a dependent one. In this case the statistical inference has been made by analyzing the original series.

The statistical inference can also be made on the whitened series. Let us set up the hypothesis that $\{X_i\}$ is a first-order autoregressive process. Then

$$X_i - u = \rho_1 (x_{i-1} - \mu) + \sigma e_i \quad . \qquad 1.7$$

The whitened series is

$$e_i = \frac{1}{\sigma} \left[X_i - u - \rho_1 (X_{i-1} - \mu) \right] \quad . \qquad 1.8$$

Under the given hypothesis, $\{e_i\}$ is a sequence of independent variables. The parameters μ , σ and ρ_1 are estimated from the observed sequence. Then the whitened series is tested for independence as was done with the original series.

The problem of statistical inference is basically the same as when the other techniques, namely the variance spectrum, ranges, and runs are used. A hypothesis is set up and the significance of the parameter is question is tested under the given hypothesis at a $1 - \alpha$ confidence level. The distribution of the parameter involved is different in each case however.

Chapter II

REVIEW OF LITERATURE

2.1 Introductory Statement

Two main aspects are reviewed; first, the distribution theory of runs for both independent and dependent random variables. Second, the multivariate normal integral, which is a base for the mathematical developments in Chapter III.

2.2 Distribution Theory of Number of Runs of Independent Random Variables

A run is defined, in probability theory, as a succession of similar events preceded and succeeded by different events. The number of elements in a run is usually referred to as its length. The classical distribution theory of runs has been mainly concerned with independent arrangements of a fixed or random number of several kinds of elements. In the case of two different kinds of elements, let us assume that the number of elements N_o and N_1 of each kind are randomly drawn from a binomial population with probabilities p and q. Let K_i^o denote the number of runs of kind (o) of length i , and let $K_i^!$ denote the number of runs of kind (1) of length i. Finally let

$K^o = \sum_i K_i^o$ designate the total number of runs of elements N_o ,

$K' = \sum_i K_i'$ the total number of runs of elements N_1, with

$K = K^o + K'$ and $N = N_o + N_1$.

Wishart and Hirshfeld (1936) obtained and tabulated the following joint probabilities:

$$P(K^o = k^o \ , \ N_o = n_o \ , \ N_1 = n_1) = \binom{n_o - 1}{k_o - 1} \binom{n_1 + 1}{k_o} p^{n_o} q^{n_1} \ , \qquad 2.1$$

$$P(K' = k' \ , \ N_o = n_o \ , \ N_1 = n_1) = \binom{n_1 - 1}{k_1 - 1} \binom{n_o + 1}{k_1} p^{n_1} q^{n_o} \ , \qquad 2.2$$

$$P(K = 2m \ , \ N_o = n_o \ , \ N_1 = n_1) = 2 \binom{n_o - 1}{m - 1} \binom{n_1 - 1}{m - 1} p^{n_o} q^{n_1} \ , \qquad 2.3$$

and

$$P(K = 2m+1 \ , \ N_o - n_o \ , \ N_1 = n_1) = \left[\binom{n_o - 1}{m - 1} \binom{n_1 - 1}{m} + \binom{n_o - 1}{m} \binom{n_1 - 1}{m - 1} \right] p^{n_o} \ q^{n_1} \ . \qquad 2.4$$

As the sample size increases, K is asymptotically normally distributed with

$$EK = 2npq + p^2 + q^2 = 2(n-1)pq + 1 \qquad 2.5$$

and

$$\text{var } K = 4npq(1-3pq) - 2pq(3-10pq) \quad . \tag{2.6}$$

Cochran (1936) gives the following relations

$$EK^{o} = p + (n-1)pq \tag{2.7}$$

$$EK' = q + (n-1)pq \tag{2.8}$$

$$EN_o = np \ , \quad \text{and} \quad EN_1 = nq \quad . \tag{2.9}$$

Stevens (1939) gives the distribution of the total number of runs (without a regard to their length) from arrangements of two kinds of elements. He also developed a χ^2 criterion for tests of significance. Wald and Wolfowitz (1940) published the same distribution as Stevens and showed that it was asymptotically normal. The conditional distributions of K are

$$P(K=2m \mid n_o \ , \ n_1) = \frac{2\binom{n_o-1}{m-1}\binom{n_1-1}{m-1}}{\binom{n}{n_o}} \tag{2.10}$$

and

$$P(K=(2m+1) \mid n_o \ , \ n_1) = \frac{\binom{n_o-1}{m-1}\binom{n_1-1}{m} + \binom{n_o-1}{m}\binom{n_1-1}{m-1}}{\binom{n}{n_o}} \ , \tag{2.11}$$

which are independent of the parameter p .

For $n_o = \alpha n_1$, $\alpha > 0$ and $n_o \to \infty$, they are given by Wald and Wolfowitz (1940) as a normal distribution with

$$EK = \frac{2n_o}{1+\alpha} \quad , \qquad \text{var}K = \frac{4\alpha n_o}{(1+\alpha)^3} \quad . \tag{2.12}$$

For $\alpha = 1$, the statistic

$$z = \frac{K-n_o}{\sqrt{\frac{n_o}{2}}} \tag{2.13}$$

is a standard normal variable.

Mood (1942) derived distributions of runs of a given length for independent arrangements of both the fixed number of elements of two or more kinds, and the binomial and multinomial populations. He also showed these distributions become asymptotically normal as the sample size increases. He gives the following results

$$EK_i^o = p^i q \; [(n-i-1)q + 2] \tag{2.14}$$

and

$$EK'_i = q^i p \; [(n-i-1)p + 2] \quad . \tag{2.15}$$

The statistic $$x = \frac{k - 2npq}{2\sqrt{npq(1-3pq)}} \tag{2.16}$$

is asymptotically normally distributed with mean zero and variance unity. Comparing Eqs. 2.5 and 2.6 with Eq. 2.16 the mean and variance given by Mood and the mean and variance given by Wishard and Hirshfeld are slightly different. The latter equation is an approximation of the former ones. Bendat and Piersol (1966) give tables for the conditional distribution of K when $N_o = N_1 = N/2$.

2.3 Distribution Theory of Run-Length of Independent Random Variables

Let N_j^+ and N_j^- denote the positive and negative j-th run-length for the given truncation level u. Also, let $\{X_n\}$ be the sequence of independent random variables with the common distribution $F(x)$, with $F(u) = q$, and $1-F(u) = p$ and let

$$\{N_j\} = \{N_j^+ + N_j^-\}$$

be the sequence of total runs.

The probability mass function of N_1 is given by Heiny (1968) as,

$$P(N_1=k) = \frac{pq^k-qp^k}{q-p} \qquad 2.17$$

for $k = 2,3,\ldots$ with

$$EN_1 = \frac{1}{pq} \quad \text{and} \quad varN_1 = \frac{1-3pq}{p^2q^2} \quad . \qquad 2.18$$

The number of total runs $K(N)$ in a discrete time series of length N is given by Feller (1957) and has the following properties

$$EK(N) = (N-1)pq \qquad \text{for} \quad N \geq 1 \quad , \qquad 2.19$$

$$varK(M) = Npq(1-3pq - \frac{1}{N} + \frac{5}{N}pq) \quad \text{for} \quad N \geq 4 \qquad 2.20$$

and its distribution is asymptotically normal. Downer, Siddiqui and Yevjevich (1967) studied the distribution of positive and negative run-lengths (runs above and runs below a specified truncation level) for sequences of independent identically distributed random variables and applied it to the normal case. They have shown that $\{N_j^+\}$ is a sequence of independent identically distributed random variables with probability mass functions

$$P(N_j^+ = k) = qp^{k-1} \ , \ P(N_j^- = k) = pq^{k-1} \qquad 2.21$$

and moments are

$$EN_j^+ = \frac{1}{q} \quad , \quad EN_j^- = \frac{1}{p} \qquad 2.22$$

$$\text{var } N_j^+ = \frac{p}{q^2} \quad , \quad \text{var } N_j^- = \frac{q}{p^2} \quad . \qquad 2.23$$

For the case with $p = q = 1/2$,

$$P(N_j^+ = k) = P(N_j^- = k) = \frac{1}{2^k} \quad , \qquad 2.24$$

$$EN_j^+ = EN_j^- = 2 \qquad 2.25$$

and

$$\text{var } N_j^+ = \text{var } N_j^- = 2 \quad . \qquad 2.26$$

Llamas (1968) studied the case of standard one-parameter Gamma random variables with probability density function

$$f(x) = \int_{-\sqrt{\alpha}}^{x} \frac{\alpha}{\Gamma(\alpha)} \exp(-\alpha-t\sqrt{\alpha})(\alpha+t\sqrt{\alpha})^{\alpha-1}\, dt \quad . \qquad 2.27$$

He obtained for $u = 0$, $p = F(0) = P(\alpha,\alpha)$ and $q = 1-P(\alpha,\alpha)$ where $P(\alpha,\alpha)$ is the incomplete Gamma function

$$P(\alpha,\alpha) = \frac{1}{\Gamma(\alpha)} \int_{o}^{\alpha} e^{-t}\, t^{\alpha-1}\, dt \quad . \qquad 2.28$$

Llamas and Siddiqui (1969) studied the case of a sequence of a two-dimensional process (X_n, Y_n) , $n = 1, 2, \ldots,$ where these two variates are actually independent and have a common distribution function $F(x,y)$. Given two levels, u_1 and u_2, such that $0 < F(u_1,u_2) < 1$, they defined four possible events

$$A_n = \{X_n \leq u_1, Y_n \leq u_2\}, \quad B_n = \{X_n \leq u_1, Y_n > u_2\}$$

$$C_n = \{X_n > u_1, Y_n \leq u_2\}, \text{ and } D_n = \{X_n > u_1, Y_n > u_2\}.$$

The n-th value is associated with the sign minus for both sequences, if A_n occurs, and with the sign plus if D_n occurs. A sequence of k consecutive A's followed and preceded by any other event is a negative run of length k. A sequence of k consecutive D's, followed and preceded by any other event, is a positive run of length k, and for the initial run the requirement of "preceded by" is dropped. Then if A_n^c is the set complement of A_n and

$$P(A_n) = F(u_1,u_2) = q \quad ,$$

$$P(A_n^c) = p$$

they have shown

$$P(N_j^- = k) = pq^{k-1} \qquad 2.29$$

$$EN_j^- = \frac{1}{p} \quad \text{and} \quad \text{var } N_j^- = \frac{q}{p^2} \tag{2.30}$$

analogous relationships hold for N_j^+ for its corresponding values p and q.

2.4 Distribution Theory of Runs of Dependent Random Variables

For a Markov chain with two states (0) and (1) Cox and Miller (1965) give the distribution of the recurrence time of state 0, designated by N^o, as

$$P(N^o = k) = \alpha\beta(1-\beta)^{k-2} \quad , \quad \text{for} \quad k = 2,3,\ldots \tag{2.31}$$

and

$$P(N^o = 1) = 1-\alpha \quad . \tag{2.32}$$

The transition probability matrix is

$$P = \begin{array}{c} 0 \\ 1 \end{array} \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \quad . \tag{2.33}$$

The mean run-length of the state 0 is

$$EN^o = \frac{\alpha+\beta}{\beta} \quad . \tag{2.34}$$

Similar relationships hold for the recurrence time of state 1, designated by N', by interchanging α and β.

Heiny (1968) defines the following variables

$$P(X_j > u | X_{j-1} > u) = r$$

$$P(X_j \leq u | X_{j-1} > u) = s$$

with $r + s = 1$. He found the following relationships to be valid for a Markov Gaussian provess $\{X_i\}$

$$P(N^+ = k | X_1 > 0) = sr^{k-1}\left[1 + 0(\rho^2)\right], \; k = 1,2,3,.. \qquad 2.35$$

$$E(N^+ | X_1 > 0) = \frac{1}{s}\left[1 + 0(\rho^2)\right] \qquad 2.36$$

and

$$\text{var } (N^+ | X_1 > 0) = \frac{r}{s^2}\left[1 + 0(\rho^2)\right] \qquad 2.37$$

where $0(\rho^2)$ indicates an expression which becomes negligible for small values of ρ. He also found an approximation for the conditional joint probability mass function of the first j positive and negative runs, given $X_1 > 0$, as follows,

$$P(N_j^+ = n_j, N_j^- = m_j, N_{j-1}^+ = n_{j-1}, \ldots, N_1^+ = n_1, N_1^- = m_1 | X_1 > 0)$$

$$= sr^{n_1-1} tv^{m_1-1} sr^{n_2-1} tv^{m_2-1} \ldots sr^{n_j-1} tv^{m_j-1} \left[1+0(\rho^2)\right] \qquad 2.38$$

where

$$t = P(X_j > u | X_{j-1} \leq u), \quad v = P(X_j \leq u | X_{j-1} \leq u), \text{ and}$$

$$t + v = 1 \quad .$$

This treatment, however, has the disadvantages that it is based on a conditional probability that $X_1 > u$, and it is applicable only to very small values of ρ since errors $0(\rho^2)$ may be large for larger values of ρ.

2.5 The Multivariate Normal Integral

Gupta (1963a) presents an exhaustive bibliography on the multinormal integral and related topics, and also gave a survey (Gupta 1963b) of the work on the same topic. For a detailed review, the reader is referred to these two references. Only that part, which is not overlapping with these references, and related to mathematical developments in the following chapters is reviewed here. The multinormal integral is involved in the theory of runs of dependent normal variables because it is directly related to the following problem of h autocorrelated random variables, $Z_1, Z_2, \ldots, Z_h$. If these variables follow the standard multivariate normal distribution, the problem to solve is the probability that all h variables are simultaneously positive. Let us define a new sequence of random variables as follows

$$x_i = \begin{cases} 1 & \text{for} \quad z_i \geq 0 \\ -1 & \text{for} \quad z_i < 0 \end{cases} .$$

Now the problem is the probability that all h variables are simultaneously positive. Let this probability be $P_m(h^+)$ where the subindex m indicates that the truncation level of definition of $\{x_i\}$ corresponds to the median of the distribution of $\{Z_i\}$. For $r_{ij} = EX_iX_j$ McFadden (1955) gives for any $h \geq 4$

$$P_m(h^+) = 2^{-h} \Big[1 + \sum_{j>i\geq 1} r_{ij} + \sum_{\ell>k>j>i\geq 1} (r_{ij}r_{k\ell} + r_{ik}r_{j\ell}$$

$$+ r_{i\ell}r_{jk}) + O(r^3)\Big] . \qquad 2.39$$

For $\rho_{ij} = E\, Z_iZ_j$,

$$r_{ij} = \frac{2}{\pi} \arcsin \rho_{ij} = \frac{2}{\pi} \left[\rho_{ij} + O(\rho_{ij}^3)\right] . \qquad 2.40$$

If Eq. 2.40 is substituted into Eq. 2.39, Eq. 2.39 becomes

$$P_m(h^+) = 2^{-h}\left[1 + \frac{2}{\pi}\sum_{j>i\geq 1} \arcsin \rho_{ij} + \frac{4}{\pi^2}\sum_{\ell>k>j>i\geq 1} (\rho_{ij}\rho_{k\ell}\right.$$

$$\left. + \rho_{ik}\rho_{j\ell} + O(\rho^3)\right] . \qquad 2.41$$

Obviously for the univariate case, Eq. 2.41 becomes

$$P_m(1^+) = \frac{1}{2} \quad . \qquad 2.42$$

For the bivariate case the result is known as Sheppard's (1898) theorem of median dichotomy and is expressed as

$$P_m(2^+) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho. \qquad 2.43$$

This Eq. 2.43 is tabulated by the National Bureau of Standards (1959) for ρ varying from 0 to 1, with increments of 0.01. For the trivariate case the following result is given by David (1953) as

$$P_m(3^+) = \frac{1}{8} + \frac{1}{4\pi}(\arcsin \rho_{12} + \arcsin \rho_{13} + \arcsin \rho_{23}) \quad .$$

2.44

Chapter III

MATHEMATICAL FORMULATION OF THE PROBLEM

3.1 Definition of Runs

A series is cut at many places by an arbitrary horizontal truncation level u, and the relationship of this constant u to all other values of the process serves as a basis for the definition of runs in this study.

The number of values of a discrete sequence between an upcrossing of the truncation level and the following downcrossing is defined in this study as a positive run. Similarly, a negative run is defined as the number of values of a discrete series between a downcrossing and the next upcrossing as shown in the lower graph of Fig. 3.1. They are designated as N_j^+ = the length of the j-th positive run, and N_j^- = the length of the j-th negative run. The j-th total run is defined as

$$N_j = N_j^+ + N_j^- \quad , \quad \text{with} \quad j = 1, 2, \ldots$$

The definitions for T_j^+, T_j^- and T_j as intervals of positive and negative runs of a continuous process are analogous to the definitions of runs of discrete time series. Other parameters used in literature as definitions of various runs of discrete time series are:

1. Sum of deviations associated with positive runs.
2. Sum of deviations associated with negative runs.

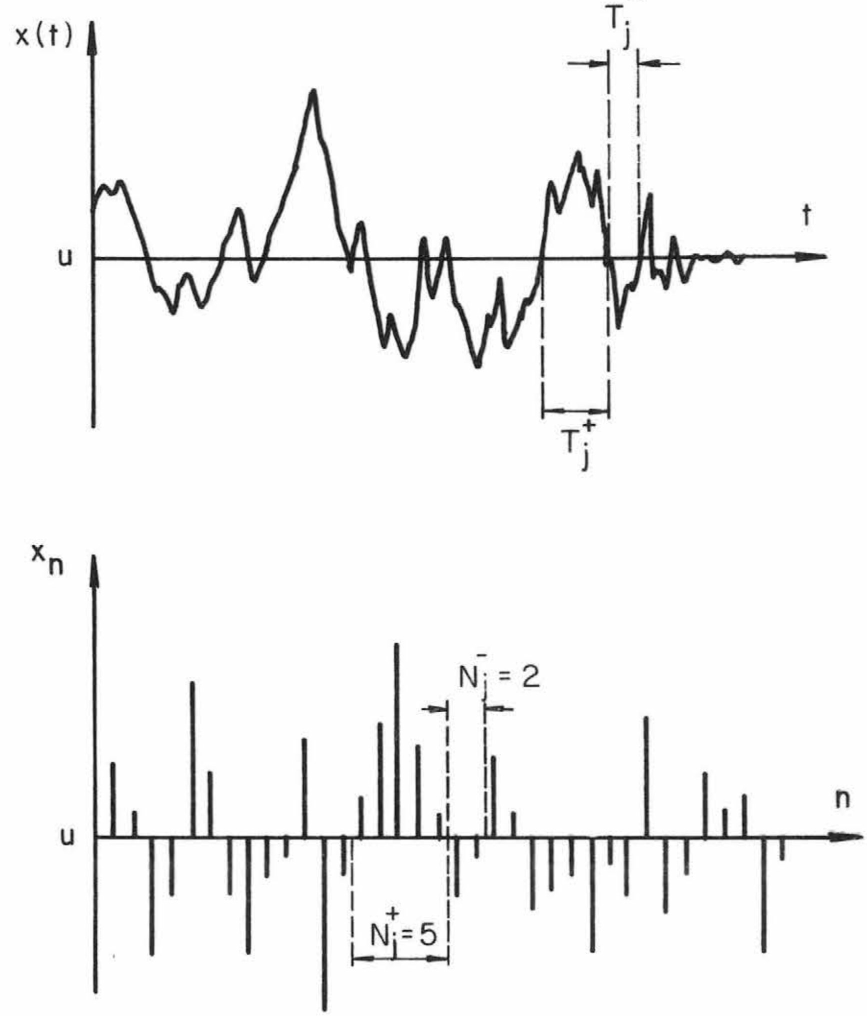

Fig. 3.1 Definition of positive and negative runs for a given truncation level. Upper graph refers to a continuous series and lower graph to a discrete series.

3. Number of positive runs.

4. Number of negative runs.

5. Number of total runs.

The continuous time series are treated most extensively in related literature and the following parameters, besides T_j^+, T_j^- and T_j are used:

1. Areas above truncation level.

2. Areas below truncation level.

3. Number of positive runs.

4. Number of negative runs.

5. Number of total runs.

6. Time interval between successive peaks.

7. Time interval between successive troughs.

All of them are random variables as random functions of the process $\{X_i\}$ and the truncation level u. It should be noted that u does not need to be a constant. It may be a periodic function, or even a stochastic variable. When u is not a constant the determination of the characteristics of runs becomes complex. The properties of runs can be directly used in many water resources problems. If u determines the level of demand, and if this level is not reached, a drought occurs. If a flooded area begins for $x > u$ and flood damage is a function of the time during which $x > u$ then the distribution of positive run-length determines the character of flooding. If a given type of run is regionalized, or shown over an area, with its isolines, then the regional phenomena of droughts, floods and similar phenomena may be studied for their probabilities of recurrence (Yevjevich 1967).

3.2 Probabilities of Run-Lengths

For purposes of simplicity the following notation is adopted by the writer.

$$P\{X_1 \le u,\ X_2 \le u, \ldots, X_k \le u,\ X_{k+1} > u,\ X_{k+2} > u, \ldots X_{k+j} > u\}$$

$$= P(k^-, j^+)$$

with $k = 1, 2, \ldots$, $j = 1, 2, \ldots,$ and

$$P\{X_1 > u,\ X_2 > u, \ldots,\ X_j > u\} = P(j^+).$$

The probability of N_1^+ being not less than j is derived by the writer as follows,

$$P(N_1^+ \ge j) = P(j^+) + \sum_{k=1}^{\infty} P(k^-, j^+). \qquad 3.1$$

The probability mass function of N_1^+ is derived as

$$P(N_1^+ = j) = P(N_1^+ \ge j) - P(N_1^+ \ge j+1) \qquad 3.1a$$

The evaluation of the joint probabilities $P(k^-, j^+)$ requires the joint probability distribution of the variables

X_1, X_2,..., which is considered in following sections. This joint distribution is assumed to be multivariate normal.

3.3 Stationary Gaussian Processes

Consider an arbitrary Gaussian process, i.e., a process for which the joint distribution of $X_{n_1}, \ldots X_{n_\ell}$ is multivariate normal. These multivariate normal distributions, of all orders, are completely characterized by

(a) the mean $E(X_n)$, as a function of n, and

(b) the covariance matrix, $\mathrm{cov}(X_{n_1}, X_{n_2})$, as a function of n_1, n_2.

A Gaussian process is strictly stationary if and only if the mean is constant and the covariance function depends only on the lag time $n_2 - n_1$. For any stationary process

$$E(X_n) = \mu \ , \text{ and } \ \mathrm{cov}(X_{n+k}, X_n) = \gamma(k) \ .$$

In particular, $\gamma(0) = \mathrm{var}\ (X_n)$ is a constant. The function $\gamma(k)$ is the autocovariance function and

$$r(k) = \frac{\gamma(k)}{\gamma(0)} \qquad 3.2$$

is the autocorrelation function that specifies the correlation coefficient between values of the process k time intervals apart.

Let $\{X_i\}$ be a stationary Gaussian process with zero mean and variance unity. Then the probability density function of X is given by,

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) . \qquad 3.3$$

The bivariate probability density of X_i and X_j with $EX_i = EX_j = 0$ and $\operatorname{var} X_i = \operatorname{var} X_j = 1$ is

$$f(x_i,x_j) = \frac{1}{2\pi\sqrt{1-\rho_{ij}}} \exp\left[-\frac{1}{2}(x_i^2-2\rho x_i x_j + x_j^2)\right] \qquad 3.4$$

where ρ_{ij} is the correlation coefficient between X_i and X_j. The multivariate normal probability density function of $X_1, X_2,\ldots,X_p$ takes a more complex form and is considered in section 3.5.

A correlation matrix of random variables $X_1, X_2,\ldots, X_p$ is a p by p matrix with the elements ρ_{ij} representing the correlation coefficients between any two variables, X_i and X_j. It is a symmetric matrix since $\rho_{ji} = \rho_{ij}$ and obviously all elements of the main diagonal are unity. For a stationary process we have

$$\rho_{ij} = \rho_{|j-i|} \quad , \qquad 3.5$$

and therefore all elements of a given diagonal are identical. The correlation matrix of a stationary process can be written as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \cdot & \rho_{p-1} \\ \rho_1 & & & & & & \cdot \\ \rho_2 & & & & & & \cdot \\ \cdot & & & & & & \\ \cdot & & & & & & \rho_2 \\ \cdot & & & & & & \rho_1 \\ \rho_{p-1} & & & & \rho_2 & \rho_1 & 1 \end{bmatrix} .$$

3.4 Ergodic Processes

If the random process $\{X_n\}$ is weakly stationary and if the expected values and crossproduct functions which are defined by ensemble averages as

$$EX = \int_{-\infty}^{\infty} x \, dF(x) , \qquad 3.6$$

$$E(X_i X_{i+K}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i \, x_{i+k} \, dF(x_i, x_{i+k}) \qquad 3.7$$

may be calculated by performing corresponding time averages as

$$EX_i = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i \qquad 3.8$$

$$E(X_i X_{i+k}) = \lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{N} X_i X_{i+k} , \qquad 3.9$$

then the process is said to be weakly ergodic. If the process is Gaussian and is weakly stationary and weakly ergodic, then it is also strictly stationary and strictly ergodic, i.e., all ensemble averaged statistical properties are deducible from the corresponding time averages.

Hence the verification of self-stationarity for a single time series justifies an assumption of stationarity.

3.5 The Multivariate Normal Probability Density Function

A generalization of the normal distribution to p variables is

$$dF = \frac{1}{(2\pi)^{p/2}\sqrt{|R|}} \exp\{-\frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} a_{jk}x_j x_k\} \prod_{j=1}^{p} dx_j, \qquad 3.10$$

where the variables $X_1, X_2, \ldots X_p$ have means zero and variances unity. Also, $|R|$ is the determinant of the correlation matrix of these variables and a_{jk} are the elements of the inverse of the correlation matrix. The characteristic function of this distribution, however, is not expressed in terms of the inverse of the correlation matrix, but in terms of the elements of the correlation matrix itself; this is an interesting property which can be taken advantage of for finding the probabilities of runs. The characteristic function is

$$\phi(t) = \exp\{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} \rho_{ij}t_i t_j\} \qquad 3.11$$

3.6 Joint Probabilities of Eq. 3.1

In order to find an expression for the joint probabilities involved in Eq. 3.1, the following assumptions are made:

1. It is assumed the annual hydrologic time series of precipitation and runoff are stationary. However some of these hydrologic time series have a small degree of non-stationarity. It comes either from man-made changes in river basins and around the precipitation gaging stations, or from the inconsistency in data (Yevjevich 1964). These latter series are made stationary by corrections before the theory of runs, as discussed here, is applied.

