APPLICATION OF RUN-LENGTHS TO HYDROLOGIC SERIES

by

J. Saldarriaga and V. Yevjevich

April 1970

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

40

APPLICATION OF RUN-LENGTHS

TO HYDROLOGIC SERIES

by

J. Saldarriaga and V. Yevjevich

HYDROLOGY PAPERS
COLORADO STATE UNVIERSITY
FORT COLLINS, COLORADO 80521

April 1970

No. 40

ACKNOWLEDGEMENTS

The financial support of the U.S. National Science Foundation under grant GK-11444 (Hydrologic Stochastic Processes) and grant GK-11564 (Large Continental Droughts) for the research leading to this Hydrology Paper is gratefully acknowledged.

Acknowledgement is also made to the National Center for Atmospheric Research, sponsored by the National Science Foundation, for some use of its CDC 6600 computer in the investigations leading to this paper.

Acknowledgement goes to Dr. M. M. Siddiqui, Professor, and Dr. P. Todorovic, Associate Professor, at Colorado State University for their advice in connection with some mathematical problems of run theory.

The cooperation of two M.S. graduate students in the computing phase of the study, Mr. V. K. Gupta and Mr. P. C. Tao, is appreciated. The bulk of computations was done on the CDC 6400 computer at the Colorado State University Computing Center.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES AND TABLES

LIST OF SYMBOLS

Symbol	Meaning
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
$0(\rho^3)$	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
N_0, N_1	Number of elements of kinds 0 and 1 respectively in a binomial population
N	$N_0 + N_1$
α	N_0/N_1. Also level of significance
x_o	Truncation level
q	Probability of drawing an element of kind 1 in a binomial population. Also $F(x_o)$
p	Probability of drawing an element of kind 0 in a binomal population. Also $1-F(x_o) = 1-q$
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
T_j^+	j-th positive run in continuous time
T_j^-	j-th negative run in continuous 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
$P_m(k^-, j^+)$	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=0.5
$P(k^-, j^+)$	Same as previous one for any q.

LIST OF SYMBOLS (continued)

Symbol	Meaning
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
$H_i(X)$	Hermite polynomial of order i.
k^+	Number of positive runs
k^-	Number of negative runs
k	Number of total runs
k^*	$(k^+ + k^-)/2$.
$\bar{N}^+$	Mean positive run-length
$\bar{N}^-$	Mean negative run-length
$\bar{N}$	Mean total run-length
$\bar{N}^*$	Mean run-length when two levels q and p = 1-q are considered.

ABSTRACT

A method is developed for investigating time series structure by using the mean run-length parameter. This method is distribution-free. Applications to selected annual precipitation series and annual runoff series demonstrate the feasibility of this method.

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 is either an independent or a dependent stationary series of a variable which follows the first-order linear autoregressive model.

Numerical values of probabilities of run-lengths are obtained by the digital computer integration of expansion equations for run-length probabilities of the first-order linear autoregressive model. A set of tables and a set of graphs are presented to make the numerical values readily useable. Probabilities of run-lengths of dependent variables with a common distribution are also distribution free.

The significance of this investigation, and several applications in the text, are based on the premise that run-lengths, as statistical properties of time series, represent attractive parameters in studying droughts and surpluses.

APPLICATION OF RUN-LENGTHS TO HYDROLOGIC SERIES

by

Jaime Saldarriaga* and Vujica Yevjevich**

Chapter I

DEFINITION OF PROBLEMS INVESTIGATED

1.1 Stationary Hydrologic Series. Annual precipitation, annual effective precipitation (precipitation minus evaporation), and annual river flow vary from year to year. This variation is generally referred to as the sequence of wet and dry years. These sequences are hydrologic time processes. For all practical purposes in water resources development, they can be assumed to be approximately stationary time series [1,2]. The hydrologic stationary processes of annual river flow and annual effective precipitation are dependent time series. This means that successive values are linked in some persistent manner, or sequences of annual river flow and annual effective precipitation are stationary dependent processes [2]. Sequences of annual precipitation are very near to being stationary and independent stochastic processes [2].

Hydrologic continuous time processes, such as river flow discharge, intensity of precipitation and similar variables, and hydrologic discrete time series of time intervals, which are fractions of the day or year, or multiples of the day or the month, usually are non-stationary. They are periodic-stochastic processes with various weights of periodic and stochastic components [3,4,5]. Therefore, they are non-stationary processes.

The theory and properties of run-lengths, either already known or developed in this paper, are applicable only to stationary processes of annual time series of various hydrologic variables. The application of the theory of run-lengths of periodic-stochastic processes to hydrologic complex periodic-stochastic time series is not feasible at the present time for the simple reason that this theory has not yet been developed in the form to be applicable to discrete hydrologic time series composed of periodic and stochastic components.

1.2 Practical Significance. Sequences of annual values of many hydrologic variables have several practical connotations. The behavior of severe and prolonged droughts, with their properties, may not be known with sufficient accuracy to allow the probabilistic prediction of their occurrence, duration and areal coverage with a sufficient degree of reliability. Statistical properties of runs of time series may represent one of the best ways for an objective definition of drought [6]. This investigation of run-lengths and their application to series of wet and dry years is related to some significant problems of hydrology and water resource development.

Apart from determining the probabilities of droughts of various durations and severity at one point or over a region, the probability of droughts occurring in adjacent regions have significant economic implications. If two or more regions produce an important crop, or are supplying water to the producers of the same industrial product, then the conditional probabilities of droughts covering simultaneously these regions may be of importance to various plans.

The probability of an extended period of wet years is similar to the problem of the probability of droughts. It may be important for restoration of biological cover in semi-arid or arid regions, or for the fight of prolonged pollution produced during dry years in soils and various water environments.

1.3 Two Problems Related to the Application of the Theory of Run-Lengths to Hydrologic Processes. 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. Therefore, the successions are called run-lengths. Two types of events must be appropriately defined, either as greater, or smaller values than a given value.

The application of the theory of run-lengths to hydrologic stationary processes may be viewed from two basic standpoints:

(1) Some parameters of the run-lengths, as functions of another parameter, may be used for the investigation of stochastic hydrologic processes, particularly whether the series are stationary or not, and if so, whether they are serially independent or dependent. If found to be dependent, the interest is, what are the best mathematical models to describe this dependence.

(2) To determine, in the most reliable way, the properties of run-lengths of a hydrologic series whenever it is found to be stationary, independent, or dependent, and the mathematical model of dependence is found to describe well the empirical dependence, if the series is dependent.

Before these two standpoints are discussed in detail, the two classical methods and the two new potential methods, including runs, are briefly reviewed in order to better define the problems investigated in this paper.

1.4 Methods for Investigation of Hydrologic Sequences. Four methods based on specific statistical parameters, as they change with other parameters,

1. Autocorrelation analysis. Parameters involved are the autocorrelation coefficients, ρ_k, as a function of the lag k between the correlated

*Former Ph.D. Graduate of Colorado State University, Civil Engineering Department, Fort Collins, Colorado, now temporary Research Associate, Civil Engineering Department, Colorado State University, Fort Collins, Colorado.

**Professor of Civil Engineering and Professor-in-Charge of Hydrology and Water Resources Program, Civil Engineering Department, Colorado State University,

are or may be effectively used for the investigation of hydrologic processes:

values, or $\rho_k = f(k)$, with ρ_k defined by

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

for a discrete time series. The values ρ_k are estimated by the sample values r_k .

The use of autocorrelation analysis as an investigative technique of hydrologic time series is based on the concept of analogy. One should know the correlograms of particular processes, and then by statistical inference determine whether a computed correlogram of a hydrologic process is well approximated by the correlogram of a known process. To read the type of process that results from a correlogram, the alphabet of correlograms must be known.

2. Variance spectrum analysis. Basically this is the Fourier series analysis where an infinite number of elementary periodic components, with a continuous distribution of frequencies, is fitted to an observed series. The parameters involved are the variance densities, ν_f , of various harmonics fitted to this series, represented against the frequencies f as the parameter. The variance of a harmonic is equal to the half of its squared amplitude. This type of analysis is a representation of the process in frequency domain,

$$\nu_f = \psi(f) \quad , \tag{1.2}$$

while the autocorrelation is a representation in time domain, or any other dimension on which the process occurs (say, the length). It might be noted that the variance density spectral function is the Fourier transform of the correlogram. The variance densities ν_f are estimated by the sample variance densities, $\mathbf{v}_f$.

The use of variance spectrum analysis as an investigative technique of hydrologic processes is also based on the concept of analogy. Statistical inference should be performed to find whether a computed variance spectrum of a hydrologic process is well approximated by the variance spectrum of a known process. A reading knowledge of the alphabet of variance spectra should be known to advance hypotheses on the kind of mathematical model for the process investigated.

3. Ranges. The ranges, R_n , are defined in terms of differences between maximum and minimum on the cumulative sums of departures of values from the average, or from any other value, for given subseries sizes, n . The expected ranges, $E(R_n)$, or similar parameters, as random variables, are related to the subsample size, or

$$E(R_n) = f(n) \quad . \tag{1.3}$$

Let $\{x_i ;\ i=1,\ldots, N\}$ be the observed sequence, and let x_o be a specified truncation level which in general represents the reference level. Then the sum is

$$S_i = \sum_{i=1}^{i} (x_i - x_o) \quad , \tag{1.4}$$

for $i=1,2,\ldots, n$. The surplus is defined by

$$S_n^+ = \max\ \{0,S_i\} \quad \text{for} \quad i=1,2,\ldots, n \quad , \tag{1.5}$$

and the deficit by

$$S_n^- = \min\ \{0,S_i\} \ , \quad i=1,2,\ldots, n \quad , \tag{1.6}$$

where n represents the size of a subsample taken from $\{x_i\}$.

The range is defined by

$$R_n = S_n^+ - S_n^- = \max\{0,S_i\} - \min\{0,S_i\} \quad , \tag{1.7}$$

for $i=1,2,\ldots, n$.

As in the case of the autocorrelation and variance spectrum analyses, the use of the expected range (or of a similar parameter), as a function of n , may be conceived as an investigative technique of hydrologic series. It should be based also on the concept of analogy. The parameters $E(R_n)$ are estimated by the sample mean ranges, $\overline{R}_n$. The comparison of the function $\overline{R}_n = f(n)$ with the function of the same parameter of a known process allows the advancement of hypotheses about mathematical models describing dependence of a stochastic process. The statistical inference of the goodness of fit of these theoretical and hypothetical models decides whether they should be accepted or rejected. The alphabet of these range functions for various types of processes should be known before hypotheses are advanced.

4. Runs. Various properties of runs, clearly defined, have parameters α , which may be used as function of another parameter β , so that $\alpha = f(\beta)$ is a characteristic of a process of independent or dependent sequences. For the purposes of this paper, the run is identical to the concept of run-length. Basically, both are the number of consecutive positive or negative departures from a specified constant value called here the truncation level. In this narrower definition of runs, positive runs are associated with positive departures and negative runs with negative departures. The structure of a series may be analyzed by studying the properties of runs at different truncation levels. Parameters of runs have practical meanings in hydrology, because a positive run can be associated with the duration of a wet period or with a water surplus interval, while a negative run can be associated with the duration of a drought, or with a water deficit interval.

5. Comparison of four techniques. The two classical techniques for the investigation of time series are autocorrelation analysis and variance spectrum analysis. The way they are used in exploring the internal structure of a process depends to some extent on the purpose of inquiry and prior knowledge of the generating system of the process. The correlogram tells something about the linear relation between the consecutive values of a series. The spectrum exhibits the extent the series is in step with certain fundamental rhythms, measured at various frequencies [7]. These two techniques offer no particular advantage over other parameters for the task of investigating the properties of various sequences. One fact seems clear, namely that it is difficult to use the two functions $\rho_k = f(k)$ or

$\nu_f = \psi(f)$, respectively for these two techniques, directly in the solution of various water resources problems.

Ranges and runs are two techniques that can be used advantageously in water resources problems and at the same time, they may be used to investigate hydrologic processes. They can be readily associated with concepts of storage and drought, or with concepts of surplus and deficit, which are of interest to the solution of various water resources problems. This is one of the main reasons for investigating properties of run-length for both objectives: the investigation of hydrologic stochastic processes, and the direct computation of properties of runs, from the information in samples of these processes.

6. Runs as the technique. If a truncation level is specified, the run-length associated with a negative run represents the duration of a deficit relative to this level. The probability of length of the deficit periods is relevant for the planning, design, and operation of water resources systems.

The structure of a stochastic process 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 variables. Identically distributed variables with a highly positive first 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 serial correlation coefficient, are characterized by a mean run-length smaller than two at the same level. These properties, which are investigated in detail in the following chapters, should justify the use of runs not only in the making of water resources decisions, but also as a technique for the investigation of series, and more specifically for the testing of stationarity and of mathematical dependence models of hydrologic processes.

1.5 Two Approaches to Investigations of Stochastic Sequences. Regardless of which of the four methods of investigation of hydrologic sequences is used, a sequence of a parameter as a function of another parameter characterizes a stochastic process, like the functions $\rho_k = f(k)$, $\nu_f = \psi(f)$, $ER_n = f(n)$, or $EN_q = f(q)$. This last case is an example of runs, where EN_q is the expected value of run-length, estimated by the sample mean run-length $\bar{N}_q$, as it changes with the probability q of all values of a variable not greater than the truncation level. These four functions, related to autocorrelation coefficients, spectral densities, expected ranges, and expected run-lengths, should have well-defined mathematical expressions for various stochastic dependence models, or for processes composed of the periodic and stochastic components. Particularly, these four functions for the population of a stochastic stationary and independent process are well defined.

Two approaches for investigating time series may be used. The first approach consists of the analysis of original data. It is here referred to as the use of the original sample series. In this case, anyone of the four above functions is computed from the sample series, and compared with the family of corresponding population functions for various mathematical dependence models. Then a model is selected, its parameters estimated, and the population function compared with the sample function in such a way that their differences are or are not statistically significant. If they are significant, new models are selected as hypotheses, their parameters estimated, and the comparison repeated. The knowledge of shapes of above functions, $\rho_k = f(k)$, $\nu_f = \psi(f)$, $ER_n = f(n)$, or $EN_q = f(q)$, for various hydrologic mathematical dependence models is a prerequisite, so that sight comparison with the sample function of any of the above four functions may lead to the most likely hypotheses for the population models.

The second approach assumes a mathematical model for the dependence of a process that is composed of a systematic dependence component(s), and an independent stochastic component. A residual series is obtained by separating the systematic dependence component(s) from the original series. Under the hypothesis that the assumed model is an adequate representation of the process, the residual series after this separation should be a sequence of independent stochastic variables. The independence of the residual series is then tested. The assumed dependence 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 a "white noise," or independent series.

It is perhaps interesting to emphasize a basic difference between these two approaches. The first approach does not assume a model a priori for the process, but rather the curve of the sample function leads to the hypothesis about the structure of the process, so that eventually a mathematical dependence model can be fitted to it. The second approach may start a priori by assuming a dependence model for the process, without computing the sample function, and after the model parameters are estimated, the supposed independent stochastic component (white noise) is computed and tested. Logically, the sample function in any of the four above methods helps advance a more realistic hypothesis about the model. However, if previous knowledge about these models is already available for the similar processes in a region, the hypothesis can be advanced a priori, and the whitening and testing performed in an appropriate way.

In order to use the methods of run-length for investigating hydrologic sequences, run properties should be known for various mathematical dependence models of hydrologic sequences, regardless of the two approaches used. Therefore, the objective of investigation in this paper is to add knowledge about the properties of run-lengths for some mathematical dependence models of stationary hydrologic processes.

As an example, let the hypothesis be that $\{X_i\}$ is a first-order linear autoregressive process in the form

$$X_i - \mu = \rho_1(X_{i-1}-\mu) + \sigma\sqrt{1-\rho_1^2}\,\varepsilon_i \qquad (1.8)$$

with μ the expected value and σ^2 the variance of X , ε_i is a standardized stationary independent variable (0,1), while ρ_1 is the first autocorrelation coefficient. The parameters, μ , σ^2 and ρ_1 are estimated by sample parameters $\overline{X}$, $\hat{S}^2$ and r_1 .

The "whitened" series is

$$\varepsilon_i = \frac{1}{\sigma\sqrt{1-\rho_1^2}} [X_i-\mu-\rho_1(X_{i-1}-\mu)] \quad . \tag{1.9}$$

Under the given hypothesis, $\{\varepsilon_i\}$ is a sequence of standardized, independent random variables. Then the whitened series is tested for independence.

1.6 Objectives for Determinating Properties of Run-Length. The first objective of this study is to develop a method for investigating stationary independent and dependent hydrologic time series by using statistical parameters of runs. Four phases must be involved in this investigation:

(a) Mathematical formulation of the problem;

(b) Selection of suitable parameters for testing hypotheses of stationarity and time dependence;

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

(d) Tests of application of the method to some selected time series.

The second objective of this study is to develop, in an approximate analytical procedure, the properties of run-lengths of the stationary, first-order, and linear autoregressive mathematical model of time dependence, as defined by Equation (1.8). This objective has a significant, practical aspect, as shown by the following example.

For a river with large storage capacities, what is the probability of a drought to occur with a duration of n or more years, if the drought is defined as a run of all annual inflows into reservoir of above capacity, which are not greater than a given annual runoff. In this case, it is possible to determine the truncation level of the series of annual runoff and from it the probability q . If the dependence in the series of runoff can be well approximated by the model of Equation (1.8), and μ , σ , and ρ_1 are estimated from it, then the results of investigations in this study should answer readily and accurately the above classical problem. The available runoff series may not include even a drought of the duration of n/2 or of a shorter duration, so that the current empirical methods cannot give an answer to this problem. There are two reasons for concentration on the model of Equation (1.8): (1) It is often the most appropriate model for dependence of series of annual river flows, and (2) It is simple for an analytical treatment.

1.7 Definition of Runs. A series of the variable x is cut at many places by an arbitrary horizontal truncation level, x_o , and the relation of this constant x_o to all other values of x of the process serves as a basis for the definition of runs in this study. Basically, there must be two processes intersecting each other in order to define runs. Because these two processes cross each other, the theory of runs is often called the crossing theory. The term "theory of runs" is used in the case of discrete series [7], and the term "crossing theory" in the case of continuous series [8].

One of the two processes must be the original process. The second process may be a constant x_o , the process of a random variable y , or any other type of deterministic, combined deterministic - stochastic, or pure stochastic process. When this second process is not a constant, the development of properties of runs becomes complex. In the case of runs to be used in this study, the main assumptions are:

1. Only discrete series are investigated, so that the expression "runs" is used;

2. The variable x may have discrete, continuous, or fixed probability distribution;

3. The second process is a constant x_o , or any constant value in the range of fluctuation of the variable x ;

4. The probability $P(x \leq x_o) = q$ may replace the constant x_o , in order to make some properties of runs independent of the type of distribution of x.

The number of values of a discrete sequence between an upcrossing of the truncation level and the following downcrossing is defined as a positive run-length, or briefly, for this study, the positive run. Similarly, a negative run-length, or the negative run, is defined as the number of values of a discrete series between a downcrossing and the next upcrossing. They are shown in the upper graph of Figure 1.1, and are designated by

N_j^+ for the length of the j-th positive run, and

by N_j^- for the length of the j-th negative run. The j-th total run is defined as

$N_j = N_j^+ + N_j^-$, with $j=1,2,\ldots,$ where j is counted from the origin of a time series.

These may be extended to definitions T_j^+ , T_j^- , and T_j , as the positive, the negative and the total run of a continuous process, respectively. This is analogous to the definitions of runs of discrete time series, as shown in the lower graph of Figure 1.1.

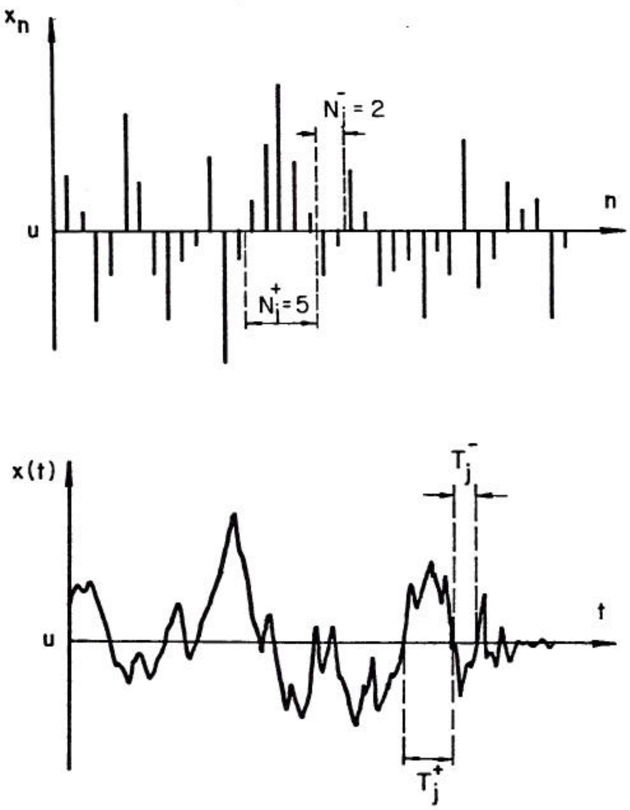


Fig. 1.1 Definition of positive and negative runs for a given truncation level. Upper graph refers to a discrete series and lower graph to a continuous series.