2. The process of annual values is assumed to be a Gaussian process or approximately so. This assumption is justified from the point of view that some runs are distribution free (independent of the underlying distributions of $\{X_i\}$). This point will be treated in more detail in chapter IV.

3. It is also assumed that the stationary processes of annual values are standardized for a simpler treatment of various problems. Under these assumptions the joint probabilities involved in Eq. 3.1 can be expressed as

$$P(k^-,j^+) = \underbrace{\int_{-\infty}^{u}\cdots\int_{-\infty}^{u}}_{k}\ \underbrace{\int_{u}^{\infty}\cdots\int_{u}^{\infty}}_{j}\ dF \quad . \qquad 3.12$$

Substitution of dF by its equivalent in Eq. 3.10 gives

$$P(k^-,j^+) = \frac{1}{(2\pi)^{p/2}|R|^{1/2}} \underbrace{\int_{-\infty}^{u}\cdots\int_{-\infty}^{u}}_{k} \underbrace{\int_{u}^{\infty}\cdots\int_{u}^{\infty}}_{j} \exp\left\{-\frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} a_{jk}x_j x_k\right\} \prod_{j=1}^{p} dx_j \,, \qquad 3.13$$

where $p = j + k$.

3.13

Equation 3.12 is then a multinormal integral. No explicit expression exists for the general case of the multinormal integral. A great deal of effort has been dedicated in this study to finding expressions for several cases of this multinormal integral, so that specific numbers can be assigned to probabilities of runs. This is discussed in the next chapter.

3.7 Joint Probabilities of Eq. 3.1 and Probabilities of Runs for $q = .5$.

The mathematical developments in the following sections are applicable to both independent and dependent processes. The independent case is considered as a particular case of the dependent case with $\rho_{ij} = 0$. Throughout the text the subscript m refers to a probability truncation level of the median.

Probabilities of the type $P_m(j^+)$ - McFadden's expression, given by Eq. 2.41, is used as a starting basis for developing an expression for evaluation of these probabilities. For a stationary dependent process the following equations are derived by the writer, after renaming variables for convenience of notation.

$$\sum_{s>r\geq 1}^{j} \arcsin \rho_{rs} = (\arcsin \rho_{12} + \arcsin \rho_{13} + \dots +$$

$$\arcsin \rho_{1j}) + (\arcsin \rho_{23} + \arcsin \rho_{24} + \dots + \arcsin \rho_{2j})$$

$$+ \dots + \arcsin \rho_{j-1,j} \ .$$

But it is known that the condition of stationarity implies

$$\rho_{rs} = \rho_{|s-r|} \quad .$$

Hence,

$$\sum_{s>r\geq 1}^{j} \arcsin \rho_{rs} = (\arcsin \rho_1 + \arcsin \rho_2 + \dots$$

$$+ \arcsin \rho_{j-1}) + (\arcsin \rho_1 + \arcsin \rho_2 + \dots$$

$$+ \arcsin \rho_{j-2}) + \dots + \arcsin \rho_1$$

$$= (j-1)\arcsin\rho_1 + (j-2)\arcsin\rho_2 + \dots +$$

$$\arcsin\rho_{j-1}$$

$$= \sum_{r=1}^{j-1} (j-r)\arcsin\rho_r \,. \qquad 3.15$$

The second summation in Eq. 2.41 can be rewritten for a stationary process, after renaming variables for convenience of notation, as follows,

$$\sum_{u>t>s>r\geq 1}^{j} \rho_{rs}\rho_{tu} + \rho_{rt}\rho_{su} + \rho_{ru}\rho_{st} = (\rho_{12}\rho_{34} + \rho_{13}\rho_{24}$$

$$+ \rho_{14}\rho_{23}) + (\rho_{23}\rho_{45} + \rho_{24}\rho_{35} + \rho_{25}\rho_{34}) + \dots +$$

$$(\rho_{j-3,j-2}\,\rho_{j-1,j} + \rho_{j-3,j-1}\rho_{j-2,j} + \rho_{j-3,j}\,\rho_{j-2,j-1})$$

$$= (\rho_1\rho_1 + \rho_2\rho_2 + \rho_3\rho_1) + (\rho_1\rho_1 + \rho_2\rho_2 + \rho_3\rho_1) + \dots$$

$$+ (\rho_1\rho_1 + \rho_2\rho_2 + \rho_3\rho_1) \,. \qquad 3.16$$

Equation 3.16 can be rewritten as

$$\sum_{u>t>s>r\geq 1}^{j} \rho_{rs}\rho_{tu} + \rho_{rt}\rho_{su} + \rho_{ru}\rho_{st} = \rho_1^2\,\frac{1+(j-3)}{2}\,(j-3)$$

$$+ 0(\rho_1^3) = (j-2)(j-3)\,\frac{\rho_1^2}{2} + 0(\rho_1^3) \qquad 3.17$$

for a first order autoregressive process, since $\rho_2^2 < \rho_1^3$. ?
This is the basic dependence model that will be considered in this study for fitting sequences of annual precipitation and runoff. The selection of the first order autoregressive process as a basic dependence model is justified on the basis of an autocorrelation analysis made on annual sequences of runoff from a global sample and annual sequences of precipitation from the Western United States by Yevjevich (1964). According to this study the sequences of annual precipitation are close to independent time series and the sequences of annual runoff can be fitted by using a first order autoregressive model. The reader is referred to this study for a comprehensive analysis of this problem. Substitution of 3.15 and 3.16 into 2.41 gives

$$P_m(j^+) = \frac{1}{2^j} + \frac{1}{2^{j-1}\pi} \sum_{r=1}^{j-1} (j-4) \arcsin \rho_r + \frac{1}{2^{j-1}\pi^2}$$

$$\cdot (j-2)(j-3)\rho_1^2 + 0(\rho_1^3) \, , \qquad 3.18$$

and finally, neglecting terms containing ρ_j for $j \geq 3$

$$P_m(j^+) = \frac{1}{2^j} + \frac{1}{2^{j-1}\pi} \left[(j-1)\arcsin \rho_1 + (j-2)\arcsin \rho_2\right]$$

$$+ \frac{1}{2^{j-1}\pi^2} (j-2)(j-3)\, \rho_1^2 + 0(\rho^3) \, . \qquad 3.19$$

Probabilities of the type $P_m(1^-,j^+)$ - By definition

$$P_m(1^-,j^+) = \int_{-\infty}^{0} \underbrace{\int_0^{\infty} \cdots \int_0^{\infty}}_{j} dF \quad . \tag{3.20}$$

Equation (3.20) can be written as

$$\int_{-\infty}^{0} \underbrace{\int_0^{\infty} \cdots \int_i^{\infty}}_{j} dF = \int_{-\infty}^{\infty} \underbrace{\int_0^{\infty} \cdots \int_0^{\infty}}_{j} dF - \underbrace{\int_0^{\infty} \cdots \int_0^{\infty}}_{j+1} dF$$

$$= P_m(j^+) - P_m\left[(j+1)^+\right] \tag{3.21}$$

Hence, for j=1

$$P_m(1^-,1^+) = P_m(1^+) - P_m(2^+) = \frac{1}{4} - \frac{1}{2\pi} \arcsin \rho . \tag{3.22}$$

For j=2

$$P_m(1^-,2^+) = P_m(2^+) - P_m(3^+) \tag{3.23}$$

$$= \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_1 - \left[\frac{1}{8} + \frac{1}{4\pi}(2 \arcsin \rho_1 + \arcsin \rho_2)\right]$$

$$= \frac{1}{8} - \frac{1}{4\pi} \arcsin \rho_2 \quad . \tag{3.24}$$

For $j \geq 3$, values of $P_m(1^-, j^+)$ can be found by means of Eq. 3.21.

Probabilities of the type $P_m(k^-, j^+)$ with $k \geq 2$ -

By definition

$$P_m(k^-, j^+) = \underbrace{\int_{-\infty}^{0} \cdots \int_{-\infty}^{0}}_{k} \underbrace{\int_{0}^{\infty} \cdots \int_{0}^{\infty}}_{j} dF \qquad 3.25$$

where the sequence of random variables involved is $\{X_i\}$, and the correlation matrix for the corresponding stationary process is

$$R = \begin{array}{c|c} & \begin{array}{ccccccc} 1 & 2 & 3 & \cdot & \cdot & \cdot & k+j \end{array} \\ \hline \begin{array}{c} 1 \\ 2 \\ 3 \\ \cdot \\ \cdot \\ \cdot \\ k+j \end{array} & \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \cdot & \cdot \\ \rho_1 & \ddots & \ddots & \ddots & & & \cdot \\ \rho_2 & \ddots & \ddots & \ddots & \ddots & & \cdot \\ \cdot & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \cdot & & \ddots & \ddots & \ddots & \ddots & \rho_2 \\ \cdot & & & \ddots & \ddots & \ddots & \rho_1 \\ \cdot & \cdot & \cdot & \rho_2 & \rho_1 & & 1 \end{bmatrix} \end{array} . \qquad 3.26$$

The multiple integral given by Eq. 3.25 can be transformed to another multiple integral having all limits of integration between zero and infinity by means of the following transformation; define a new sequence of random variables $\{X_i^*\}$ as follows,

$$X_i^* = -X_i \quad \text{for} \quad i = 1,2,\ldots,k \tag{3.27}$$

$$X_i^* = X_i \quad \text{for} \quad i = k+1,\ k+2,\ldots,\ k+j \tag{3.28}$$

then it is clear that

$$EX_i^* = 0 \quad \text{for } i = 1,2,\ldots,k+j \tag{3.29}$$

$$\begin{aligned}
EX_i^*X_\ell^* &= EX_iX_\ell = \rho_{|\ell-i|} \quad \text{for } \{i=1,2,\ldots,k;\ell=1,2,\ldots k\} \\
&= EX_iX_\ell = \rho_{|\ell-i|} \\
&\qquad \text{for } \{i = k+1,k+2,\ldots,k+j;\ \ell = k+1,k+2,\ldots k+j\} \\
&= -\,EX_iX_\ell = -\,\rho_{|\ell-i|} \\
&\qquad \text{for } \{i = 1,\ldots,k;\ \ell = k+1,\ldots,k+j\}\,.
\end{aligned} \tag{3.30}$$

The correlation matrix of $\{X_i^*\}$ is then

$$R^* = \begin{array}{cc}
 & \begin{matrix} 1 & 2 & 3 & \cdots & & k & k+1 & k+2 & & \cdots & k+j \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ 3 \\ \\ \\ k \\ k+1 \\ k+2 \\ \\ \\ k+j \end{matrix} &
\left[\begin{matrix}
1 & \rho_1 & \rho_2 & \ddots & & & & & & & \rho_{k+j-1} \\
\rho_1 & \ddots & \ddots & \ddots & & & & & & & \\
\rho_2 & \ddots & \ddots & \ddots & \ddots & & & & & & \\
 & \ddots & \ddots & \ddots & \ddots & \rho_2 & & & & & \\
 & & \ddots & \ddots & \ddots & \rho_1 & -\rho_2 & & & & \\
 & & & \ddots & \rho_1 & \ddots & -\rho_1 & -\rho_2 & & & \\
 & & & & -\rho_2 & -\rho_1 & \ddots & \rho_1 & \rho_2 & & \\
 & & & & & -\rho_2 & \rho_1 & \ddots & \ddots & \ddots & \\
 & & & & & & \ddots & \ddots & \ddots & \ddots & \rho_2 \\
 & & & & & & & \ddots & \ddots & \ddots & \rho_1 \\
\rho_{k+j-1} & & & & & & & & \rho_2 & \rho_1 & 1
\end{matrix}\right]
\end{array} \tag{3.31}$$

The distribution function of $X_1^*, X_2^*, \ldots, X_{k+j}^*$ is standard multivariate normal with matrix of correlation coefficients R^*, denoted by F^*. Also by definition of the sequence $\{X_i^*\}$,

$$\{X_1 \leq u, X_2 \leq u, \ldots, X_k \leq u, X_{k+1} > u, X_{k+2} > u, \ldots,$$

$$X_{k+j} > u\},$$

if and only if

$$X_1^* > u, X_2^* > u, \ldots, X_{k+j}^* > u \quad ,$$

and the probability in which we are interested can be written as

$$P_m(k^-, j^+) = \underbrace{\int_0^\infty \ldots \int_0^\infty}_{k+j} dF^* \quad . \qquad 3.32$$

Notice that $P_m(k^-, j^+)$ has exactly the same expression that $P_m[(k+j)^+]$ has, with the only difference that the former one involves F^*, whereas the latter one involves F. Hence $P_m(k^-, j)$ has an equation similar to Eq. 2.41:

$$P_m(k^-, j^+) = \frac{1}{2^{k+j}} + \frac{1}{2^{k+j-1}\pi} \sum_{s>r\geq 1} \arcsin \rho_{rs}^* + \frac{1}{2^{k+j-2}\pi^2}$$

$$\cdot \sum_{u>t>s>r\geq 1} (\rho_{rs}^* \rho_{tu}^* + \rho_{rt}^* \rho_{su}^* + \rho_{ru}^* \rho_{st}^*) + 0(\rho^{*3}) \qquad 3.33$$

with

$$\sum_{s>r\geq 1} \arcsin \rho^*_{rs} = \sum_{r\geq 1} \arcsin \rho^*_{r,r+1} + \sum_{r\geq 1} \arcsin \rho^*_{r,r+2}, \qquad 3.34$$

if the terms that involve the serial correlation of lag equal to or greater than three are neglected. Also it can be easily verified that for $k \geq 2$

$$\sum_{r\geq 1}^{j} \arcsin \rho^*_{r,r+1} = (k+j-3) \arcsin \rho_1 \qquad 3.35$$

$$\sum_{r\geq 1}^{j} \arcsin \rho_{r,r+2} = (k+j-6) \arcsin \rho_2. \qquad 3.36$$

The terms ρ^*_{rt} ρ^*_{su} ρ^*_{ru} ρ^*_{st} can be neglected because they are, at most, of the order of $(\rho^*_2)^2$ and $(\rho^*_3\ \rho^*_1)$ respectively, and the second summation in Eq. 3.33 can be expressed as

$$\sum_{u>t>s>r\geq 1} (\rho^*_{rs}\ \rho^*_{tu}) = \sum_{s>r\geq 1} \rho^*_{r,r+1}\ \rho^*_{s,s+1} = (k+j-s)\ \rho_1^2\ . \qquad 3.37$$

Substitution of equations 3.34, 3.35, 3.36, 3.37, into Eq. 3.33 gives

$$P_m(k^-,j^+) = \frac{1}{2^{k+j}} + \frac{1}{2^{k+j-1}\pi}[(k+j-3)\arcsin \rho_1 +$$

$$+ (k+j-6) \arcsin \rho_2] + \frac{1}{2^{k+j-2}\pi^2} (k+j-2)(k+j-5) \rho_1^2$$

$$+ 0(\rho^3) \qquad 3.38$$

Notice that $P_m(k^-,j^+)$ depends only on $k+j$ and not on the specific values of k and j. It follows then

$$P_m(3^-,j^+) = P(2^-,(j+1)^+)$$

$$P_m(4^-,j^+) = P(2^-,(j+2)^+)$$

$$\cdots\cdots\cdots$$

$$P_m(k^-,j^+) = P(2^-,(j+k)^+) \ . \qquad 3.39$$

This is an important property because it implies it is not necessary to compute all sequences $P_m(k^-,j^+)$ but it is enough to find $P_m(j^+)$, $P_m(1^-,j^+)$, $P_m(2^-,j^+)$ in order to find the probabilities of runs. These are given by

$$P_m(N_1^+ \geq j) = P_m(j^+) + \sum_{k=1}^{\infty} P_m(k^-,j^+)$$

$$= P_m(j^+) + P_m(1^-,j^+) + \sum_{k=0}^{\infty} P_m[2^-,(j+k)^+] \qquad 3.40$$

This equation in combination with Eq. 3.1a gives the probability mass function of N_1^+ .

3.8 Distribution of Runs $\{N_i^+\}$ and $\{N_i^-\}$

So far the formulas for the distribution of N_1^+ as given by Eq. 3.40 have been developed by the writer. The distribution of N_1^- is the same because of the symmetry of the normal distribution about the truncation level of the median. In the following sections it will be established that the sequence of runs are identically distributed at least within the degree of approximation that have been worked out.

The distribution of N_2^+ and N_2^- - The joint distribution of the variables N_1^+ , N_1^- , N_2^+ and N_2^- is by definition

$$\text{prob}\{N_1^+ = j_1,\ N_1^- = i_1,\ N_2^+ = j_2,\ N_2^- \geq i_2\} = P_m(j_1^+, i_1^-, j_2^+, i_2^-)$$

$$+ P_m(i_1^-, j_1^+, i_2^-, j_2^+). \qquad 3.41$$

The result expressed by Eq. 3.32 can be easily extended for the probabilities involved in Eq. 3.41, following the same procedure explained in Section 3.7. The result is

$$P_m(j_1^+, i_1^-, j_2^+, i_2^-) = P_m^*((j_1+i_1+j_2+i_2)^+) \quad , \qquad 3.42$$

$$R^* = \begin{array}{c|c} & 1 \quad 2 \quad \cdots \quad j_1 \quad \cdots \quad j_1+i_1 \quad \cdot \quad j_1+i_1+j_2 \\ \hline \begin{array}{c} 1 \\ 2 \\ \vdots \\ j_1 \\ \vdots \\ j_1+i_1 \\ \vdots \\ j_1+i_1+j_2 \\ \vdots \\ j_1+i_1+j_2+i_2 \end{array} & \begin{bmatrix} 1 & \rho_1 & \rho_2 & & & & & & & & \\ \rho_1 & \ddots & \ddots & \ddots & & & & & & & \\ \rho_2 & \ddots & \ddots & \ddots & \rho_2 & & & & & & \\ & \ddots & \ddots & \ddots & \rho_1 & \rho_2 & & & & & \\ & & \rho_2 & \rho_1 & -\rho_1-\rho_2 & \rho_1 & \rho_2 & & & & \\ & & & -\rho_2-\rho_1 & \ddots & \ddots & \ddots & \rho_2 & & & \\ & & & -\rho_2 & \rho_1 & \ddots & \rho_1-\rho_2 & \ddots & & & \\ & & & & \rho_2 & \ddots & \ddots & \ddots & \rho_2 & & \\ & & & & \rho_2 & \rho_1 & -\rho_1-\rho_2 & \rho_1 & \rho_2 & & \\ & & & & & -\rho_2-\rho_1 & \ddots & \ddots & \ddots & \rho_2 & \\ & & & & & -\rho_2 & \rho_1 & \ddots & \rho_1-\rho_2 & \ddots & \\ & & & & & & \rho_2 & \ddots & \ddots & \ddots & \rho_2 \\ & & & & & & \rho_2 & \rho_1 & -\rho_1-\rho_2 & \rho_1 & \rho_2 \\ & & & & & & & -\rho_2-\rho_1 & \ddots & \ddots & \rho_2 \\ & & & & & & & -\rho_2 & \rho_1 & \ddots & \rho_1 \\ & & & & & & & & \rho_2 & \rho_2 \; \rho_1 & 1 \end{bmatrix} \end{array} \qquad 3.43$$

An analogous result is obtained for

$$P_m(i_1^-, j_1^+, i_2^-, j_2^+)$$

Substitution in Eq. 3.41 gives

$$\text{prob}\{N_1^+ = j_1,\ N_1^- = i_1,\ N_2^+ = j_2,\ N_2^- \geq i_2\} = P_m^* \,((j_1+i_1 + j_2+i_2)^+) + P_m^* \,((i_1+j_1+i_2+j_2)^+) \quad . \quad 3.44$$

Notice that the right side of the equation does not vary if j_1 and j_2 are interchanged. This implies that the marginal distributions of N_1^+ and N_2^+ are identical. Also the distributions of N_2^- and N_2^+ are identical. The same argument applies to the sequence $N_1^+, N_1^-, N_2^+, N_2^-, \ldots N_k^+$, N_k^-. The conclusion is that runs are therefore a sequence of identically distributed random variables. The runs are also a stationary process in the mean and the variance if $\{X_n\}$ is a stationary process.

3.9 Probabilities of Runs of Stationary Dependent Gaussian Processes for Any Truncation Level

Throughout this subchapter the concern is with the evaluation of probabilities of the type

$$P_q(k^-, j^+) = \text{prob}\ (X_1 \leq u, \ldots, X_k \leq u, X_{k+1} > u, \ldots,$$
$$X_{k+j} > u)$$

where u is, in general, a different value from zero of the standardized variable, X, and $q = F(u)$. For simplicity of notation, the subindex q is dropped, and it will be used only when it is necessary to refer to it.

<u>Probabilities $P(2^+)$ and $P(1^-, 1^+)$</u> - In the case of one variable the following expression obviously holds

$$P(1^+) = \int_u^\infty dF = 1 - F(u) \qquad 3.45$$

where $F(u)$ is the standard normal distribution function. In the case of two variables

$$P(2^+) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_u^\infty \int_u^\infty \exp - \frac{x_i^2 - 2\, x_i x_{i+1} + x_{i+1}^2}{2(1-\rho^2)} \, dx_i dx_{i+1}$$

3.46

$$P(1^-, 1^+) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^u \int_u^\infty \exp - \frac{x_i^2 - 2\rho x_i x_{i+1} + x_{i+1}^2}{2(1-\rho^2)} \, dx_i dx_{i+1}.$$

3.47

These two probabilities are related as follows

$$P(1^-, 1^+) = \int_{-\infty}^u \int_u^\infty dF = \int_{-\infty}^\infty \int_u^\infty dF - \int_u^\infty \int_u^\infty dF = 1 - F(u) - P(2^+). \quad 3.48$$

Bivariate tables are given by the National Bureau of Standards (1959) for $\pm\, \rho$ = from 0 to .95, with intervals 0.05; and from 0.95 to 1, with intervals, 0.01; and variates in the range from 0 to 4, with intervals 0.1, to 6 or 7 decimal places. Also Zelen and Severo (1960) give charts for reading the bivariate normal integral value with an error of 1 per cent or less.

Probability $P(3^+)$ - In the case of three variables

$$P(3^+) = \int_u^\infty \int_u^\infty \int_u^\infty dF \qquad 3.49$$

this integral has been evaluated in terms of the tetrachoric series expansion by Kendall (1941) and is expressed as

$$P(3^+) = \sum_{j,k,\ell} \frac{\rho_{12}^{j}\rho_{13}^{k}\rho_{23}^{\ell}}{j!k!\ell!} H_{j+k-1}(u)\ H_{j+\ell-1}(u)$$

$$\cdot\ H_{k+\ell-1}(u)\ f^3(u) \qquad 3.50$$

where $F(u)$ is the standard normal probability density function, $H_r(x)$ is the rth Hermite polynomial defined by

$$H_r(x) = -\left(\frac{d}{dx}\right)^r f(x) = (-D)^r f(x), \qquad 3.51$$

and j, k, ℓ can take the values $0, 1, 2, \ldots$

The first five Hermite polynomials are

$H_0 = 1$

$H_1 = u$

$H_2 = u^2-1$

$H_3 = u^3-3u$

$H_4 = u^4-6u^2 + 3$

$H_5 = u^5-10u^3 + 15u$

Probabilities of the type $P(j^+)$ - The tetrachoric series expansion for the trivariate case given by Kendall (1941), can be generalized to the multivariate case by following the procedure indicated below. At this point, recall that the multinormal probability density function is expressed in terms of the elements of the inverse of the correlation matrix. A direct integration of the multinormal p.d.f. would imply an inversion of this correlation matrix if the integral is evaluated in terms of the correlation coefficients. This can be avoided, if the Fourier transform of the multinormal characteristic function is expressed in terms of the correlation coefficients themselves and this expression is integrated. This is a parallel procedure to the one followed by Kendall (1941) in his tetrachoric expansion for the trivariate case. In the following the writer develops a generalization of the tetrachoric expansion to the multivariate case. By definition

$$P(j^+) = \underbrace{\int_u^\infty \cdots \int_u^\infty}_{j} dF$$

$$= \frac{1}{(2\pi)^j} \int_u^\infty dx_1 \cdots \int_u^\infty dx_j \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \phi(t) \exp(-i\bar{t}'\bar{x})\, dt_1 \ldots dt_j \qquad 3.53$$

where

$$\bar{t}'\bar{x} = [t_1 t_2 \ldots t_j] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_j \end{bmatrix} = t_1 x_1 + t_2 x_2 + \ldots + t_j x_j \ . \qquad 3.54$$

Also, $\phi(t)$ can be rewritten as

$$\phi(t) = \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{j} t_i^2 + 2\sum_{k>i\geq 1} \rho_{ik}t_it_k\right)\right] . \qquad 3.55$$

But using the exponential series expansion

$$\exp\left(-\sum_{k>i\geq 1}^{k=j} \rho_{ik}t_it_k\right) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!}\left(\sum_{k>i\geq 1} \rho_{ik}t_it_k\right)^r \qquad 3.56$$

Then the substitution in Eq. 3.53 gives

$$\phi(t) = \exp\left[-\frac{1}{2}\sum_{i=1}^{j} t_i^2\right] \exp\left\{\sum_{r=u}^{\infty} \frac{(-1)^r}{r!}\left[\sum_{k>i\geq 1} \rho_{ik}t_it_k\right]^r\right\} \qquad 3.57$$

where

$$\left[\sum_{k>i\geq 1} \rho_{ik}t_it_k\right]^r = \Big[(\rho_{12}t_1t_2 + \dots + \rho_{1n}t_1t_n) + (\rho_{23}t_2t_3 +$$

$$\dots + \rho_{2n}t_2t_n + \dots + \rho_{n-1,n}t_{n-1}t_n\Big]^r \qquad 3.58$$

$$= r!\sum \frac{\rho_{12}^{i_{12}}\,\rho_{13}^{i_{13}}\dots\rho_{1n}^{i_{1n}}\,\rho_{23}^{i_{23}}\,\rho_{24}^{i_{24}}\dots\rho_{2n}^{i_{2n}}\dots\rho_{n-1,n}^{i_{n-1,n}}}{i_{12}!\,i_{13}!\dots i_{1n}!\,i_{23}!\,i_{24}!\dots i_{2n}!\dots i_{n-1,n}!}\, t_1^{s_1}t_2^{s_2}\dots t_n^{s_n}$$

3.59

$$\begin{aligned} s_1 &= i_{12} + i_{13} + \dots + i_{1n} \\ s_2 &= i_{12} + i_{23} + \dots + i_{2n} \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ s_n &= i_{1n} + i_{2n} + \dots + i_{n-1,n} \end{aligned} \tag{3.60}$$

Substitution of Eq. 3.54 and Eq. 3.57 into Eq. 3.53 gives

$$P(j^+) = \frac{1}{(2\pi)^j} \int_u^\infty dx_1 \dots \int_u^\infty dx_j \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \exp - \frac{1}{2} \sum_{i=1}^j t_i^2$$

$$\cdot \sum_{r=0}^\infty (-1)^r \quad \frac{\rho_{12}^{i_{12}} \rho_{13}^{i_{13}} \dots \rho_{1n}^{i_{1n}} \rho_{23}^{i_{23}} \rho_{24}^{i_{24}} \dots \rho_{2n}^{i_{2n}} \rho_{n-1,n}^{i_{n-1,n}}}{i_{12}! i_{13}! \dots i_{1n}! i_{23}! i_{24}! \dots i_{2n}! i_{n-1,n}!}$$

$$t_1^{s_1} t_2^{s_2} \dots t_n^{s_n} \quad \exp(-i\overline{t}'\overline{x}) \; dt_1 \dots dt_j \; , \tag{3.61}$$

By adopting the notation

$$\frac{\rho_{12}^{i_{12}} \rho_{13}^{i_{13}} \dots \rho_{1n}^{i_{1n}} \rho_{23}^{i_{23}} \rho_{24}^{i_{24}} \dots \rho_{2n}^{i_{2n}} \dots \rho_{n-1,n}^{i_{n-1,n}}}{i_{12}! i_{13}! \dots i_{1n}! i_{23}! i_{24}! \dots i_{2n}! \dots i_{n-1,n}!} = A(\rho, i) . \tag{3.62}$$

Equation 3.61 can be written as

$$P(j^+) = \frac{1}{(2\pi)^j} \sum_{r=0}^\infty (-1)^r \sum A(\rho, i) \underbrace{\int_u^\infty dx_1 \dots \int_u^\infty dx_j}_{j}$$

$$\cdot \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{j} \exp\left[-\frac{1}{2} \sum_{i=1}^{j} t_1^2\right]$$

$$\cdot \exp(-i\bar{t}'\bar{x})\ t_1^{s_1} \ldots t_j^{s_j}\ dt_1 \ldots dt_j \ . \qquad 3.63$$

This is the product of j integrals, the first of which is

$$\frac{1}{2\pi} \int_u^{\infty} dx_1 \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} t_1^2\right] \exp(it_1 x_1)\ dt_1 \ , \qquad 3.64$$

and the remaining $j-1$ integrals are similar expressions in X_i and t_i. Since,

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} t^2\right] t^s \exp\ (-itx)dt = \frac{d^r}{d(-ix)^r} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} t^2\right)$$

$$\cdot \exp\ (-itx)dt = \sqrt{2\pi}\ i^r D^r \exp\left(-\frac{1}{2} x^2\right) = 2\pi(-i)^r H_r(x) f(x) \ . \qquad 3.65$$

Eq. 3.64 is equal to

$$(-i)^r \int_u^{\infty} dx\ H_r(x) f(x) = (-i)^r\ H_{r-1}(u) f(u) \ , \qquad 3.66$$

and Eq. 3.63 is equal to

$$P(j^+) = f^j(u) \sum_{\substack{r=0\\ \{i\}}}^{\infty} A(\rho,i)\, H_{s_1-1}(u)\ldots H_{s_j-1}(u)$$

$$= f^j(u) \sum_{\substack{r=0\\ \{i\}}}^{\infty} \frac{\rho_{12}^{i_{12}} \ldots \rho_{n-1,n}^{i_{n-1,n}}}{i_{12}! \ldots i_{n-1,n}!}\, H_{s_1-1}(u)\ldots H_{s_j-1}(u) . \qquad 3.67$$

It is convenient to notice at this point that the definition of the Hermite polynomials, given by Eq. 3.57, applies only to $r = 0, 1, 2, \ldots$ For $r = -1$, $H_{-1}(x)$ is defined by means of Eq. 3.66 as follows

$$H_{-1}(u) f(u) = \int_u^{\infty} H_o(x) f(x)\, dx = 1 - F(u) \quad .$$

Define

$$I = \sum i \quad , \qquad 3.68$$

and

$$H_{s_1-1}(u) \ldots H_{s_j-1}(u) = \Pi(H) \quad , \qquad 3.69$$

then Eq. 3.67 can be rewritten as

$$P(j^+) = f^j(u) \sum_{I=0}^{\infty} A(\rho,i)\, \Pi(H)$$

$$=f^j(u)\left[\sum_{I=0} A(\rho,i)\Pi(H) + \sum_{I=1} A(\rho,i)\Pi(H) + \sum_{I=2} A(\rho,i)\Pi(H)\right] + 0(\rho^3)$$

$$=f^j(u)\ \Big[H_{-1}^{j}(u) + H_{-1}^{j-2}(u)\ \sum\rho + \frac{1}{2}\ H_1^2(u)H_{-1}^{j-2}(u)$$

$$.\sum\rho^2\ H_{-1}^{j-4}(u) \sum_{\rho\neq\rho'} \rho\rho' + H_1(u)H_{-1}^{j-3}(u)\Big(\sum_{\ell>k>i} \rho_{ik}\rho_{i\ell}$$

$$+ \sum_{\ell>k>i} \rho_{ik}\ \rho_{k\ell}\Big) + 0(\rho^3)\Big]$$

$$=f^j(u)H_{-1}^{j}(u)\ \Big[1+H_{-1}^{-2}(u)\{\sum\rho + \frac{1}{2}H_1^2(u)\sum\rho^2\} + H_{-1}^{-4}(u) \sum_{\rho\neq\rho'} \rho\rho'$$

$$+ H_1(u)H_{-1}^{-3}(u)\{\sum_{\ell>k>i} \rho_{ik}\rho_{i\ell} + \sum_{\ell>k>i} \rho_{ik}\rho_{k\ell}\} + 0(\rho^3)\Big]$$

$$= \Big[1-F(u)\Big]^j \Big[1 + \frac{f^2(u)}{[1-F(u)]^2}\{\sum\rho + \frac{u^2}{2}\sum\rho^2\} + \frac{f^4(u)}{[1-F(u)]^4}\sum_{\rho\neq\rho'}\rho\rho'$$

$$+ \frac{uf^3(u)}{[1-F(u)]^3}\{\sum_{\ell>k>i} \rho_{ik}\rho_{i\ell} + \sum_{\ell>k>i} \rho_{ik}\rho_{k\ell}\} + 0(\rho^3)\Big]$$

3.69

At this point recall the stationarity condition

$$\rho_{ik} = \rho_{|k-i|}\quad ,$$

hence,

$$\sum_{\ell>k>i} \rho_{ik}\rho_{i\ell} = (j-2)\rho_1\rho_2 \tag{3.70}$$

$$\sum_{\ell>k>i}^{j} \rho_{ik}\rho_{k\ell} = (j-2)\rho_1^2 + (j-3)\rho_2^2 \tag{3.71}$$

$$\sum_{i=1}^{j} \rho_i^2 = (j-1)\rho_1^2 + (j-2)\rho_2^2 \quad . \tag{3.72}$$

Finally, substitution of Eq. 3.70, 3.71 and 3.72 into Eq. 3.69 gives,

$$\begin{aligned} P(j^+) = \left[1-F(u)\right]^j \Bigg[1 &+ \frac{f^2(u)}{[1-F(u)]^2} \{(j-1)\rho_1 + (j-2)\rho_2 \\ &+ \frac{u^2}{2}\left[(j-1)\rho_1^2 + (j-2)\rho_2^2\right] + \frac{f^4(u)}{[1-F(u)]^4} \{\frac{(j-2)(j-3)}{2}\rho_1^2 \\ &+ \frac{(j-3)(j-4)}{2}\rho_1\rho_2\} + \frac{uf^3(u)}{[1-F(u)]^3} \{(j-2)\rho_1\rho_2 \\ &+ (j-2)\rho_1^2 + (j-3)\rho_2^2\} + 0(\rho^3)\Bigg] \quad . \end{aligned} \tag{3.73}$$

<u>Probabilities of the type $P(1^-j^+)$</u> - By definition

$$P(1^-,j^+) = \underbrace{\int_{-\infty}^{u} \int_{u}^{\infty} \cdots \int_{u}^{\infty}}_{j} dF$$

$$= \int_{-\infty}^{\infty} \underbrace{\int_{u}^{\infty} \cdots \int_{u}^{\infty}}_{j} dF - \underbrace{\int_{u}^{\infty} \cdots \int_{u}^{\infty}}_{j+1} dF$$

$$= P(j^{+}) - P\left[(j+1)^{+}\right] ; \qquad 3.74$$

these can be evaluated by means of Eq. 3.26. 3.23 I imagine

<u>Probabilities of the type $P(k^{-},j^{+})$</u> - A similar procedure to the one followed previously follows. By definition

$$P(k^{-},j^{+}) = \underbrace{\int_{-\infty}^{u} \cdots \int_{-\infty}^{u}}_{k} \underbrace{\int_{u}^{\infty} \cdots \int_{u}^{\infty}}_{j} dF$$

$$= \frac{1}{(2\pi)^{k+j}} \underbrace{\int_{-\infty}^{u} \cdots \int_{-\infty}^{u}}_{k} \underbrace{\int_{u}^{\infty} \cdots \int_{u}^{\infty}}_{j} \phi(t) \exp(-i\bar{t}'\bar{x}) . \qquad 3.75$$

Using the expansion of the multinormal characteristic function given by Eq. 3.57,

$$P(k^{-},j^{+}) = \frac{1}{(2\pi)^{k+j}} \sum_{r=0}^{\infty} (-1)^{r} \sum A(\rho,i) \int_{-\infty}^{u} dx_{1} \cdots \int_{-\infty}^{u} dx_{k}$$

$$\int_{u}^{\infty} dx_{k+1} \cdots \int_{u}^{\infty} dx_{k+j} \exp\left(-\frac{1}{2} \sum_{i=1}^{j+k} t_{1}^{2}\right) \exp(-i\bar{t}'\bar{x})$$

$$t_{1}^{s_{1}} \cdots t_{j+k}^{s_{j+k}} \; dt_{1} \cdots dt_{j} . \qquad 3.76$$

This is the product of k integrals of the type

$$\frac{1}{2\pi} \int_{-\infty}^{u} dx_1 \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}t_1^2\right) t_1^{s_1} \exp(-it_1x_1)\, dt_1 \ , \qquad 3.77$$

and j integrals of the type

$$\frac{1}{2\pi} \int_{u}^{\infty} dx_{k+1} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}t_{k+1}^2\right) t_{k+1}^{s_{k+1}} \exp(-it_{k+1}x_{k+1})\, dt_{k+1}$$

$$3.78$$

Taking into account Eq. 3.65, 3.77 is equal to

$$(-r)^r \int_{-\infty}^{u} dx H_r(x) f(x) \equiv \alpha_r^c(u) \ , \qquad 3.79$$

and Eq. 3.78 is equal to

$$(-i)^r \int_{u}^{\infty} dx H_r(x) f(x) \equiv \alpha_r(u) \ . \qquad 3.80$$

Thus Eq. 3.76 is equal to

$$P(k^-, j^+) = \sum_{I=0}^{\infty} A(\rho, i)\ \alpha_{s_1}^c(u) \ldots \alpha_{s_k}^c(u)\, \alpha_{s_{k+1}}(u) \ldots \alpha_{s_{k+j}}(u) \ .$$

$$3.81$$

The sequences $\{\alpha(u)\}$ and $\{\alpha^c(u)\}$ can be expressed as

$$\begin{aligned} \alpha_o(u) &= 1-F(u) \\ \alpha_r(u) &= H_{r-1}(u) f(u) \quad \text{for} \quad r = 1,2,\ldots \\ \alpha_o^c(u) &= F(u) \\ \alpha_r^c(u) &= -\ H_{r-1}(u) f(u) \quad \text{for} \quad r = 1,2,\ldots, \end{aligned} \qquad 3.82$$

and I is defined as in Eq. 3.68.

Let us define

$$\alpha^c_{s_1}(u) \ldots \alpha^c_{s_k}(u) = \Pi^c(\alpha) \qquad 3.83$$

$$\alpha_{s_{k+1}}(u) \ldots \alpha_{s_{k+j}}(u) = \Pi(\alpha) \quad , \qquad 3.84$$

then

$$P(k^-,j^+) = \sum_{I=0}^{\infty} A(\rho,i)\Pi^c(\alpha)\Pi(\alpha) \quad . \qquad 3.85$$

But for $I = 0$,

$$A(\rho,i) = 1 \qquad 3.86$$

$$\Pi^c(\alpha) = \left[\alpha^c_o(u)\right]^k = F^k(u) \qquad 3.87$$

$$\Pi(\alpha) = \left[\alpha_o(u)\right]^j = \left[1-F(u)\right]^j \quad . \qquad 3.88$$

Hence,

$$P(k^-,j^+) = F^k(u)\left[1-F(u)\right]^j + \sum_{I=1,2} A(\rho,i)\Pi^c(\alpha)\Pi(\alpha) + 0(\rho^3) . \qquad (3.89)$$

<u>Distribution of N_1^+</u> - This distribution can be written now for any truncation level provided the expressions are available for the joint probabilities, or

$$P(N_1^+ \geq j) = \sum_{k=0}^{\infty} P(k^-, j^+)$$

$$P\{N_1^+ \geq j\} = P(j^+) + \sum_{k=1}^{\infty} P(k^-, j^+) \quad . \qquad 3.90$$

These probabilities are evaluated by means of equations 3.73, 3.74, and 3.89 respectively. Equation 3.90 in combination with Eq. 3.1a gives the probability mass function of N_1^+ .

<u>Distribution of N_1^-</u> - By definition

$$P(N_1^- \geq j) = P(j^-) + \sum_{k=0}^{\infty} P(k^+, j^-) \qquad 3.91$$

where

$$P(j^-) = P(X_1 \leq u, \ldots, x_j \leq u)$$

$$P(k^+, j^-) = P\{X_1 > u, \ldots X_k > u, X_{k+1} \leq u, \ldots X_{k+j} \leq u\} \quad .$$

3.92

Consider

$$P(X_1 \leq u', \ldots, X_k \leq u', X_{k+1} > u, \ldots, X_{k+j} > u) = P_p(k^-, j^+)$$

3.93

where u' and p are defined as

$$p = F(u') = 1-F(u) = 1-q \qquad 3.94$$

Then because of the symmetry of the normal distribution,

$$P_q(j^-) = P_p(j^+)$$

$$P_q(k^+,j^-) = P_p(k^-,j^+) \quad . \qquad 3.95$$

The distribution of N_1^- is

$$P(N_1^- \geq j) = P_p(j^+) = \sum_{k=1}^{\infty} P_p(k^-,j^+) \qquad 3.96$$

and these probabilities are evaluated by means of equations 3.73, 3.74, and 3.89 respectively. The probability mass function of N_1^- is derived as

$$P(N_1^- = j) = P(N_1^- \geq j) - P(N_1^- \geq j+1) \qquad 3.96a$$

<u>The joint distribution of N_1^+, N_1^-, N_2^+, N_2^-</u> - By definition

$$\text{prob}\{N_1^+ = j_1, N_1^- = i_1, N_2^+ = j_2, N_2^- = i_2\} = P(j_1^+, i_1^-, j_2^+, i_2^-, 1^+)$$

$$+ P(i_1^-, j_1^+, i_2^-, j_2^+, 1^-) \quad . \qquad 3.97$$

The result expressed by equations 3.81, 3.85, and 3.89 can be extended to the joint probabilities involved in Eq. 3.97. Following the same procedure used for the probabilities of the type $P(k^-,j^+)$ the result is

$$P(j_1^+, i_1^-, j_2^+, i_2^-, 1^+) = \sum_{I=0}^{\infty} A(\rho, i)\ \alpha_{s_1} \cdots \alpha_{s_{j_1}}\ \alpha^c_{s_{j_1+1}} \cdots \alpha^c_{s_{j_1+i_1}}$$

$$\cdot \alpha_{s_{j_1+i_1+1}} \cdots \alpha_{s_{j_1+i_1+j_2}}\ \alpha^c_{s_{j_1+i_1+j_2+1}} \cdots \alpha^c_{s_{j_1+i_1+j_2+i_2}}$$

$$\cdot \alpha_{s_{j_1+i_1+j_2+i_2+1}} \qquad 3.98$$

a completely analogous result is obtained for

$$P(i_1^-, j_1^+, i_2^-, j_2^+, 1^-) \quad .$$

In Eq. 3.98, if we interchange the places of j_1^+ and j_2^+ the result does not vary because we are considering all possible permutations of $\{i\}$. This implies that the marginal distribution of N_1^+ and N_2^+ are identical. This also holds for N_1^- and N_2^- .

<u>The distributions of N_k^+ and N_k^-</u> - The same argument explained in the previous sub-section applies to the sequence $\{N_i ;\ i = 1,\ldots,k\}$ and the result is that it is a sequence of identically distributed random variables with distribution function given by Eq. 3.90. The same is true for the sequence $\{N_i ;\ i = 1,\ldots,k\}$ which is also a sequence of identically distributed random variables with distribution fucntion given by Eq. 3.96.

<u>Joint probabilities with errors $0(\rho^2)$</u> - Some equations for joint probabilities, developed so far, have been of an infinite series expansion type. Some others

have been expressions with errors $0(\rho^3)$. In this section, expressions are developed with errors $0(\rho^2)$ for a stationary Gaussian process. From Eq. 3.73

$$P(j^+) = \left[1-F(u)\right]^j \left[1 + \frac{f^2(u)}{[1-F(u)]^2}(j-1)\rho_1\right] + 0(\rho^2) \qquad 3.99$$

where ρ_2 has been assumed to be of the order of ρ^2.
Also, from Eq. 3.89,

$$P(k^-,j^+) = F^k(u)\left[1-F(u)\right]^j + \sum_{I=1} A(\rho,i)\Pi^c(\alpha)\Pi(\alpha) + 0(\rho^2)$$

3.100

$$\sum_{I=1} A(\rho,i)\Pi^c(\alpha)\Pi(\alpha) = \sum\rho_1\alpha^c_{s_1}(u)\ldots\alpha^c_{s_k}(u)\alpha_{s_{k+1}}(u)\ldots\alpha_{s_{k+j}}(u)$$

$$= \rho_1(\alpha^c_o)^{k-2}(\alpha_o)^{j-2}\left[(k-1)(\alpha_1\alpha^c_o)^2 + \alpha_1\alpha_o\alpha^c_1\alpha^c_o + (j-1)(\alpha^c_o\alpha_1)^2\right]$$

$$= \rho_1\left[F(u)\right]^{k-2}\left[1-F(u)\right]^{j-2} f^2(u)\{(k-1)\left[1-F(u)\right]^2 - F(u)\left[1-F(u)\right]$$

$$+ (j-1)\left[1-F(u)\right]^2\}$$

$$P(k^-,j^+) = F^k(u)\left[1-F(u)\right]^j + \rho_1 F^{k-2}(u)\left[1-F(u)\right]^{j-2}f^2(u)$$

$$\{(k-1)\left[1-F(u)\right]^2 - F(u)\left[1-F(u)\right] + (j-1)\left[1-F(u)\right]^2\} + 0(\rho^2) .$$

3.101

For $k = 1$ the following equation is used

$$P(1^-, j^+) = P(j^+) - P(j+1^+) \quad ,$$

in combination with Eq. 3.99.

Chapter IV

RUNS OF STATIONARY DEPENDENT GAUSSIAN PROCESSES

4.1 Introduction

The usual linear regression prediction models, namely the moving averages and the autoregressive models, can be shown to be Gaussian processes when the independent component is normally distributed. Among these models the first and second order normal autoregressive processes are considered because of their broad applicability in hydrology.

First order autoregressive process - Suppose that the process $\{X_n\}$ is defined by the recurrence relation

$$X_n = \rho X_{n-1} + Z_n \quad . \qquad 4.1$$

It can be solved formally by successive substitutions and rewritten as

$$X_n = \sum_{i=0}^{\infty} \rho^i Z_{n-i} \qquad 4.2$$

then for $EX_n = 0$

$$\text{var } X_n = \frac{\sigma_z^2}{1-\rho^2} \qquad 4.3$$

where $|\rho| < 1$ is required for the process to be stationary.

It is a well known result that

$$\rho_k = \rho^k \quad . \tag{4.4}$$

It is apparent from Eq. 4.2 that the first order autoregressive model is a type of moving average of infinite extent. Therefore, $\{X_n\}$ as a linear combination of $\{Z_{n-i}\}$ is also a sequence of normal variables and a Gaussian process.

Second order autoregressive process - The process is defined as

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + Z_n \quad . \tag{4.5}$$

Then for $EX_n = 0$,

$$\text{var } X = \frac{1-a_2}{(1+a_2)\{(1-a_2)^2 - a_1^2\}} \quad \sigma_z^2 \tag{4.6}$$

$$\rho_1 = \frac{a_1}{1-a_2} \tag{4.7}$$

$$\rho_k = a_1 \, \rho_{k-1} + a_2 \, \rho_{k-2} \tag{4.8}$$

are well known results in literature.

4.2 Probability Mass Function and Moments of Runs

The following relationship holds between the probility mass of a given run $N^+ = j$ and the probability distribution functions of runs N^+:

$$P(N^+ = j) = P(N^+ \geq j) - P(N^+ \geq j+1) \quad . \qquad 4.9$$

The moments of runs are derived by the writer in the following way. By definition the first moment of N^+ is

$$EN^+ = \sum_{j=1}^{\infty} jP(N^+ = j) \quad . \qquad 4.10$$

Substitution of Eq. 4.9 into Eq. 4.10 yields

$$EN^+ = \sum_{j=1}^{\infty} j\left[P(N^+ \geq j) - P(N^+ \geq j+1)\right]$$

$$= P(N^+ \geq 1) - P(N^+ \geq 2) + 2\left[P(N^+ \geq 2) - P(N^+ \geq 3)\right]$$

$$+ \ldots$$

$$= P(N^+ \geq 1) + P(N^+ \geq 2) + \ldots$$

$$= \sum_{j=1}^{\infty} P(N^+ \geq j) \quad . \qquad 4.11$$

Also, by definition, the second moment of N^+ is

$$E(N^+)^2 = \sum_{j=1}^{\infty} j^2 P(N^+ = j) \qquad 4.12$$

Substitution of Eq. 4.9 into Eq. 4.12 yields

$$E(N^+)^2 = \sum_{j=1}^{\infty} j^2 \left[P(N^+ \geq j) - P(N^+ \geq j+1)\right]$$

$$= P(N^+ \geq 1) - P(N^+ \geq 2) + 4\left[P(N^+ > 2) - P(N^+ \geq 3)\right]$$

$$+ 9\left[P(N_1^+ \geq 3) - P(N_1^+ \geq 4)\right] + \ldots .$$

$$= \sum_{j=0}^{\infty} (2j+1)\ P(N^+ \geq j) \quad . \qquad 4.13$$

The third moment of N^+ is

$$E(N^+)^3 = \sum_{j=1}^{\infty} j^3\ P(N^+ = j)$$

$$= P(N^+ \geq 1) + (2^3-1^3)\ P(N^+ \geq 2) + (3^3-2^3)\ P(N^+ \geq 3) + ..$$

$$= P(N^+ \geq 1) + 7P(N^+ \geq 2) + 19P(N^+ \geq 3) + \ldots$$

$$= \sum_{j=1}^{\infty} \left[j^3-(j-1)^3\right] P(N^+ \geq j) \quad . \qquad 4.14$$

More general the r-th moment of X is

$$E(N^+)^r = \sum_{j=1}^{\infty} j^r\ P(N^+ = j) = \sum_{j=1}^{\infty} \left[j^r-(j-1)^r\right] P(N^+ \geq j)$$

$$= P(N^+ \geq 1) + (2^r-1)\ P(N^+ \geq 2) + (3^r-2^r)\ P(N^+ \geq 2)+\ldots$$

$$4.15$$

All equations given in this subchapter analogously are applicable to N^- .

4.3 Properties of $N = N^+ + N^-$

The statistical properties of this variable are rather complex due to the fact that N^+ and N^- are not independent in autoregressive models and their bivariate distribution is unknown. The only property of N that can be calculated on basis of the univariate distribution of N^+ and N^- is the mean

$$EN = EN^+ + EN^- \quad . \qquad 4.16$$

This equation can be rewritten in the form

$$EN(q) = EN^+(q) + EN^-(q) \qquad 4.17$$

where

$$q = F(u) \quad . \qquad 4.18$$

But

$$EN^-(q) = EN^+(p) \qquad 4.19$$

because of the symmetry of normal distribution. Substitution into Eq. 4.17 yields

$$EN(q) = EN^+(q) + EN^+(p) \quad . \qquad 4.20$$

4.4 General Procedure for Evaluation of Properties of Runs

The statistical properties of runs of stationary linear Gaussian processes can now be evaluated for any truncation level by using the obtained relationships in this chapter and in Chapter III.

The general procedure can be summarized in the 4 following steps. The equations involved are presented in Table 4.1.

1. The expressions for probabilities $P(j^+)$ give the probability that, starting at arbitrary time, at least the first j values of X are above the truncation level specified by q. However, nothing is specified about the values of X preceding or following the occurrence of these j values. They may be either above or below the truncation level. Therefore it must be kept in mind that $P(j^+)$ are not probabilities of runs, but they are needed for their computation.

2. Calculation of probabilities $P(k^-,j^+)$. These expressions give us the probability that starting at arbitrary time, the first k values of X are below and the j subsequent values of X are above the truncation level specified by q. However, nothing is specified about the values of X preceding or following the occurrence of these $k+j$ values. These expressions are not probabilities of runs but they are needed for their computation.

TABLE 4.1

EQUATIONS FOR THE EVALUATION OF PROPERTIES OF RUNS

Step	Expression	Truncation level of the median	Any truncation level
1	$P(1^+)$	2.42	
	$P(2^+)$	2.43	Bivariate
	$P(3^+)$	2.44	tables
	$P(j^+)$, $j \geq 3$	3.19	3.26 3.73
2	$P(1^-,1^+)$	3.22	
	$P(1^-,2^+)$	3.24	
	$P(1^-,j^+)$, $j \geq 3$	3.21	3.74
	$P(2^-,j^+)$	3.38	
	$P(k^-,j^+)$, $k \geq 2$	3.39	3.82, 3.83, 3.84
			3.89
3	$P(N^+ \geq j)$	3.42	3.96
	$P(N^+ = j)$	4.9	4.9
4	EN^+	4.11	
	$E(N^+)^2$	4.13	
	$E(N^+)^3$	4.14	
	$E(N^+)^r$	4.15	
	EN^-	4.19	
	EN	4.20	

3. Using the expressions calculated in steps 1 and 2, the probability distribution and the probability mass functions of the run-length may be calculated.