Other parameters used in literature as definitions of various runs of discrete time series, besides N_j^+ , N_j^- , and N_j , are:

1. Sum of deviations associated with positive runs, as the positive run-sum, or the run-surplus,

2. Sum of deviations associated with negative runs, as the negative run-sum, or the run-deficit,

3. Number of positive runs for a given series of size N ,

4. Number of negative runs for a given series of size N ,

5. Number of total runs for a given series of size N .

For the continuous time series, the following parameters other than T_j^+ , T_j^- , and T_j are used:

1. Area above truncation level for T_j^+ , as the positive run-sum, or the run-surplus;

2. Area below truncation level for T_j^- , as the negative run-sum, or the run-deficit;

3. Number of positive runs for a given series length, T ;

4. Number of negative runs for a given series length, T ;

5. Number of total runs for a given series length, T ;

6. Time interval between successive peaks;

7. Time interval between successive troughs.

All of these runs are random variables, and are functions of the process $\{x_i\}$ and the truncation level x_o .

Properties of runs relating to these functions can be directly used in many water resources problems. If x_o determines the level of demand, and if this level is not reached, a drought occurs. If a flooded area begins for $x > x_o$, and the flood damage is a function of the time during which $x > x_o$, then the distribution of positive run-length and/or run-sum determines the character of flooding. If a given type of run is regionalized, or shown over an area with its isolines, the regional phenomena of drought, flood, and similar phenomena may be studied for their probabilities of recurrence [6].

Chapter II

SUMMARY AND STATUS OF KNOWLEDGE ON DISCRETE RUNS

2.1 Introductory Statement. Two main aspects are reviewed, the distribution theory of runs for both independent and dependent random variables, and the multivariate normal integral which serves as a base for the mathematical developments in Chapter III. The summary is related only to those properties of runs, which are relevant to investigations in this paper.

2.2 Distribution Theory of the Number of Various Runs of Independent Random Variables. The classical distribution theory of runs has been mainly concerned with independent arrangements of a fixed or a random number of several kinds of elements. This is not particularly relevant for this study, but is summarized for the sake of completeness. In the case of two different kinds of elements, it is assumed that the number of elements of each kind are N_o and N_1, and that they are all randomly drawn without replacement. This is equivalent to sampling a binomial population, with probabilities of elements, p and $q = 1 - p$, respectively. Let K_i^o denote the number of runs of kind (o) of the length i, and let K_i^1 denote the number of runs of kind (1) of the length i. Finally, let $K^o = \sum_i K_i^o$ designate the number of all runs of elements N_o, $K^1 = \sum_i K_i^1$ the number of all runs of elements N_1, and $K = K^o + K^1$ the total number of runs, and $N = N_o + N_1$ the total number of elements, or the sample size, with $i = 1,2,\ldots$.

Wishart and Hirshfeld [9] obtained and tabulated the joint probabilities of the number of runs

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

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

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

and

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

In Equations (2.1) through (2.4), the capital letters designate the random variables, and the small letters the values those variables can take.

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

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

and

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

However, Cochran [10] gives expressions for the expected values of the number of positive and negative runs as

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

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

and

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

with N_o and N_1 being also the random variables in this case.

Stevens [11] gives the distribution of the total number of runs, without a regard to their length, from the arrangements of two kinds of elements. He develops a χ^2-criterion for the test of significance. Wald and Wolfowitz [12] study the same distribution as Stevens [11], and show that it is asymptotically normal. The conditional distributions of K are

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

and

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

where n_o and n_1 are values that N_o and N_1 can take. These probabilities are independent of the parameter p. For $n_o = \alpha n_1$, with $\alpha > 0$, and $n_o \to \infty$, Wald and Wolfowitz [12] give the above distributions of Equation (2.10) as a normal asymptotic distribution with

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

For $\alpha = 1$, the statistic

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

is a standard normal variable.

Mood [13] derives distributions of the number of runs of a given length for the independent arrangements of the fixed number of elements of two or more kinds of the binomial and multinomial populations. He shows these distributions as asymptotically normal with an increase in the sample size. Their expected values are:

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

and

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

The statistic

$$x = \frac{K-EK}{\sqrt{\text{var}K}} = \frac{K-2npq}{2\sqrt{npq(1-3pq)}} \tag{2.16}$$

is asymptotically normal with the mean of zero and the variance of unity. Comparing Equations (2.5) and (2.6) with the mean 2npq , and the variance 4npq(1-3pq) of Equation (2.16), the mean and variance given by Mood [13], and the mean and variance given by Wishard and Hirshfeld [9], are different. Parameters in Equation (2.16) are approximations to those of Equations (2.5) and (2.6). Bendat and Piersol [14] give tables for the conditional distribution of K when $N_o = N_1 = N/2$.

2.3 Distribution Theory of Run-Lengths of Independent Random Variables. Let N_j^+ and N_j^- denote the positive and negative j-th run-length for the given truncation level, x_o . Also let $\{X\}$ be the sequence of independent random variables of the common distribution, $F(x)$, with $F(x_o) = q$, and $1-F(x_o) = p$, and let

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

be the random sequence of the total j-th run-length.

The probability mass function of N_1 is given by Feller [15] as

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

for k=2,3,..., with

$$EN_1 = \frac{1}{pq} \quad , \text{ and } \quad \text{var } N_1 = \frac{1-3pq}{p^2q^2} \quad . \tag{2.18}$$

The distribution of the number of total runs, k(N) , in a discrete time series of length N has the following parameters

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

and

$$\text{var}k(N) = Npq \; (1-3pq - \frac{1}{N} + \frac{5}{N} pq) \quad , \text{ for } N \geq 4; \tag{2.20}$$

this distribution is asymptotically normal. Downer, Siddiqui and Yevjevich [16] studied the distribution of positive and negative run-lengths for a sequence of independent identically distributed random variables, and applied it to the normal variable. They have shown that $\{N_j^+\}$ is also a sequence of independent identically distributed random variables with the probability mass functions

$$P(N_j^+ = k) = qp^{k-1} \quad , \text{ and } \quad P(N_j^- = k) = pq^{k-1} \quad , \tag{2.21}$$

and their moments are

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

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

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

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

$$EN_j^+ = EN_j^- = 2 \quad , \tag{2.25}$$

and

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

Llamas [17] studied the case of standard, one-parameter Gamma random variables, with the probability distribution function

$$F(x) = \int_{-\sqrt{\alpha}}^{x} \frac{\alpha(\alpha+t\sqrt{\alpha})^{\alpha-1}}{\Gamma(\alpha)} \; e^{-\alpha-t\sqrt{\alpha}} \; dt \quad . \tag{2.27}$$

For $x_o = 0$, he obtained $p = F(0) = P(\alpha,\alpha)$, and $q = 1 - P(\alpha,\alpha)$, where $P(\alpha,\alpha)$ is the incomplete Gamma function, or

$$P(\alpha,\alpha) = \frac{1}{\Gamma(\alpha)} \int_0^{\alpha} e^{-t} \; t^{\alpha-1} \; dt \quad . \tag{2.28}$$

Llamas and Siddiqui [18] studied the case of a sequence of a two-dimensional random process $\{x,y\}$, where the two variables are independent and have a common distribution function, $F(x,y)$. Given the two levels, x_o and y_o, such that $o < F(x_o,y_o) < 1$, the four possible events are defined as

$$A = \{x \leq x_o,\ y \leq y_o\} \quad , \quad B = \{x \leq x_o,\ y > x_o\}$$

$$C = \{x > x_o,\ y \leq y_o\} \quad , \quad D = \{x > x_o,\ y > y_o\} \quad .$$

Both sequences are associated with the sign minus if A occurs, and with the sign plus if D occurs. The sequence of k consecutive A events followed and preceded by any other event is a negative run of the length k. The sequence of k consecutive D events followed and preceded by any other event, is a positive run of the length k, and for the initial run the requirement of "preceded by" is dropped. If A^c is the complement set of A, then

$$P(A) = F(x_o,y_o) = q \quad , \text{ and } \quad P(A^c) = p \quad .$$

Llamas and Siddiqui have shown [18] that

$$P(N_j^- = k) = p\, q^{k-1} \quad , \tag{2.29}$$

with

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

the analogous relations hold for N_j^+ for its corresponding values of p and q.

2.4 Distribution Theory of Run-Lengths of Dependent Random Variables. For a Markov chain with two states (0) and (1), Cox and Miller [19], give the transition probability matrix of this chain, which is

$$P = \begin{array}{c} 0 \\ \\ 1 \end{array} \begin{bmatrix} 1-\alpha & \alpha \\ & \\ \beta & 1-\beta \end{bmatrix} \quad . \tag{2.31}$$

They give the distribution of the recurrence time of state (0), designated by N^o, which is equal to the run-length of state (1) plus unity, as

$$P(N^o = k) = \alpha\beta(1-\beta)^{k-2} \quad , \text{ for } \quad k=2,3,\ldots, \tag{2.32}$$

and

$$P(N^o = k) = 1 - \alpha \quad , \text{ for } \quad k=1 \quad . \tag{2.33}$$

The mean recurrence time of the state 0 is then

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

Similar relations hold for the recurrence time of the state (1), which is equal to the run-length of state (0) plus unity, designated by N^1, by interchanging α and β.

Heiny [20] defines the two states with their transition probabilities of the Markov chain as

$$P(x_j > x_o | x_{j-1} > x_o) = r \quad , \text{ and}$$

$$P(x_j \leq x_o | x_{j-1} > x_o) = s \quad ,$$

with $r + s = 1$. The following relations are valid for this Markov Gaussian process $\{x\}$:

$$P(N^+ = k | x_1 > 0) = s r^{k-1}\, [1 + 0(\rho^2)],\ k=1,2,3,\ldots, \tag{2.35}$$

with

$$E(N^+ | x_1 > 0) = \frac{1}{s}\, [1 + 0(\rho^2)] \quad , \tag{2.36}$$

and

$$\operatorname{var}\ (N^+ | x_1 > 0) = \frac{r}{s^2}\, [1 + 0(\rho^2)] \quad , \tag{2.37}$$

where $0(\rho^2)$ indicates an expression that 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 the first j negative runs, given $x_1 > 0$, as follows:

$$P(N_j^+ = n_j,\ N_j^- = m_j,\ N_{j-1}^+ = n_{j-1},\ldots,\ N_1^+ = n_1,\ N_1^- = m_1 | x_1 > 0) = $$
$$= s r^{n_1 - 1}\ t v^{m_1 - 1}\ s r^{n_2 - 1}\ t v^{m_2 - 1} \ldots s r^{n_{j-1}}\ t v^{m_{j-1}} [1 + 0(\rho^2)], \tag{2.38}$$

where

$$t = P(x_j > x_o | x_{j-1} \leq x_o) \quad , \quad v = P(x_j \leq x_o | x_{j-1} \leq x_o) \quad ,$$

and

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

This treatment, however, has two disadvantages: (a) it is based on a conditional probability that $x_1 > 0$, and (b) it is applicable only to very small values of ρ, since the errors $0(\rho^2)$ may be significant for larger values of ρ.

2.5 The Multivariate Normal Integral. Gupta [21] presents an exhaustive bibliography on the multinormal integral and related topics, and gives a review [22] of these works. Only works that do not overlap with references in [21] and [22], but are related to mathematical developments in the following chapters are reviewed here.

The multinormal integral is involved in the theory of runs of dependent normal variables because it is directly related to the problem of h autocorrelated random variables, $Z_1, Z_2, \ldots, Z_h$. If these variables have a standard multivariate normal distribution, the problem to solve is the probability that all h variables are simultaneously positive. A new sequence of random variables $\{X\}$ is defined as follows:

$$X = \begin{cases} 1 & \text{for } Z \geq 0 \\ -1 & \text{for } Z < 0 \end{cases} .$$

The probability that all h variables are simultaneously positive is $P_m(h^+)$, where the index m indicates that the truncation level x_o of the random process $\{X\}$ is the median of the distribution of $\{Z\}$. For $r_{ij} = EX_iX_j$, McFadden [23] gives, for any $h \geq 4$,

$$P_m(h^+)=2^{-h}[1 + \sum_{j>i\geq 1} r_{ij} + \sum_{\ell>k>j>i\geq 1} (r_{ij}r_{k\ell} + r_{ik}r_{j\ell} + r_{i\ell}r_{jk} + O(r^3)] \quad . \tag{2.39}$$

For $\rho_{ij} = EZ_iZ_j$,

$$r_{ij} = \frac{2}{\pi} \text{ arc } \sin\rho_{ij} = \frac{2}{\pi} [\rho_{ij} + O(\rho_{ij}^3)] \quad . \tag{2.40}$$

If Equation (2.40) is substituted into Equation (2.39),

$$P_m(h^+) = 2^{-h}\left[1 + \frac{2}{\pi} \sum_{j>i\geq 1} \text{arc } \sin\rho_{ij} + \frac{4}{\pi^2} \sum_{\ell>k>j>i\geq 1} \rho_{ij}\rho_{k\ell} + \rho_{ik}\rho_{j\ell} + O(\rho^3)\right] \quad . \tag{2.41}$$

Obviously, and for the univariate case, Equation (2.41) becomes

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

For the bivariate case, the result is known as Sheppard's theorem [24] of the median dichotomy; it is

$$P_m(2^+) = \frac{1}{4} + \frac{1}{2\pi} \text{ arc } \sin\rho \quad . \tag{2.43}$$

This equation is tabulated in the Tables of Mathematical Functions of the National Bureau of Standards [28] for ρ, which varies from 0 to 1, with increments of 0.01. For the trivariate case, the following result is given by David [26]

$$P_m(3^+) = \frac{1}{8} + \frac{1}{4\pi} (\arcsin\rho_{12} + \arcsin\rho_{13} + \arcsin\rho_{23}) \quad . \tag{2.44}$$

Chapter III

PROBABILITIES OF RUN-LENGTH OF THE FIRST-ORDER LINEAR AUTOREGRESSIVE MODEL OF NORMAL VARIABLES

3.1 General Notations and Expressions for Probabilities of Run-Length. For purposes of simplicity, the following notation is adopted:

$$P\{X_1 \leq x_o, X_2 \leq x_o, \ldots, X_k \leq x_o, X_{k+1} > x_o, X_{k+2} > x_o, \ldots, X_{k+j} > x_o\} = P(k^-, j^+) \ ,$$

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

with $k=1,2,\ldots$ and $j=1,2,\ldots$.
The probability of the first positive run-length N_1^+ from the beginning of a series being equal to or greater than j , is

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

The probability mass function of N_1^+ is

$$P(N_1^+ = j) = P(N_1^+ \geq j) - P(N_1^+ \geq j+1) \quad . \qquad (3.2)$$

The computation of joint probabilities $P(k^-, j^+)$ requires the joint probability distribution of the variables $x_1, x_2, \ldots$. This joint distribution for the purposes of this study is assumed multivariate normal.

3.2 Stationary and Ergodic Multidimensional Gaussian Processes. An arbitrary Gaussian random process $\{x_i\}$, or $x_1, x_2, \ldots, x_n$, where $i=1,2,\ldots,n$ at arbitrary or equally spaced positions in time, has the multivariate normal distribution in n dimensions. This process is completely described by the parameters of this distribution: the expected values $E(x_i)$, $i=1,2,\ldots,n$, and the covariance matrix, $\mathrm{cov}(x_i, x_j)$ as a function of i and j .

A multivariate Gaussian process is stationary if, and only if, the expected value is constant and the covariances depend only on the lag $|j-i|$, and are independent of i . For any stationary process Ex_i is equal to μ , and $\mathrm{cov}(x_i, x_{i+k})$ is equal to $C(k)$. In particular, $C(o)$ is equal to var x and is a constant independent of i . The function $C(k)$ is the autocovariance function, while

$$\rho(k) = \frac{C(k)}{C(o)} \qquad (3.3)$$

is the autocorrelation function. It specifies the correlation coefficient between values of the process, which are k intervals apart, and it is the k-th autocorrelation coefficient.

Let $\{x\}$ be a stationary Gaussian process with zero expected value and variance unity. Its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \quad . \qquad (3.4)$$

The bivariate probability density function of x_i and x_j , with $Ex_i = Ex_j = 0$, and var x_i = var $x_j = 1$, is

$$f(x_i, x_j) = \frac{1}{2\pi\sqrt{1-\rho_{ij}}} \exp\left[-\frac{1}{2}\,(x_i^2 - 2\rho_{ij}x_i x_j + x_j^2)\right] \ , \qquad (3.5)$$

where ρ_{ij} is the correlation coefficient between s_i and x_j . The multivariate normal probability density function of $x_1, x_2, \ldots, x_n$ takes a more complex form, but is analogous to Equation (3.5) and given by Equation (3.9). In this case, the correlation matrix of random vairables $x_1, x_2, \ldots, x_n$ is the n by n matrix with the elements ρ_{ij} representing the correlation coefficients between any two variables x_i and x_j, $i=1,2,\ldots,n$ and $j=1,2,\ldots,n$. It is a symmetrical matrix since $\rho_{ji} = \rho_{ij}$, and all elements of the main diagonal are one. For a stationary process

$$\rho_{ij} = \rho_{|j-i|} = \rho_k \quad , \qquad (3.6)$$

with $k = |j-i|$; therefore, all elements of any diagonal are identical. The correlation matrix of a stationary process is

$$\pi = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \rho_{n-2} & \rho_{n-1} \\ \rho_1 & \ddots & \ddots & \ddots & & & \rho_{n-2} \\ \rho_2 & \ddots & \ddots & \ddots & \ddots & & \cdot \\ \cdot & & \ddots & \ddots & \ddots & \ddots & \cdot \\ \cdot & & & \ddots & \ddots & \ddots & \rho_2 \\ \rho_{n-2} & & & & \ddots & \ddots & \rho_1 \\ \rho_{n-1} & \rho_{n-2} & \cdot & \cdot & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

If the random process $\{x\}$ is second-order stationary, as described above, and if the expected values and crossproduct functions defined by averages of individual realizations (sample functions) as

$$Ex_i = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i \quad , \qquad (3.7)$$

and

$$Ex_ix_{i+k} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_ix_{i+k} \quad , \tag{3.8}$$

then the process is ergodic. A second-order stationary and ergodic Gaussian process is also strictly stationary and ergodic, or higher-order stationary and ergodic. This means that all ensemble averaged statistical properties are equal to the corresponding time averages. Hence, the verification of self-stationarity for a single time series justifies the assumption of stationarity and ergodicity.

3.3 Multivariate Normal Probability Density Function. The normal distribution of n variables is

$$dF = \frac{1}{(2\pi)^{n/2}\sqrt{|R|}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_jx_k\right\}\prod_{j=1}^{n} dx_j \quad ,$$

where the variables $x_1, x_2, \ldots, x_n$, have expected values of zero and variances of unity. Also, $|R|$ is the determinant of the correlation matrix of these variables, while a_{jk} are the elements of the inverse of the correlation matrix. The characteristic function of this distribution is not expressed in terms of the inverse of the correlation matrix, but in terms of the elements of the correlation matrix itself. This property helps in computing probabilities of run-lengths. The characteristic function is

$$\phi(t) = \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}t_it_j\right\} \quad . \tag{3.10}$$

3.4 General Expression for Joint Probability of at Least k Subsequent Values Below Truncation Level, Followed by at Least j Subsequent Values Above Truncation Level. In order to find an expression for the joint probabilities, $P(k^-,j^+)$, involved in Equation (3.1), the following assumptions are made:

1. The hydrologic time series of annual precipitation and annual runoff are second-order stationary. Some of these series may have, however, a small degree of non-stationarity, which comes from either man-made changes in river basins and around the precipitation gauging stations, or from the inconsistency in data [27]. These series should be made stationary by corrections before the theory of runs, as discussed here, is applied.

2. The process of annual values is a Gaussian process or approximately so. This assumption is justified from the point of view that some runs are distribution free, or independent of the underlying distributions of $\{X_i\}$. It is also justified from the point of view that many non-Gaussian hydrologic processes can be reduced to Gaussian processes through appropriate transformations. This point will be treated in detail in Chapter IV.