4. Finally, moments of the run-length are calculated by using the probability distribution previously obtained.

4.5 Properties of Runs of the First Order Autoregressive Process

Positive runs - Following the procedure indicated in the previous section, the properties of runs of the first-order linear autoregressive model are obtained by a digital computer. Five values of the probability truncation level $q = F(u)$ and five values of ρ were used, as indicated in Tables 4.2, 4.3 and 4.4. For each possible combination of q and ρ the probabilities $P(k^-,j^+)$ were calculated for $j = 1,2,\ldots,10$ and $k = 0,1,2,\ldots,10$. Theoretically, the probabilities of runs are given by an infinite series of $P(k^-,j^+)$ as given by Eq. 3.90. Actually these terms become very small for sufficiently high values of k. In order to determine the sufficiently high value of k, the following procedure is used

$$P(N^+ \geq 1) = 1$$

$$\simeq \sum_{k=0}^{\ell} P(k^-,1^+) + P(1^+) \quad . \tag{4.21}$$

The value of ℓ is chosen in such a way that

$$P(1^+) + \sum_{k=0}^{\ell} P(k^-,1^+) \geq .99 \quad . \tag{4.22}$$

With this criterion, the error in computing $P(N^+ \geq 1)$ is set, at most, 0.01. The following properties of runs

TABLE 4.2

MEAN OF N^+ OF THE FIRST ORDER LINEAR AUTOREGRESSIVE GAUSSIAN PROCESS

q \ ρ	0	.1	.2	.3	.4	.5
.3	3.33	3.45	3.60	3.81	4.04	4.28
.4	2.50	2.64	2.82	3.02	3.24	3.48
.5	2.00	2.14	2.29	2.47	2.68	2.92
.6	1.67	1.77	1.89	2.04	2.20	2.37
.7	1.43	1.50	1.60	1.72	1.84	1.99

TABLE 4.3

VARIANCE OF N^+ OF THE FIRST ORDER LINEAR AUTOREGRESSIVE GAUSSIAN PROCESS

q \ ρ	0	.1	.2	.3	.4	.5
.3	7.77	7.77	8.37	9.35	10.42	11.55
.4	3.75	4.19	4.89	5.68	6.54	7.45
.5	2.00	2.43	2.97	3.59	4.29	5.04
.6	1.11	1.36	1.67	2.08	2.49	2.93
.7	.61	.80	1.00	1.22	1.47	1.71

TABLE 4.4

SKEWNESS OF N^+ OF THE FIRST ORDER LINEAR AUTOREGRESSIVE GAUSSIAN PROCESS

q \ ρ	0	.1	.2	.3	.4	.5
.3	44.10	114.52	124.84	144.63	166.69	190.68
.4	15.00	48.75	59.92	73.10	88.12	104.75
.5	6.00	23.52	30.78	39.81	50.58	62.97
.6	2.59	10.51	13.66	18.65	23.38	28.93
.7	1.14	5.13	6.85	8.90	11.28	13.91

of the first order autoregressive process are given in Appendix A: $P(k^-,j^+)$, $P(N^+ \geq j)$ and $P(N^+ = j)$, with $k = 1,2,\ldots,10$ and $j = 1,2,\ldots,10$; EN^+; $\mathrm{var}(N^+)$ and coefficient of skewness of N^+ . Tables 4.2, 4.3 and 4.4 summarize the results for the first three moments of N^+. Figures 4.1 and 4.2 give the probability distributions of positive runs of the first order linear autoregressive Gaussian process for values of ρ varying from 0 to 0.5, with the increment 0.1, and values of q from 0.3 to 0.7, with the increment 0.1

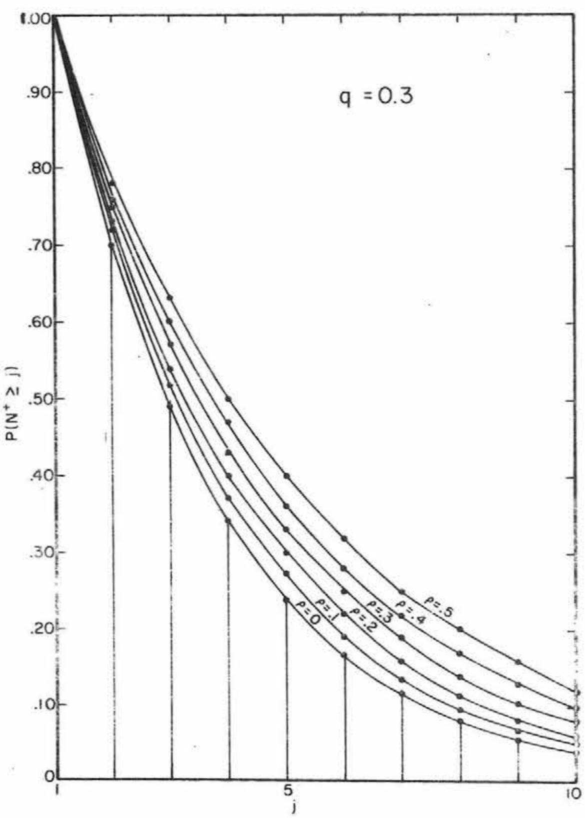

Fig. 4.1 Probability distributions of positive runs of the first order linear autoregressive Gaussian process for $q = 0.3$

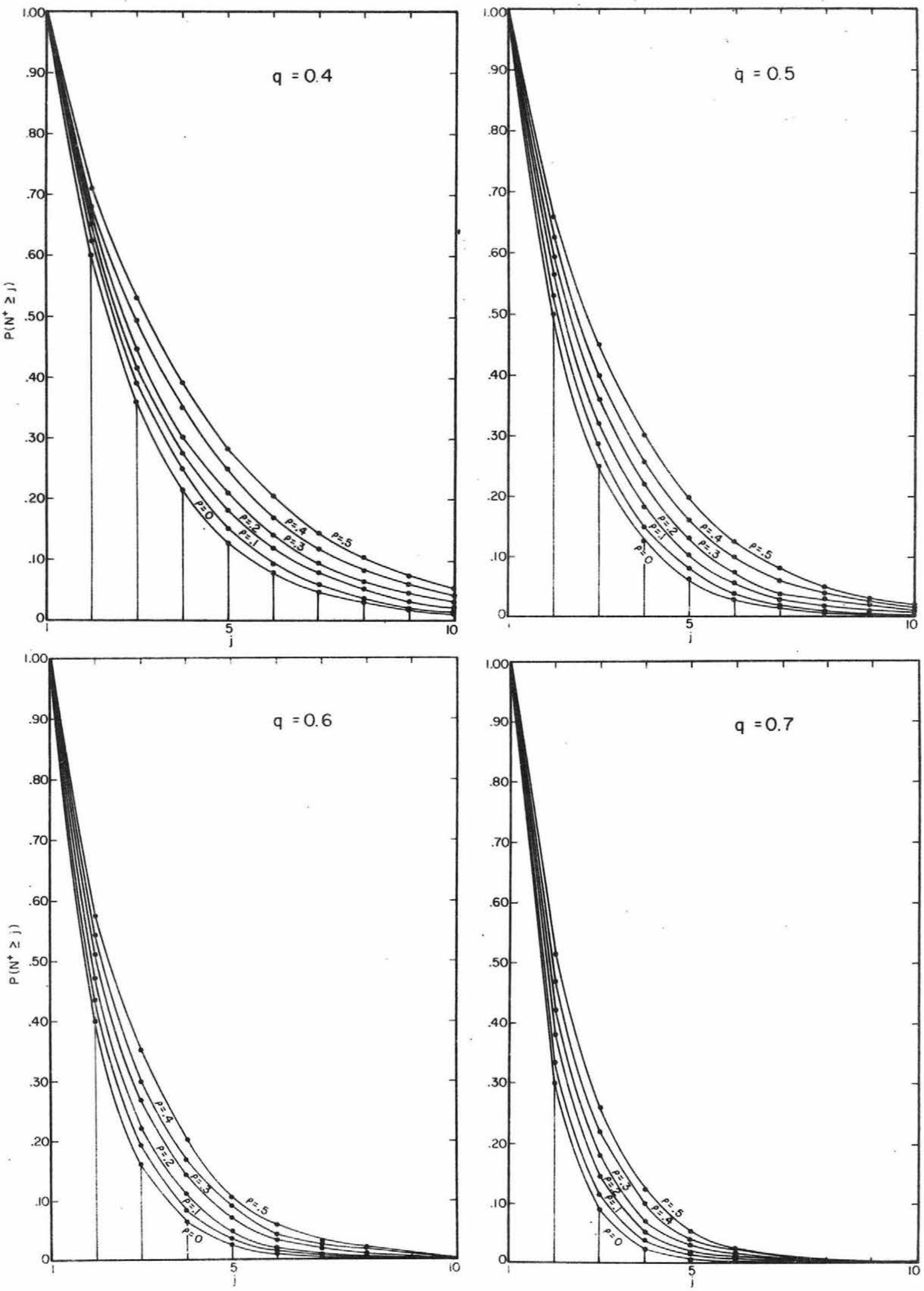

Fig. 4.2 Probability distributions of positive runs of the first order linear autoregressive Gaussian process for $q = 0.4, 0.5, 0.6, 0.7$.

Negative runs - Properties of negative runs are found by means of Eq. 3.95: $P_q(k^+,j^-) = P_p(k^-,j^+)$. It then follows

$$EN_1^-(q) = EN_1^+(p) \qquad 4.23$$

$$\text{var } N_1^-(q) = \text{var } N_1^+(p) \qquad 4.24$$

$$\text{skew } N_1^-(q) = \text{skew } N_1^+(p) \quad . \qquad 4.25$$

Total runs - The process of total runs has been defined as $\{N_j\} = \{N_j^+ + N_j^-\}$. Then if the subscript j is dropped it is obvious that

$$EN = EN^+ + EN^- \qquad 4.26$$

Figure 4.3 shows a graph with three different sets of curves, namely, EN^+, EN^- and EN against q for values of ρ from 0 to 0.5 with the increment 0.1.

4.6 Reduction of Non-Normal Case to Normal

Consider the case in which the marginal distribution of X_i is not normal. Let F' be its multivariate distribution function, then

$$P'(k^-,j^+;v) = P(X_1' \leq v,..,X_k' \leq v,\ X_{k+1}' > v,...,X_{k+j}' > v)$$

4.27

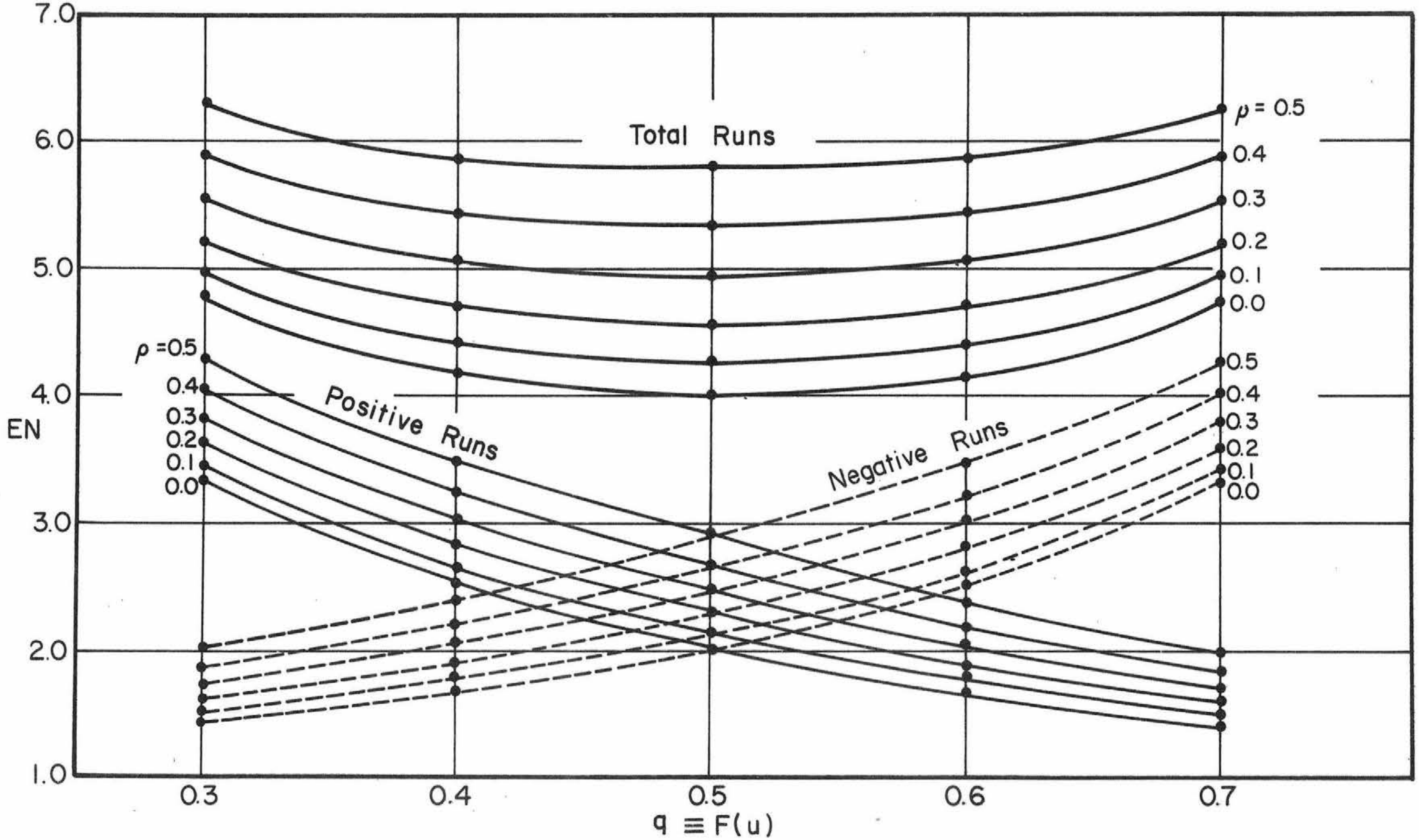

Fig. 4.3 Mean run-lengths of the first order autoregressive linear Gaussian process for positive, negative, and the sum of the two run-lengths.

$$P(k^-,j^+;u) = P(X_1 \leq u,..,X_k \leq u,X_{k+1} > u,...,X_{k+j} > u)$$

4.28

and the superscript (') is reserved for the non-normal case. For a strictly increasing distribution function F', it is always possible to find a unique value u such that

$$F'(v) = F(u) = q \quad . \tag{4.29}$$

Now let us assume that the multivariate distribution F' is such that for all k and j

$$P'(k^-,j^+;v) = P(k^-,j^+;u) \quad . \tag{4.30}$$

If the joint probabilities are identified with a subindex denoting the probability level, then

$$P'_q(k^-,j^+) = P_q(k^-,j^+) \quad . \tag{4.31}$$

This equation implies that the joint probabilities are dependent only on the probability level q and not on the underlying distribution, as long as the condition given by Eq. 4.30 is satisfied.

As an example, consider the case of $\{X'_i\}$ log-normally distributed, with

$$P(X'_i \leq v) = F'(v) \quad . \tag{4.32}$$

By definition of the log-normal distribution the following relationship holds

$$F'(v) = F(\log v) = q \quad . \qquad 4.33$$

Let

$$P(k^-,j^+) = P(X_1' \leq v,\ldots,X_k' \leq v, X_{k+1}' > v,\ldots,X_{k+j}' > v) \qquad 4.34$$

$$= P(\log X_1' \leq \log v,\ldots, \log X_k' \leq \log v, \log X_{k+1}' > v,\ldots \log X_{k+j}' > v) \qquad 4.35$$

$$= P(X_1 \leq \log v,\ldots,X_k \leq \log v, X_{k+1} > \log v\ldots,X_{k+j} > \log v)$$

$$= P_q(k^-,j^+) \quad . \qquad 4.36$$

It follows immediately that all properties of run-length depend only on q in this case, which is the general case as long as Eq. 4.30 holds.

Chapter V

APPLICATION TO INVESTIGATION OF TIME SERIES

5.1 Introduction

The hydrologist is concerned with two basic types of variables, namely serially independent and serially dependent variables. The interest is first in testing whether they are stationary or not. If they are stationary, further tests are concerned with the significance of various models of serial dependence. The model is a representation of the process and it reflects the statistical characteristics of the sequence in terms of parameters which are related to the physical properties of the system.

In the case of investigation of time series by autocorrelation analysis the parameters involved are the serial correlation coefficients, r_k . A comparison of the computed correlogram with correlograms of theoretical models allows one to make inference about the mathematical structure of the observed time series. Figure 5.1 shows the expected correlogram of an independent series $Er_k = -1/(N-k)$ with $\text{var } r_k = (N-k-1)/(N-k)^2$ and with the 95% confidence limits of r_1 for a sample size $N = 30$, obtained by means of Eq. 1.6. In order to test the independence of the observed time series a null hypothesis is established as follows. The observed time series in an independent sequence. Under this hypothesis the 95% confidence limits of r_1 for normal

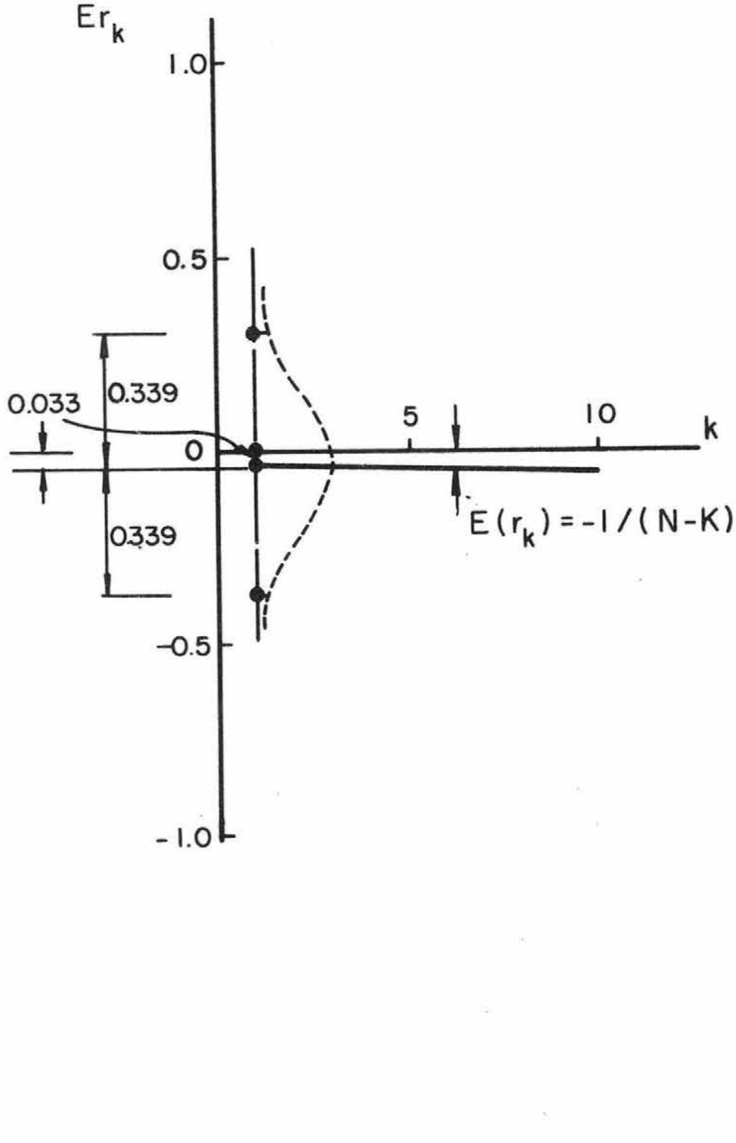

Fig. 5.1 Expected correlogram of an independent series, with sample size N = 30 .

variables are given in Fig. 5.1. If the calculated value of r_1 falls inside the confidence interval, this hypothesis is accepted. Otherwise it is rejected.

In the case of investigation of time series by the spectral analysis, the parameters involved are the spectral densities V_f, as functions of frequencies, f. Again a comparison of the observed variance density spectrum with spectra of the known theoretical models permits one to infere the mathematical structure of dependence in the observed time series. For a discrete time series with equal time intervals Δt, the maximum frequency is given by $f_{max} = 1/2\Delta t$, and if the time series is finite with sample size N, the minimum frequency is given by $f_{min} = 1/N\Delta t$. The spectrum has the property that the area under the variance density graph represents the total variance of the variable in question. Also, the spectrum of an independent stochastic series is a straight horizontal line between f_{min} and f_{max}.

The distribution of the variance density for independent variables is approximated by a χ^2 distribution with the number of degrees of freedom given by

$$\nu = \frac{2N}{m} - \frac{1}{2} \quad , \qquad 5.1$$

where m is the maximum number of lags used in computing the serial correlation coefficients.

For $N = 30$, $m = 4$, $\Delta t = 1$, the number of degrees of freedom is $\nu = 7$. The maximum and minimum frequencies are given by $f_{max} = .5$ and $f_{min} = .0333$. The mean variance density is

$$EV_f = \frac{\sigma^2}{0.4667} = 2.143\ \sigma^2 \quad . \qquad 5.2$$

Then $\frac{\nu V_f}{\sigma^2}(f_{max} - f_{min})$ is a χ^2_ν random variable. The 95% confidence limits are

$$\chi^2_{.025} < \frac{\nu V_f (f_{max}-f_{min})}{\sigma^2} < \chi^2_{.975} \quad . \qquad 5.3$$

For $\nu = 7$, $\chi^2_{.025} = 1.69$, and $\chi^2_{.975} = 16.0$. Hence, the 95% confidence limits are

$$.517\ \sigma^2 < V_f < 4.898\ \sigma^2 \ . \qquad 5.4$$

Figure 5.2 shows the 95% confidence limits for V_f for independent variables.

In the case of investigation of time series by runs, the basic parameter selected here is the run-length for the following reasons.

1. If a given time series is cut at many levels and for each level the sequences of positive and negative run-lengths are obtained, it is theoretically possible to reproduce the original time series at least at a finite number of points by using the sequences of runs. The greater

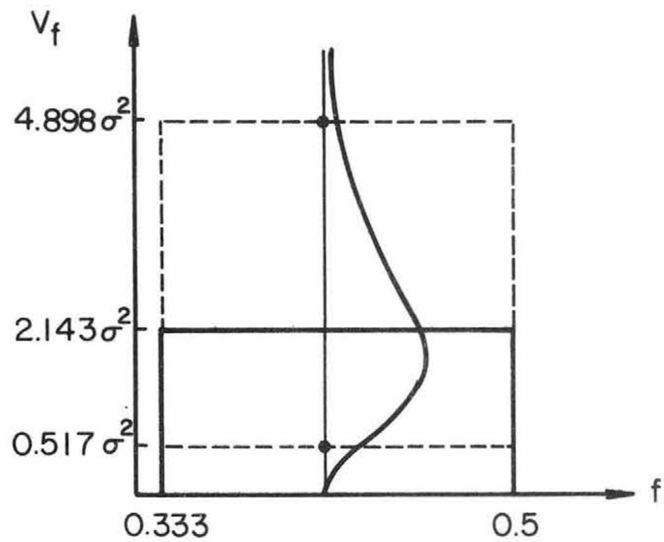

Fig. 5.2 The 95% confidence limits of variance spectrum of an independent series with sample size N = 30.

the number of levels selected, the larger the number of points of the original time series that can be reproduced. If all sequences of positive and negative runs at all possible truncation levels are known, the whole original time series can be reproduced.

2. The properties of run-lengths based on a probability truncation level are distribution-free for independent variables. Furthermore they are independent of the underlying univariate distribution of the original variables for the case of dependent variables. This is an important result because it allows the use of the theory developed for Gaussian processes to other types of processes.

3. The physical significance of positive and negative run-lengths is obvious in hydrology. They can immediately be associated with periods of duration of deficit and surplus, or with the durations of droughts and floods.

4. A parallel technique to autocorrelation analysis and variance density spectrum for investigation of time series can be developed on the basis of run-lengths.

It is possible to compare the properties of observed runs with the same properties of runs of theoretical models. The mean positive run-length $\bar{N}^+$, the mean negative run-length $\bar{N}^-$ and the mean total run-length $\bar{N}$ are possible parameters to be compared with those of theoretical models. Because $\bar{N}$ contains more information that $\bar{N}^-$ or $\bar{N}^+$, this parameter is selected here as basic parameter for investigation of hydrologic time series by runs.

The autocorrelation analysis essentially compares r_k as a function of k of the observed time series with the correlograms of theoretical models. The variance spectrum analysis analogously compares V_f as a function of f of the observed time series with the variance density spectra of theoretical models. The technique by runs developed here compares the statistic $\overline{N}_q$ as a function of q of the observed time series with the expected value of the total run-length N of theoretical models.

5.2 Properties of Run-Length for Sequences of Independent Identically Distributed Random Variables

Let $\{X_n\}$ be a sequence of independent random variables with a common distribution and $\{N_j\}$ be the associated process of the total run. Then $\{N_j\}$ is a renewal process and as such it is a sequence of independent identically distributed, random variables. Let

$$\overline{N}_k = \frac{N_1 + \ldots + N_k}{k} , \qquad 5.5$$

Then by the central limit theorem for large k, $\overline{N}_k$ is asymptotically normally distributed with

$$E\overline{N}_k = EN \qquad 5.6$$

$$\text{var } \bar{N}_k = \frac{\text{var } N}{k} \quad . \qquad 5.7$$

Substitution of equation 2.18 into equations 5.6 and 5.7 gives

$$E\bar{N}_k = \frac{1}{pq} \qquad 5.8$$

$$\text{var } \bar{N}_k = \frac{p^3+q^3}{k\ p^2q^2} \quad . \qquad 5.9$$

This result holds when k is a fixed number. Now consider the case when the time series length, N is fixed. Then the number of total runs in the interval $(0,N)$ becomes a random variable $k(N)$. Let us define

$$\bar{N}_{k(N)} = \frac{N_1 + \ldots + N_{k(N)}}{k(N)}$$

and consider the ratio $\frac{k(N)}{N}$. 5.10

Feller (1966) has shown that this ratio is asymptotically normal with the mean equal to the mean recurrence time of the completion of a total run. Therefore, it converges in probability to a positively valued random variable. In virtue of the central limit theorem for a sum of a random number of independent random variables (Blum, Hanson and Rosenblatt 1963), the result obtained for $\bar{N}_k$ also holds for $\bar{N}_{k(N)}$.

Table 5.1 gives values of the mean and variance of N^+, N^-, and N respectively for a range of values of q between 0.1 and 0.9. Figure 5.3 shows a graph of EN^+, EN^- and EN versus q for the independent case.

TABLE 5.1

STATISTICAL PROPERTIES OF RUN-LENGTH FOR INDEPENDENT IDENTICALLY DISTRIBUTED VARIABLES

q	N^+		N^-		N	
	Mean	Variance	Mean	Variance	Mean	Variance
0.1	10.00	90.00	1.11	.12	11.11	90.12
0.2	5.00	20.00	1.25	.31	6.25	20.31
0.3	3.33	7.78	1.43	.61	4.77	8.39
0.4	2.50	3.75	1.67	1.11	4.17	4.86
0.5	2.00	2.00	2.00	2.00	4.00	4.00
0.6	1.67	1.11	2.50	3.75	4.17	4.86
0.7	1.43	.61	3.33	7.78	4.77	8.39
0.8	1.25	.31	5.00	20.00	6.25	20.31
0.9	1.11	.12	10.00	90.00	11.11	90.12

5.3 Properties of Run-Length for Sequences of Dependent Variables

For the purpose of investigating time sequences by using runs, it is desirable to obtain functional relationships between the mean run-lengths and the probability level q for sequences of dependent variables. The exact type of functional relationships is not known, but some conditions that they must satisfy are known, namely:

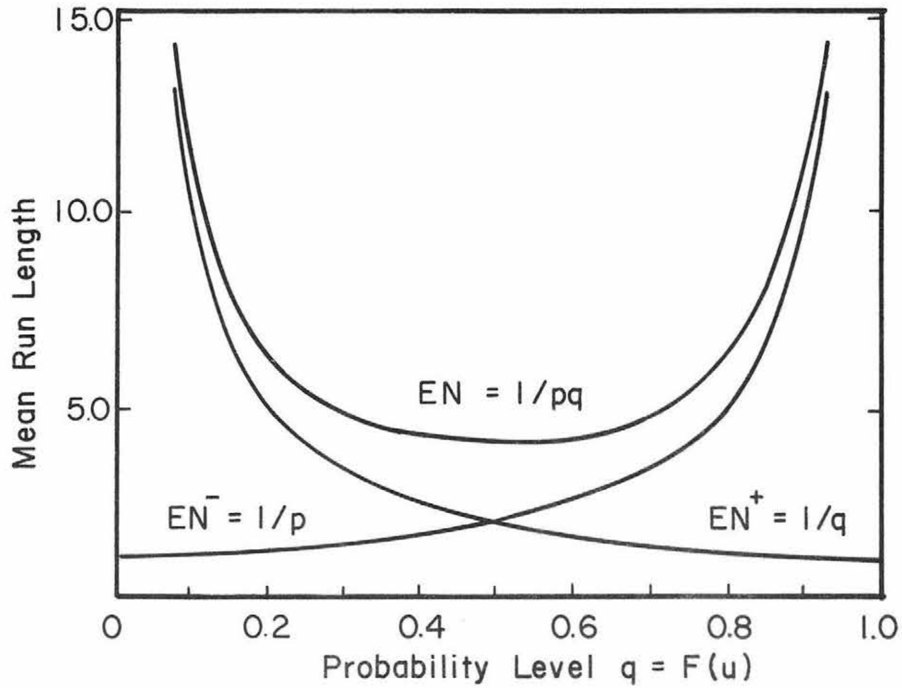

Fig. 5.3 Mean run-lengths for independent variables with a common distribution.

1. $\lim_{q \to 1} EN^+ = \sum_{j=1}^{\infty} P(N^+ \geq j) = 1.0$.

2. $\lim_{q \to 0} EN^+ = \infty$, i.e., the curve is asymptotic to the axis EN^+ at $q = 0^+$.

3. It must be such that the equation

$$EN^+ = \frac{1}{q}$$

is a particular case of it.

Consider the exponential function

$$q = \exp\left[- \frac{1}{1-\delta} \quad (EN^+-1) \right] , \tag{5.11}$$

where ρ is a parameter, and expand the exponential function by using the Taylor expansion:

$$\frac{1}{q} = \exp\left[\frac{1}{1-\delta} \ (EN^+-1) \right] = 1 + \frac{1}{1-\delta} \ (EN^+-1) + \ldots \tag{5.12}$$

Truncation of the series after the second term gives

$$EN^+ = 1 + \frac{1}{1-\delta} \left(\frac{1}{q} - 1 \right) = 1 + \frac{p}{(1-\delta)q} . \tag{5.13}$$

For $\delta = 0$, $EN^+ = \frac{1}{q}$, which is the relationship for the independent case. Equation 5.11 satisfies the three conditions stated above and it is adopted here as a mathematical model for the approximate values of mean run-length of the dependent case. A parallel analysis leads to

$$EN^- = 1 + \frac{q}{(1-\delta)p} \qquad 5.14$$

$$EN = EN^+ + EN^- = 2 + \frac{1}{1-\delta}\left(\frac{p}{q} + \frac{q}{p}\right) . \qquad 5.15$$

For $\delta = 0$ $\quad EN^- = \frac{1}{p}$ $\qquad$ 5.16

and

$$EN = 2 + \frac{p^2+q^2}{pq} = \frac{(p+q)^2}{pq} = \frac{1}{pq} , \qquad 5.17$$

which refer to the independent case.

5.4 Run-Length Test

The properties of $\overline{N}_k$ derived in previous subchapter allows the construction of a test. The null hypothesis is that $\{X_n\}$ is a sequence of independent identically distributed variables. Then $\overline{N}_k$ is approximately normally distributed for large k with the mean and the variance given by equations 5.6 and 5.7. At the $1-\alpha$ confidence level the region of acceptance of the hypothesis is

$$EN - t_{\alpha/2}\,(\mathrm{var}\ \overline{N}_k)^{1/2} \leq \overline{N}_k \leq EN + t_{\alpha/2}\,(\mathrm{var}\ \overline{N}_k)^{1/2}$$

5.18

or

$$\frac{1}{pq} - \frac{t_{\alpha/2}}{pq}\left(\frac{p^3+q^3}{k}\right)^{1/2} \leq \overline{N}_k \leq \frac{1}{pq} + \frac{t_{\alpha/2}}{pq}\left(\frac{p^3+q^3}{k}\right)^{1/2} . \qquad 5.19$$

Now, for a median as the truncation level

$$p = q = 1/2; \quad EN = 4 \qquad 5.20$$

$$\text{var } N = 4 \quad \text{and} \quad \text{var } \overline{N}_k = \frac{4}{k} . \qquad 5.21$$

The 95% confidence limits, with $\alpha = .05$ and $t_{\alpha/2} = 1.96$, are

$$4 - \frac{3.92}{\sqrt{k}} \leq \overline{N}_k \leq 4 + \frac{3.92}{\sqrt{k}} \qquad 5.22$$

If $\overline{N}_k$ falls outside the limits of Eq. 5.22 the hypothesis is rejected. The test is illustrated by Fig. 5.4 for the case of the confidence level $1-\alpha = 0.95$ and the truncation level being the median. For a truncation level $q \neq 1/2$, the confidence limits are different than those given in Fig. 5.4, as indicated by Fig. 5.5. For a right-sided run-test the 95% confidence limit is

$$\frac{1}{pq} + \frac{t_\alpha}{pq}\left(\frac{p^3+q^3}{k}\right)^{1/2} .$$

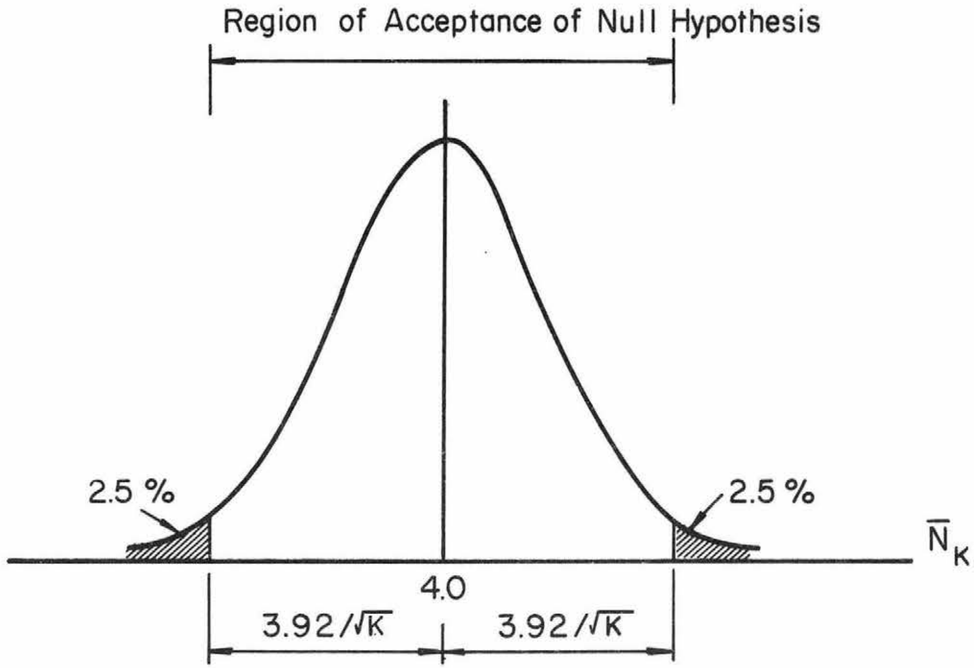

Fig. 5.4 Two-sided run-length test for the truncation level of the median.

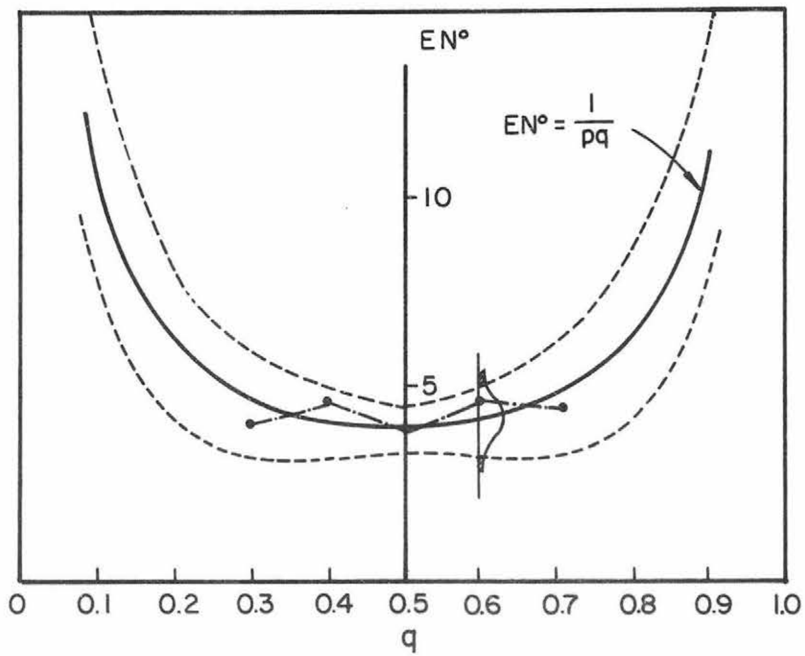

Fig. 5.5 Confidence region for $\overline{N}_k$ and observed $\overline{N}_k$ for an independent time series.

For a left-sided run-length test the 95% confidence limit is

$$\frac{1}{pq} - \frac{t_\alpha}{pq}\left(\frac{p^3+q^3}{k}\right)^{1/2} .$$

5.5 Preliminary Analysis

Before proceeding with a detailed analysis it would be desirable to take a quick look at the time series to see if the sample record is obviously nonstationary. The sequence may be nonstationary because of non-homogeneity in data.

5.6 Run-Length Test for Stationarity

The first assumption is that if the time series are stationary then they are also ergodic. In actual practice, random data representing stationary physical phenomena are generally ergodic (Bendat and Piersol 1966). This assumption permits one to measure the properties of a physical process from a single observed time series. It also permits one to accept the proof of selfstationarity of a single sequence as a proof of stationarity for the entire process.

The second assumption is that the verification of weak stationarity is sufficient for the analysis. In fact, random data of physical phenomena generally are strongly stationary if they are weakly stationary (Bendat and Piersol

1966). On the other hand if the process is Gaussian this assumption is automatically satisfied.

The third assumption is that the series is sufficiently long compared to its sampling fluctuation, so that short time averages truly reflect the average properties of data, and not just the sampling fluctuation.

The fourth assumption is, if the variance is stationary then the autocorrelation function is also stationary. The stationarity of time sequences is tested as follows:

a. Divide the time series $\{X_n\}$ into m equal time intervals, such that the statistical properties of the data for each of these intervals are independent among themselves.

b. Calculate the mean and the variance for each interval and obtain two new sequences: $\{\overline{X}_m\}$ and $\{\overline{X}^2{}_m\}$.

c. Set up the null hypothesis that

$$\{X_n\} \text{ is a stationary sequence.}$$

Then under this hypothesis

$$\{X_n\} \text{ is stationary if } \{\overline{X}_m\} \text{ and } \{\overline{X}^2{}_m\}$$

are independent random variables with a common distribution.

d. Select a probability level (say $q = 0.5$)

e. Apply the run-length test to the sequence $\{\overline{X}_m\}$ and the sequence $\{\overline{X}^2{}_m\}$.

f. If the hypothesis is accepted by both sequences, the time series is accepted as stationary.

g. This test could be performed at any arbitrary level, with $q \neq 0.5$.

5.7 Run-Length Test for Independence

If a sequence is not accepted as stationary, a non-stationary analysis is required which is beyond the scope of this study. It usually involves the investigation of the type of non-stationarity, and/or in many cases the transformation of the series to some kind of stationary series by decomposing the process and removing the non-stationary component. If a sequence is accepted as stationary the test for independence is as follows:

a. Set up the null hypothesis of $\{X_n\}$ being a sequence of independent random variables with a common distribution.

b. Select a probability level (say $q = .5$).

c. Apply the run-length test to the sequence $\{X_n\}$.

d. If the hypothesis is accepted, $\{X_n\}$ is considered as an independent sequence. Otherwise it is considered as a dependent one.

5.8 Theoretical Relationship between Mean Run-Length and Truncation Level

After the test for independence has been performed and the result is that the sequence is independent, the relationship between the mean run-length and the truncation level is

$$EN = \frac{1}{pq} \quad . \tag{5.25}$$

If the hypothesis of independence is rejected, the relationship may be approximated by

$$EN = 2 + \frac{1}{1-\delta} \left(\frac{1-q}{q} + \frac{q}{1-q} \right) \tag{5.26}$$

where the parameter δ must be estimated from the observed data. For a set of values $\{q\}$, a set of the sample mean total run-lengths $\overline{N}_q$ is obtained. The difference between the observed and the theoretical values of the mean total run-length is

$$\Delta N(q) = N(q) - 2 - \frac{1}{1-\delta} \left(\frac{p}{q} + \frac{q}{p} \right) . \tag{5.27}$$

As a criterion to estimate δ, the sum of the squares of the deviations from the theoretical line should be minimized. In that case,

$$\frac{\partial}{\partial \delta} \sum_{\{q\}} \left\{ \overline{N}(q) - 2 - \frac{1}{1-\delta} \left(\frac{p}{q} + \frac{q}{p} \right) \right\}^2 = 0$$

$$- 2 \sum_{\{q\}} \left\{ \overline{N}(q) - 2 - \frac{1}{1-\delta} \left(\frac{p}{q} + \frac{q}{p} \right) \right\} \left(\frac{p}{q} + \frac{q}{p} \right) = 0$$

$$\frac{1}{1-\delta} = \frac{\sum_{\{q\}} [\overline{N}(q)-2] \left(\frac{p}{q} + \frac{q}{p} \right)}{\sum_{\{q\}} \left(\frac{p}{q} + \frac{q}{p} \right)^2} \quad . \tag{5.28}$$

Chapter VI

APPLICATION OF RUNS TO INVESTIGATION OF HYDROLOGIC ANNUAL TIME SERIES

6.1 Introduction

The application of runs to hydrologic annual time series is two-fold. First, they are used to investigate the structure of a series by testing whether or not a particular time series is a sequence of independent random variables with a common distribution. If it is so, then this distribution is sufficient to generate synthetic sequences of annual values. If it is not so, then the first-order or the second-order autoregressive model is assumed, the stochastic component of the original series is computed from these models and tested to determine if it is really a sequence of independent variables with a common distribution. If this hypothesis is accepted then the first-order or the second-order autoregressive model may be used for the generation of large samples of annual values by the Monte Carlo method. Second the runs are used for prediction of durations of multiannual periods of surpluses and deficits at a particular station. Once the structure of the time series is known, i.e., whether it is stochastically independent or the first-order or the second-order autoregressive models, the derived properties of runs are used to make probability statements about durations of multi-annual periods of surpluses and deficits. The target level is specified by the probability truncation level, q.

6.2 Application to Annual Precipitation Series

A computed value of annual precipitation has random errors, systematic errors (inconsistency) and nonhomogeneity in data. Inconsistency is caused primarily by changes in instruments, method and time of measurements, etc. Non-homogeneity derives primarily from two sources: a) movement of precipitation station by a substantial distance; and b) changes in the environment of a station, i.e., tree growth, building of houses, or any other substantial change around the station which affects the flow pattern of air. Precipitation data must be considered, therefore, as often having relatively large errors and non-homogeneity in its annual values (Yevjevich, 1963).

The presence of inconsistency and/or non-homogeneity in the observed time series implies that it does not come from a sequence of variables with a common distribution. Therefore, when applying the run-length test to an observed precipitation series with inconsistency and/or non-homogeneity in the data, the null hypothesis may be rejected. This means that it is not possible to generate synthetic values of annual precipitation at that station simply by computing the probability distribution and then using the Monte Carlo method. It is necessary to first remove inconsistency and/or non-homogeneity in data.

The presence of inconsistency and/or non-homogeneity in data is reflected in the run-length test by the fact that $\bar{N}$ is greater than four, and may be outside the right tail 95% confidence limit of the distribution of $\bar{N}$.

On the other hand, if no significant inconsistency and/or non-homogeneity is present in the data and the observed time series is stochastically independent the null hypothesis is accepted in applying the run-length test. In this case $\bar{N}$ is inside the region of acceptance of Ho at the 95% confidence level. Under these circumstances it is possible to compute the probability distribution of observed data and then use it as a technique for generating large samples of annual precipitation by the Monte Carlo method.

6.3 Examples of Run-Length Test with Annual Precipitation Series

In order to show the applicability of the run-length test to annual precipitation series, a limited number of stations were selected in the United States. It is important to notice that the applicability of this test is limited because it was developed on the assumption that the number of total runs is large. Therefore, the observed time series must be long enough to satisfy this assumption. For this reason only a few of the longest precipitation series were selected.

Example 1. The annual precipitation data of the station No. 4.1715, Chico Experimental Station, California, is considered. Data are given in Table B.1, Appendix B. Sequences of observed runs are given below in Table 6.1 for $q = 0.5$. The null hypothesis is that the time series is a sequence of independent identically distributed variables. Since $\bar{N} > 4$ the alternative hypothesis which is associated with a high value of $\bar{N}$ is that inconsistency and/or non-homogeneity and/or positive serial correlation are present in the data. The 95% confidence limit using a right-sided test is computed as

$$4 + \frac{2t_\alpha}{\sqrt{K}} = 4.740$$

with $\alpha = 0.05$, $t_\alpha = t_{.05} = 1.645$ and $K = 21$ which is the maximum value of j given in Table 6.1. The observed mean total run-length is $\bar{N} = 4.058$ which is less than 4.740. Therefore the null hypothesis is accepted and the annual precipitation time series is considered as a sequence of independent variables with a common distribution.

Example 2. A similar analysis was performed on the annual precipitation data of the station No. 25.6335 Ord, Nebraska. Data are given in Table B.2 Appendix B. The results obtained are given below in Table 6.2 for $q = 0.5$. The annual precipitation series is accepted as a sequence

TABLE 6.1

RUN-LENGTH PROPERTIES OF ANNUAL PRECIPITATION OF CHICO EXPERIMENTAL STATION, CALIFORNIA

j	N_j^-	N_j^+	$\overline{N}^+$	$\overline{N}^-$	$\overline{N}$	95% confidence limit
1	1	1	2.190	2.238	4.428	4.718
2	1	1				
3	3	2				
4	9	1				
5	2	5				
6	2	1				
7	2	1				
8	1	4				
9	1	1				
10	3	4				
11	3	1				
12	1	1				
13	3	2				
14	7	4				
15	1	6				
16	2	1				
17	1	3				
18	1	2				
19	1	2				
20	1	1				
21	1	2				
Σ	47	46				

$4.428 < 4.718$

$\therefore$ The time series is a sequence of independent identically distributed variables.

TABLE 6.2

RUN-LENGTH PROPERTIES OF ANNUAL PRECIPITATION OF ORD, NEBRASKA

j	N_j^+	N_j^-	$\overline{N}^+$	$\overline{N}^-$	$\overline{N}$	95% confidence limit
1	2	2				
2	4	1	1.882	2.176	4.058	4.798
3	2	1				
4	1	4				
5	1	1				
6	1	4				
7	5	1				
8	1	2				
9	2	4				
10	1	6				
11	1	1				
12	1	1				
13	6	1				
14	1	3				
15	1	2				
16	1	1				
17	1	2				
Σ	32	37				

$4.058 < 4.798$

$\therefore$ The time series is a sequence of independent identically distributed variables.

of independent variables with a common distribution.

Example 3. In order to show the applicability of the run-length test for detecting non-homogeneity in annual precipitation data, a non-homogeneous station was selected, station No. 2.5825, Natural Bridge N. M., Arizona. The run-lngth properties of the annual series of this station are given below in Table 6.3.

TABLE 6.3

RUN-LENGTH PROPERTIES OF ANNUAL PRECIPITATION OF NATURAL BRIDGE, N. M., ARIZONA

J	N_j^+	N_j^-	$\overline{N}^+$	$\overline{N}^-$	$\overline{N}$	95% confidence limit
1	1	14	2.429	2.500	4.929	4.880
2	5	1				
3	1	1				
4	8	1				
5	2	2				
6	2	2				
7	2	3				
8	2	3				
9	2	2				
10	3	2				
11	1	1				
12	2	1				
13	2	1				
14	1	1				
Σ	34	35				

$4.929 > 4.880$

$\therefore$ The time series is not a sequence of independent identically distributed variables.

In this case, the null hypothesis is rejected and the alternative hypothesis is accepted at $\alpha = 0.05$ level of significance. This result is consistent with the fact that this particular time series has non-homogeneity in data.

Examples 4 and 5. Two more precipitation stations having long records are considered: No. 4.0227, Antioch F. Mills, California, and No. 25.7040, Ravenna, Nebraska. The annual precipitation data of these stations are given in Tables B.4 and B.5, Appendix B. The run-length properties of the observed time series are given in Tables 6.4 and 6.5 for $q = 0.5$. The computed mean run-length in both of these cases is smaller than four, indicating a possible negative serial correlation in the series. Therefore, the alternative hypothesis is that the series is negatively serially correlated. Then the run-length test, in this case is a left sided test and the 95% confidence limit is given by

$$4 - \frac{2t_\alpha}{\sqrt{K}} \quad \text{with} \quad t_\alpha = 1.645.$$

In both cases, the null hypothesis is accepted and therefore the series are considered as sequences of independent random variables with common distributions.

6.4 Application to Annual River Flows Series

The river flow essentially integrates the precipitation received on large areas but also includes the effects of storage and evaporation as other important physical factors. The water carryover from year to year and especially the change in this carryover from year to year usually introduces time dependence into the sequences of annual flows. For the purpose of generating large samples of annual river flow data by the Monte Carlo method, it is not sufficient to use only the probability distribution of annual runoff, because the annual flow series are more

TABLE 6.4

RUN-LENGTH PROPERTIES OF ANNUAL PRECIPITATION OF ANTIOCH F. MILLS, CALIFORNIA

j	N_j^+	N_j^-	$\overline{N}^+$	$\overline{N}^-$	$\overline{N}$	95% confidence limit
1	1	1	1.833	1.708	3.541	3.329
2	1	3				
3	6	4				
4	1	1				
5	1	2				
6	1	1				
7	1	2				
8	2	1				
9	1	1				
10	1	1				
11	3	2				
12	1	1				
13	2	2				
14	2	3				
15	1	3				
16	3	4				
17	3	1				
18	2	1				
19	1	2				
20	3	1				
21	2	1				
22	2	1				
23	1	1				
24	2	1				
Σ	44	41				

$3.541 > 3.329$

$\therefore$ The time series is a sequence of independent identically distributed variables.

TABLE 6.5

RUN-LENGTH PROPERTIES OF ANNUAL PRECIPITATION OF RAVENNA, NEBRASKA

j	N_j^-	N_j^+	$\overline{N}^+$	$\overline{N}^-$	$\overline{N}$	95% confidence limit
1	1	1	1.826	1.870	3.696	3.315
2	1	3				
3	1	5				
4	1	2				
5	3	2				
6	1	1				
7	1	6				
8	1	1				
9	4	1				
10	1	1				
11	2	3				
12	2	1				
13	2	1				
14	1	1				
15	1	1				
16	1	1				
17	2	1				
18	1	1				
19	4	1				
20	3	1				
21	2	1				
22	1	1				
23	5	6				
Σ	42	43				

$3.696 > 3.315$

$\therefore$ The time series is a sequence of independent identically distributed variables.

often than not time dependent series. The first-order or the second-order autoregressive models often fit well the time dependence of annual river flows.

The question of whether or not a sequence of annual flows can be simulated by using the first-order or the second-order autoregressive models is answered. It may happen, however, that for a particular gaging station the carryover effect is small, the serial correlation is not significant. In that case, the distribution of annual flows is sufficient for generating large samples of annual flows.

The general procedure is as follows. First, the run-length test is applied to the original series $\{X_i\}$ to determine if it is a sequence of independent variables with a common distribution. If so, no further tests are needed. If $\overline{N} > 4$, there may be significant positive serial correlation in data, or possibly inconsistency or non-homogeneity. In this case, the alternative hypothesis is that positive serial correlation, inconsistency or non-homogeneity may be present in the data. If $\overline{N} < 4$, the alternative hypothesis is that there may be significant negative serial correlation in the data. The 95% confidence limits are then

$$4 \pm \frac{3.29}{\sqrt{K}} \quad ,$$

where the plus sign is used when $\overline{N} > 4$ and the minus sign is used when $\overline{N} < 4$.