3. The stationary Gaussian processes are standardized for a simpler treatment of various problems.

With the above three assumptions, the joint probabilities $P(k^-,j^+)$ can be expressed as

$$P(k^-,j^+) = \underbrace{\int_{-\infty}^{x_o}\cdots\int_{-\infty}^{x_o}}_{k}\ \underbrace{\int_{x_o}^{\infty}\cdots\int_{x_o}^{\infty}}_{j} dF \quad . \tag{3.11}$$

Substituting dF by its equivalent into Equation (3.9) gives

$$P(k^-,j^+) = \frac{1}{(2\pi)^{n/2}|R|^{1/2}} \underbrace{\int_{-\infty}^{x_o}\cdots\int_{-\infty}^{x_o}}_{k}\ \underbrace{\int_{x_o}^{\infty}\cdots\int_{x_o}^{\infty}}_{j}$$

$$\cdot \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_jx_k\right\}\prod_{j=1}^{n} dxj \quad , \tag{3.12}$$

where $n = j + k$. Equation (3.11) is the multinormal integral. No explicit expression exists for the general solution of the multinormal integral. Efforts are devoted to finding expressions for several cases of this multinormal integral in this study, so that specific numbers can be assigned to probabilities in Equation (3.12). These probabilities will be called joint probabilities to distinguish them from the probabilities of runs.

3.5 Probabilities of Runs for Any Truncation Level. Throughout this subchapter concern is with the evaluation of probabilities of the type

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

where $q = F(x_o)$. To simplify notation, the sub-index q is dropped, and it will be used only when it is necessary to refer to it.

Probabilities $P(2^+)$ and $P(1^-,1^+)$. In the univariate case, the following expression obviously holds

$$P(1^+) = \int_{x_o}^{\infty} dF = 1 - F(x_o) \tag{3.13}$$

where $F(x_o)$ is the standard normal distribution function. In the bivariate case (x_i, x_{i+1}) ,

$$P(2^+) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x_o}^{\infty}\int_{x_o}^{\infty} \exp\left[-\frac{x_i^2 - 2\rho x_ix_{i+1} + x_{i+1}^2}{2(1-\rho^2)}\right]$$

$$\cdot\, dx_i dx_{i+1} \quad , \tag{3.14}$$

and

$$P(1^-,1^+) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{x_o}\int_{x_o}^{\infty} \exp\left[-\frac{x_i^2 - 2\rho x_ix_{i+1} + x_{i+1}^2}{2(1-\rho^2)}\right]$$

$$\cdot\, dx_i dx_{i+1} \quad . \tag{3.15}$$

These two probabilities are related as

$$P(1^-,1^+) = \int_{-\infty}^{x_o}\int_{x_o}^{\infty} dF = \int_{-\infty}^{x_o}\int_{x_o}^{\infty} dF$$

$$-\int_{x_o}^{\infty}\int_{x_o}^{\infty} dF = 1-F(x_o) - P(2^+) \quad . \tag{3.16}$$

Bivariate tables are given by the National Bureau of Standards [28] 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. Zelen and Severo [29] give charts for the bivariate normal integral with an error of 1 percent or less.

<u>Probability $P(3^+)$</u>. For three variables,

$$P(3^+) = \int_{x_o}^{\infty}\int_{x_o}^{\infty}\int_{x_o}^{\infty} dF \quad .$$

The integral of this equation has been evaluated in terms of the tetrachoric series expansion by Kendall [20]. It is

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

$$H_{k+\ell-1}(x_o) f^3(x_o) \quad ,$$

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

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

and j, k, ℓ can take the values 0, 1, 2, ... The first three Hermite polynomials are $H_o(x)=1$, $H_1(x)=x$ and $H_2(x)=x^2-1$.

<u>Probabilities of the type $P(j^+)$</u>. The tetrachoric series expansion for the trivariate case [30] can be generalized to the multivariate case by the following procedure. As discussed previously, the multinormal probability density function can be expressed in terms of 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 integrated. This is a parallel procedure to the one followed by Kendall [30] for the trivariate case. By definition

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

$$= \frac{1}{(2\pi)j}\int_{x_o}^{\infty} dx_1 \ldots \int_{x_o}^{\infty} dx_j \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(t)$$

$$\cdot \exp(-i\bar{t}'\bar{X})dt_1 \ldots dt_j \quad , \tag{3.20}$$

where

$$\bar{t}'\bar{X} = [t_1\ _2 \ldots t_j]\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \end{bmatrix} = t_1X_1+t_2X_2+\ldots+t_jX_j \quad . \tag{3.21}$$

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

$$\phi(t) = \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{j} t_i^2 + 2\sum_{k>i\geq 1}\rho_{ik}t_it_k\right)\right] \quad . \tag{3.22}$$

In using the exponential series expansion

$$\exp\left(-\sum_{k>i\geq 1}^{k=j}\rho_{ik}t_it_k\right) = \sum_{r=0}^{\infty}\frac{(-1)^r}{r!}\left(\sum_{k>i\geq 1}\rho_{ik}t_it_k\right)^r \quad . \tag{3.23}$$

Substituting Equation (3.23) into Equation (3.22),

$$\phi(t) = \exp\left[-\frac{1}{2}\sum_{i=1}^{j}t_i^2\right]\sum_{r=0}^{\infty}\frac{(-1)^r}{r!}\left[\sum_{k>i\geq 1}^{j}\rho_{ik}t_it_k\right]^r \tag{3.24}$$

where

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

$$\ldots+\rho_{2n}t_2t_n+\ldots+\rho_{n-1,n}t_{n-1}t_n)\Big]^r$$

$$= r!\sum\frac{\rho_{12}^{i_{12}}\rho_{13}^{i_{13}}\ldots\rho_{1n}^{i_{1n}}\rho_{23}^{i_{23}}\rho_{24}^{i_{24}}\ldots\rho_{2n}^{i_{2n}}\ldots\rho_{n-1,n}^{i_{n-1,n}}}{i_{12}!i_{13}!\ldots i_{1n}!i_{23}!i_{24}!\ldots i_{2n}!\ldots i_{n-1,n}!}$$

$$\cdot\, t_1^{S_1}t_2^{S_2}\ldots t_n^{S_n} \tag{3.25}$$

with $S_1=i_{12}+i_{13}+\ldots+i_{1n}$; $S_2=i_{12}+i_{23}+\ldots+i_{2n};\ldots;$

$S_n=i_{1n}+i_{2n}+\ldots+i_{n-1,n}$.

Substituting Equation (3.21) and Equation (3.24) into Equation (3.20) gives

$$P(j^+)=\frac{1}{(2\pi)j}\int_{x_o}^{\infty} dx_1 \ldots \int_{x_o}^{\infty} dx_j \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\sum_{i=1}^{j} t_i^2\right]$$

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

$$\cdot\ t_1^{S_1} t_2^{S_2}\ldots t_n^{S_n} \exp(-i\bar{t}'\bar{X})dt_1\ldots dt_j \quad . \tag{3.26}$$

By adopting the notation

$$\frac{\rho_{12}^{i_{12}}\rho_{13}^{i_{13}}\ldots\rho_{1n}^{i_{1n}}\rho_{23}^{i_{23}}\rho_{24}^{i_{24}}\ldots\rho_{2n}^{i_{2n}}\ldots\rho_{n-1,n}^{i_{n-1,n}}}{i_{12}!i_{13}!\ldots i_{1n}!i_{23}!i_{24}!\ldots i_{2n}!\ldots i_{n-1,n}!} = A(\rho,i) \ , \tag{3.27}$$

Equation (3.26) becomes

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

$$\cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{j} t_i^2\right] \exp(-i\bar{t}'\bar{X}) t_1^{S_1}\ldots t_j^{S_j}\ dt_1\ldots dt_j \quad . \tag{3.28}$$

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

$$\frac{1}{2\pi}\int_{x_o}^{\infty} dx_1 \int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}t_1^2\right]\exp(-it_1x_1)dt_1 \quad , \tag{3.29}$$

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

$$\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}t^2\right]t^S\exp(-itx)dt$$

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

$$= \sqrt{2\pi}\ i^r D^r \exp(-\frac{1}{2}x^2) = 2\pi(-i)^r\ H_r(x)f(x) \quad , \tag{3.30}$$

Equation (3.29) is

$$(-i)^r\int_{x_o}^{\infty} dx H_r(x)f(x) = (-i)^r H_{r-1}(x_o)f(x_o) \quad , \tag{3.31}$$

and Equation (3.28) becomes

$$P(j^+) = f^j(x_o)\sum_{\substack{r=0\\ \{i\}}}^{\infty} A(\rho,i)H_{S_1-1}(x_o)\ldots H_{S_j-1}(x_o)$$

$$= f^j(x_o)\sum_{\substack{r=0\\ \{i\}}}^{\infty}\frac{\rho_{12}^{i_{12}}\ldots\rho_{n-1,n}^{i_{n-1,n}}}{i_{12}!\ldots i_{n-1,n}!} H_{S_1-1}(x_o)\ldots H_{S_j-1}(x_o) \quad . \tag{3.32}$$

It is important to notice at this point that the definition of the Hermite polynomials applies only to $r=0,1,2,\ldots$. For $r=-1$, $H_{-1}(x)$ is defined by means of Equation (3.31) as

$$H_{-1}(x_o)f(x_o) = \int_{x_o}^{\infty} H_o(x)f(x)dx = 1 - F(x_o) \quad .$$

For $I = \Sigma i$, and $H_{S_1-1}(x_o)\ldots H_{S_j-1}(x_o)=\pi(H)$,

Equation (3.32) becomes

$$P(j^+) = f^j(x_o)\sum_{I=0}^{\infty} A(\rho,i)\pi(H) \quad . \tag{3.33}$$

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

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

$$= \int_{-\infty}^{\infty}\underbrace{\int_{x_o}^{\infty}\ldots\int_{x_o}^{\infty}}_{j} dF - \underbrace{\int_{x_o}^{\infty}\ldots\int_{x_o}^{\infty}}_{j+1} dF = P(j^+)-P[(j+1)^+] \quad . \tag{3.34}$$

The probabilities $P(j^+)$ and $P[(j+1)^+]$ can be evaluated by Equation (3.33).

Probabilities of the type $P(k^-,j^+)$. By a similar procedure,

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

$$= \frac{1}{(2\pi)^{k+j}} \underbrace{\int_{-\infty}^{x_o} \cdots \int_{-\infty}^{x_o}}_{k} \underbrace{\int_{x_o}^{\infty} \cdots \int_{x_o}^{\infty}}_{j} \phi(t)\exp(-i\overline{t}'\overline{X}) \quad .$$

Using the expansion of the multinormal characteristic function given by Equation (3.24),

$$P(k^-,j^+) = \frac{1}{(2\pi)^{k+j}} \sum_{r=0}^{\infty} (-1)^r \sum A(\rho,i) \int_{-\infty}^{x_o} dx_1 \cdots \int_{-\infty}^{x_o} dx_k$$

$$\int_{x_o}^{\infty} dx_{k+1} \cdots \int_{x_o}^{\infty} dx_{k+j} \exp\left(-\frac{1}{2}\sum_{i=1}^{j+k} t_i^2\right) \exp(i\overline{t}'\overline{X})$$

$$t_1^{S_1} \cdots t_{j+k}^{S_{j+k}} dt_1 \cdots dt_j \quad . \tag{3.36}$$

This is the product of k integrals of the type

$$\frac{1}{2\pi}\int_{-\infty}^{x_o} dx_1 \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}t_1\right) t_1^{S_1} \exp(-it_1x_1)dt_1 \ ,$$

and j integrals of the type

$$\frac{1}{2\pi}\int_{x_o}^{\infty} dx_{k+1} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}t_{k+1}^2\right) t_{k+1}^{S_{k+1}}$$

$$\cdot \exp(-it_{k+1}x_{k+1})dt_{k+1} \quad .$$

Taking into account Equation (3.30), the product of k integrals is

$$(-i)^r \int_{-\infty}^{x_o} dxH_r(x)f(x) \equiv \alpha_r^c(x_o) \quad ,$$

and the product of j integrals is

$$(-i)^r \int_{x_o}^{\infty} dxH_r(x)f(x) \equiv \alpha_r(x_o) \quad .$$

Thus Equation (3.36) becomes

$$P(k^-,j^+) = \sum_{I=0}^{\infty} A(\rho,i)\alpha_{S_1}^c(x_o)\cdots\alpha_{S_k}^c(x_o)\alpha_{S_{k+1}}(x_o)$$

$$\cdots \alpha_{S_{k+j}}(x_o) \quad . \tag{3.37}$$

The sequences $\{\alpha(x_o)\}$ and $\{\alpha^c(x_o)\}$ can be expressed in functions of Hermite polynomials as $\alpha_o(x_o) = 1-F(x_o)$ and $\alpha_r(x_o)=H_{r-1}(x_o)f(x_o)$, for $r=1,2,\ldots$; and $\alpha_o^c(x_o)=F(x_o)$ and $\alpha_r^c(x_o)=-H_{r-1}(x_o)f(x_o)$, for $r=1,2,\ldots$. In Equation (3.37), $I=\Sigma i$.

Let us define

$$\alpha_{S_1}^c(x_o)\cdots\alpha_{S_k}^c(x_o) = \pi^c(\alpha)$$

$$\alpha_{S_{k+1}}(x_o)\cdots\alpha_{S_{k+j}}(x_o) = \pi(\alpha) \quad ,$$

then

$$P(k^-,j^+) = \sum_{I=0}^{\infty} A(\rho,i)\pi^c(\alpha)\pi(\alpha) \quad . \tag{3.38}$$

For $I = 0$,

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

$$\pi^c(\alpha) = [\alpha_o^c(x_o)]^k = F^k(x_o)$$

$$\pi(\alpha) = [\alpha_o(x_o)]^j = [1 - F(x_o)]^j \quad \text{, so that}$$

$$P(k^-,j^+) = F^k(x_o)[1-F(x_o)]^j + \sum_{I=1,2,\ldots} A(\rho,i)\pi^c(\alpha)\pi(\alpha) \tag{3.39}$$

Equation (3.40) is an infinite series. However, in practice it is only necessary to include a finite number of terms of this series to compute numerical values of $P(k^-,j^+)$. A truncation of this series after $I=2$ implies that terms containing ρ_1^3 or higher powers of ρ_1 are neglected. For values of ρ_1 less than 0.30, the error introduced by this truncation is negligible. However for values of ρ_1 greater than or equal to 0.40, this truncation may introduce a significant error. In this case it is necessary to include more terms in Equation (3.40), and truncate the series at a higher value of I .

For values of ρ equal to or greater than 0.50, the truncation of this series after $I=3$ implies that terms containing ρ_1^4 or higher powers of ρ_1 are neglected. The error introduced in this case may be negligible. In other words, the higher ρ_1 the larger power m of ρ_1 should be included.

Distribution of N_1^+. This distribution can be obtained for any truncation level, provided the expressions are available for the joint probabilities $P(k^-,j^+)$, so that

$$P(N_1^+ \geq j) = P(j^+) + \sum_{k=1}^{\infty} P(k^-,j^+) \quad . \tag{3.40}$$

Probabilities $P(j^+)$, $P(1^-,j^+)$, and $P(k^-,j^+)$ for $k=2,3,\ldots$ may be evaluated by Equations (3.32), (3.34), and (3.39), respectively. Equation (3.40) combined with Equation (3.2) gives the probability mass function of N_1^+.

Distribution of N_1^-. By definition

$$P(N_1^- \geq j) = P(j^-) + \sum_{k=1}^{\infty} P(k^+,j^-) \tag{3.41}$$

where

$$P(j^-) = P\{x_1<x_o, x_2<x_o, \ldots, x_j<x_o\} \tag{3.42}$$

and

$$P(k^+,j^-) = P\{x_1>x_o, \ldots, x_k>x_o, x_{k+1}<x_o, \ldots, x_{k+j}<x_o\} \quad . \tag{3.43}$$

Consider

$$P(x_1 \leq x_o^c, \ldots, x_k \leq x_o^c, x_{k+1}>x_o, \ldots, x_{k+j}>x_o) = P_p(k^-,j^+) \quad ,$$

where x_o^c and p are defined by $P = F(x_o^c) = 1-F(x_o) = 1-q$. Because of the symmetry of the normal distribution,

$$P_q(k^+,j^-) = P_p(k^-,j^+) \quad . \tag{3.44}$$

The distribution of N_1^- is

$$P(N_1^- \geq j) = P_p(j^+) + \sum_{k=1}^{\infty} P_p(k^-,j^+) \quad , \tag{3.45}$$

and probabilities $P_p(j^+)$, $P_p(1^-,j^+)$, and $P_p(k^-,j^+)$ for $k=2,3,\ldots$ may be also evaluated by Equations (3.32), (3.34), and (3.39), respectively.

The probability mass function of N_1^- is then

$$P(N_1^- = j) = P(N_1^- \geq j) - P(N_1^- \geq j+1) \quad . \tag{3.46}$$

Joint distribution of N_1^+, N_1^-, N_2^+, N_2^-. By definition

$$P\{N_1^+=j_1, N_1^-=i_1, N_2^+=j_2, N_2^-=i_2\} = P(j_1^+,i_1^-,j_2^+,i_2^-,1^+)$$
$$+ P(i_1^-,j_1^+,i_2^-,j_2^+,1^-) \quad . \tag{3.47}$$

The result expressed in Equations (3.37), (3.38), and (3.39) can be extended to joint probabilities involved in Equation (3.47). In this case

$$P(j_1^+,i_1^-,j_2^+,i_2^-,1^+) = \sum_{I=0}^{\infty} A(\rho,i)\alpha_{S_1}\cdots\alpha_{S_{j_1}}\alpha^c_{S_{j_1+1}}\cdots\alpha^c_{S_{j_1+i_1}}$$
$$\cdot\alpha_{S_{j_1+i_1+1}}\cdots\alpha_{S_{j_1+i_1+j_2}}\alpha^c_{S_{j_1+i_1+j_2+i_2}}$$
$$\cdot\alpha_{S_{j_1+i_1+j_2+i_2+1}} \quad . \tag{3.48}$$

A completely analogous result is obtained for $P(i_1^-,j_1^+,i_2^-,j_2^+,1^-)$. If the places of j_1^+ and j_2^+ are interchanged in Equation (3.48), the result does not vary because all possible permutations of $\{i\}$ are considered. This implies that the marginal distribution of N_1^+ and N_2^+ are identical. This also holds for N_1^- and N_2^-.

Distributions of N_k^+ and N_k^-. The approach in the previous subsection also applies to the sequence $\{N_i^+ \,;\, i=1,\ldots,k\}$; the result is a sequence of identically distributed random variables, with distribution function of Equation (3.40). The same is true for the sequence $\{N_i^- \,;\, i=1,\ldots,k\}$. It is also a sequence of identically distributed random variables with distribution function of Equation (3.45). For a more detailed analysis of this chapter and particularly for a special treatment of the case of a truncation level of the median, the reader is referred to Saldarriaga [6].

Chapter IV

RUNS OF STATIONARY DEPENDENT GAUSSIAN PROCESSES

4.1 First-Order Linear Autoregressive Process. 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-order linear autoregressive processes of normal variables are considered only, because of its broad application in hydrology, and its simplicity.

Suppose that the process $\{x_i\}$ is defined by the recurrence relation

$$x_i = \rho x_{i-1} + \varepsilon_i \quad , \tag{4.1}$$

for $E\varepsilon_i=0$ and $D\varepsilon_i=\text{var } \varepsilon_i=1$. It can be solved formally by successive substitutions, and be rewritten as

$$x_i = \sum_{j=0}^{\infty} \rho^j \varepsilon_{i-j} \quad . \tag{4.2}$$

Then for $Ex_i = 0$,

$$Dx_i = \text{var } x_i = \frac{1}{1-\rho^2} \quad , \tag{4.3}$$

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 Equation (4.2) that the first-order linear autoregressive model is a moving average scheme of an infinite extent, with monotonically decreasing weights, $\rho, \rho^2, \rho^3, \ldots$. Therefore, if $\{\varepsilon_i\}$ are normal variables, $\{x_i\}$ as a linear combination of $\{\varepsilon_{i-j}\}$ is also a sequence of normal variables and is a Gaussian process.

4.2 Probability Mass Function, and Moments of Run-Lengths N^+ and N^-. The following relation holds between the probability mass $P(N^+=j)$ of a given run-length $N^+=j$, and the probability distribution functions $P(N^+\geq j)$ of run-lengths N^+ ,

$$P(N^+= j) = P(N^+\geq j) - P(N^+\geq j + 1) \quad . \tag{4.5}$$

By definition, the first moment of N^+ is

$$EN^+ = \sum_{j=1}^{\infty} jP(N^+=j) \quad . \tag{4.6}$$

Substituting Equation (4.5) into Equation (4.6) gives

$$EN^+ = \sum_{j=1}^{\infty} P(N^+ \geq j) \quad . \tag{4.7}$$

By definition, the second moment of N^+ is

$$E(N^+)^2 = \sum_{j=1}^{\infty} j^2P(N^+ = j) \quad . \tag{4.8}$$

Substituting Equation (4.5) into Equation (4.8) gives

$$E(N^+)^2 = \sum_{j=1}^{\infty} (2j-1)\ P(N^+ \geq j) \quad . \tag{4.9}$$

In general, the r-th moment of X is

$$E(N^+)^r = \sum_{j=1}^{\infty} j^rP(N^+=j) = \sum_{j=1}^{\infty} [j^r-(j-1)^r]P(N^+\geq j) . \tag{4.10}$$

Equations (4.5) to (4.10) analogously apply to N^-.