If the null hypothesis is rejected, the first-order or the second-order autoregressive models are fitted to the data. Then a new test is set up with a null hypothesis that the sequence $\{X_i\}$ is the first-order or the second-order autoregressive process. Under the hypothesis of the first-order autoregressive model,

$$e_i = \frac{x_i - \overline{x}}{s} - \frac{r_1 (x_{i-1} - \overline{x})}{s}$$

is a sequence of independent variables with a common distribution. The alternative hypothesis is that $\{X_i\}$ is not a first-order autoregressive process. In this case the run-length test is a two-sided test and the 95% confidence limits are

$$4 \pm \frac{2t_{\alpha/2}}{\sqrt{K}}$$

with $\alpha = 0.05$ and $t_{\alpha/2} = 1.96$. If the null hypothesis is accepted no further analysis is needed and large samples of annual flows may be simulated by means of this first-order autoregressive model. If it is rejected, it may be that there is inconsistency and/or non-homogeneity in the data. It may also occur that a higher-order autoregressive process is needed to fit the data. Finally it is theoretically possible that an autoregressive model is not adequate. Experience shows that the possibility of this latter alternative is remote and may occur only when there is inconsistency and/or non-homogeneity in the data. This assertion is supported on the basis of previous investigations made on annual river flows (Yevjevich, 1964).

6.5 Examples of Run-Length Test with Annual River Flows Series

As in the case of annual precipitation series the applicability of the tests is limited to long series for the same reason. Five series of annual river flows with long records were selected as examples:

1. Mississippi River at Saint Louis, Missouri
2. St. Lawrence River at Ogdensburg, New York
3. Mississippi River at Keokuk, Iowa
4. Gota River at Sjotorp, Vanersburg, Sweden
5. Rhine River at Basle, Switzerland.

The data of these rivers taken from the research data assembly of the Hydrology Program at Colorado State University are given in Tables B.6 - B.10 in Appendix B. The analysis by runs is given in Table 6.6. For the first four rivers, the first-order autoregressive model is accepted as the proper model. Annual flow series of the Rhine River is accepted as being a sequence of independent variables with a common distribution.

TABLE 6.6

RESULTS OF RUN-LENGTH TESTS OF DEPENDENCE MODELS FOR THE ANNUAL RIVER FLOW SERIES

Station	x_i					$x_i - r_1 x_{i-1}$						Result
	$\bar{N}^+$	$\bar{N}^-$	$\bar{N}$	k	95% C.L.	r_1	$\bar{N}^+$	$\bar{N}^-$	$\bar{N}$	k	95% C.L.	
1. Mississippi River at St. Louis, Missouri	2.389	2.500	4.889	18	4.775	.283	1.783	2.130	3.913	23	3.184	FAM
2. St. Lawrence River at Ogdensburg	3.583	3.917	7.500	12	4.950	.705	2.045	2.136	4.181	22	4.835	FAM
3. Mississippi River at Keokuk, Iowa	2.500	2.857	5.357	14	4.878	.410	1.714	1.619	3.333	21	3.146	FAM
4. Gota River at Sjotorp, Vanersborg	2.704	2.778	5.482	27	4.634	.461	2.242	2.212	4.454	33	4.682	FAM
5. Rhine River at Basle, Switzerland	2.027	2.027	4.054	37	4.540							I

CL = Confidence limit; FAM = First order autoregressive model; I = independent identically distributed variables.

Chapter VII

CONCLUSIONS

In consideration of the foregoing results, the writer ventures the following conclusions;

1) A technique has been advanced for testing the structure of time series, with the basic statistical parameter being the mean total run-length. More specifically it is used for testing:

a. Whether or not an annual precipitation time series is a sequence of stochastically independent variables with a common distribution. In this case this distribution is sufficient for generating samples of annual precipitation by the Monte Carlo method.

b. Whether or not an annual river flows time series is independent, the first-order, the second-order or higher-order autoregressive process, with these various models being used for generation of large samples by the Monte Carlo method.

2) This technique does not depend on the underlying distribution of variables that are being tested. In other words, it is a distribution-free technique. In this sense, it has advantage over other techniques for the investigation of time series which depend on the distribution of the variable of a given time series.

3) Autoregressive models and the moving average schemes, which are widely used in hydrology, usually refer to stationary Gaussian processes, if the independent stochastic component is normally distributed. The properties of runs of these models are relevant for the investigations of multiannual periods of surplus and deficit, and for the study of hydrologic droughts.

4) An analytical expression is developed here, by which the probabilities of sequences of wet and dry years of specified lengths can be calculated when the basic hydrologic time series are either independent or stationary Gaussian processes, and the truncation level is specified. Numerical values of these probabilities are obtained for a first order autoregressive process by means of the digital computer. The range of r_1 (the first serial correlation coefficient) values is between 0 and 0.5, with increments 0.1, and the range of q values (probability of truncation level) is between 0.3 and 0.7 with increments of 0.1. These probabilities can be readily used for making probability statements about the multiannual periods of surpluses and deficits, with respect to a specified target level. They are presented in a set of graphs in order to make them useable.

5) It has been shown that the probabilities of runs of dependent variables with a common distribution do not depend on the underlying univariate distribution of the

variable. They depend on the probability q of a truncation level and the time dependence model. Therefore, the same probabilities of runs obtained for the stationary Gaussian processes may be used for non-Gaussian processes for the same probability truncation level q.

REFERENCES

Bendat, J. S. and Piersol, A. G., 1966, Measurement and analysis of random data. Wiley and Sons, New York.

Blum, J. R., Hanson, D. L. and Rosenblatt, J. I., 1963, On the central limit theorem for the sum of a random number of independent random variables. Z. Wahrscheinlichkeitstheoric, 1, pp. 389-393.

Cochran, W. G., 1936, The statistical analysis of field counts of diseased plants. J. R. Statistical Society Suppl., 3: 49-67.

Cox, D. R., Miller, H. D., 1965, The theory of stochastic processes. Methuen and Co. Ltd., London.

Cramer, H. and Leadbetter, M. R., 1967, Stationary and related stochastic processes. Wiley and Sons, New York.

David, F. N., 1953, A note on the evaluation of the multivariate normal integral. Biometrika 40, pp. 458-459.

Downer, R. N., Siddiqui, M. M., and Yevjevich, V., 1967, Application of runs to hydrologic droughts. Proceedings of the International Hydrology Symposium, Fort Collins, Colorado.

Feller, W., 1957, An introduction to probability theory and its applications. Wiley and Sons, New York.

Gupta, S. S., 1963a, Bibliography on the multivariate normal integrals and related topics. Annals of Mathematical Statistics, Vol. 34, pp. 829-838.

Gupta, S. S., 1963b, Probability integrals of multivariate normal and multivariate t. Annals of Mathematical Statistics, Vol. 34, pp. 792-828

Kendall, M. G., 1941, Proof of relations connected with the tetrachoric series and its generalization. Biometrika, Vol. 32, pp. 196-198.

Heiny, R., 1968, Stochastic variables of surplus and deficit. Unpublished Ph.D. Dissertation, Colorado State University.

Kendall, M. G. and Stuart, A., 1958, The advanced theory of statistics. Charles Griffin and Co., London. Vol. I.

Kendall, M. G. and Stuart, A., 1966, The advanced theory of statistics. Charles Griffin and Co., London, Vol. III.

Llamas, J., 1968, Deficit and surplus in precipitation series. Unpublished Ph.D. Dissertation, Colorado State University.

Llamas, J. and Siddiqui, M. M., 1969, Runs of precipitation series. Hydrology Paper No. 33, Colorado State University.

Mc Fadden, J. A., 1955, Urn models of correlation and a comparison with the multivariate normal integral. Annals of mathematical statistics Vol. 26, pp. 478-489.

Mood, A. M., 1940, The distribution theory of runs. Annals of mathematical statistics, p. 367.

National Bureau of Standards, 1959, Tables of the bivariate normal distribution and related functions. Applied Mathematics Series 50, U. S. Government Printing Office, Washington.

Sheppard, W. F., 1900, On the calculation of the double integral expressing normal correlation. Transactions of the Cambridge Philosophical Society. Vol. 19, pp. 23-66.

Siddiqui, M. M., 1957, Distribution of some serial correlation coefficients. Unpublished thesis, University of North Carolina.

Stevens, W. L., 1939, Distribution of groups in a sequence of alternatives. Annals of Eugenics, Vol. 9, p. 10.

Wald, A. and Wolfowitz, J., 1940, On a test whether two samples are from the same population. Annals of Mathematical Statistics. Vol. II.

Wishart, J. and Hirshfeld, H. O., 1936, A theory concerning the distribution of joins between line segments. London Mathematical Society Journal, Vol. 11, p. 227.

Yevjevich, V., 1963, Fluctuations of wet and dry years, Part I Research data assembly and mathematical models. Hydrology Paper No. 1, Colorado State University.

Yevjevich, V., 1964, Fluctuations of wet and dry years, Part II, Analysis by serial correlation. Hydrology Paper No. 4, Colorado State University.

Yevjevich, V., 1965, The application of surplus, deficit and range in hydrology. Hydrology Paper No. 10, Colorado State University.

Yevjevich, V., 1967, An objective approach to definitions and investigations of continental hydrologic droughts. Hydrology Paper No. 23, Colorado State University.

Yevjevich, V., 1969, Stochastic processes in hydrology. Unpublished class notes of a course given at Colorado State University.

Zalen, M. and Severo, N. C., 1960, Graphs for bivariate normal probabilities. Annals of Mathematical Statistics, Vol. 31, pp. 619-624.

APPENDIX A

Joint Probabilities, Probability Distribution and Probability Mass of Positive Runs of a First Order Autoregressive Model

JOINT PROBABILITIES, PROBABILITY DISTRIBUTION AND PROBABILITY MASS
OF POSITIVE RUNS OF A FIRST ORDER AUTOREGRESSIVE MODEL

F(U) = .300

ρ = .1

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
3.447	7.766	114.522

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.700000	.197744	.067782	.022847	.007705	.002601	.000877	.000295	.000099	.000033	.000011	.999998	.283161
2	.502256	.141493	.048273	.016403	.005569	.001886	.000637	.000214	.000072	.000024	.000008	.716838	.201899
3	.360763	.101656	.034698	.011785	.003998	.001353	.000457	.000154	.000051	.000017	.000006	.514939	.145079
4	.259108	.073018	.024936	.008465	.002870	.000971	.000327	.000110	.000037	.000012	.000004	.369859	.104217
5	.186089	.052448	.017918	.006079	.002060	.000696	.000234	.000079	.000026	.000009	.000003	.265643	.074862
6	.133641	.037672	.012873	.004365	.001478	.000499	.000168	.000056	.000019	.000006	.000002	.190780	.053775
7	.095969	.027058	.009247	.003134	.001060	.000358	.000120	.000040	.000013	.000004	.000001	.137006	.038625
8	.068911	.019433	.006641	.002249	.000760	.000256	.000086	.000029	.000010	.000003	.000001	.098380	.027742
9	.049477	.013956	.004769	.001614	.000545	.000184	.000062	.000021	.000007	.000002	.000001	.070638	.019924
10	.035521	.010022	.003424	.001158	.000391	.000132	.000044	.000015	.000005	.000002	.000001	.050713	.014308

ρ = .2

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
3.604	8.366	124.837

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.700000	.185156	.072537	.026685	.009846	.003653	.001350	.000494	.000179	.000064	.000023	.999996	.267111
2	.514844	.134717	.051846	.019683	.007430	.002773	.001021	.000371	.000133	.000047	.000016	.732895	.191512
3	.380127	.099663	.038396	.014532	.005464	.002031	.000745	.000269	.000096	.000034	.000012	.541373	.141861
4	.280464	.073585	.028402	.010716	.004014	.001486	.000543	.000196	.000070	.000025	.000009	.399512	.104777
5	.206879	.054332	.020986	.007893	.002945	.001086	.000395	.000142	.000051	.000018	.000006	.294735	.077383
6	.152547	.040111	.015489	.005807	.002159	.000793	.000288	.000103	.000037	.000013	.000004	.217352	.057141
7	.112435	.029605	.011420	.004267	.001581	.000579	.000209	.000075	.000026	.000009	.000003	.160211	.042181
8	.082830	.021843	.008411	.003133	.001154	.000422	.000152	.000054	.000019	.000007	.000002	.118030	.031126
9	.060987	.016110	.006188	.002298	.000845	.000307	.000111	.000039	.000014	.000005	.000002	.086905	.022957
10	.044877	.011875	.004548	.001683	.000617	.000224	.000080	.000028	.000010	.000003	.000001	.063948	.016923

ρ = .3

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
3.812	9.354	144.628

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.700000	.172235	.077380	.030391	.012019	.004835	.001927	.000753	.000288	.000108	.000040	.999992	.251759
2	.527765	.126763	.054818	.023070	.009553	.003849	.001508	.000576	.000215	.000079	.000028	.748233	.178930
3	.401002	.097013	.041964	.017503	.007176	.002866	.001114	.000423	.000157	.000057	.000021	.569303	.137560
4	.303989	.073699	.032007	.013235	.005374	.002129	.000822	.000310	.000115	.000042	.000015	.431743	.104587
5	.230290	.056000	.024330	.009977	.004018	.001579	.000606	.000227	.000084	.000030	.000011	.327156	.079519
6	.174290	.042540	.018437	.007501	.002996	.001149	.000446	.000166	.000061	.000022	.000008	.247637	.060431
7	.131750	.032291	.013930	.005624	.002229	.000864	.000327	.000122	.000044	.000016	.000006	.187205	.045885
8	.099459	.024487	.010497	.004207	.001655	.000637	.000240	.000089	.000032	.000012	.000004	.141350	.034801
9	.074972	.018545	.007890	.003140	.001227	.000470	.000176	.000065	.000023	.000008	.000003	.106519	.026358
10	.056427	.014025	.005916	.002339	.000908	.000345	.000129	.000047	.000017	.000006	.000002	.080162	.019933

ρ = .4

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
4.037	10.423	166.688

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.700000	.158982	.082414	.033964	.014148	.006120	.002605	.001074	.000428	.000166	.000063	.999907	.237044
2	.541018	.117692	.057191	.026565	.011938	.005116	.002098	.000829	.000318	.000119	.000044	.762943	.164159
3	.423325	.093825	.045402	.020697	.009134	.003858	.001565	.000613	.000234	.000087	.000032	.598704	.132432
4	.329500	.073338	.035749	.016022	.006957	.002901	.001165	.000453	.000172	.000064	.000023	.466352	.103674
5	.256163	.057386	.027950	.012334	.005278	.002175	.000865	.000334	.000126	.000046	.000017	.362678	.081208
6	.198777	.044886	.021715	.009447	.003989	.001626	.000641	.000246	.000092	.000034	.000012	.281469	.063561
7	.153891	.035058	.016778	.007204	.003005	.001212	.000474	.000181	.000067	.000025	.000009	.217908	.049663
8	.118833	.027323	.012900	.005472	.002257	.000902	.000350	.000133	.000049	.000018	.000006	.168245	.038711
9	.091509	.021239	.009874	.004141	.001691	.000670	.000258	.000097	.000036	.000013	.000005	.129534	.030090
10	.070270	.016462	.007527	.003124	.001263	.000496	.000190	.000071	.000026	.000009	.000003	.099444	.023319

S = .5

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	4.278	11.549	190.680

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+\geq j)$	$P(N^+=j)$
1	.700000	.145397	.087729	.037424	.016155	.007476	.003377	.001457	.000601	.000239	.000092	.999990	.222930
2	.554603	.107541	.058363	.030167	.014584	.006572	.002792	.001132	.000442	.000148	.000062	.777050	.147124
3	.447062	.090241	.048710	.024115	.011338	.005009	.002098	.000842	.000326	.000123	.000045	.629926	.126772
4	.356821	.072495	.039531	.019078	.008757	.003800	.001571	.000624	.000240	.000090	.000033	.503154	.102086
5	.284326	.058426	.031845	.014962	.006725	.002873	.001174	.000462	.000177	.000066	.000024	.401068	.082397
6	.225899	.047080	.025326	.011646	.005138	.002164	.000875	.000342	.000130	.000048	.000018	.318671	.066450
7	.178819	.037847	.019364	.009008	.003909	.001625	.000650	.000252	.000095	.000035	.000013	.252222	.053437
8	.140972	.030311	.015520	.006928	.002961	.001217	.000482	.000186	.000070	.000026	.000009	.198785	.042798
9	.110661	.024167	.012140	.005302	.002236	.000909	.000357	.000137	.000051	.000019	.000007	.155987	.034116
10	.086494	.019175	.009381	.004039	.001682	.000677	.000264	.000100	.000037	.000014	.000005	.121870	.027060

F(U) = .400

ρ = .1

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.645	4.191	48.749

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.600000	.225022	.098559	.043058	.018801	.008211	.003584	.001563	.000681	.000296	.000128	.999958	.375733
2	.374978	.140269	.061300	.026847	.011750	.005137	.002242	.000977	.000425	.000184	.000080	.624224	.233443
3	.234709	.087828	.038406	.016812	.007353	.003212	.001401	.000610	.000265	.000115	.000050	.390792	.146208
4	.146881	.054972	.024055	.010524	.004599	.002007	.000874	.000380	.000165	.000071	.000031	.244574	.091525
5	.091909	.034407	.015061	.006585	.002875	.001254	.000546	.000237	.000103	.000044	.000019	.153049	.057292
6	.057502	.021534	.009426	.004118	.001797	.000783	.000340	.000148	.000064	.000028	.000012	.095757	.035859
7	.035968	.013475	.005897	.002575	.001122	.000488	.000212	.000092	.000040	.000017	.000007	.059898	.022441
8	.022493	.008431	.003688	.001609	.000701	.000305	.000132	.000057	.000025	.000011	.000005	.037457	.014040
9	.014062	.005274	.002305	.001005	.000437	.000190	.000082	.000036	.000015	.000007	.000003	.023417	.008782
10	.008788	.003298	.001440	.000627	.000273	.000118	.000051	.000022	.000010	.000004	.000002	.014634	.005492

ρ = .2

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.823	4.890	59.924

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.600000	.209947	.100320	.047473	.022381	.010546	.004975	.002330	.001085	.000501	.000230	.999914	.352677
2	.390053	.135103	.064040	.030625	.014575	.006888	.003230	.001502	.000693	.000318	.000144	.647236	.223833
3	.254950	.088551	.042061	.020044	.009498	.004468	.002086	.000966	.000444	.000203	.000092	.423403	.146973
4	.166398	.057860	.027561	.013085	.006173	.002891	.001344	.000620	.000284	.000129	.000058	.276430	.096098
5	.108538	.037811	.018017	.008520	.004003	.001867	.000864	.000397	.000181	.000082	.000037	.180332	.062823
6	.070727	.024697	.011750	.005534	.002589	.001202	.000554	.000254	.000115	.000052	.000023	.117509	.041043
7	.046030	.016117	.007544	.003586	.001671	.000773	.000355	.000162	.000073	.000033	.000015	.076466	.026785
8	.029913	.010505	.004962	.002318	.001074	.000496	.000227	.000103	.000047	.000021	.000009	.049601	.017457
9	.019408	.006837	.003213	.001495	.000691	.000317	.000145	.000066	.000030	.000013	.000006	.032224	.011359
10	.012571	.004443	.002076	.000963	.000443	.000203	.000092	.000042	.000019	.000008	.000004	.020865	.007379

ξ = .3

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
3.021	5.681	73.105

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.600000	.194777	.101354	.051648	.026016	.013175	.006628	.003294	.001615	.000781	.000373	.999848	.330745
2	.405223	.128579	.065819	.034373	.017690	.008941	.004438	.002166	.001041	.000494	.000231	.669153	.211274
3	.276643	.088616	.045524	.023518	.011964	.005982	.002941	.001423	.000679	.000320	.000149	.457858	.146471
4	.188027	.060421	.031255	.015976	.008040	.003980	.001939	.000931	.000441	.000207	.000096	.311358	.100035
5	.127606	.041230	.021310	.010782	.005372	.002635	.001273	.000607	.000286	.000133	.000061	.211323	.068317
6	.086376	.028099	.014436	.007233	.003570	.001736	.000833	.000394	.000185	.000086	.000039	.143006	.046571
7	.058277	.019102	.009721	.004826	.002362	.001140	.000543	.000256	.000119	.000055	.000025	.096435	.031653
8	.039174	.012942	.006509	.003204	.001555	.000745	.000353	.000165	.000076	.000035	.000016	.064702	.021435
9	.026233	.008735	.004337	.002117	.001020	.000485	.000228	.000106	.000049	.000022	.000010	.043348	.014457
10	.017498	.005871	.002876	.001393	.000667	.000315	.000147	.000068	.000031	.000014	.000006	.028831	.009710

ξ = .4

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
3.240	6.543	88.124

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.600000	.179512	.101766	.055605	.029621	.015999	.008539	.004463	.002281	.001141	.000560	.999759	.309864
2	.420488	.120769	.066637	.038092	.021095	.011295	.005866	.002968	.001469	.000713	.000341	.689894	.195841
3	.299719	.088090	.048796	.027236	.014752	.007752	.003965	.001981	.000970	.000467	.000221	.494053	.144912
4	.211630	.062577	.035136	.019199	.010200	.005273	.002660	.001314	.000637	.000304	.000143	.349141	.103266
5	.149053	.044579	.024939	.013371	.006984	.003558	.001774	.000867	.000417	.000198	.000093	.245875	.073657
6	.104474	.031686	.017484	.009215	.004741	.002385	.001176	.000570	.000272	.000128	.000060	.172219	.052356
7	.072789	.022406	.012126	.006294	.003195	.001589	.000776	.000373	.000177	.000083	.000038	.119863	.036997
8	.050383	.015739	.008330	.004265	.002139	.001053	.000510	.000243	.000114	.000053	.000025	.082866	.025960
9	.034644	.010976	.005576	.002870	.001424	.000694	.000333	.000158	.000074	.000034	.000016	.056906	.018080
10	.023668	.007598	.003339	.001919	.000943	.000456	.000217	.000102	.000048	.000022	.000010	.038826	.012499

S = .5

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	3.477	7.453	104.753

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.600000	.164150	.101600	.059378	.033111	.018994	.010700	.005842	.003090	.001589	.000797	.999643	.289968
2	.435850	.111735	.066494	.041781	.024792	.013951	.007515	.003908	.001976	.000976	.000473	.709675	.177574
3	.324115	.087058	.051876	.031198	.017862	.009779	.005158	.002639	.001317	.000643	.000309	.532100	.142543
4	.237057	.064254	.039205	.022753	.012652	.006770	.003507	.001769	.000872	.000422	.000201	.389558	.105731
5	.172803	.047771	.028906	.016286	.008837	.004638	.002365	.001178	.000575	.000276	.000131	.283827	.078727
6	.125032	.035397	.020894	.011481	.006102	.003148	.001584	.000780	.000377	.000180	.000094	.205099	.058309
7	.089635	.026001	.014859	.007991	.004171	.002121	.001054	.000514	.000247	.000117	.000054	.146790	.042768
8	.063634	.018892	.010425	.005502	.002828	.001419	.000698	.000337	.000161	.000075	.000035	.104022	.031015
9	.044742	.013570	.007230	.003754	.001902	.000944	.000460	.000220	.000104	.000049	.000022	.073008	.022231
10	.031173	.009639	.004965	.002540	.001272	.000624	.000301	.000143	.000067	.000031	.000014	.050777	.015758

F(U) = 0.5

ρ = 0

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.000	2.000	17.990

J	P0(J)	P(1,J)	P(2,J)	$P(N^+=j)$	$P(N^+\leq j)$
1	.500	.250	.125	.500	.500
2	.250	.125	.063	.250	.750
3	.125	.063	.031	.125	.875
4	.063	.031	.016	.063	.937
5	.031	.016	.008	.031	.969
6	.016	.008	.004	.016	.984
7	.008	.004	.002	.008	.992
8	.004	.002	.001	.004	.996
9	.002	.001	.000	.002	.998
10	.001	.000	.000	.001	.999
11	.000	.000	.000	.000	1.000
12	.000	.000	.000	.000	1.000
13	.000	.000	.000	.000	1.000
14	.000	.000	.000	.000	1.000
15	.000	.000	.000	.000	1.000
16	.000	.000	.000	.000	1.000
17	.000	.000	.000	.000	1.000
18	.000	.000	.000	.000	1.000
19	.000	.000	.000	.000	1.000
20	.000	.000	.000	.000	1.000

ρ = .1

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.136	2.433	23.520

J	P0(J)	P(1,J)	P(2,J)	$P(N^+=j)$	$P(N^+\leq j)$
1	.500	.234	.124	.468	.468
2	.266	.124	.066	.248	.716
3	.142	.066	.035	.132	.849
4	.075	.035	.019	.071	.919
5	.040	.019	.010	.038	.957
6	.021	.010	.005	.020	.977
7	.011	.005	.003	.011	.987
8	.006	.003	.002	.006	.993
9	.003	.002	.001	.003	.996
10	.002	.001	.000	.002	.998
11	.001	.000	.000	.001	.999
12	.000	.000	.000	.000	.999
13	.000	.000	.000	.000	.999
14	.000	.000	.000	.000	.999
15	.000	.000	.000	.000	.999
16	.000	.000	.000	.000	.999
17	.000	.000	.000	.000	1.000
18	.000	.000	.000	.000	1.000
19	.000	.000	.000	.000	1.000
20	.000	.000	.000	.000	1.000

δ = .2

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.293	2.968	30.776

J	$P_0(J)$	P(1,J)	P(2,J)	$P(N^+=j)$	$P(N^+\leq j)$
1	.500	.218	.122	.436	.436
2	.282	.122	.069	.243	.679
3	.160	.069	.039	.139	.818
4	.091	.039	.022	.079	.896
5	.051	.022	.013	.045	.941
6	.029	.013	.007	.025	.966
7	.016	.007	.004	.014	.980
8	.009	.004	.002	.008	.988
9	.005	.002	.001	.004	.993
10	.003	.001	.001	.003	.995
11	.002	.001	.000	.001	.997
12	.001	.000	.000	.001	.998
13	.000	.000	.000	.000	.998
14	.000	.000	.000	.000	.998
15	.000	.000	.000	.000	.998
16	.000	.000	.000	.000	.998
17	.000	.000	.000	.000	.998
18	.000	.000	.000	.000	.998
19	.000	.000	.000	.000	.998
20	.000	.000	.000	.000	.998