4.3 Properties of Total Run-Length, $N=N^++N^-$. Statistical properties of this parameter are rather complex because 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 the basis of a univariate distribution of N^+ and N^- is the mean,

$$EN = EN^+ + EN^- \quad . \tag{4.11}$$

This equation can be written also in the form

$$EN(q) = EN^+(q) + EN^-(q) \quad , \tag{4.12}$$

where $q = F(x_o)$. Because

$$EN^-(q) = EN^+(p) \tag{4.13}$$

due to the symmetry of normal distributions,

$$EN(q) = EN^+(q) + EN^+(p) \quad . \tag{4.14}$$

4.4 General Procedure for Evaluating Properties of Runs. Statistical properties of run-lengths of stationary Gaussian processes can be evaluated for any truncation level, x_o , by using the relations obtained in this chapter and in Chapter III. The general procedure of this evaluation can be made in four steps. Equation numbers to be used for various expressions in these four steps are given in Table 4.1. These steps are:

1. Starting at an arbitrary time, probabilities $P(j^+)$ for at least the first j values of X, being above the truncation level specified by q, are first computed. For these probabilities, 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; $P(j^+)$ are not probabilities of runs but are needed for their computation.

2. Starting at an arbitrary time, probabilities $P(k^-,j^+)$ for the first k values of X, being below and the j subsequent values of X being above the truncation level specified by q, are next computed. For these probabilities, nothing is specified about the values of X preceding or following the occurrence of these $k+j$ values. They are not probabilities of runs but they are necessary for the computation of these probabilities or runs.

3. The probability distribution and the probability mass function of the run-length are calculated by using probabilities obtained in the two previous steps.

4. Moments of run-lengths may then be calculated, when needed, by using the computed probability distribution. It might be noted that the probabilities $P(j^+)$ and $P(k^-,j^+)$ also have practical meaning by themselves, besides being used for the computations of probabilities of runs. In fact, $P(j^+)$ is associated with the probability that starting at an arbitrary year at least the first j years are wet. Similarly, $P(k^-,j^+)$ is associated with the probability that starting at an arbitrary year the first k years are dry and are followed by at least j wet years. This analogously applies to $P(j^-)$ and $P(k^+,j^-)$.

TABLE 4.1

EQUATIONS FOR THE EVALUATION OF PROPERTIES OF RUNS

Step	Expression	Equation
1	$P(1^+)$	(3.13)
	$P(2^+)$	(3.14)
	$P(j^+)$, $j \geq 3$	(3.33)
2	$P(1^-,j^+)$	(3.34)
	$P(k^-,j^+)$	(3.39)
3	$P(N^+ \geq j)$	(3.40)
	$P(N^+ = j)$	(3.2)
4	EN^+	(4.7)
	$E(N^+)^2$	(4.9)
	$E(N^+)^r$	(4.10)

4.5 Probabilities of the Non-Normal Case. Let $F_1(y)$ be the distribution function of a non-normal variable Y, while $F_2(x)$ for the variable X is normal. In this case, probabilities of multivariate events are

$$P_r(k^-,j^+;y_o)=P(Y_1\leq y_o,\ldots,Y_k\leq y_o,Y_{k+1}>y_o,\ldots,Y_{k+j}>y_o) \tag{4.15}$$

and

$$P_X(k^-,j^+;x_o)=P(X_1\leq x_o,\ldots,X_k\leq x_o,X_{k+1}>x_o,\ldots,X_{k+j}>x_o) , \tag{4.16}$$

respectively, for $F_1(y)$ and $F_2(x)$. For strictly increasing distribution functions, $F_1(y)$ and $F_2(x)$, it is always possible to find such unique values y_o and x_o that satisfy

$$F_1(y_o) = F_2(x_o) = q \quad . \tag{4.17}$$

If the probabilities of multivariate events are equal, with Equation (4.17) satisfied, then

$$P_Y(k^-,j^+;y_o) = P_X(k^-,j^+;x_o) \quad . \tag{4.18}$$

Equation (4.18) implies that the joint probabilities are dependent only on the probability level q for given x_o and y_o and not on the underlying distribution.

As an example, consider the case of $\{y\}$ lognormally distributed with

$$P(Y_i \leq y_o) = F_Y(y_o) = q \quad . \tag{4.19}$$

By definition of the log-normal distribution the following relation holds,

$$F_Y(y_o) = F_X(\ell n x_o) = q \quad . \tag{4.20}$$

The probability $P_y(k^-,j^+)$ can be expressed in terms as

$$P_Y(k^-,j^+)=P(Y_1\leq y_o,\ldots,Y_k\leq y_o,Y_{k+1}>y_o,\ldots,Y_{k+j}>y_o)$$

$$=P(\ell nY_1\leq \ell n x_o,\ldots,\ell nY_k\leq \ell n x_o,\ell nY_{k+1}\leq \ell n x_o,\ldots,\ell nY_{k+j}\leq \ell n x_o),$$

$$=P(X_1\leq \ell n x_o,\ldots,X_k\leq \ell n x_o,X_{k+1}\leq \ell n x_o,\ldots,X_{k+j}\leq \ell n x_o) \quad .$$

This last expression is $P_q(k^-,j^+)$, so that

$$P_Y(k^-,j^+) = P_q(k^-,j^+) \quad . \tag{4.21}$$

It follows that all properties of run-lengths depend only on q.

4.6 <u>Properties of Runs of the First-Order, Autoregressive Linear Process</u>. <u>Positive Runs</u>. Properties of run-lengths of the first-order linear autoregressive model are computed on a digital computer by using the appropriate equations of Table 4.1. Five values of the probability truncation level $q = F(x_0)$, and five values of ρ were used, as shown by Tables 4.2 and 4.3. For each possible combination of q and ρ , the probabilities $P(j^+)$ and $P(k^-,j^+)$ were first calculated for $j=1,2,\ldots,10$ and $k=1,2,\ldots,10$. Theoretically, probabilities of runs are given by an infinite series of $P(k^-,j^+)$, as shown by Equation (3.40). Actually these terms become very small for sufficiently large values of k and only a finite number of terms need to be considered. Tables of computed probabilities and parameters of runs of the first-order linear autoregressive process are given in Appendix A; they are: $P(k^-,j^+)$, $P(N^+ \geq j)$, and $P(N^+ = j)$, with $k=1,2,\ldots,10$ and $j=1,2,\ldots,10$. The parameters of distributions are EN^+ , var N^+ . Figures 4.1 and 4.2 give these probability distributions of positive run-lengths of the first-order linear autoregressive process for values of ρ , varying from 0 to 0.50, with increments of 0.10, and for values of q , from 0.30 to 0.70, with increments of 0.10. In these figures, points are used for computed probabilities of distributions of discrete (integer) random events, while curves are used for the visualization of these distributions. Tables 4.2 and 4.3 summarize the results for the first two moments of N^+ .

TABLE 4.2

EXPECTED VALUES OF N^+ OF THE FIRST-ORDER LINEAR AUTOREGRESSIVE PROCESS

q \ ρ	0	.1	.2	.3	.4	.5
.3	3.33	3.45	3.65	3.87	4.29	4.69
.4	2.50	2.66	2.84	3.05	3.32	3.60
.5	2.00	2.14	2.29	2.47	2.66	2.88
.6	1.67	1.78	1.90	2.05	2.20	2.33
.7	1.43	1.51	1.61	1.72	1.86	2.00

TABLE 4.3

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

q \ ρ	0	.1	.2	.3	.4	.5
.3	7.77	7.77	8.74	9.83	12.78	15.46
.4	3.75	4.32	5.10	5.98	7.34	8.66
.5	2.00	2.42	2.94	3.53	4.24	4.98
.6	1.11	1.38	1.71	2.08	2.50	2.94
.7	0.61	0.79	0.99	1.22	1.47	1.72

<u>Negative Runs</u>. Properties of negative runs are found by means of the equation $P_q(k^+,j^-)=P_p(k^-,j^+)$. It then follows $EN^-(q)=EN^+(p)$, var $N^-(q)=$ var $N^+(p)$, and, in general, $E[N^-(q)]^r = E[N^+(p)]^r$.

<u>Total Runs</u>. The process of the total run-length has been defined as $\{N_j\}=\{N_j^+ + N_j^-\}$. Then $EN_j = EN_j^+ + EN_j^-$ is its expected value. Figure 4.3 shows three different sets of curves: EN_j^+ , EN_j^- , and EN_j plotted against q for ρ values from 0 to 0.5 with increments of 0.1.

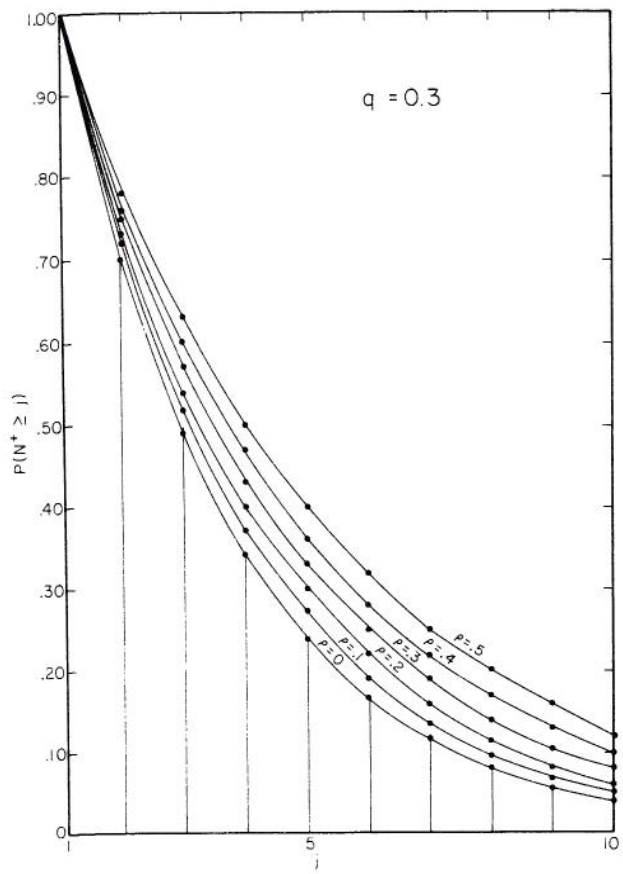


Fig. 4.1 Probability distributions of positive run-lengths of the first-order linear autoregressive process for $q = 0.3$

4.7 <u>Properties of runs of the first-order linear autoregressive Gaussian process obtained by the data generation method</u>. Probabilities of run-lengths are given by an infinite series as shown by Equation (3.40). At the same time, each term contained in this series is given by an expansion of the tetrachoric series

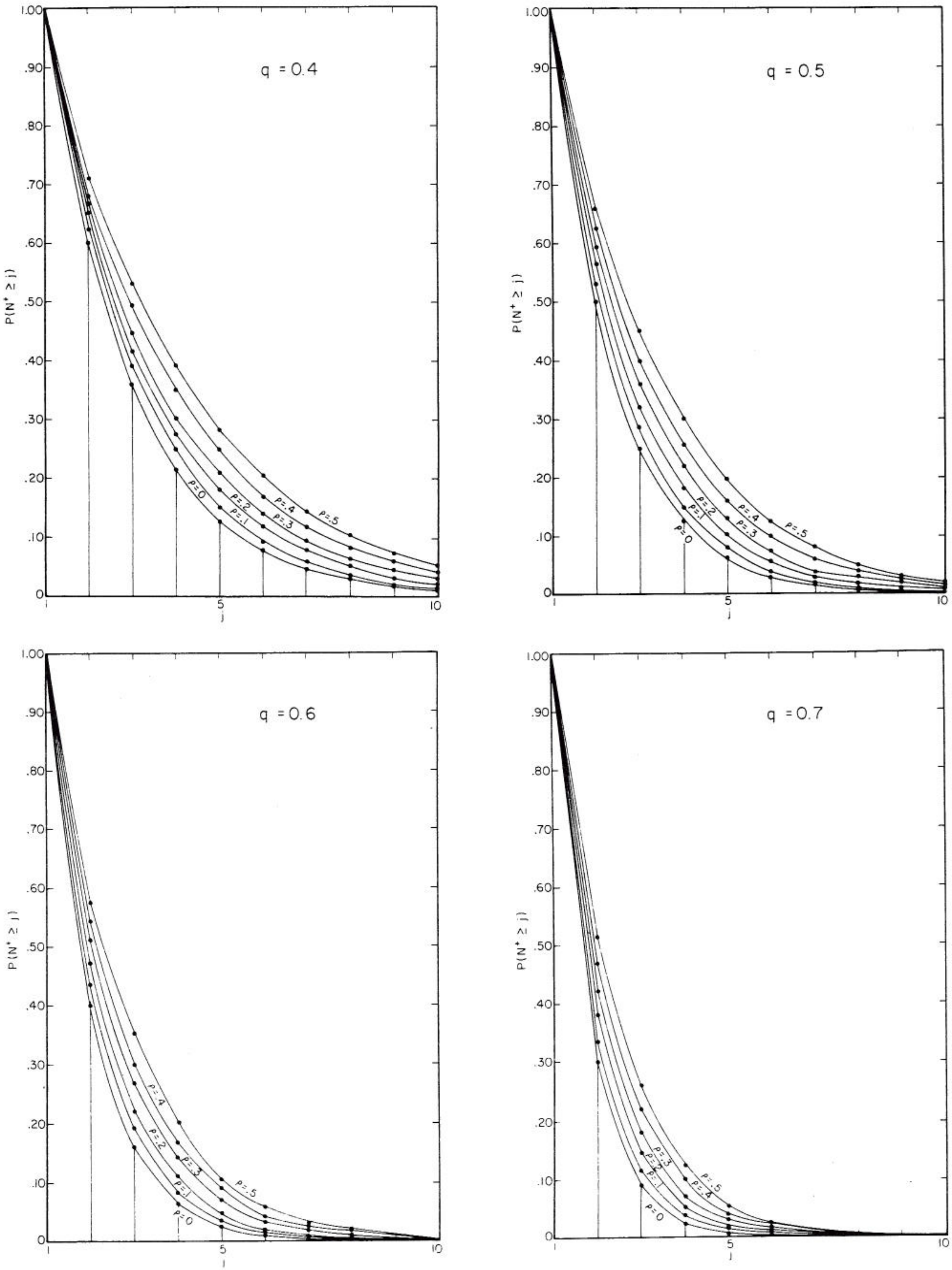


Fig. 4.2 Probability distributions of positive run-lengths of the first-order linear autoregressive process for $q = 0.4, 0.5, 0.6, 0.7$

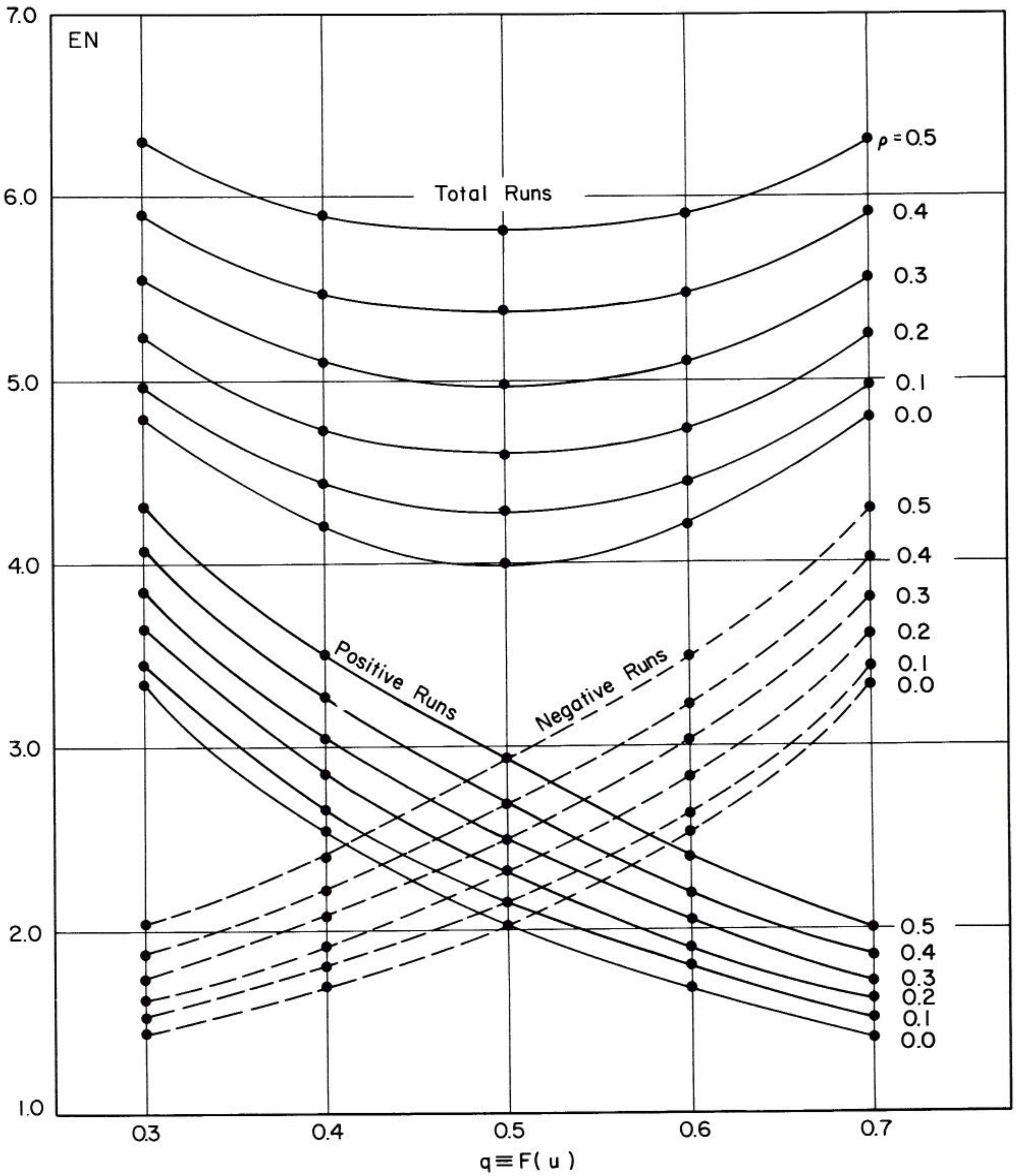


Fig. 4.3 Mean positive, negative, and total run-lengths of the first-order x_o linear autoregressive process

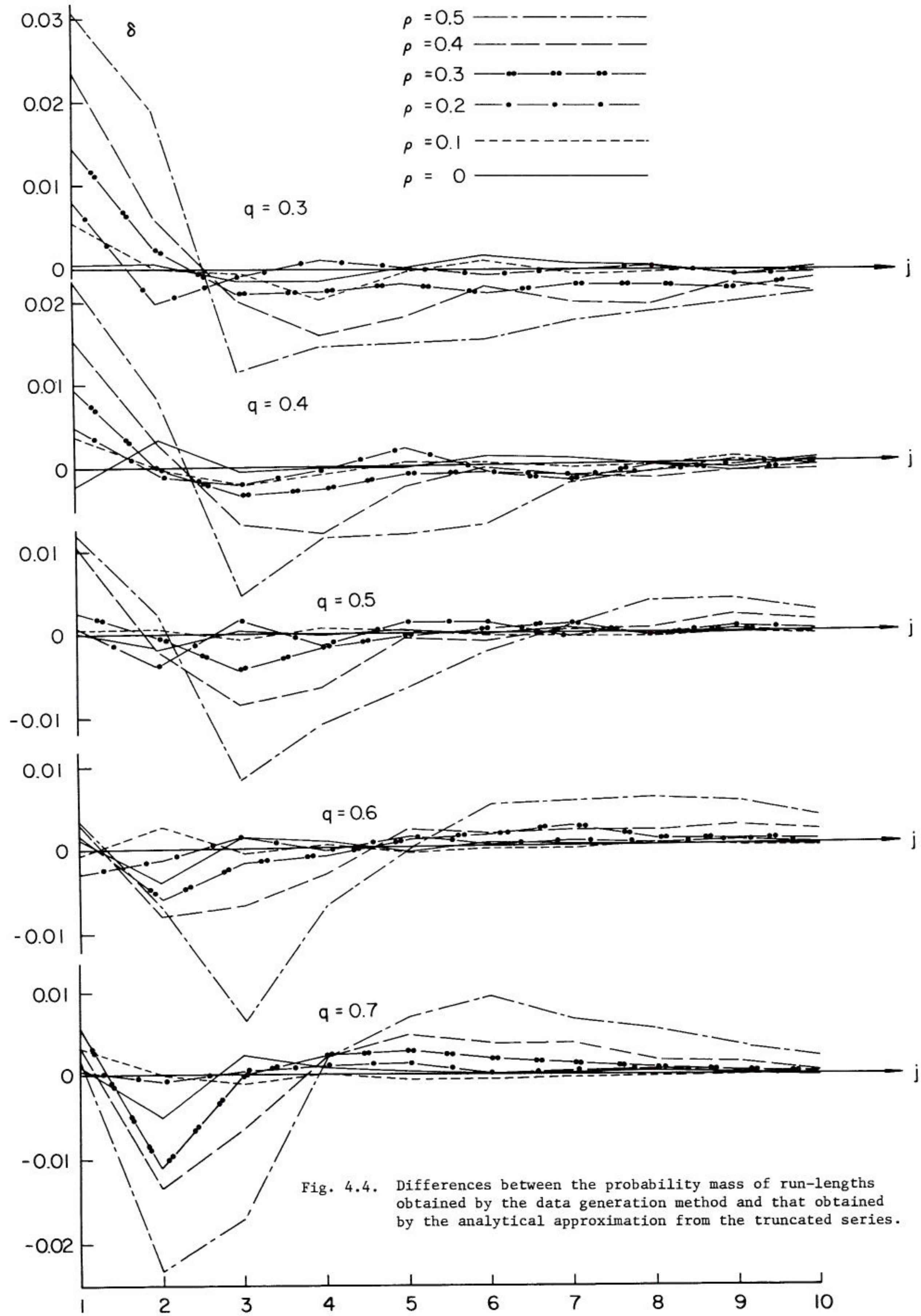


Fig. 4.4. Differences between the probability mass of run-lengths obtained by the data generation method and that obtained by the analytical approximation from the truncated series.

which is also an infinite series. Both of these series are convergent. However, the rate of convergence of the series used for calculating $P(j^{+})$ and $P(k^{-},j^{+})$, given by Equation (3.33) and Equation (3.39) respectively, varies with q, ρ, k, and j. The rate of convergence of the series given by Equation (3.40) varies also with q, ρ, and j. In other words, the problem of the rate of convergence of each of these two series is a complex mathematical problem in itself, and to the writer's knowledge, to date, it has never been solved.

In order to assess the accuracy of the computed probabilities of run-lengths, obtained in this study by using the truncated series, the data generation method (Monte Carlo method) is used to experimentally compute the properties of run-lengths, and to compare them with the probabilities obtained from the truncated series. The procedure is the following.

(a) Normal random numbers were generated following the first-order autoregressive model: $x_i = \rho x_{i-1} + \varepsilon_i$, where ε_i are standard normal random numbers and ρ is given the values 0, 0.1, 0.2, 0.3, 0.4, and 0.5.

(b) The probability truncation level, q, was given the values 0.3, 0.4, 0.5, 0.6, and 0.7, with corresponding x_o values equal to -0.524002, -0.252933, 0, 0.252933, and 0.524002, respectively.

(c) First a value of ρ was selected, then for each value of q, x_i's are generated until 30,000 positive run-lengths were obtained. Absolute frequencies are calculated for the runs having run-lengths 1, 2, 3, ..., and the probability mass of each run-length is estimated by computing the relative frequency as the absolute frequency divided by the sample size of 30,000.

(d) Then another value of ρ is selected, and step (c) is repeated until all values of ρ are used.

A comparison of the probability mass of run-length, obtained from the truncated series, as an analytical approximation of the data generation method and probabilities obtained by the analytical approximation from the truncated series are shown in Figure 4.3.

Chapter V

APPLICATION OF RUN-LENGTHS TO INVESTIGATION OF SERIES

5.1 Introduction. The hydrologist is concerned with two basic types of variables, namely serially independent and serially dependent variables. He is interested in first testing whether they are stationary or not. If they are stationary, further tests are used to determine the goodness of fit of various mathematical models of serial dependence. The mathematical model is a representation of the process. It reflects statistical characteristics of the sequence in terms of parameters related to the physical properties of the system.

Before discussing the application of run-lengths to the investigation of series, the application of autocorrelation coefficients and the variance densities to the investigation of series is briefly reviewed. This shows an analogy, and points to the necessity of carrying out various tests.

In the case of investigation of series by autocorrelation, the parameters involved are the serial correlation coefficients, r_k, as estimates of population coefficients, ρ_k. A comparison of the computed sample correlogram $r_k = f(k)$ is made with correlograms of theoretical models, or with the expected correlograms, provided these models are good after the model parameters are estimated. This comparison allows making inference about the mathematical structure of this dependence. For an independent series of size N, $Er_k = -1/(N-k)$, var $r_k = (N-k-1)/(N-k)^2$, and the confidence limits at 95% probability level for the distribution of r_1 are given by $(r_1)_{1,2} = -1/N \pm 1.96 [(N^3 - 3N^2+\Delta)/N^2(N^2-1)]^{\frac{1}{2}}$. For r_k, N is replaced by N-k+1, or by similar expressions [2]. Figure 5.1 shows the expected correlogram and the 95% tolerance limits of an independent series with N = 30.

In order to test the structure of an observed series, the null hypothesis, that the observed time series is an independent sequence, is used. If the value of r_k fall within the 95% tolerance interval, this hypothesis is accepted. Or if 5% of k values of r_k are outside the limits, but 95% are inside, the hypothesis is accepted; otherwise it is rejected.

In the case of investigation of series by spectral analysis, the parameters involved are the spectral densities $v_f = \Psi(f)$, with the frequencies f. These v_f values are estimates of population spectral densities, ν_f. Again a comparison of the observed variance density spectrum with spectra of theoretical models permits an inference about the matheamtical structure of dependence in a series. For a discrete series with equal intervals Δt, in the case of time series, the maximum frequency is $f_{max} = 1/2\Delta t$, and for finite series of size N the minimum frequency is $f_{min} = 1/N\Delta t$, which is usually extended to f = o. The spectrum has the property that the area under the variance density graph represents the total variance of the variable. The expected spectrum of an independent stochastic series is a straight horizontal line between f_{min} and f_{max}, or between o and $1/2\Delta t$. 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 $\nu = 2N/m - \frac{1}{2}$, where m is the number of serial correlation coefficients used in computing the variance densities. For N = 30, m = 4, and $\Delta t = 1$, the number of degrees of freedom is $\nu = 7$. The maximum frequency is 0.50, assuming $f_{min} = 0$. The mean variance density is $\overline{v}_f = 2\sigma^2$. Then $\nu \overline{v}_f f_{max}/\sigma^2$ is a χ^2_ν random variable. The 95% tolerance limits are

$$\chi^2_{0.025} < \frac{\nu v_f f_{max}}{\sigma^2} < \chi^2_{975} .$$

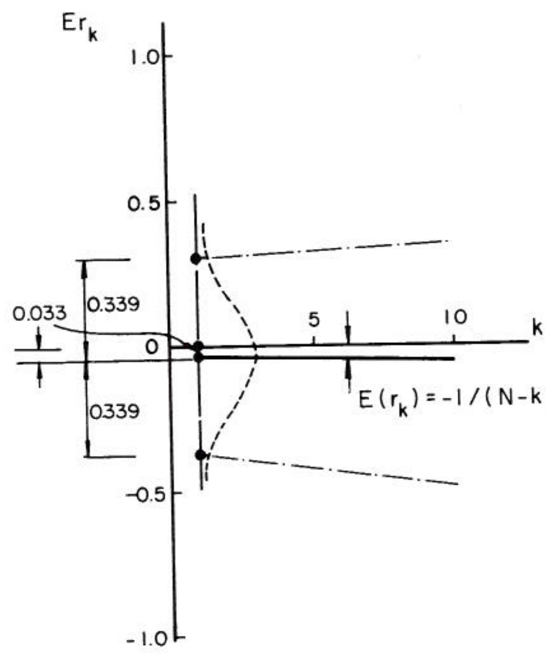


Fig. 5.1 Expected correlogram of an independent series, of the sample size N = 30, with tolerance limits at the 95% level.

For $\nu = 7$, $\chi^2_{0.025} = 1.69$, and $\chi^2_{0.975} = 16.0$, the 95% confidence limits are $0.483\sigma^2 < v_f < 4.571\sigma^2$. Figure 5.2 shows the expected values and the 95% tolerance limits for v_f, for independent variables.

5.2 Using Runs for Investigation of Series. In the use of investigation of series by using runs, the basic parameter selected here is the run-length, selected for the following reasons:

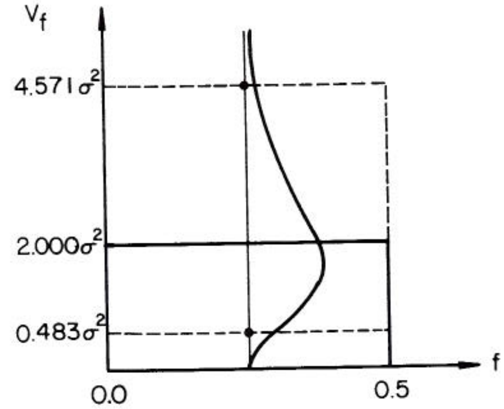


Fig. 5.2 The expected values and the 95% tolerance limits, of variance densities in a spectrum of an independent series with the sample size N = 30.

(a) If a given series is cut at many truncation levels, and for each level the sequences of positive and negative run-length are obtained, theoretically it is possible to reproduce the original time series, at least at a finite number of points, by using these sequences. The larger 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.

(b) Properties of run-lengths based on a probability truncation level are distribution-free, both for independent variables and linearly dependent processes. This is an important property, because the results obtained for Gaussian processes can also be applied to other types of stationary processes.

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

(d) A parallel technique to autocorrelation analysis and analysis by the variance density spectrum can be developed for run-lengths to investigate hydrologic series.

(e) A comparison of properties of observed run-lengths with the corresponding properties of run-lengths of population theoretical models is similar to other techniques of investigation. The mean positive run-length $\overline{N}^+$, the mean negative run-length $\overline{N}^-$, and the mean total run-length $\overline{N}$, are attractive parameters for this comparison. Because $\overline{N}$ contains information of both $\overline{N}^-$ and $\overline{N}^+$, this parameter $\overline{N}$ is used in this study as a basic parameter for investigating hydrologic time series by runs. The parameter $\overline{N}^+$ is also used as the alternative.

The technique by runs as developed in this study compares the statistic $\overline{N}_q$ as a function of q of observed time series with the expected value $E(N_q)$ of the total run-length of theoretical models, or the statistic $\overline{N}^+_q$ is compared with its expected value $E(N^+_q)$.

5.3 Properties of Run-Length for Sequence 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 also a sequence of independent identically distributed, random variables.

For a given q,

$$\overline{N}_k = \frac{N_1 + \ldots + N_k}{k} \quad , \tag{5.1}$$

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

$$E\overline{N}_k = EN \tag{5.2}$$

$$\text{var } \overline{N}_k = \frac{\text{var } N}{k} \quad . \tag{5.3}$$

Substituting Equations (2.22) and (2.23) into Equations (5.2) and (5.3) gives, for a given q and $p = 1-q$,

$$E\overline{N}_k = \frac{1}{pq} \tag{5.4}$$

$$\text{var } \overline{N}_k = \frac{p^3+q^3}{kp^2q^2} \quad . \tag{5.5}$$

This result holds when k is a fixed number. In the case the series length n = being a fixed value, the number k of total runs in the interval (o,n) is a random variable, k(n), and the mean $\overline{N}_{k(n)}$ is

$$\overline{N}_{k(n)} = \frac{N_1 + \ldots + N_{k(n)}}{k(n)} \quad .$$

Feller [15] shows that the ratio k(n)/n is asymptotically normal with the mean equal to the mean recurrence time of the completion of a total run. It converges in probability to a positive-valued random variable. In virtue of the central limit theorem for the sum of a random number of independent random variables [31], the result obtained for $\overline{N}_k$ also holds for $\overline{N}_{k(n)}$.

Table 5.1 gives values of the mean and variance of N^+, N^-, and N, respectively, for a range of values of q between 0.10 and 0.90. Figure 5.3 shows a graph of EN^+, EN^-, and EN versus q for the independent case. It is apparent from this graph that these functions are symmetrical about the line q = 0.5.

5.4 Run-Length Test for Stationary Independent Variables. Properties of $\overline{N}_k$ derived in the previous subchapter allows the construction of a test. The null hypothesis is that $\{X_n\}$ is a sequence of independent, identically distributed, variables, either for the original or "whitened" series. Then $\overline{N}_k$ is approximately

TABLE 5.1

PROPERTIES OF RUN-LENGTHS FOR INDEPENDENT IDENTICALLY DISTRIBUTED VARIABLES

q	N^+ Mean	N^+ Variance	N^- Mean	N^- Variance	N Mean	N 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

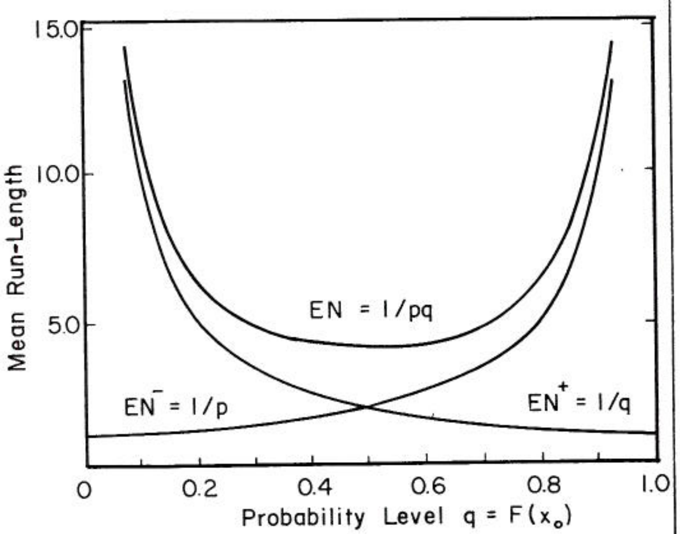


Fig. 5.3 Mean run-length for independent variables, with a common distribution, versus q.

normally distributed for large k with the mean and the variance given by Equations (5.4) and (5.5). At the 1-α tolerance level, the region of acceptance of the hypothesis is, for a two-sided test,

$$EN - t_{\alpha/2}\ (\text{var}\ \overline{N}_k)^{\frac{1}{2}} \le \overline{N}_k \le EN + t_{\alpha/2}\ (\text{var}\ N_k)^{\frac{1}{2}}, \quad (5.6)$$

or

$$\frac{1}{pq} - \frac{t_{\alpha/2}}{pq}\left(\frac{p^3+q^3}{k}\right)^{\frac{1}{2}} \le \overline{N}_k \quad \frac{1}{pq} + \frac{t_{\alpha/2}}{pq}\left(\frac{p^3+q^3}{k}\right)^{\frac{1}{2}} . \quad (5.7)$$

Now, for a median as the truncation level, or for $p = q = 0.50$,

$$EN = 4 \quad (5.8)$$

$$\text{var}\ N = 4, \quad \text{and} \quad \text{var}\ \overline{N}_k = \frac{4}{k} . \quad (5.9)$$

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

$$4 - \frac{3.92}{\sqrt{k}} \le N_k \le 4 + \frac{3.92}{\sqrt{k}} . \quad (5.10)$$

If $\overline{N}_k$ falls outside the limits of Equation (5.10) the hypothesis is rejected. The test is illustrated by Figure 5.4 for the case of the tolerance level 1-α = 1.95 and the truncation level being the median, q = 0.50.

For a truncation level $q \neq 1/2$, the tolerance limits are different than those given in Figure 5.4, as shown in Figure 5.5. For any value of q, the right-sided run-length test, for the 95% tolerance limit, is

$$\frac{1}{pq} + \frac{1^{t_\alpha}}{pq}\left(\frac{p^3+q^3}{k}\right)^{\frac{1}{2}} . \quad (5.11)$$

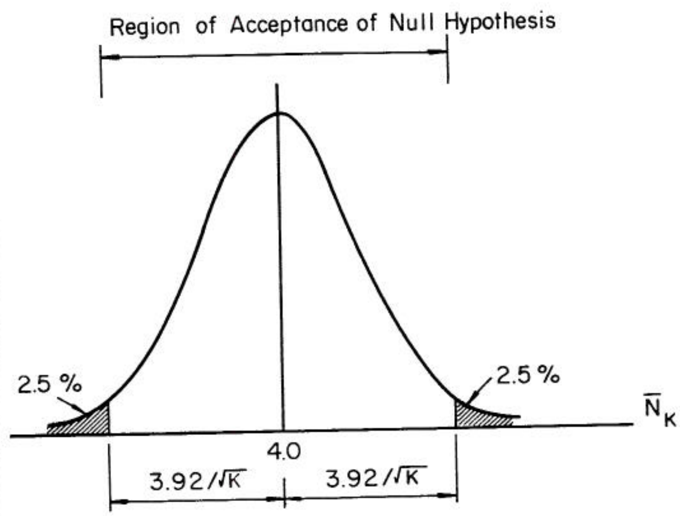


Fig. 5.4 Two-sided run-length test for the truncation level of the median, q = 0.50, for independent variables.

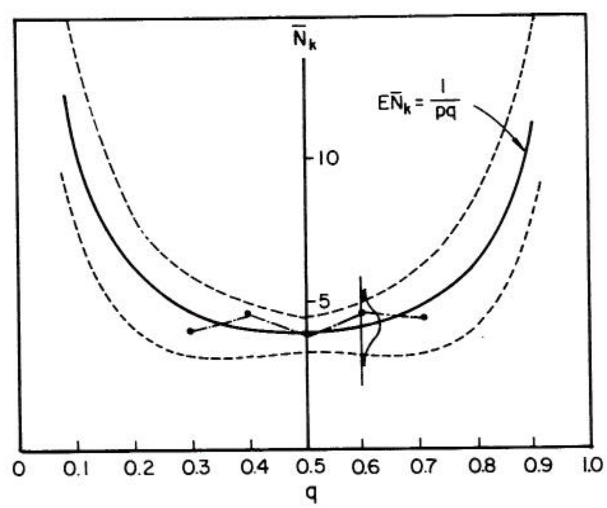


Fig. 5.5 Tolerance region for $\overline{N}_k$, of an observed independent time series.

The left-sided run-length test, for the 95% tolerance limit, is

$$\frac{1}{pq} + \frac{t_\alpha}{pq} \left(\frac{p^3+q^3}{k}\right)^{\frac{1}{2}} \quad . \tag{5.12}$$

Figure 5.6 shows a graph with $E\overline{N}$ and the 95% tolerance limits of $\overline{N}(q)$ for $0.2 \leq q \leq 0.8$ and $k = 10, 15, 20, 25, 30, 35$. The q values are given as abscissas, and the $\overline{N}(q)$ are given as ordinates in the upper graph. The k values are given by the ordinates downwards versus q in the lower graph. The upper graph shows a family of curves. The central curve is $E\overline{N}(q) = 1/pq$. The upper and lower sets of curves are the 95% tolerance limits of $\overline{N}(q)$. This is a convenient plotting graph that can be used readily for the analysis of time series by the mean total run-length $\overline{N}(q) = f(q)$. The sample function $\overline{N}(q)$ is calculated from the observed time series and plotted in the upper graph. The sample function of k is plotted in the lower graph. Finally, by using the upper and lower families of curves of the upper graph and the k values of the lower graph, the upper and lower 95% tolerance limits of $\overline{N}(q)$ are drawn in the upper graph. If the whole sample function is confined inside the tolerance region, the analyzed series is considered as not being significantly different from an independent series at the 95% level.