δ = .3

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.474	3.594	39.806

J	$P_0(J)$	P(1,J)	P(2,J)	$P(N^+=j)$	$P(N^+\leq j)$
1	.500	.202	.118	.405	.405
2	.298	.118	.071	.234	.639
3	.180	.072	.043	.144	.783
4	.108	.043	.026	.087	.869
5	.064	.026	.016	.052	.922
6	.038	.016	.009	.031	.953
7	.022	.009	.005	.019	.971
8	.013	.005	.003	.011	.982
9	.008	.003	.002	.006	.989
10	.004	.002	.001	.004	.993
11	.002	.001	.001	.002	.995
12	.001	.001	.000	.001	.996
13	.001	.000	.000	.001	.997
14	.000	.000	.000	.000	.997
15	.000	.000	.000	.000	.997
16	.000	.000	.000	.000	.997
17	.000	.000	.000	.000	.997
18	.000	.000	.000	.000	.997
19	.000	.000	.000	.000	.998
20	.000	.000	.000	.000	.998

ρ =	.4	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS	
		2.680	4.295	50.581	

J	P0(J)	P(1,J)	P(2,J)	$P(N^+ = j)$	$P(N^+ \leq j)$
1	.500	.186	.112	.373	.373
2	.314	.112	.073	.223	.596
3	.201	.074	.047	.148	.744
4	.127	.048	.030	.095	.839
5	.080	.030	.019	.061	.899
6	.049	.019	.012	.038	.937
7	.030	.012	.007	.024	.961
8	.018	.007	.004	.015	.976
9	.011	.004	.003	.009	.985
10	.006	.003	.002	.005	.990
11	.004	.002	.001	.003	.993
12	.002	.001	.001	.002	.995
13	.001	.001	.000	.001	.996
14	.001	.000	.000	.001	.996
15	.000	.000	.000	.000	.997
16	.000	.000	.000	.000	.997
17	.000	.000	.000	.000	.997
18	.000	.000	.000	.000	.997
19	.000	.000	.000	.000	.997
20	.000	.000	.000	.000	.997

ρ =	.5	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS	
		2.916	5.045	62.971	

J	P0(J)	P(1,J)	P(2,J)	$P(N^+=j)$	$P(N^+ \leq j)$
1	.500	.170	.105	.341	.341
2	.330	.105	.075	.209	.550
3	.224	.076	.051	.152	.702
4	.148	.051	.035	.103	.804
5	.097	.035	.023	.069	.874
6	.062	.023	.015	.046	.920
7	.039	.015	.009	.030	.950
8	.024	.010	.006	.019	.969
9	.015	.006	.004	.012	.981
10	.009	.004	.002	.007	.988
11	.005	.002	.001	.004	.992
12	.003	.001	.001	.003	.995
13	.002	.001	.000	.002	.996
14	.001	.000	.000	.001	.997
15	.001	.000	.000	.000	.998
16	.000	.000	.000	.000	.998
17	.000	.000	.000	.000	.998
18	.000	.000	.000	.000	.998
19	.000	.000	.000	.000	.998
20	.000	.000	.000	.000	.998

F(U) = .600

ρ = .1	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	1.769	1.364	10.508

J	P0(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.400000	.225022	.139947	.087762	.055010	.034477	.021604	.013534	.008476	.005307	.003321	.996538	.560844
2	.174978	.098238	.061390	.038406	.024055	.015061	.009426	.005897	.003688	.002305	.001440	.435694	.244739
3	.076741	.043121	.025347	.016812	.010524	.006585	.004118	.002575	.001609	.001005	.000627	.190955	.107354
4	.033620	.018904	.011750	.007353	.004599	.002875	.001797	.001122	.000701	.000437	.000273	.083602	.047036
5	.014715	.008283	.005137	.003212	.002007	.001254	.000783	.000488	.000305	.000190	.000118	.036565	.020595
6	.006432	.003625	.002242	.001401	.000874	.000546	.000340	.000212	.000132	.000082	.000051	.015971	.009007
7	.002807	.001584	.000977	.000610	.000380	.000237	.000148	.000092	.000057	.000036	.000022	.006964	.003933
8	.001223	.000691	.000425	.000265	.000165	.000103	.000064	.000040	.000025	.000015	.000010	.003031	.001714
9	.000532	.000301	.000184	.000115	.000071	.000044	.000028	.000017	.000011	.000007	.000004	.001317	.000746
10	.000231	.000131	.000080	.000050	.000031	.000019	.000012	.000007	.000005	.000003	.000002	.000571	.000324

ρ = .2	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	1.888	1.673	13.664

J	P0(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.400000	.209947	.133700	.088243	.058023	.038117	.025006	.016378	.010707	.006985	.004548	.994608	.522975
2	.190053	.098917	.064140	.042061	.027561	.018017	.011750	.007644	.004962	.003213	.002076	.471633	.245975
3	.091135	.047734	.030525	.020044	.013085	.008520	.005534	.003586	.002318	.001495	.000963	.225658	.118421
4	.043401	.022834	.014575	.009498	.006173	.004003	.002589	.001671	.001076	.000691	.000443	.107237	.056549
5	.020567	.010886	.006888	.004468	.002891	.001867	.001202	.000773	.000496	.000317	.000203	.050688	.026894
6	.009481	.005158	.003230	.002086	.001344	.000864	.000554	.000355	.000227	.000145	.000092	.023794	.012708
7	.004522	.002425	.001502	.000966	.000620	.000397	.000254	.000162	.000103	.000066	.000042	.011096	.005960
8	.002096	.001132	.000593	.000444	.000284	.000181	.000115	.000073	.000047	.000030	.000019	.005126	.002773
9	.000964	.000524	.000318	.000203	.000129	.000082	.000052	.000033	.000021	.000013	.000008	.002353	.001280
10	.000441	.000241	.000144	.000092	.000058	.000037	.000023	.000015	.000009	.000006	.000004	.001172	.000587

F(U) = .600

ρ = .3

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
2.044	2.082	18.651

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+\geq j)$	$P(N^+=j)$
1	.400000	.194777	.125150	.087790	.060803	.041985	.028869	.019753	.013446	.009105	.006135	.999027	.489003
2	.205223	.097944	.065819	.045524	.031255	.021310	.014436	.009721	.006509	.004337	.002876	.510025	.244667
3	.107278	.052239	.034373	.023518	.015976	.010782	.007233	.004826	.003204	.002117	.001393	.265358	.129807
4	.055041	.027129	.017590	.011964	.008040	.005372	.003570	.002362	.001555	.001020	.000667	.135551	.067176
5	.027911	.013971	.008941	.005982	.003980	.002635	.001736	.001140	.000745	.000485	.000315	.068375	.034407
6	.013940	.007087	.004438	.002941	.001939	.001273	.000833	.000543	.000353	.000228	.000147	.033968	.017351
7	.006853	.003534	.002165	.001423	.000931	.000607	.000394	.000256	.000165	.000106	.000068	.016617	.008605
8	.003319	.001734	.001041	.000679	.000441	.000286	.000185	.000119	.000076	.000049	.000031	.008012	.004200
9	.001585	.000837	.000494	.000320	.000207	.000133	.000086	.000055	.000035	.000022	.000014	.003812	.002020
10	.000748	.000399	.000231	.000149	.000096	.000061	.000039	.000025	.000018	.000010	.000006	.001792	.000959

ρ = .4

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
2.196	2.487	23.381

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+\geq j)$	$P(N^+=j)$
1	.400000	.179512	.114222	.086331	.063273	.046047	.033197	.023691	.016737	.011711	.008121	.996563	.451690
2	.220488	.095220	.066637	.048796	.035136	.024939	.017484	.012126	.008330	.005676	.003839	.544872	.237485
3	.125269	.056616	.038092	.027236	.019199	.013371	.009215	.006294	.004265	.002870	.001919	.307387	.140016
4	.068653	.031776	.021095	.014752	.010200	.006984	.004741	.003195	.002139	.001424	.000943	.167372	.078178
5	.036877	.017561	.011295	.007752	.005273	.003558	.002385	.001589	.001053	.000694	.000456	.089193	.042827
6	.019316	.009444	.005966	.003965	.002660	.001774	.001176	.000776	.000510	.000333	.000217	.046366	.022828
7	.009872	.004937	.002968	.001981	.001314	.000867	.000570	.000373	.000243	.000158	.000102	.023539	.011840
8	.004935	.002515	.001469	.000970	.000637	.000417	.000272	.000177	.000114	.000074	.000048	.011699	.005991
9	.002420	.001253	.000713	.000467	.000304	.000198	.000128	.000083	.000053	.000034	.000022	.005708	.002968
10	.001166	.000612	.000341	.000221	.000143	.000093	.000060	.000038	.000025	.000016	.000010	.002740	.001443

F(U)= .600

ρ = .5	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	2.368	2.929	28.935

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.400000	.164150	.100775	.083770	.065350	.050269	.038001	.028226	.020625	.014848	.010549	.995081	.415520
2	.235850	.090640	.066494	.051876	.039205	.028906	.020894	.014859	.010425	.007230	.004965	.579561	.226121
3	.145210	.060840	.041781	.031198	.022753	.016286	.011481	.007991	.005502	.003754	.002540	.353440	.149876
4	.084370	.036764	.024792	.017862	.012652	.008837	.006102	.004171	.002828	.001902	.001272	.203564	.089909
5	.047607	.021683	.013951	.009779	.006770	.004638	.003148	.002121	.001419	.000944	.000624	.113655	.052360
6	.025923	.012267	.007515	.005158	.003507	.002365	.001584	.001054	.000698	.000460	.000301	.061295	.029267
7	.013657	.006665	.003908	.002639	.001769	.001178	.000780	.000514	.000337	.000220	.000143	.032028	.015742
8	.006992	.003498	.001976	.001317	.000872	.000575	.000377	.000247	.000161	.000104	.000067	.016286	.008193
9	.003494	.001783	.000976	.000643	.000422	.000276	.000180	.000117	.000075	.000049	.000031	.008092	.004149
10	.001711	.000887	.000473	.000309	.000201	.000131	.000084	.000054	.000035	.000022	.000014	.003943	.002053

F(U)= .700

ρ = .1

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	1.502	.795	5.129

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	P($N^+ \geq j$)	P($N^+ = j$)
1	.300000	.197744	.141008	.101458	.072983	.052496	.037756	.027152	.019524	.014037	.010091	.990458	.652641
2	.102256	.067297	.048273	.034698	.024936	.017918	.012873	.009247	.006641	.004769	.003424	.337818	.222713
3	.034959	.023040	.016403	.011785	.008465	.006079	.004365	.003134	.002249	.001614	.001158	.115105	.075965
4	.011919	.007866	.005569	.003998	.002870	.002060	.001478	.001060	.000760	.000545	.000391	.039140	.025863
5	.004053	.002680	.001886	.001353	.000971	.000696	.000499	.000358	.000256	.000184	.000132	.013277	.008788
6	.001374	.000910	.000637	.000457	.000327	.000234	.000168	.000120	.000086	.000062	.000044	.004490	.002977
7	.000464	.000308	.000214	.000154	.000110	.000079	.000056	.000040	.000029	.000021	.000015	.001512	.001005
8	.000156	.000104	.000072	.000051	.000037	.000026	.000019	.000013	.000010	.000007	.000005	.000508	.000338
9	.000052	.000035	.000024	.000017	.000012	.000009	.000006	.000004	.000003	.000002	.000002	.000170	.000113
10	.000017	.000012	.000008	.000006	.000004	.000003	.000002	.000001	.000001	.000001	.000001	.000056	.000038

ρ = .2

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	1.602	.996	6.847

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	P($N^+ \geq j$)	P($N^+ = j$)
1	.300000	.1[illegible]5156	.132601	.098767	.073418	.054536	.040479	.030020	.022244	.016467	.012179	.990020	.610170
2	.114844	.070421	.051846	.038396	.028402	.020986	.015489	.011420	.008411	.006188	.004548	.379850	.234634
3	.044423	.027510	.019683	.014532	.010716	.007893	.005807	.004267	.003133	.002298	.001683	.145216	.090347
4	.016913	.010553	.007430	.005464	.004014	.002945	.002159	.001581	.001156	.000845	.000617	.054869	.034376
5	.006360	.004003	.002773	.002031	.001486	.001086	.000793	.000579	.000422	.000307	.000224	.020493	.012943
6	.002357	.001496	.001021	.000745	.000543	.000395	.000288	.000209	.000152	.000111	.000080	.007550	.004807
7	.000861	.000551	.000371	.000269	.000196	.000142	.000103	.000075	.000054	.000039	.000028	.002743	.001760
8	.000310	.000200	.000133	.000096	.000070	.000051	.000037	.000026	.000019	.000014	.000010	.000984	.000635
9	.000110	.000071	.000047	.000034	.000025	.000018	.000013	.000009	.000007	.000005	.000003	.00[illegible]3[illegible]	.000226
10	.000034	.000025	.000016	.000012	.000009	.000006	.000004	.000003	.000002	.000002	.000001	.00[illegible]	.000080

F(U) = .700

ρ = .3

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
1.718	1.219	8.902

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.300000	.172235	.121604	.094700	.073237	.056488	.043449	.033325	.025488	.019441	.014788	.992055	.567888
2	.127765	.072222	.054818	.041964	.032007	.024330	.018437	.013930	.010497	.007890	.005916	.424167	.244085
3	.055542	.032347	.023070	.017503	.013235	.009977	.007501	.005624	.004207	.003140	.002339	.180082	.105844
4	.023195	.013764	.009553	.007176	.005376	.004018	.002996	.002229	.001655	.001227	.000908	.074237	.044389
5	.009432	.005706	.003849	.002866	.002129	.001579	.001169	.000864	.000637	.000470	.000345	.029849	.018167
6	.003725	.002293	.001538	.001114	.000822	.000606	.000446	.000327	.000240	.000176	.000129	.011682	.007223
7	.001432	.000894	.000576	.000423	.000310	.000227	.000166	.000122	.000089	.000065	.000047	.004459	.002794
8	.000538	.000340	.000215	.000157	.000115	.000084	.000061	.000044	.000032	.000023	.000017	.001665	.001055
9	.000196	.000126	.000079	.000057	.000042	.000030	.000022	.000016	.000012	.000008	.000006	.000610	.000390
10	.000072	.000046	.000028	.000021	.000015	.000011	.000008	.000006	.000004	.000003	.000002	.000220	.000142

ρ = .4

MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
1.844	1.467	11.284

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.300000	.158982	.107817	.089085	.072330	.058298	.046657	.037091	.029301	.023012	.017975	.991454	.522785
2	.141018	.072539	.057191	.045402	.035749	.027950	.021715	.016778	.012900	.009874	.007527	.468669	.249512
3	.068479	.037575	.026565	.020697	.016022	.012334	.009447	.007204	.005472	.004141	.003124	.219157	.122003
4	.030904	.017542	.011938	.009134	.006957	.005278	.003989	.003005	.002257	.001691	.001263	.097154	.055774
5	.013361	.007831	.005116	.003858	.002901	.002175	.001626	.001212	.000902	.000670	.000496	.041380	.024452
6	.005530	.003326	.002098	.001565	.001165	.000865	.000641	.000474	.000350	.000258	.000190	.016928	.010242
7	.002204	.001353	.000829	.000613	.000453	.000334	.000246	.000181	.000133	.000097	.000071	.006685	.004122
8	.000851	.000531	.000318	.000234	.000172	.000126	.000092	.000067	.000049	.000036	.000026	.002564	.001604
9	.000320	.000202	.000119	.000087	.000064	.000046	.000034	.000025	.000018	.000013	.000009	.000959	.000608
10	.000118	.000075	.000044	.000032	.000023	.000017	.000012	.000009	.000006	.000005	.000003	.000352	.000225

F(U)= .700

ρ = .5

	MEAN	VARIANCE	COEFFICIENT OF SKEWNESS
	1.992	1.714	13.907

J	PU(J)	P(1,J)	P(2,J)	P(3,J)	P(4,J)	P(5,J)	P(6,J)	P(7,J)	P(8,J)	P(9,J)	P(10,J)	$P(N^+ \geq j)$	$P(N^+ = j)$
1	.300000	.145397	.091008	.081699	.070567	.059912	.050104	.041353	.033738	.027247	.021807	.996170	.480248
2	.154603	.071196	.058963	.048710	.039631	.031845	.025326	.019964	.015620	.012140	.009381	.515922	.252247
3	.083407	.043210	.030167	.024115	.019078	.014962	.011646	.009008	.006928	.005302	.004039	.263675	.139449
4	.040197	.021941	.014584	.011338	.008757	.006725	.005138	.003909	.002961	.002236	.001682	.124225	.068837
5	.018256	.010427	.006572	.005009	.003800	.002873	.002164	.001625	.001217	.000909	.000677	.055389	.031960
6	.007829	.004627	.002792	.002098	.001571	.001174	.000875	.000650	.000482	.000357	.000264	.023428	.013944
7	.003202	.001941	.001132	.000842	.000624	.000462	.000342	.000252	.000186	.000137	.000100	.009485	.005779
8	.001261	.000779	.000442	.000326	.000240	.000177	.000130	.000095	.000070	.000051	.000037	.003706	.002299
9	.000482	.000302	.000168	.000123	.000090	.000066	.000048	.000035	.000026	.000019	.000014	.001407	.000885
10	.000180	.000114	.000062	.000045	.000033	.000024	.000018	.000013	.000009	.000007	.000005	.000522	.000332

APPENDIX B

Data of Annual Precipitation and Annual River Flow Stations

TABLE B.1 ANNUAL PRECIPITATION DATA OF STA. N^{o} 4.1715
CHICO EXPERIMENTAL STATION, CALIFORNIA

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1871	22.58	.92	-.26
1872	26.48	1.08	.29
1873	19.38	.79	-.71
1874	24.34	.99	-.01
1875	15.41	.63	-1.27
1876	21.86	.89	-.36
1877	17.54	.71	-.97
1878	31.36	1.28	.98
1879	25.05	1.02	.08
1880	17.38	.71	-.99
1881	14.56	.59	-1.39
1882	17.69	.72	-.95
1883	17.00	.69	-1.05
1884	23.19	.94	-.17
1885	20.41	.83	-.56
1886	15.91	.65	-1.20
1887	15.44	.63	-1.27
1888	19.92	.81	-.63
1889	29.82	1.22	.76
1890	21.78	.89	-.37
1891	19.79	.81	-.65
1892	36.24	1.48	1.67
1893	25.49	1.04	.15
1894	30.61	1.25	.87
1895	27.35	1.11	.41
1896	33.78	1.38	1.32
1897	20.84	.85	-.50
1898	12.31	.50	-1.71
1899	27.30	1.11	.40
1900	20.14	.82	-.60
1901	20.27	.83	-.58
1902	28.04	1.14	.51
1903	22.76	.93	-.23
1904	30.39	1.24	.84
1905	24.11	.98	-.04
1906	37.27	1.52	1.81
1907	24.15	.98	-.03
1908	17.92	.73	-.92
1909	36.57	1.49	1.72
1910	14.06	.57	-1.46
1911	23.63	.96	-.11
1912	21.95	.89	-.34
1913	28.10	1.15	.52
1914	28.37	1.16	.55
1915	34.49	1.41	1.42
1916	32.09	1.31	1.08
1917	17.61	.72	-.96
1918	21.43	.87	-.42
1919	21.75	.89	-.37
1920	31.74	1.29	1.03

TABLE B.1 (continued)

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1921	23.57	.96	-.12
1922	27.41	1.12	.42
1923	15.08	.61	-1.32
1924	22.36	.91	-.29
1925	22.51	.92	-.27
1926	33.16	1.35	1.23
1927	30.01	1.22	.79
1928	22.03	.90	-.33
1929	17.01	.69	-1.04
1930	15.94	.65	-1.20
1931	21.81	.89	-.36
1932	13.70	.56	-1.51
1933	20.03	.82	-.62
1934	19.82	.81	-.65
1935	29.01	1.18	.64
1936	24.54	1.00	.01
1937	37.43	1.53	1.84
1938	32.13	1.31	1.09
1939	13.34	.54	-1.56
1940	44.02	1.80	2.77
1941	45.54	1.86	2.99
1942	32.94	1.34	1.20
1943	24.84	1.01	.05
1944	28.90	1.18	.63
1945	32.15	1.31	1.09
1946	13.35	.54	-1.56
1947	20.87	.85	-.50
1948	25.77	1.05	.19
1949	14.39	.58	-1.42
1950	29.62	1.21	.73
1951	27.17	1.11	.38
1952	33.64	1.37	1.30
1953	19.95	.81	-.63
1954	29.40	1.20	.70
1955	24.48	1.00	0.00
1956	19.36	.79	-.71
1957	25.36	1.03	.13
1958	34.33	1.40	1.40
1959	18.71	.76	-.80
1960	24.60	1.00	.02
1961	23.77	.97	-.10
1962	28.89	1.18	.64
1963	29.09	1.19	.67
1964	21.28	.87	-.46
1965	22.24	.91	-.32

TABLE B.2 ANNUAL PRECIPITATION DATA OF STA. N^{o}25.6335
ORD, NEBRASKA

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1896	25.54	1.09	.39
1897	24.31	1.04	.17
1898	15.95	.68	-1.33
1899	13.70	.58	-1.74
1900	27.88	1.19	.82
1901	27.07	1.16	.67
1902	31.02	1.32	1.39
1903	25.57	1.09	.40
1904	21.64	.92	-.30
1905	35.87	1.53	2.27
1906	34.98	1.49	2.11
1907	19.88	.85	-.62
1908	28.39	1.21	.91
1909	19.21	.82	-.74
1910	21.85	.93	-.26
1911	23.29	.99	0.00
1912	18.45	.79	-.88
1913	27.88	1.19	.82
1914	19.44	.83	-.70
1915	33.78	1.44	1.89
1916	19.64	.84	-.66
1917	22.80	.97	-.09
1918	22.49	.96	-.15
1919	21.63	.92	-.30
1920	36.21	1.55	2.33
1921	25.21	1.08	.34
1922	24.06	1.03	.13
1923	33.28	1.42	1.80
1924	25.27	1.08	.35
1925	21.06	.90	-.41
1926	24.29	1.04	.17
1927	22.17	.95	-.21
1928	19.63	.84	-.67
1929	25.89	1.10	.46
1930	31.18	1.33	1.42

TABLE B.2 (continued)

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1931	16.37	.70	-1.26
1932	20.89	.89	-.44
1933	21.37	.91	-.35
1934	10.98	.47	-2.23
1935	27.32	1.17	.72
1936	15.17	.65	-1.47
1937	17.27	.74	-1.09
1938	19.47	.83	-.69
1939	15.07	.64	-1.49
1940	19.08	.81	-.77
1941	23.19	.99	-.02
1942	24.75	1.06	.25
1943	16.47	.70	-1.24
1944	26.80	1.14	.62
1945	22.17	.95	-.21
1946	26.35	1.12	.54
1947	25.01	1.07	.30
1948	25.54	1.09	.39
1949	24.15	1.03	.14
1950	26.31	1.12	.53
1951	28.28	1.21	.89
1952	18.37	.78	-.89
1953	23.48	1.00	.02
1954	16.82	.72	-1.18
1955	16.14	.69	-1.30
1956	16.90	.72	-1.16
1957	32.29	1.38	1.62
1958	23.26	.99	-.01
1959	21.03	.90	-.41
1960	25.81	1.10	.44
1961	22.13	.94	-.24
1962	30.76	1.31	1.33
1963	20.05	.86	-.62
1964	21.29	.91	-.40
1965	31.39	1.34	1.44

TABLE B.3 ANNUAL PRECIPITATION DATA OF STA. N° 2.5825
NATURAL BRIDGE, N.M., ARIZONA

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1890	30.45	1.27	.86
1891	13.69	.57	-1.36
1892	17.85	.74	-.80
1893	20.71	.86	-.42
1894	15.86	.66	-1.07
1895	21.60	.90	-.31
1896	17.93	.74	-.79
1897	22.33	.93	-.21
1898	19.56	.81	-.58
1899	17.63	.73	-.83
1900	12.28	.51	-1.54
1901	14.54	.60	-1.24
1902	16.12	.67	-1.03
1903	14.56	.60	-1.24
1904	15.13	.63	-1.16
1905	50.17	2.09	3.47
1906	28.26	1.18	.57
1907	29.02	1.21	.67
1908	33.14	1.38	1.21
1909	24.88	1.03	.12
1910	15.44	.64	-1.12
1911	31.62	1.32	1.01
1912	23.32	.97	-.08
1913	23.53	.98	-.05
1914	27.42	1.14	.46
1915	31.59	1.31	1.01
1916	33.94	1.41	1.32
1917	27.04	1.12	.41
1918	26.14	1.09	.29
1919	33.12	1.38	1.21
1920	25.29	1.05	.17
1921	22.54	.94	-.18
1922	26.89	1.12	.39
1923	30.09	1.25	.81
1924	16.44	.68	-.99
1925	19.48	.81	-.59

TABLE B.3 (continued)

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1926	30.50	1.27	.86
1927	37.62	1.57	1.81
1928	18.50	.77	-.72
1929	22.38	.93	-.20
1930	34.33	1.43	1.37
1931	36.40	1.51	1.65
1932	23.48	.98	-.06
1933	18.25	.76	-.75
1934	16.01	.66	-1.05
1935	27.82	1.16	.51
1936	26.02	1.08	.27
1937	20.78	.86	-.42
1938	23.28	.97	-.08
1939	20.33	.84	-.47
1940	31.75	1.32	1.03
1941	39.41	1.64	2.05
1942	16.71	.69	-.96
1943	20.86	.87	-.40
1944	26.17	1.09	.29
1945	25.50	1.06	.20
1946	24.27	1.01	.04
1947	18.54	.77	-.71
1948	16.55	.69	-.98
1949	27.88	1.16	.52
1950	13.69	.57	-1.36
1951	31.48	1.31	.99
1952	29.14	1.21	.68
1953	10.99	.45	-1.71
1954	23.85	.99	-.01
1955	26.88	1.12	.38
1956	10.41	.43	-1.79
1957	34.09	1.42	1.34
1958	21.83	.91	-.28
1959	26.19	1.09	.29
1960	18.86	.78	-.67