5.5 Two-Levels Run-Length Test for Stationary Independent Variables. An alternative statistic to $\overline{N}(q)$ is also considered in this study, and is defined as

$$\overline{N}^*(q) = \frac{1}{k^+ + k^-}\left[\sum_{i=1}^{k^+} N_i^+(q) + \sum_{i=1}^{k^-} N_i^-(p)\right] = \frac{k^+\overline{N}^+(q) + k^-\overline{N}^-(p)}{k^+ + k^-} \tag{5.13}$$

For

$$k^* = \frac{k^+ + k^-}{2} \quad , \tag{5.14}$$

rewriting Equation (5.13) in the form

$$\overline{N}^*(q) = \frac{1}{2}\left[\frac{k^+}{k^*}\overline{N}^+(q) + \frac{k^-}{k^*}\overline{N}^-(p)\right] \tag{5.15}$$

and taking into account Equation (2.22) and Equation (4.13), the expected value of Equation (5.13) becomes

$$E\overline{N}^*(q) = \frac{1}{2}\left[\frac{k^+}{k^*}E\overline{N}^+(q) + \frac{k^-}{k^*}E\overline{N}^-(p)\right] = \frac{1}{q} \quad . \tag{5.16}$$

Figure 5.3 shows $EN^*(q)$ as a function of q for independent variables. The random variables $N_i^+(q)$ and $N_j^-(p)$ are independent for $q \geq 0.5$. Furthermore, it is assumed here that they are also independent for $0.2 \leq q \leq 0.5$. Equation (5.17) gives

$$\text{var } \overline{N}^*(q) = \frac{1}{4}\left[\left(\frac{k^+}{k^*}\right)^2 \text{var } \overline{N}^+(q) + \left(\frac{k^-}{k^*}\right)^2 \text{var } \overline{N}^-(p)\right] \quad . \tag{5.17}$$

However,

$$\text{var } \overline{N}^+(q) = \frac{\text{var } N^+(q)}{k^+} = \frac{p}{k^+q^2} \quad , \tag{5.18}$$

and

$$\text{var } \overline{N}^-(p) = \frac{\text{var } N^-(p)}{k^-} = \frac{p}{k^-q^2} \quad , \tag{5.19}$$

since $\{N^+\}$ and $\{N^-\}$ are sequences of independent variables with variances given by Equation (2.23). Substituting Equation (5.20) and Equation (5.21) into Equation (5.19) gives

$$\text{var } \overline{N}^*(q) = \frac{(k^+ + k^-)p}{4(k^*)^2q^2} \quad . \tag{5.20}$$

Finally, taking into account Equation (5.16),

$$\text{var } \overline{N}^*(q) = \frac{p}{2k^*q^2} \quad . \tag{5.21}$$

Since the central limit theorem applies under rather broad conditions, it is evident that it also applies in the case of Equation (5.17), so that $\overline{N}^*(q)$ is approximately normally distributed for large values of k^* . For a two-sided test, the $1-\alpha$ tolerance region for independent variables is

$$\frac{1}{q}\left(1 - t_{\alpha/2}\sqrt{\frac{p}{2k^*}}\right) \leq \overline{N}^* \leq \frac{1}{q}\left(1 + t_{\alpha/2}\sqrt{\frac{p}{2k^*}}\right) \quad . \tag{5.22}$$

Figure 5.7 shows a practical graph that can be used readily for the analysis of time series by using $\overline{N}^*(p)$, with $p = 1 - q$. This graph is constructed and used in a manner similar to the graph of Fig. 5.6. The difference is that the statistic $\overline{N}^*(q)$ is used instead of $\overline{N}(q)$.

The parameter $\overline{N}^*(p)$ is used instead of $\overline{N}^*(q)$ only with the purpose of having the curve $\overline{N}(p) = f(p)$ increasing from the left to the right, instead of from the right to the left which would be the case for $\overline{N}^*(q) = f(q)$.

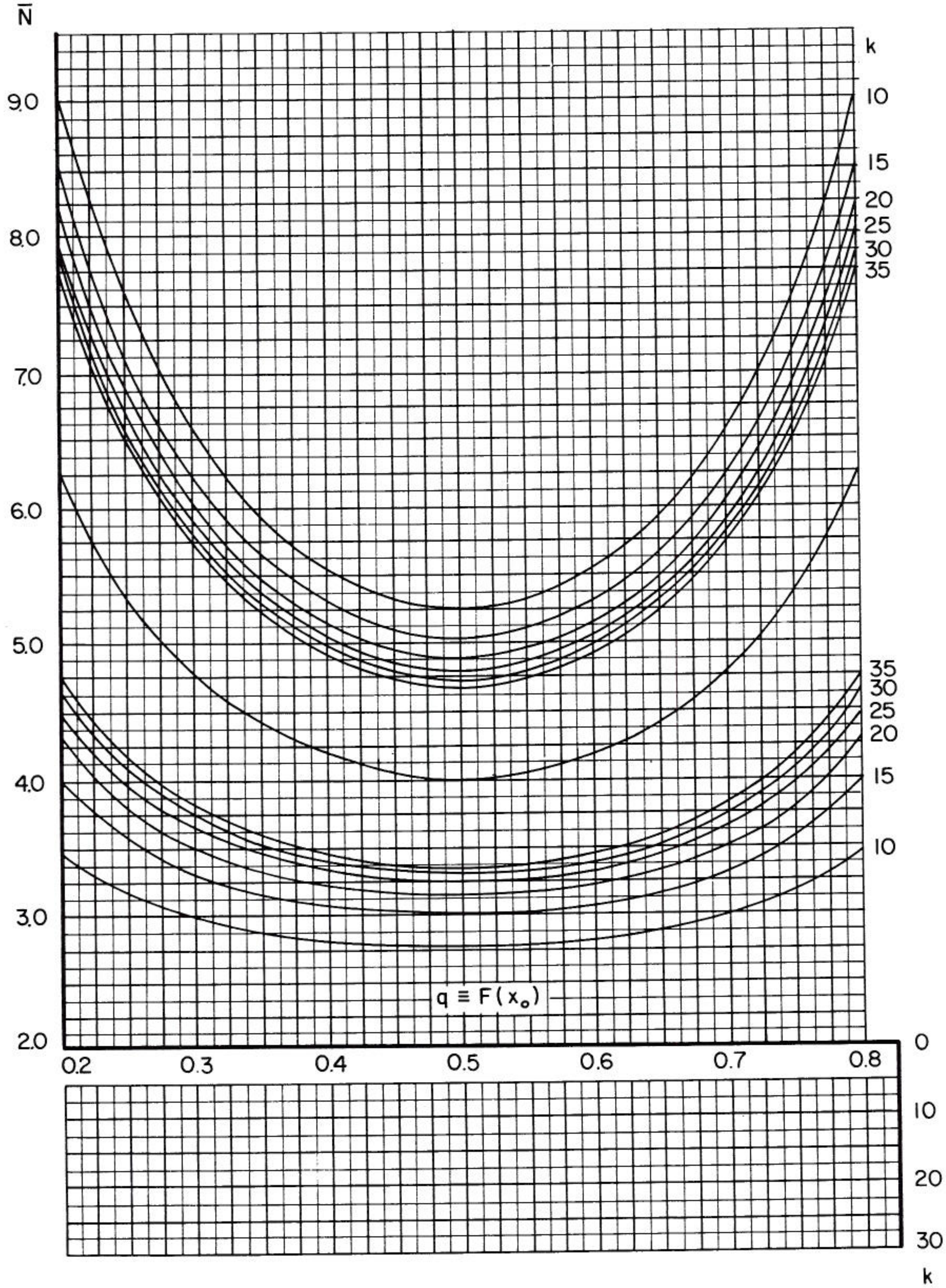


Fig. 5.6. Graph paper for plotting the mean total run-length, $\bar{N}(q)$. The upper graph serves for plotting $\bar{N}(q)$ versus q, the truncation level, as well as its tolerance limits at the 95% level versus q. The lower graph serves for plotting the number of total run-lengths k versus q, needed for computing the tolerance limits of the upper graph.

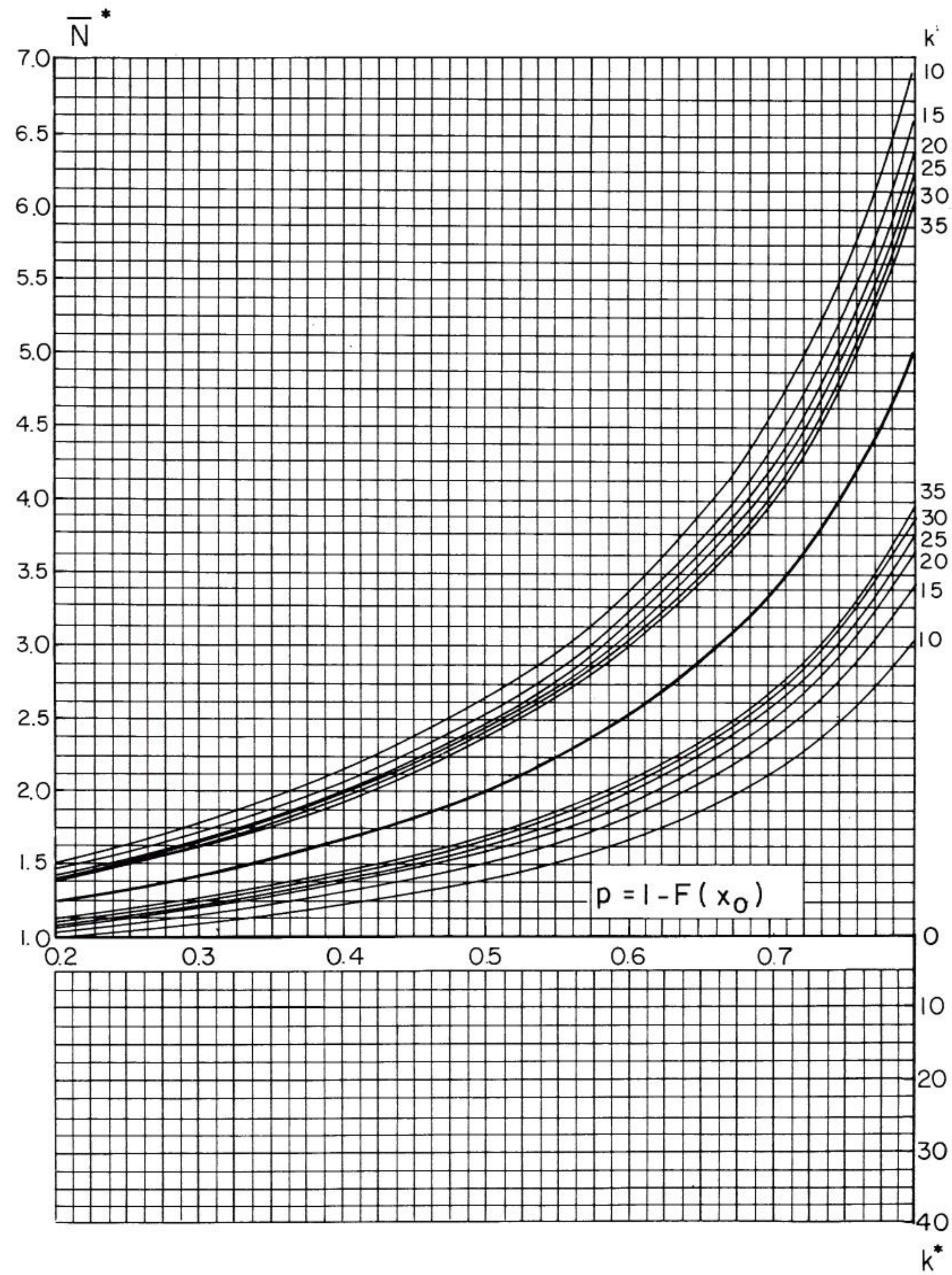


Fig. 5.7. Graph paper for plotting the mean run-length, $\bar{N}^*$, as defined in the text. The upper graph serves for plotting $\bar{N}^*(p)$ versus $p = 1 - q$, as well as its tolerance limits at the 95% level versus p. The lower graph serves for plotting the number k^*, as defined in the text, versus p, needed for computing the tolerance limits of $\bar{N}^*(p)$ of the upper graph.

Chapter VI

EXAMPLES OF INVESTIGATION OF STATIONARY HYDROLOGIC SERIES BY RUN-LENGTHS

6.1 Introduction. There is a two-fold application of runs to stationary hydrologic series. First, runs are used to investigate the structure of a series by testing whether or not a particular series is a sequence of independent random variables with a common distribution. If this is not the case, then the linear autoregressive models are assumed, the independent component in the original series is computed by using these models, and tests are performed to determine whether this component is a sequence of independent variables with a common distribution.

Three approaches are used for the investigation in this test:

1. Only the median truncation with $q = 0.50$ and the corresponding value $\bar{N}(0.50)$ are used. This case is analogous to using only the first serial correlation coefficient, r_1, in the autocorrelation investigation of whether a series is independent or dependent.

2. The total run parameter, $\bar{N}(q)$, is used, for various values of q, and a curve $\bar{N}(q)=f(q)$ is obtained.

3. The parameter $\bar{N}^*(p)$ is used as a function of p. $\bar{N}^*(p)$ is used instead of $\bar{N}^*(q)$ in order to have an increasing function as p increases.

Examples are given for each of these three approaches for investigating hydrologic series by run-lengths, using the supposed stationary time series of annual precipitation and annual river flow.

The second application of runs to stationary hydrologic series is for the prediction of durations of periods of surpluses and deficits for a variable. Once the structure of the series is known, i.e., whether it is stochastically independent or dependent of the first-order, or the second-order linear autoregressive models, or similar models, the derived properties of runs are used to make probability statements about durations of periods of surpluses and deficits. The truncation level in each case is specified by the probability q.

Chapter VI is concerned with first application, the investigation of series. Chapter VII treats the second application, the prediction of duration of surpluses and deficits (run-lengths) of stationary hydrologic series.

6.2 Application to Investigation of Annual Precipitation Series. Computed values of annual precipitation have random errors, systematic errors (inconsistency), and nonhomogeneity. Inconsistency is caused primarily by changes in instruments, methods of measurements, and so forth. Inconsistency comes mainly from two sources: moving a precipitation station a substantial distance, and changes in the environment around a station, such as growth of trees, building of houses, or any other major environmental change which affects the flow pattern of air around the station. Nonhomogeneity comes mainly from cloud seeding operations. Precipitation data must be considered, therefore, as often having relatively large random errors and inconsistency in their annual series [2, 27].

The presence of inconsistency and/or nonhomogeneity in an observed series implies that it does not come from a sequence of variables with a common distribution. When applying the run-length test to an observed precipitation series with inconsistency and/or nonhomogeneity in data, this null hypothesis may be rejected, though the series without these factors present may show the acceptance of a null hypothesis.

The presence of inconsistency and/or nonhomogeneity in data is reflected in the run-length test when $\bar{N}$ is greater than the expected value, $EN = 4$, for $q = 0.5$, and may be outside the 95% tolerance limits of the distribution of $\bar{N}$. It is therefore necessary to first remove inconsistency and/or nonhomogeneity in data.

On the other hand, if no significant inconsistency and/or nonhomogeneity are 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 the null hypothesis at a given level.

6.3 Examples of Investigation of Annual Precipitation Series by the Mean Run-Length of the Median. To show the method of run-lengths for the investigation of whether annual precipitation series are independent, identically distributed variables (null hypothesis) or not, five series are selected from around the United States. The application of run-length is developed on the assumption that the number of total runs is large. As a consequence, the observed series should be long enough to satisfy this assumption. For this reason, long precipitation series were selected.

The five annual precipitation series are:

1. Chico Experimental Station, California, for the period 1871-1965, $N = 95$;

2. Ord, Nebraska, 1896-1965, $N = 70$;

3. Natural Bridge N.M., Arizona, 1890-1960, $N = 71$;

4. Antioch F. Mills, California, 1879-1965, $N = 87$ years, and

5. Ravenna, Nebraska, 1878-1965, $N = 88$ years.

Table 6.1 gives the sequence $j = 1, 2, \ldots$ of runs, with run-lengths of N_j^+ and N_k^- for these five series and for $q = 0.50$. The incomplete first and last runs are included. They introduce a small bias, but their effects may be neglected for series with $N \geq 70$.

TABLE 6.1 RUNS OF FIVE ANNUAL PRECIPITATION SERIES

	Chico Station		Ord Station		Natural Bridge Station		Antioch F. Mills Station		Ravenna Station	
j	N_j^-	N_j^+	N_j^+	N_j^-	N_j^+	N_j^-	N_j^+	N_j^-	N_j^-	N_j^+
1	1	1	2	2	1	14	1	1	1	1
2	1	1	4	1	5	1	1	3	1	3
3	3	2	2	1	1	1	6	4	1	5
4	9	1	1	4	8	1	1	1	1	2
5	2	5	1	1	2	2	1	2	3	2
6	2	1	1	4	2	2	1	1	1	1
7	2	1	5	1	2	3	1	2	1	6
8	1	4	1	2	2	3	2	1	1	1
9	1	1	2	4	2	2	1	1	4	1
10	3	4	1	6	3	2	1	1	1	1
11	3	1	1	1	1	1	3	2	2	3
12	1	1	1	1	2	1	1	1	2	1
13	3	2	6	1	2	1	2	2	2	1
14	7	4	1	3	1	1	2	3	1	1
15	1	6	1	2			1	3	1	1
16	2	1	1	1			3	4	1	1
17	1	3	1	2			3	1	2	1
18	1	2					2	1	1	1
19	1	2					1	2	4	1
20	1	1					3	1	3	1
21	1	2					2	1	2	1
22							2	1	1	1
23							1	1	5	6
24							2	1		
$\sum$	47	46	32	37	34	35	44	41	42	43

Table 6.2 gives the following parameters of these $\overline{N}^+$, $\overline{N}^-$, $\overline{N} = \overline{N}^+ + \overline{N}^-$, and the two tolerance limits on the 95% level about $\overline{N}$. These limits are computed by

$$\overline{N}_{1,2} = EN \pm \frac{2t_{\alpha/2}}{\sqrt{k}} , \tag{6.1}$$

where $EN = 4$ for $q = 0.50$ of independent, identically distributed variables, $\alpha/2$ is the one-tail confidence probability level, $t_{\alpha/2}$ is the normal standard variate for the given $\alpha/2$, and k is the largest number j of run-lengths in Table 6.1, or $k = j_{max}$. Using the 95% tolerance level and the two-sided test, $\alpha/2 = 0.025$, then $t_{\alpha/2} = 1.96$.

In the above case, the tolerance limits are

$$\overline{N}_{1,2} = 4.00 \pm \frac{3.92}{\sqrt{k}} . \tag{6.2}$$

Table 6.2 shows that all five computed $\overline{N}$-values of the precipitation series are confined within the tolerance limits $\overline{N}_1$ and $\overline{N}_2$.

For a stronger test, with $\alpha = 80\%$ level, and $t_{\alpha/2} = 1.28$, all computed $\overline{N}$-values of five stations except Natural Bridge, which has a nonhomogeneous series, are still within these new limits, as shown in the last two columns of Table 6.2.

TABLE 6.2 PROPERTIES OF RUN-LENGTHS OF THE FIVE ANNUAL PRECIPITATION SERIES

					α = 95%		α = 80%	
Station	$\overline{N}^+$	$\overline{N}^-$	$\overline{N}$	k	$\overline{N}_1$(95%)	$\overline{N}_2$(95%)	$\overline{N}_1$(80%)	$\overline{N}_2$(80%)
Chico	2.190	2.238	4.428	21	4.858	3.142	4.560	3.440
Ord	1.882	2.176	4.058	16	4.980	3.020	4.640	3.360
Natural Bridge	2.429	2.500	4.929	14	5.050	2.950	4.685	3.315
Antioch F. Mills	1.833	1.708	3.541	24	4.800	3.200	4.522	3.478
Ravenna	1.826	1.870	3.696	23	4.820	3.180	4.535	3.465

It can be concluded that four out of the five annual precipitation series are independent variables, as it relates to the use of total run-length parameter, $\overline{N}$, for the median truncation level with $q = 0.50$, as the investigation parameter.

6.4 Application to Investigation of Annual River Flows Series. River flow essentially integrates the precipitation received over large areas, but it also includes the effects of evaporation and storage as other important physical factors. The water carryover from year to year, and especially the change in this carryover from one year to another, usually introduces time dependence into the sequence of annual flows.

To investigate the annual flow series either the original or the whitened series is used. By applying the same procedure to the original annual river flow series, as for the annual precipitation series, the hypothesis that the series is independent is accepted or rejected. If the hypothesis is rejected, the dependence models are assumed and the series is whitened. Then the same procedure used for the annual precipitation series is applied to the whitened series of annual river flow. If the hypothesis of the whitened series being independent is accepted, the postulated dependence model is also accepted. Equations (6.1) and (6.2) are used for these investigations and tests just as they are used for the annual precipitation series.

6.5 Examples of Investigation of Annual River Flows Series by the Mean Run-Length of the Median. Investigations and tests are limited to long series for the same reason they are limited in the case of annual precipitation series of annual river flows with long records [1] were selected as examples. They are:

1. The Mississippi River at Saint Louis, Missouri, 1862-1957, N = 96;

2. The St. Lawrence River at Ogdensburg, New York, 1861-1957, N = 97;

3. The Mississippi River at Keokuk, Iowa, 1879-1957, N = 79;

4. The Göta River at Sjötörp, Vanersburg, Sweden, 1808-1957, N = 150, and

5. The Rhine River at Basle, Switzerland, 1808-1957, N = 150.

Table 6.3 shows the results of this investigation. The analysis of five original series of annual river flows shows that only the Rhine River at Basle, Switzerland is an independent time series, or rather the null hypothesis is accepted for the 95% tolerance level. The whitened series are obtained by the hypothesis of the first-order, linear, autoregressive model of dependence, or

$$\varepsilon_i = x_i - r_1 x_{i-1} \quad , \tag{6.3}$$

where x_i and x_{i-1} are elements of a standardized series of annual river flows and r_1 is the first serial correlation coefficient, also given in Table 6.3.

Table 6.3 shows the results of the analysis of whitened series of the first four rivers by using the mean run-length of the median, or Equations (6.1) and (6.2). All four whitened series are shown to be independent series, or the null hypothesis for the whitened series is accepted for the 95% tolerance level. This finding means also that the hypothesis of the first-order linear autoregressive model describing well the series dependence is accepted for all four rivers.

6.6 Examples of Investigation of Annual Precipitation Series by the Relation of Mean Run-Length to Values of q . In the case of using any value of q as the truncation level, the $1-\alpha$ tolerance limits for a two-sided test are given by

$$\overline{N}_{1,2} = \frac{1}{pq}\left[1 \pm \frac{t_{\alpha/2}\ (p^3+q^3)^{\frac{1}{2}}}{\sqrt{k}}\right] \quad . \tag{6.4}$$

By using the 95% tolerance level, $= 0.05$, $t_{\alpha/2} = t_{0.025} = 1.96$, the limits are

$$\overline{N}_{1,2} = \frac{1}{pq}\left[1 \pm \frac{1.96}{\sqrt{k}}\ (p^3+q^3)^{\frac{1}{2}}\right] \quad . \tag{6.5}$$

Table 6.4 gives values of $1/pq$ and $(p^3+q^3)^{\frac{1}{2}}$ for values of q ranging between 0.20 and 0.80 with the increment of 0.05. Table 6.5 gives the tolerance limits of $\overline{N}$ at the 95% level.

The mean and the 95% tolerance limits of $\overline{N}$ are shown in Fig. 6.1 and 6.2 for values of k = 10, 15, 20, 25, 30, and 35, and q = 0.20 to q = 0.80 with increment of 0.05. The examples of annual precipitation series are analyzed by computing $\overline{N}(q)$ for values of q = 0.3, 0.4, 0.5, 0.6, 0.7, and plotting $\overline{N}(q)$ against q, as shown in Figure 6.1. For each q and the corresponding k, the tolerance limits, $\overline{N}_{1,2}$, are computed and plotted on the same graph.

If the analyzed series is a sequence of independent, identically distributed variables, the sample function $\overline{N}(q)$ should be inside the tolerance region. As can be seen from Figure 6.1, four out of the five series of the above examples have $\overline{N}(q)$ inside the tolerance region, but one series has only a point of $\overline{N}(q)$ outside the tolerance region. This particular series is a nonhomogeneous precipitation series.

6.7 Examples of Investigation of Annual Runoff Series by the Relation of Mean Run-Length to Values of q. The same procedure used for the precipitation series is applied to investigating the examples of the original series of annual runoff. The results are given in Figure 6.3 and 6.4. It is apparent from this figure that at least three out of the five series are not independent, identically distributed random variables, since their $\overline{N}(q)$ are not completely inside the tolerance region. The Mississippi River at St. Louis seems to be a case of weak serial dependence and/or of some nonstationarity, since its $\overline{N}(q)$ is so close to the upper tolerance limit. Finally, the annual runoff of the Rhine River is accepted as a sequence of independent, identically distributed variables. This result agrees with the autocorrelation analysis made on this series in a previous study [2].

For the four other series where the hypothesis that the series are independent, identically distributed variables is rejected, the first-order autoregressive model is assumed, and consequently the series are whitened. Then, the same procedure is used with both the whitened and the original series. The results are shown in Figure 6.5. The first-order autoregressive model for annual series is accepted for these four series.

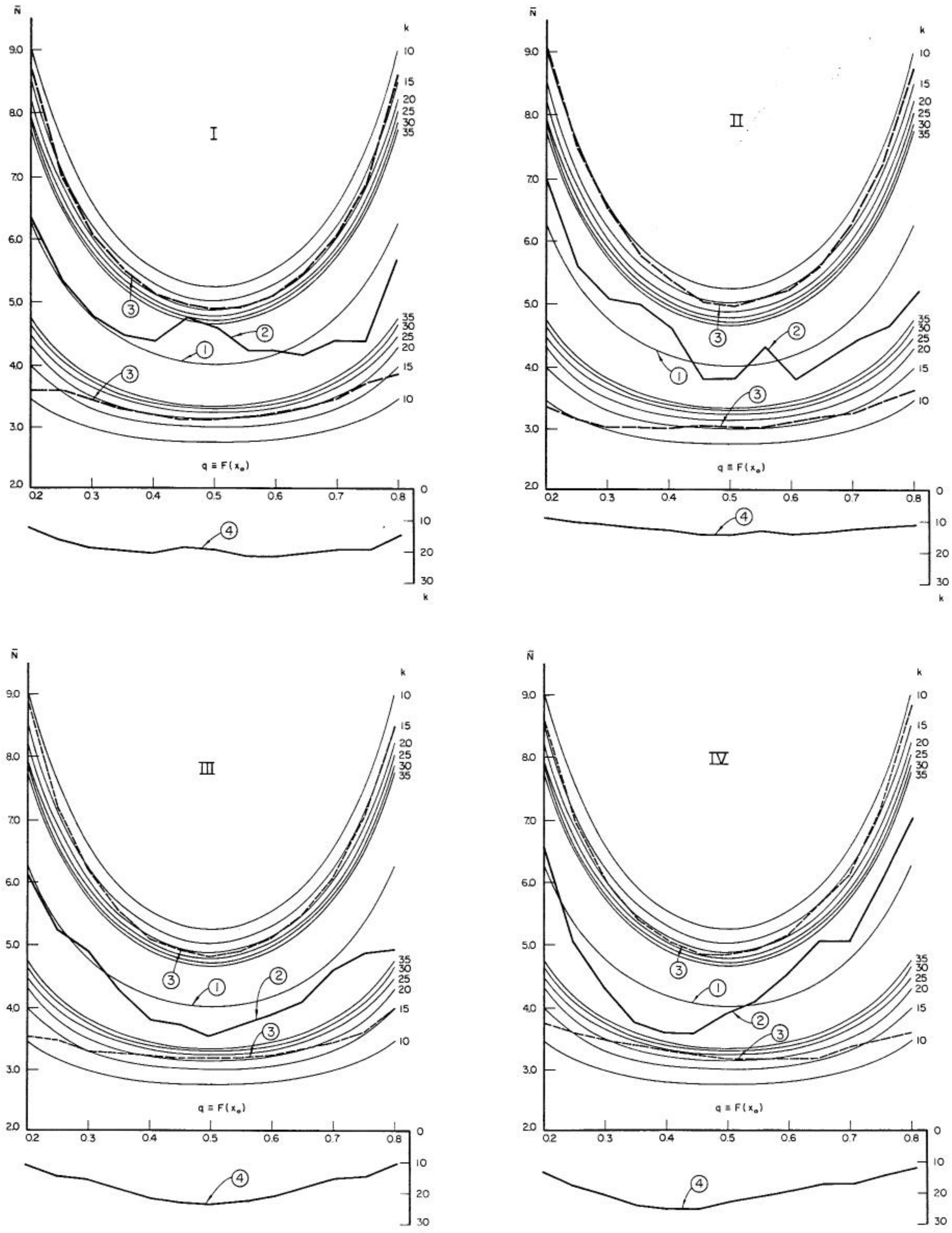


Fig. 6.1. Investigation of time series independence by the mean total run-length, $\bar{N}(q)$, as a function of the probability, q, of the truncation level, for four annual precipitation series: (I) Chico Experimental Station, California; (II) Ord, Nebraska; (III) Antioch F. Mills, California, and (IV) Ravenna, Nebraska; and for each station: (1) the expected mean total run-length of independent series $EN(q) = 1/q(1-q)$; (2) the observed $\bar{N}(q) = f(q)$; (3) the 95% tolerance limits for given k as function of q, and (4) the number of total run-lengths, k.

TABLE 6.3 PROPERTIES OF RUN-LENGTHS OF THE FIVE ANNUAL RIVER FLOW SERIES

Station	Original Series, x_i						Whitened Series, $x_i - r_1 x_{i-1}$						
	$\bar{N}^+$	$\bar{N}^-$	$\bar{N}$	k	$\bar{N}_1(95\%)$	$\bar{N}_2(95\%)$	r_1	$\bar{N}^+$	$\bar{N}^-$	$\bar{N}$	k	$\bar{N}_1(95\%)$	$\bar{N}_2(95\%)$
1. Mississippi River	2.389	2.500	4.889	18	4.775	3.225	0.301	1.783	2.130	3.913	23	4.816	3.184
2. St. Lawrence River	3.583	3.917	7.500	12	4.950	3.050	.707	2.045	2.136	4.181	22	4.835	3.165
3. Mississippi River	2.500	2.857	5.357	14	4.878	3.122	0.407	1.714	1.619	3.333	21	4.854	3.146
4. Gota River	2.704	2.778	5.482	27	4.634	3.365	0.445	2.202	2.212	4.054	33	4.682	3.317
5. Rhine River	2.027	2.027	4.054	37	4.540	3.450	---	---	---	---	--	---	---

TABLE 6.4. Values of $1/pq$ and $(p^3 + q^3)^{1/2}$

q	p	$\frac{1}{pq}$	$(p^3+q^3)^{\frac{1}{2}}$
.20	.80	6.25	.7211
.25	.75	5.33	.6614
.30	.70	4.76	.6083
.35	.65	4.40	.5635
.40	.60	4.17	.5292
.45	.55	4.04	.5074
.50	.50	4.00	.5000
.55	.45	4.04	.5074
.60	.40	4.17	.5292
.65	.35	4.40	.5635
.70	.30	4.76	.6083
.75	.25	5.33	.6614
.80	.20	6.25	.7211

TABLE 6.5 TOLERANCE LIMITS OF $\bar{N}$ AT THE 95% LEVEL

k \ q or p	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	9.04	7.51	6.55	5.91	5.54	5.32	5.24
	3.46	3.15	2.97	2.89	2.80	2.76	2.76
15	8.53	7.11	6.21	5.66	5.29	5.08	5.02
	3.97	3.55	3.31	3.14	3.05	3.00	2.98
20	8.22	6.88	6.02	5.48	5.14	4.95	4.87
	4.28	3.78	3.50	3.32	3.20	3.13	3.13
25	8.05	6.71	5.88	5.38	5.04	4.85	4.78
	4.45	3.95	3.64	3.42	3.30	3.23	3.22
30	7.85	6.58	5.78	5.29	4.96	4.78	4.71
	4.65	4.08	3.74	3.51	3.38	3.30	3.29
35	7.75	6.50	5.72	5.22	4.91	4.72	4.66
	4.75	4.16	3.80	3.58	3.43	3.36	3.34

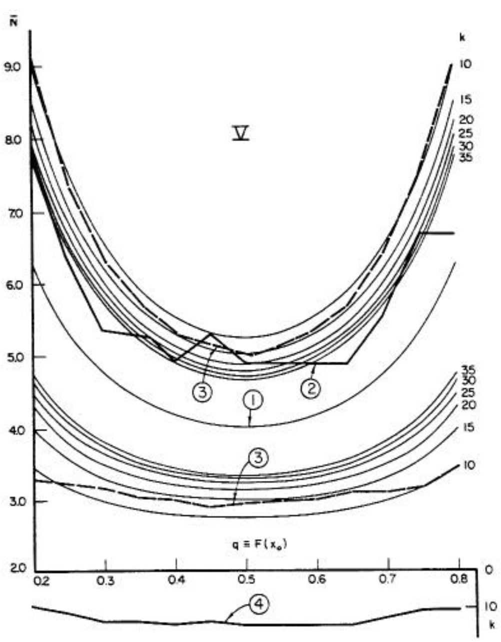


Fig. 6.2. Investigation of time independence by the mean total run-length, $\bar{N}(q)$, as a function of q, for the non-homogeneous annual precipitation series at Natural Bridge, N.M., Arizona; (1) the expected mean total run-length $EN(q) = 1/q(1-q)$; (2) the observed $\bar{N}(q) = f(q)$; (3) the 95% tolerance limits for given k and (4) the number of total run-lengths, k.

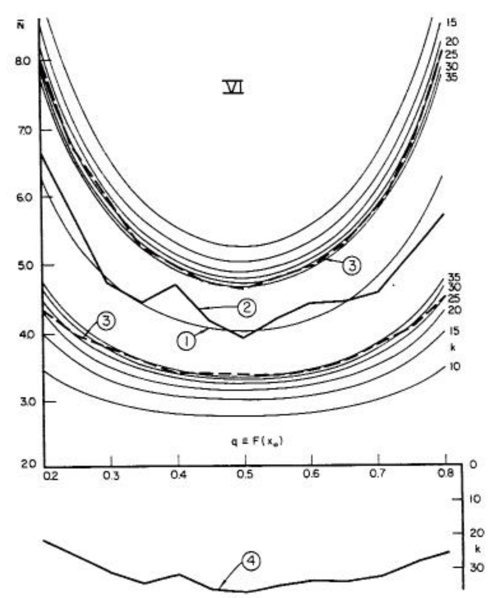


Fig. 6.3. Investigation of time independence by the mean total run-length, $\bar{N}(q)$, as a function of q, for the annual runoff series of the Rhine River at Basle, Switzerland: (1) the expected mean total run-length, $EN(q) = 1/q(1-q)$; (2) the observed $\bar{N}(q) = f(q)$; (3) the 95% tolerance limits for given k, and (4) the number of total run-lengths, k.

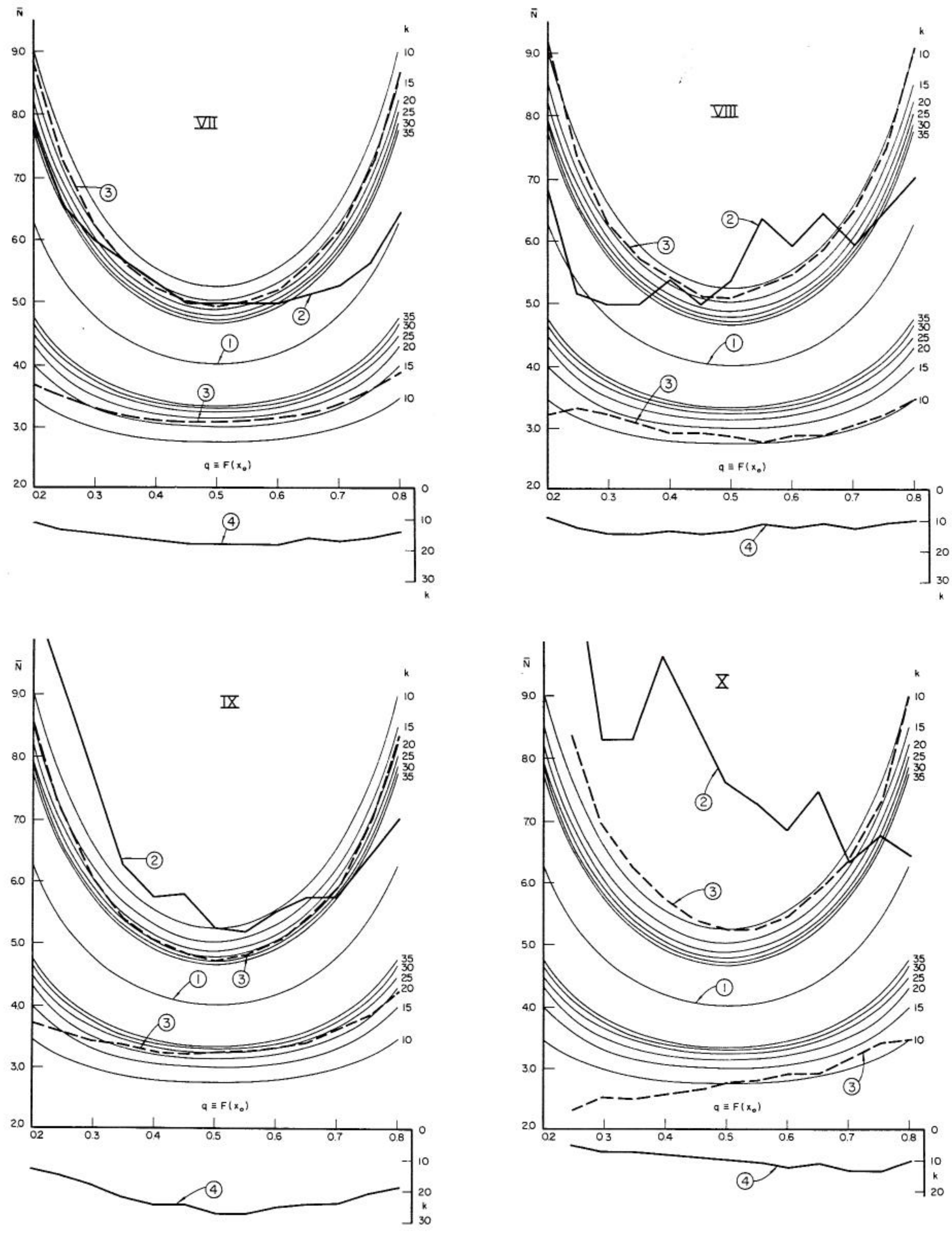


Fig. 6.4. Investigation of time series independence by the mean total run-length, $\bar{N}(q)$, as a function of the probability, q, of the truncation level, for four annual runoff series: (VII) Mississippi River at St. Louis, Missouri; (VIII) St. Lawrence River at Ogdensburg, New York; (IX) Mississippi Rover at Keokuk, Iowa, and (X) Göta River at Sjötörp, Vänersburg; and for each station: (1) The expected mean total run-length of independent series, $EN(q) = 1/q(1-q)$; (2) the observed $\bar{N}(q) = f(q)$; (3) the 95% tolerance limits for given k as function of q, and (4) the number of total run-lengths, k.

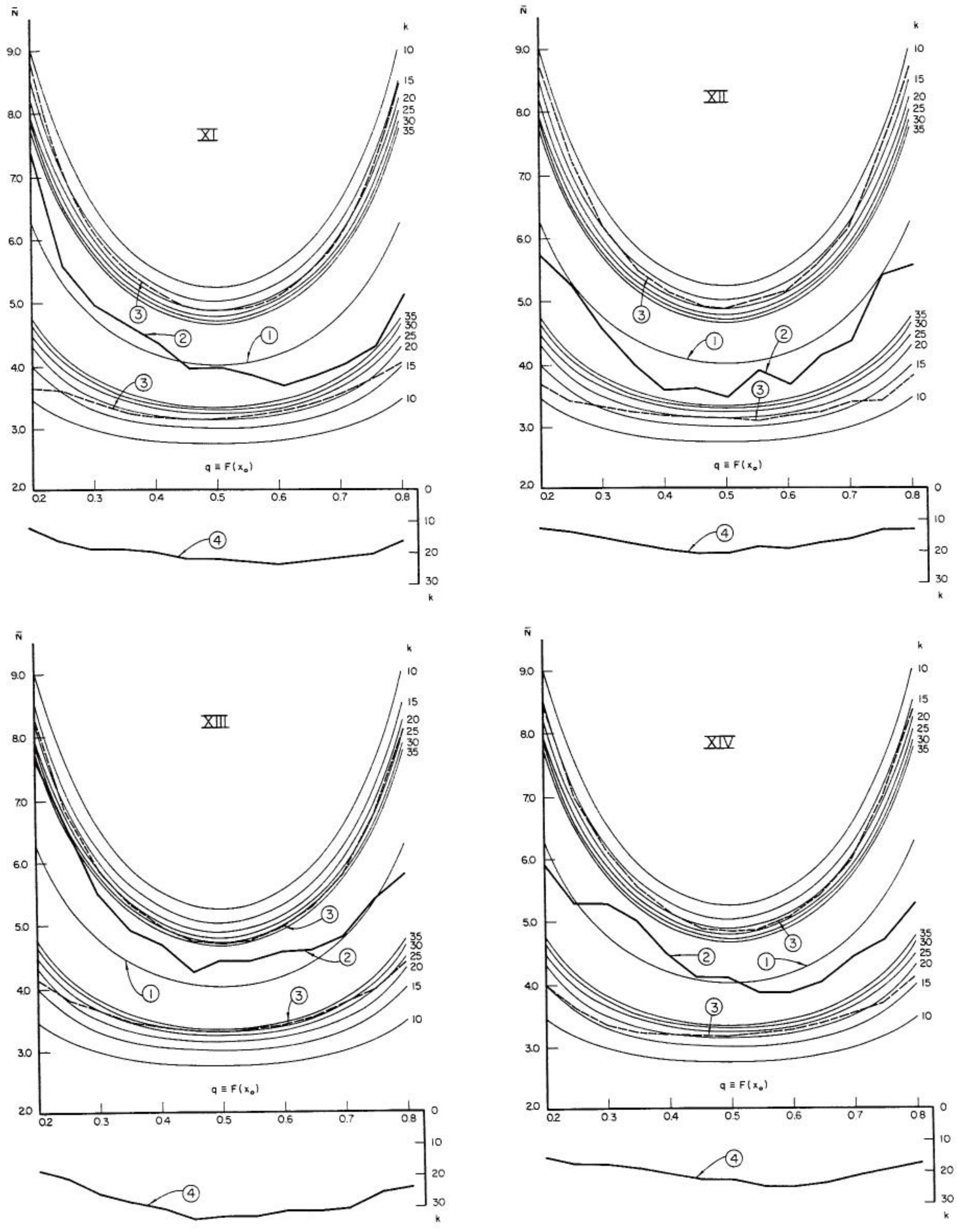


Fig. 6.5. Investigation of independence of whitened series, under the assumption of the first-order autoregressive process as a time dependence model, by using the mean total run-length, $\bar{N}(q)$, as a function of the probability, q, of the truneation level, for four annual runoff series: (XI) Mississippi River at St. Louis, Missouri; (XII) St. Lawrence River at Ogdensburg, New York; (XIII) Mississippi River at Keokuk, Iowa, and (XIV) Gota River at Sjotorp, Vauersburg; and for each station: (1) the expected mean total run-length of independent series, EN(q) $= 1/q(1-q)$; (2) the observed $\bar{N}(q) = f(q)$ of whitened series; (3) the 95% tolerance limits for given k as function q, and (4) the number of total run-lengths, k of whitened series.