TABLE B.4 ANNUAL PRECIPITATION DATA OF STA. N° 4.0227
ANTIOCH F. MILLS, CALIFORNIA

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1879	10.08	.78	-.62
1880	15.64	1.22	.67
1881	8.53	.66	-.99
1882	9.14	.71	-.85
1883	9.22	.72	-.83
1884	20.68	1.61	1.85
1885	9.65	.75	-.73
1886	7.61	.59	-1.20
1887	8.69	.68	-.95
1888	12.03	.94	-.17
1889	20.95	1.64	1.91
1890	13.72	1.07	.22
1891	15.15	1.18	.55
1892	16.02	1.25	.76
1893	12.77	1.00	0.00
1894	20.12	1.57	1.72
1895	10.94	.85	-.42
1896	16.76	1.31	.93
1897	10.60	.83	-.50
1898	4.92	.38	-1.83
1899	13.56	1.06	.18
1900	9.73	.76	-.71
1901	15.17	1.18	.56
1902	10.45	.81	-.54
1903	11.60	.90	-.27
1904	16.48	1.29	.87
1905	10.57	.82	-.51
1906	18.26	1.43	1.28
1907	16.28	1.27	.82
1908	10.11	.79	-.62
1909	17.85	1.39	1.19
1910	6.47	.50	-1.47
1911	16.13	1.26	.78
1912	7.35	.57	-1.26
1913	11.37	.89	-.32
1914	16.33	1.27	.83
1915	15.34	1.20	.60
1916	15.98	1.25	.75
1917	5.46	.42	-1.71
1918	16.46	1.28	.86
1919	10.99	.86	-.41
1920	10.45	.81	-.54
1921	12.44	.97	-.07
1922	16.32	1.27	.83
1923	6.48	.50	-1.47
1924	9.57	.74	-.74
1925	12.02	.94	-.17

TABLE B.4 (continued)

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1926	14.92	1.16	.50
1927	13.45	1.05	.15
1928	11.52	.90	-.29
1929	4.98	.39	-1.82
1930	10.02	.78	-.64
1931	13.04	1.02	.06
1932	7.96	.62	-1.12
1933	10.08	.78	-.62
1934	9.61	.75	-.74
1935	11.13	.87	-.38
1936	14.25	1.11	.34
1937	14.55	1.13	.41
1938	14.73	1.15	.45
1939	7.16	.56	-1.31
1940	23.19	1.81	2.44
1941	21.25	1.66	1.98
1942	15.05	1.17	.53
1943	11.99	.93	-.18
1944	13.79	1.08	.23
1945	13.67	1.07	.21
1946	6.81	.53	-1.39
1947	7.60	.59	-1.21
1948	12.23	.95	-.12
1949	9.24	.72	-.82
1950	17.21	1.34	1.04
1951	15.46	1.21	.63
1952	24.00	1.87	2.63
1953	6.53	.51	-1.46
1954	13.39	1.04	.14
1955	15.45	1.21	.62
1956	10.14	.79	-.61
1957	14.28	1.11	.35
1958	21.33	1.67	2.00
1959	10.77	.84	-.46
1960	13.73	1.07	.22
1961	9.66	.76	-.74
1962	14.69	1.15	.47
1963	14.86	1.17	.51
1964	11.36	.89	-.33
1965	11.68	.92	-.26

TABLE B.5 ANNUAL PRECIPITATION DATA OF STA. N° 25.7040
RAVENNA, NEBRASKA

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1878	19.64	.82	-.76
1879	25.68	1.08	.37
1880	19.71	.83	-.75
1881	37.72	1.59	2.66
1882	24.08	1.01	.07
1883	24.40	1.02	.13
1884	23.58	.99	-.02
1885	25.56	1.07	.35
1886	25.93	1.09	.42
1887	27.34	1.15	.69
1888	25.66	1.08	.37
1889	24.12	1.01	.08
1890	19.15	.80	-.86
1891	35.95	1.51	2.32
1892	25.60	1.08	.36
1893	18.13	.76	-1.05
1894	15.66	.66	-1.52
1895	20.26	.85	-.65
1896	27.50	1.16	.72
1897	32.75	1.38	1.71
1898	18.50	.78	-.98
1899	24.09	1.01	.07
1900	21.74	.91	-.37
1901	27.15	1.14	.65
1902	37.06	1.56	2.53
1903	36.89	1.55	2.50
1904	27.45	1.15	.71
1905	33.28	1.40	1.81
1906	29.70	1.25	1.13
1907	17.57	.74	-1.16
1908	28.42	1.19	.89
1909	21.90	.92	-.33
1910	19.47	.82	-.80
1911	22.33	.94	-.25
1912	16.42	.69	-1.37
1913	25.87	1.09	.41
1914	21.46	.90	-.42
1915	31.07	1.31	1.39
1916	18.23	.76	-1.03
1917	21.71	.91	-.37
1918	25.29	1.06	.30
1919	29.93	1.26	1.18
1920	27.28	1.15	.68

TABLE B.5 (continued)

YEAR	ANNUAL PRECIP.	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1921	23.51	.99	-.03
1922	19.21	.81	-.85
1923	31.47	1.32	1.47
1924	23.50	.99	-.03
1925	18.77	.79	-.93
1926	23.75	1.00	.01
1927	20.28	.85	-.64
1928	23.77	1.00	.01
1929	17.60	.74	-1.15
1930	28.74	1.21	.95
1931	20.24	.85	-.65
1932	27.19	1.14	.66
1933	21.37	.90	-.44
1934	15.50	.65	-1.55
1935	24.56	1.03	.16
1936	14.11	.59	-1.81
1937	25.69	1.08	.37
1938	21.47	.90	-.42
1939	16.55	.69	-1.35
1940	12.33	.52	-2.15
1941	22.21	.93	-.28
1942	27.22	1.14	.66
1943	20.04	.84	-.69
1944	23.40	.98	-.05
1945	21.40	.90	-.43
1946	24.75	1.04	.20
1947	21.70	.91	-.37
1948	21.96	.92	-.32
1949	25.14	1.06	.27
1950	22.85	.96	-.15
1951	25.03	1.05	.25
1952	18.70	.78	-.94
1953	23.24	.98	-.08
1954	18.38	.77	-1.00
1955	21.63	.91	-.39
1956	12.86	.54	-2.05
1957	27.37	1.15	.69
1958	24.68	1.04	.18
1959	24.69	1.04	.18
1960	28.32	1.19	.87
1961	27.20	1.15	.67
1962	25.54	1.08	.35
1963	18.63	.79	-.97
1964	18.49	.78	-1.00
1965	30.03	1.27	1.21

TABLE B.6 ANNUAL RIVER FLOW DATA OF MISSISSIPPI AT ST. LOUIS, MISSOURI

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1862.	1.1500	0.5017	1910.	0.9500	-0.1672
1863.	0.7080	-0.9766	1911.	0.5450	-1.5217
1864.	0.5020	-1.6656	1912.	1.1860	0.6221
1865.	0.9270	-0.2441	1913.	0.8330	-0.5585
1866.	1.0740	0.2475	1914.	0.6920	-1.0301
1867.	1.3410	1.1405	1915.	1.4620	1.5452
1868.	0.7120	-0.9632	1916.	1.4400	1.4716
1869.	1.1800	0.6020	1917.	0.9600	-0.1338
1870.	0.9450	-0.1839	1918.	0.7140	-0.9565
1871.	0.7870	-0.7124	1919.	0.9820	-0.0602
1872.	0.7810	-0.7324	1920.	1.2640	0.8829
1873.	1.0200	0.0669	1921.	0.8520	-0.4950
1874.	0.7240	-0.9231	1922.	1.1460	0.4883
1875.	1.0670	0.2241	1923.	0.7580	-0.8094
1876.	1.4270	1.4281	1924.	1.0670	0.2241
1877.	1.1590	0.5318	1925.	0.8580	-0.4749
1878.	1.1870	0.6254	1926.	0.9620	-0.1271
1879.	0.7840	0.7224	1927.	1.8530	2.8528
1880.	0.9690	0.1037	1928.	1.1660	0.5552
1881.	1.3250	1.0870	1929.	1.5100	1.7057
1882.	1.7420	2.4816	1930.	0.7150	-0.9532
1883.	1.3480	1.1639	1931.	0.4500	-1.8395
1884.	1.1480	0.4950	1932.	0.9120	-0.2943
1885.	1.3650	1.2207	1933.	0.7800	-0.7358
1886.	1.0700	0.2341	1934.	0.3860	-2.0535
1887.	0.8290	-0.5719	1935.	1.0710	0.2375
1888.	1.2490	0.8328	1936.	0.6490	-1.1739
1889.	0.6890	-1.0401	1937.	0.8380	-0.5418
1890.	0.7730	-0.7592	1938.	0.8950	-0.3512
1891.	0.8990	-0.3378	1939.	0.8450	-0.5184
1892.	1.5050	1.6890	1940.	0.4520	-1.8328
1893.	1.0580	0.1940	1941.	0.6850	-1.0535
1894.	0.6380	-1.2107	1942.	1.3490	1.1672
1895.	0.4650	-1.7893	1943.	1.3400	1.1371
1896.	0.8300	-0.5686	1944.	1.1940	0.6488
1897.	1.1320	0.4415	1945.	1.2740	0.9164
1898.	0.8980	-0.3411	1946.	0.9890	-0.0368
1899.	0.9760	-0.0803	1947.	1.3560	1.1906
1900.	0.7260	-0.9164	1948.	0.9300	-0.2341
1901.	0.7600	-0.8027	1949.	0.9700	-0.1003
1902.	0.8710	-0.4314	1950.	1.1360	0.4548
1903.	1.4910	1.6421	1951.	1.5120	1.7124
1904.	1.2800	0.9365	1952.	1.3320	1.1104
1905.	1.0380	0.1271	1953.	0.8060	-0.6488
1906.	1.1090	0.3645	1954.	0.6480	-1.1773
1907.	1.1850	0.6187	1955.	0.7440	-0.8562
1908.	1.2900	0.9699	1956.	0.5370	-1.5485
1909.	1.1650	0.5518	1957.	0.7030	-0.9933

TABLE B.7 ANNUAL RIVER FLOW DATA OF ST. LAWRENCE (MAIN STEM) AT OGDENSBURG, NEW YORK

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1861.	1.1420	1.6322	1910.	0.9590	-0.4713
1862.	1.1790	2.0575	1911.	0.9090	-1.0460
1863.	1.1340	1.5402	1912.	0.9510	-0.5632
1864.	1.1040	1.1954	1913.	1.0710	0.8161
1865.	1.1040	1.1954	1914.	1.0010	0.0115
1866.	1.0210	0.2414	1915.	0.9180	-0.9425
1867.	1.1290	1.4828	1916.	0.9970	-0.0345
1868.	1.0050	0.0575	1917.	1.0010	0.0115
1869.	1.0340	0.3908	1918.	1.0260	0.2989
1870.	1.1580	1.8161	1919.	1.0420	0.4828
1871.	1.0750	0.8621	1920.	0.9380	-0.7126
1872.	0.9300	-0.8046	1921.	0.9720	-0.3218
1873.	0.9840	-0.1839	1922.	0.9510	-0.5632
1874.	1.0840	0.9655	1923.	0.9010	-1.1379
1875.	0.9590	-0.4713	1924.	0.9180	-0.9425
1876.	1.0920	1.0575	1925.	0.8890	-1.2759
1877.	1.0460	0.5287	1926.	0.8640	-1.5632
1878.	1.0460	0.5287	1927.	0.9380	-0.7126
1879.	1.0750	0.8621	1928.	0.9880	-0.1379
1880.	1.0130	0.1494	1929.	1.0590	0.6782
1881.	0.9630	-0.4253	1930.	1.0920	1.0575
1882.	1.0630	0.7241	1931.	0.9050	-1.0920
1883.	1.0420	0.4828	1932.	0.9050	-1.0920
1884.	1.1210	1.3908	1933.	0.8640	-1.5632
1885.	1.0500	0.5747	1934.	0.7810	-2.5172
1886.	1.1540	1.7701	1935.	0.7600	-2.7586
1887.	1.1170	1.3448	1936.	0.7970	-2.3333
1888.	1.0010	0.0115	1937.	0.8890	-1.2759
1889.	1.0210	0.2414	1938.	0.9050	-1.0920
1890.	1.0880	1.0115	1939.	0.8970	-1.1839
1891.	1.0750	0.8621	1940.	0.8720	-1.4713
1892.	0.9510	-0.5632	1941.	0.8840	-1.3333
1893.	1.0050	0.0575	1942.	0.8890	-1.2759
1894.	1.0050	0.0575	1943.	1.0300	0.3448
1895.	0.8970	-1.1839	1944.	1.0260	0.2989
1896.	0.8840	-1.3333	1945.	1.0050	0.0575
1897.	0.9050	-1.0920	1946.	1.0670	0.7701
1898.	0.9470	-0.6092	1947.	1.0670	0.7701
1899.	0.9470	-0.6092	1948.	1.0710	0.8161
1900.	0.9380	-0.7126	1949.	1.0010	0.0115
1901.	0.9420	-0.6667	1950.	0.9800	-0.2299
1902.	0.9470	-0.6092	1951.	1.1210	1.3908
1903.	1.0050	0.0575	1952.	1.1420	1.6322
1904.	1.0130	0.1494	1953.	1.0750	0.8621
1905.	0.9880	-0.1379	1954.	1.0550	0.6322
1906.	1.0010	0.0115	1955.	1.0630	0.7241
1907.	1.0130	0.1494	1956.	1.0710	0.8161
1908.	1.0880	1.0115	1957.	1.0010	0.0115
1909.	1.0000	0.0000			

TABLE B.8 ANNUAL RIVER FLOW DATA OF MISSISSIPPI
KEOKUK, IOWA

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1879.	0.8100	-0.6441	1919.	1.0750	0.2542
1880.	1.2370	0.8034	1920.	1.1680	0.5695
1881.	1.3000	1.0169	1921.	0.7570	-0.8237
1882.	2.0760	3.6475	1922.	0.9760	-0.0814
1883.	1.3530	1.1966	1923.	0.6370	-1.2305
1884.	1.1390	0.4712	1924.	0.8340	-0.5627
1885.	1.4670	1.5831	1925.	0.6120	-1.3153
1886.	1.1280	0.4339	1926.	0.7420	-0.8746
1887.	0.8790	-0.4102	1927.	1.2310	0.7831
1888.	1.3650	1.2373	1928.	1.0400	0.1356
1889.	0.6970	-1.0271	1929.	1.3200	1.0847
1890.	0.8250	-0.5932	1930.	0.6680	-1.1254
1891.	0.8200	-0.6102	1931.	0.4040	-2.0203
1892.	1.2820	0.9559	1932.	0.8410	-0.5390
1893.	1.0390	0.1322	1933.	0.6940	-1.0373
1894.	0.7940	-0.6983	1934.	0.3520	-2.1966
1895.	0.5100	-1.6610	1935.	0.9360	-0.2169
1896.	0.7280	-0.9220	1936.	0.7790	-0.7492
1897.	1.1480	0.5017	1937.	0.8160	-0.6237
1898.	0.6960	-1.0305	1938.	1.0700	0.2373
1899.	0.9080	-0.3119	1939.	0.9590	-0.1390
1900.	0.7770	-0.7559	1940.	0.5690	-1.4610
1901.	0.9930	-0.0237	1941.	0.8930	-0.3627
1902.	0.8990	-0.3424	1942.	1.2380	0.8068
1903.	1.4200	1.4237	1943.	1.3600	1.2203
1904.	1.2060	0.6983	1944.	1.2270	0.7695
1905.	1.3250	1.1017	1945.	1.1570	0.5322
1906.	1.4340	1.4712	1946.	1.1280	0.4339
1907.	1.3200	1.0847	1947.	1.2560	0.8678
1908.	1.2570	0.8712	1948.	0.8470	-0.5186
1909.	1.1060	0.3593	1949.	0.7220	-0.9424
1910.	0.8120	-0.6373	1950.	1.0130	0.0441
1911.	0.5810	-1.4203	1951.	1.3590	1.2169
1912.	1.1810	0.6136	1952.	1.4670	1.5831
1913.	0.8850	-0.3898	1953.	1.0530	0.1797
1914.	0.8090	-0.6475	1954.	0.8910	-0.3695
1915.	1.1370	0.4644	1955.	0.8920	-0.3661
1916.	1.4420	1.4983	1956.	0.6840	-1.0712
1917.	1.0130	0.0441	1957.	0.6890	-1.0542
1918.	0.8390	-0.5458			

TABLE B.9 ANNUAL RIVER FLOW DATA OF GOTA
SJOTORP VANERSBORG

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1808.	0.8430	-0.9874	1846.	1.2060	1.2956
1809.	1.0070	0.0440	1847.	1.1510	0.9497
1810.	0.8060	-1.2201	1848.	0.8500	-0.9434
1811.	0.9930	-0.0440	1849.	0.9330	-0.4214
1812.	1.0900	0.5660	1850.	1.0520	0.3270
1813.	1.0730	0.4591	1851.	1.1440	0.9057
1814.	0.9540	-0.2893	1852.	1.0560	0.3522
1815.	1.1040	0.6541	1853.	1.2080	1.3082
1816.	1.2590	1.6289	1854.	0.8300	-1.0692
1817.	1.4290	2.6981	1855.	1.1480	0.9308
1818.	0.9410	-0.3711	1856.	0.9300	-0.4403
1819.	0.7990	-1.2642	1857.	0.7640	-1.4843
1820.	0.9370	-0.3962	1858.	0.6840	-1.9874
1821.	1.0650	0.4088	1859.	0.8500	-0.9434
1822.	0.9220	-0.4906	1860.	1.1970	1.2390
1823.	1.0120	0.0755	1861.	1.0170	0.1069
1824.	1.1150	0.7233	1862.	0.8300	-1.0692
1825.	1.0840	0.5283	1863.	0.8810	-0.7484
1826.	0.8730	-0.7987	1864.	0.9370	-0.3962
1827.	1.0480	0.3019	1865.	0.7430	-1.6164
1828.	1.0170	0.1069	1866.	0.9360	-0.4025
1829.	0.8660	-0.8428	1867.	1.2290	1.4403
1830.	1.0180	0.1132	1868.	0.9810	-0.1195
1831.	1.2390	1.5031	1869.	0.9850	-0.0943
1832.	0.7890	-1.3270	1870.	0.7750	-1.4151
1833.	0.8950	-0.6604	1871.	0.9790	-0.1321
1834.	0.9370	-0.3962	1872.	0.8920	-0.6792
1835.	0.8230	-1.1132	1873.	0.9960	-0.0252
1836.	0.8680	-0.8302	1874.	0.8150	-1.1635
1837.	1.1340	0.8428	1875.	0.8250	-1.1006
1838.	0.9620	-0.2390	1876.	1.3130	1.9686
1839.	1.0280	0.1761	1877.	1.1350	0.8491
1840.	0.8730	-0.7987	1878.	1.1110	0.6981
1841.	1.0700	0.4403	1879.	1.1100	0.6918
1842.	0.8920	-0.6792	1880.	0.8960	-0.6541
1843.	1.1500	0.9434	1881.	1.1410	0.8868
1844.	1.0910	0.5723	1882.	0.8490	-0.9497
1845.	1.0170	0.1069			

TABLE B.9 (continued)

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1883.	1.1900	1.0440	1921.	0.9940	-0.0330
1884.	1.1900	1.0440	1922.	0.7010	-1.6429
1885.	1.0090	0.0495	1923.	0.6920	-1.6923
1886.	1.0990	0.5440	1924.	1.0860	0.4725
1887.	0.8430	-0.8626	1925.	1.3060	1.6813
1888.	0.8080	-1.0549	1926.	0.8950	-0.5769
1889.	0.8990	-0.5549	1927.	1.1490	0.8187
1890.	0.9100	-0.4945	1928.	1.1970	1.0824
1891.	1.0090	0.0495	1929.	1.1680	0.9231
1892.	1.0260	0.1429	1930.	1.2180	1.1978
1893.	0.8790	-0.6648	1931.	1.2090	1.1484
1894.	1.0200	0.1099	1932.	0.9740	-0.1429
1895.	1.0880	0.4835	1933.	0.8340	-0.9121
1896.	1.1510	0.8297	1934.	0.6380	-1.9890
1897.	1.0350	0.1923	1935.	0.9910	-0.0495
1898.	1.1580	0.8681	1936.	1.1980	1.0879
1899.	1.2670	1.4670	1937.	1.0910	0.5000
1900.	1.0130	0.0714	1938.	0.8920	-0.5934
1901.	0.9350	-0.3571	1939.	1.0200	0.1099
1902.	0.6620	-1.8571	1940.	0.8690	-0.7198
1903.	0.9500	-0.2747	1941.	0.7720	-1.2527
1904.	1.1210	0.6648	1942.	0.6060	-2.1648
1905.	0.8800	-0.6593	1943.	0.7390	-1.4341
1906.	0.8020	-1.0879	1944.	0.8130	-1.0275
1907.	0.8560	-0.7912	1945.	1.1730	0.9505
1908.	1.0800	0.4396	1946.	0.9160	-0.4615
1909.	0.9590	-0.2253	1947.	0.8800	-0.6593
1910.	1.3450	1.8956	1948.	0.6010	-2.1923
1911.	1.1530	0.8407	1949.	0.7200	-1.5385
1912.	0.9290	-0.3901	1950.	0.9550	-0.2473
1913.	1.1580	0.8681	1951.	1.1860	1.0220
1914.	0.9570	-0.2363	1952.	1.1400	0.7692
1915.	0.7050	-1.6209	1953.	0.9920	-0.0440
1916.	0.9050	-0.5220	1954.	1.0480	0.2637
1917.	1.0000	0.0000	1955.	1.1230	0.6758
1918.	0.9480	-0.2857	1956.	0.7740	-1.2418
1919.	0.9070	-0.5110	1957.	0.7690	-1.2692
1920.	0.9910	-0.0495			

TABLE B.10 ANNUAL RIVER FLOW DATA OF RHINE BASEC, SWITZERLAND

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1808.	0.7670	-1.2802	1846.	1.1860	1.0220
1809.	0.7240	-1.5165	1847.	1.0630	0.3462
1810.	0.7010	-1.6429	1848.	0.9830	-0.0934
1811.	0.7800	-1.2088	1849.	1.1290	0.7088
1812.	1.0070	0.0385	1850.	0.9890	-0.0604
1813.	1.0130	0.0714	1851.	1.1400	0.7692
1814.	0.7930	-1.1374	1852.	1.2760	1.5165
1815.	0.9910	-0.0495	1853.	1.2760	1.5165
1816.	1.0440	0.2418	1854.	0.9550	-0.2473
1817.	1.3100	1.7033	1855.	0.6830	-1.7418
1818.	1.2760	1.5165	1856.	0.7440	-1.4066
1819.	0.9660	-0.1868	1857.	0.8250	-0.9615
1820.	0.8280	-0.9451	1858.	0.7140	-1.5714
1821.	0.9530	-0.2582	1859.	0.6600	-1.8681
1822.	0.9330	-0.3681	1860.	0.9050	-0.5220
1823.	1.0200	0.1099	1861.	1.3790	2.0824
1824.	1.1860	1.0220	1862.	1.1010	0.5549
1825.	1.1450	0.7967	1863.	1.1530	0.8407
1826.	1.1190	0.6538	1864.	1.0730	0.4011
1827.	0.9290	-0.3901	1865.	0.8800	-0.6593
1828.	0.9630	-0.2033	1866.	0.9420	-0.3187
1829.	1.0480	0.2637	1867.	1.3150	1.7308
1830.	1.1210	0.6648	1868.	1.2980	1.6374
1831.	1.3430	1.8846	1869.	1.1140	0.6264
1832.	1.1600	0.8791	1870.	1.0190	0.1044
1833.	1.0090	0.0495	1871.	0.9050	-0.5220
1834.	1.2160	1.1868	1872.	0.8520	-0.8132
1835.	1.0110	0.0604	1873.	1.3260	1.7912
1836.	0.9440	-0.3077	1874.	1.3430	1.8846
1837.	1.1040	0.5714	1875.	1.0470	0.2582
1838.	1.1040	0.5714	1876.	0.7760	-1.2308
1839.	1.2160	1.1868	1877.	0.9100	-0.4945
1840.	1.1380	0.7582	1878.	1.1140	0.6264
1841.	1.1790	0.9835	1879.	0.9360	-0.3516
1842.	1.3230	1.7747	1880.	0.9610	-0.2143
1843.	0.8800	-0.6593	1881.	0.7260	-1.5055
1844.	0.7460	-1.3956	1882.	1.1060	0.5824
1845.	0.9220	-0.4286			

TABLE B.10 (continued)

YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE	YEAR	MODULAR COEFFICIENT	STANDARDIZED VARIABLE
1883.	1.2250	1.4151	1921.	0.5950	-2.5472
1884.	0.8810	-0.7484	1922.	1.1230	0.7736
1885.	0.6620	-2.1258	1923.	1.0760	0.4780
1886.	0.9540	-0.2893	1924.	1.2330	1.4654
1887.	0.8280	-1.0818	1925.	0.7760	-1.4088
1888.	1.1030	0.6478	1926.	1.1490	0.9371
1889.	1.0360	0.2264	1927.	1.1770	1.1132
1890.	0.9730	-0.1698	1928.	0.9210	-0.4969
1891.	0.9810	-0.1195	1929.	0.9050	-0.5975
1892.	1.0210	0.1321	1930.	1.0060	0.0377
1893.	0.7850	-1.3522	1931.	1.3610	2.2704
1894.	0.7780	-1.3962	1932.	0.9590	-0.2579
1895.	0.8250	-1.1006	1933.	0.9020	-0.6164
1896.	1.1010	0.6352	1934.	0.7850	-1.3522
1897.	1.3200	2.0126	1935.	1.0540	0.3396
1898.	0.9480	-0.3270	1936.	1.2980	1.8742
1899.	0.9310	-0.4340	1937.	1.2110	1.3270
1900.	0.9250	-0.4717	1938.	0.9710	-0.1824
1901.	0.9910	-0.0566	1939.	1.0650	0.4088
1902.	1.0280	0.1761	1940.	1.3600	2.2642
1903.	0.9480	-0.3270	1941.	1.1420	0.8931
1904.	0.9930	-0.0440	1942.	0.8970	-0.6478
1905.	0.9790	-0.1321	1943.	0.7890	-1.3270
1906.	1.0870	0.5472	1944.	0.8460	-0.9686
1907.	0.9420	-0.3648	1945.	1.2030	1.2767
1908.	0.9250	-0.4717	1946.	1.0310	0.1950
1909.	0.7980	-1.2704	1947.	0.7400	-1.6352
1910.	1.3340	2.1006	1948.	1.0580	0.3648
1911.	0.9500	-0.3145	1949.	0.6370	-2.2830
1912.	1.0420	0.2642	1950.	0.7570	-1.5283
1913.	1.0400	0.2516	1951.	1.1510	0.9497
1914.	1.2710	1.7044	1952.	0.9530	-0.2956
1915.	1.0730	0.4591	1953.	1.1630	1.0252
1916.	1.1660	1.0440	1954.	0.9160	-0.5283
1917.	1.1200	0.7547	1955.	1.2550	1.6038
1918.	0.9130	-0.5472	1956.	1.0430	0.2704
1919.	1.0950	0.5975	1957.	1.0330	0.2075
1920.	1.0830	0.5220			