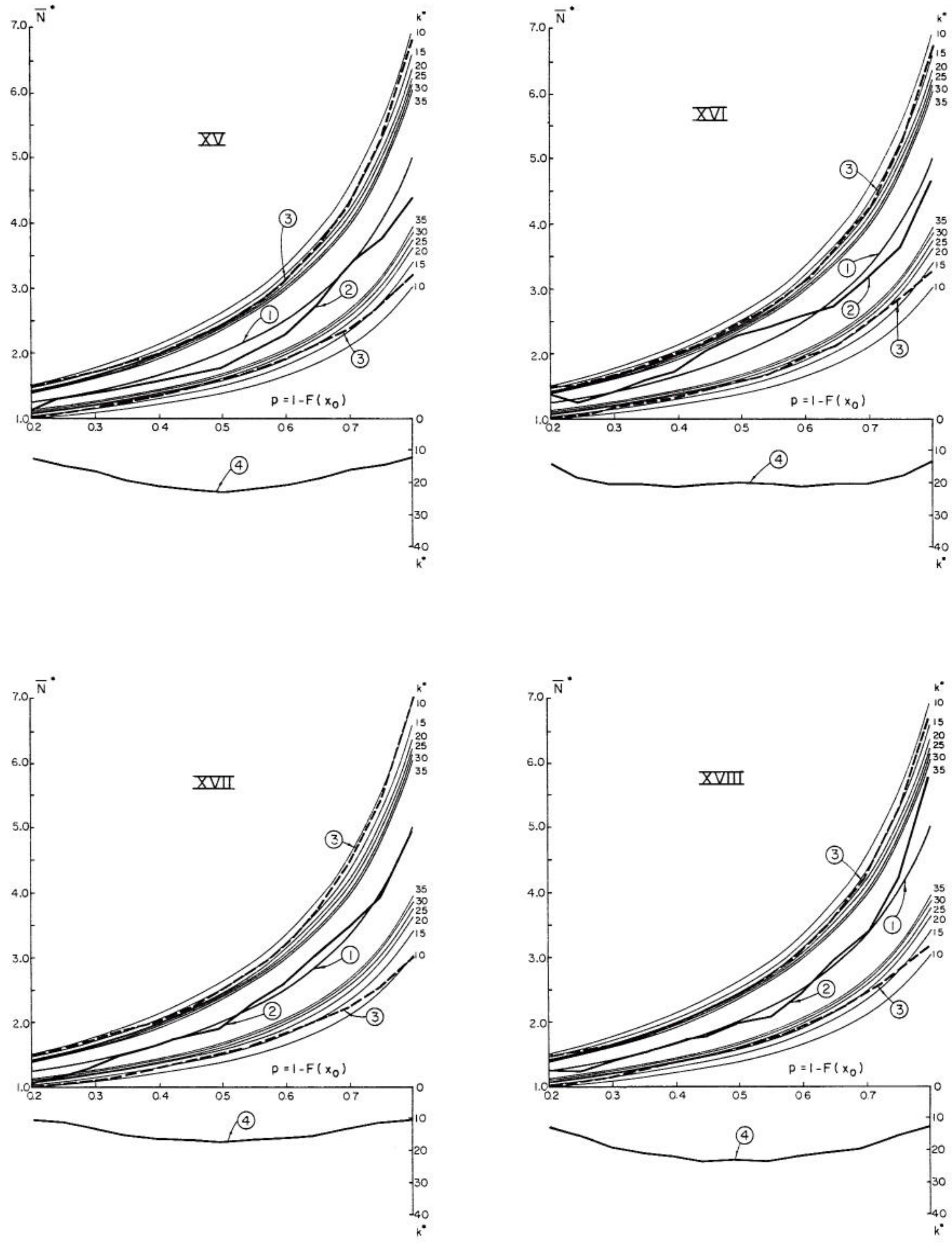


Fig. 6.6. Investigation of time series independence by the mean run-length, $\bar{N}(q)$, as a function of the probability, q, of the truncation level, for four annual precipitation series: (XV) Chico Experimental Station, California; (XVI) Ord, Nebraska; (XVII) Antioch F. Mills, California and (XVIII) Ravenna, Nebraska; and for each station: (1) the expected mean run-length, $E\bar{N}^*(q) = 1/q$; (2) the observed $\bar{N}^*(q) = f(q)$; (3) the 95% tolerance limits for given $k^* = (k^+ + K^-)/2$.

6.8 Examples of Investigation of Annual Precipitation and Runoff Series by $N^*(p)$ for Values of p. The $1-\alpha$ tolerance region for independent identically distributed variables is given by Equation (5.22). Using the 95% tolerance level, $\alpha = 0.05$, $t_{\alpha/2} = t_{0.025} = 1.96$, the tolerance limits are

$$\bar{N}^*_{1,2} = \frac{1}{q}\left(1 \pm 1.96\sqrt{\frac{p}{2k^*}}\right) . \qquad (6.6)$$

Table 6.6 gives the mean and the tolerance limits of $\bar{N}^*$ at the 95% level.

The sample of precipitation series was analyzed by computing $\bar{N}^*(p)$ for values of $q = 0.3, 0.4, 0.5, 0.6, 0.7$ and plotting $\bar{N}^*(p)$ against p as shown in Figures 6.6 and 6.7. The results are shown in these figures.

In a similar manner the sample of the original runoff series was analyzed. The results are shown in Figures 6.8 and 6.9. Finally, the whitened runoff series using a first-order autoregressive model were also analyzed in a similar manner. The results are shown in Figure 6.6. From these figures it is apparent that all precipitation series are accepted as independent time series, whereas all runoff series except Rhine River are accepted as first-order autoregressive processes.

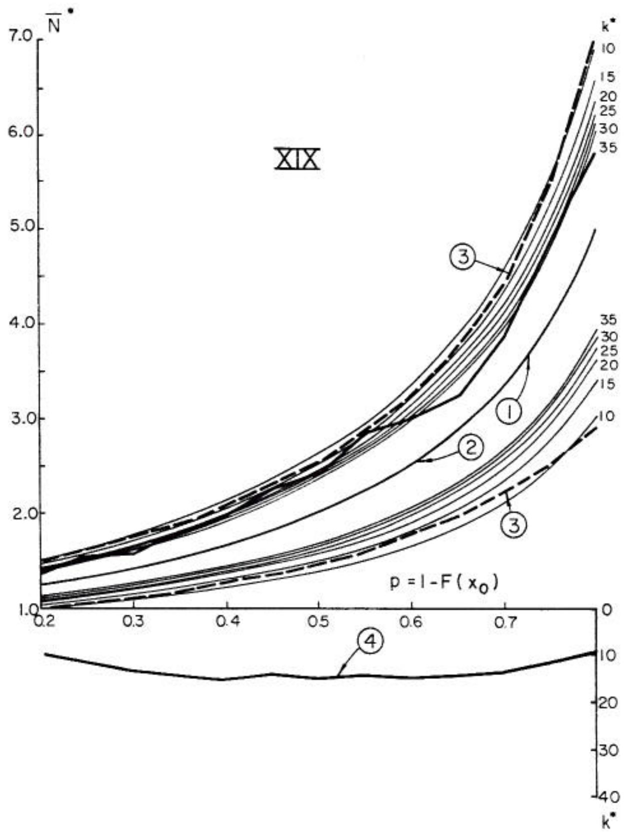


Fig. 6.7. Investigation of time independence by the mean run-length, $\bar{N}^*(q)$, as a function of q, for the non-homogeneous annual precipitation series at (XIX) Natural Bridge, N.M., Arizona: (1) the expected mean run-length $E\bar{N}^*(q) = 1/q$; (2) the observed $\bar{N}^*(q) = f(q)$; (3) the 95% tolerance limits for given k^* as function of q, and (4) $k^* = (k^+ + k^-)/2$

TABLE 6.6 MEAN AND 95% TOLERANCE LIMITS OF $\bar{N}^*$

p	k* / EN*	10	15	20	25	30	35
0.20	2.500	2.696	2.660	2.639	2.624	2.613	2.605
	---	2.304	2.340	2.361	2.376	2.387	2.395
0.25	2.666	2.885	2.845	2.821	2.805	2.793	2.783
	---	2.447	2.487	2.511	2.527	2.539	2.549
0.30	2.858	3.098	3.054	3.028	3.010	2.997	2.986
	---	2.618	2.662	2.688	2.706	2.719	2.730
0.35	3.076	3.335	3.288	3.259	3.240	3.226	3.215
	---	2.817	2.864	2.893	2.912	2.926	2.937
0.40	3.334	3.611	3.560	3.513	3.509	3.494	3.482
	---	3.057	3.108	3.155	3.159	3.174	3.186
0.45	3.636	3.930	3.876	3.844	3.822	3.806	3.793
	---	3.342	3.396	3.428	3.450	3.466	3.479
0.50	4.000	4.310	4.253	4.219	4.196	4.179	4.166
	---	3.690	3.747	3.781	3.804	3.821	3.834
0.55	4.444	4.769	4.614	4.674	4.650	4.632	4.618
	---	4.119	4.274	4.214	4.238	4.256	4.270
0.60	5.000	5.340	5.277	5.240	5.215	5.196	5.182
	---	4.660	4.723	4.760	4.785	4.804	4.818
0.65	5.714	6.067	6.003	5.964	5.938	5.918	5.903
	---	5.361	5.425	5.464	5.490	5.510	5.525
0.70	6.667	7.034	6.966	6.926	6.899	6.879	6.863
	---	6.300	6.368	6.408	6.435	6.455	6.471
0.75	8.000	8.380	8.310	8.268	8.240	8.219	8.203
	---	7.620	7.690	7.722	7.760	7.781	7.797
0.80	10.000	10.392	10.320	10.277	10.248	10.226	10.210
	---	9.608	9.680	9.723	9.752	9.774	9.790

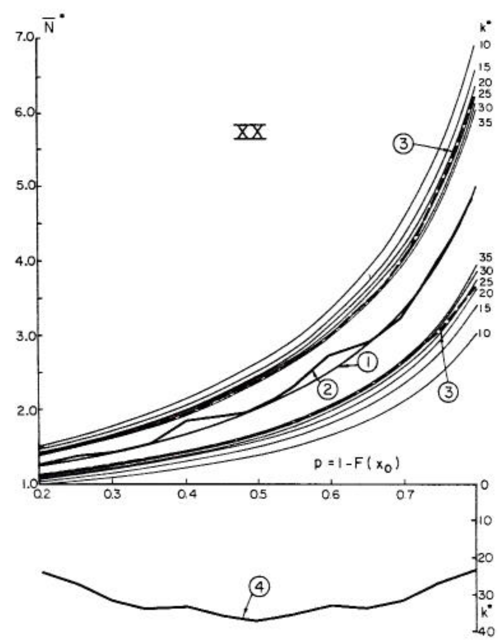


Fig. 6.8. Investigation of time independence by the mean run-length, $\bar{N}^*(q)$, as a function of q, for the annual run-off series of (XX) Rhine River at Basle, Switzerland: (1) the expected mean run-length, $EN^*(q) = 1/q$; (2) the observed $N^*(q) = f(q)$; (3) the 95% tolerance limits for given k^* as function of q, and (4) $k^* = (k^+ + k^-)/2$

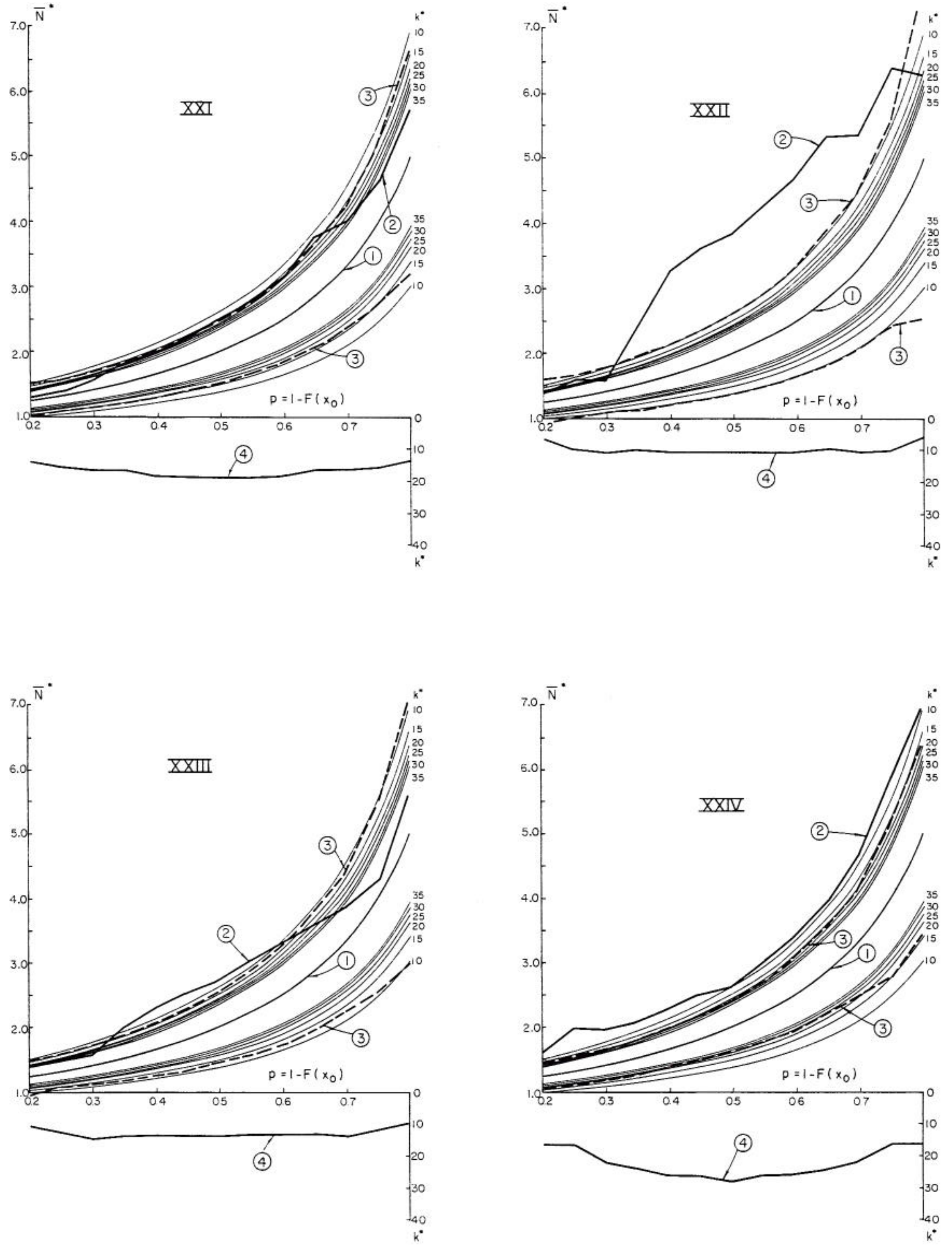


Fig. 6.9. Investigation of time series independence by the mean run-length, $\bar{N}^*(q)$, as a function of the probability, q, of the truncation level, for four annual runoff series: (XXI) Mississippi River at St. Louis, Missouri; (XXII) St. Lawrence River at Ogdensburg, New York; (XXIII) Mississippi River at Keskuk, Iowa, and (XXIV) Gota River at Sjotorp, Vanersburg; and for each station: (1) the expected mean run-length of independent series $EN^*(q) = 1/q$; (2) the observed $\bar{N}^*(q) = f(q)$: (3) the 95% tolerance limits for given k^* as a function of q, and(4) $k^* = (k^+ + k^-)/2$

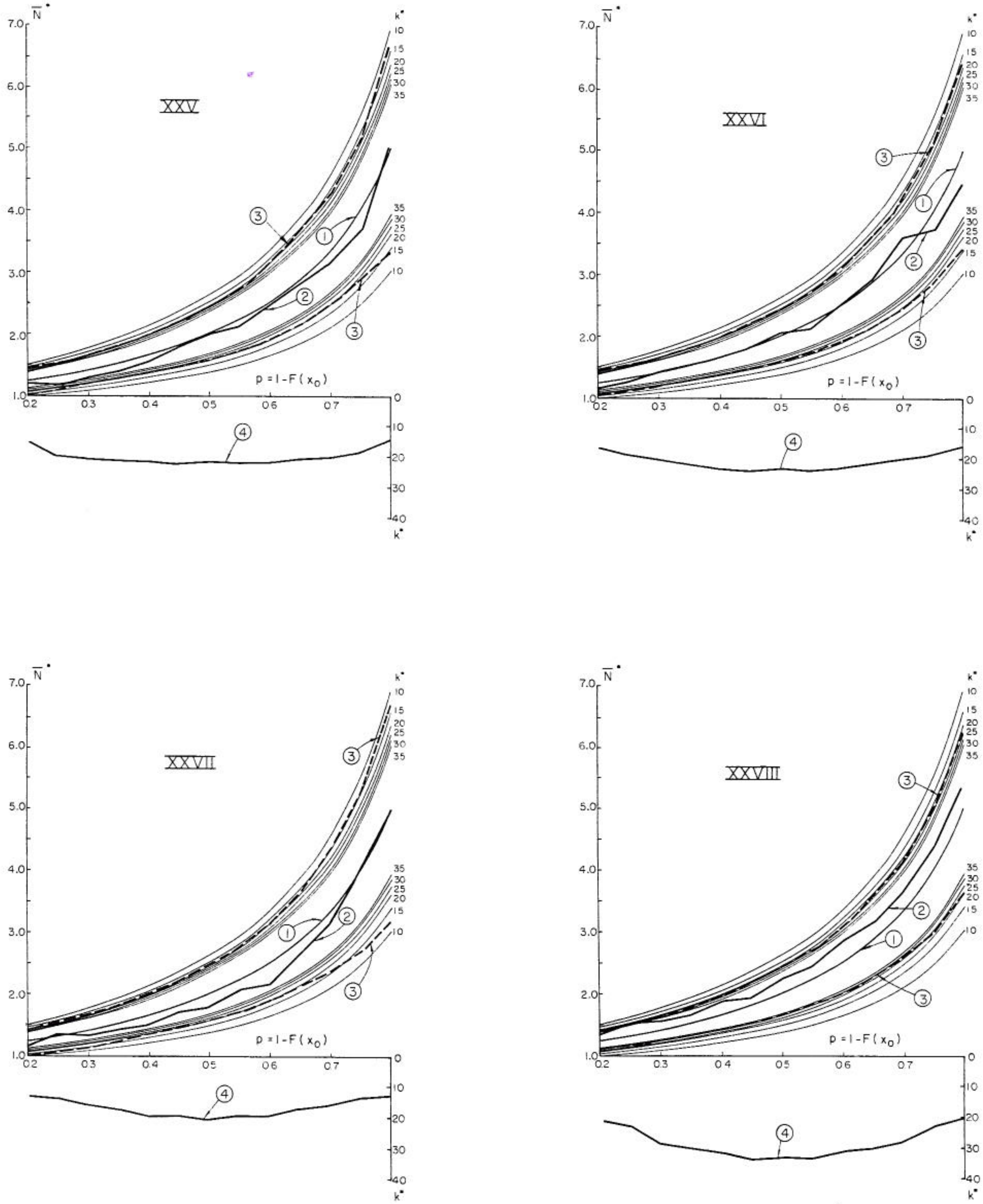


Fig. 6.10. Investigation of independence of whitened series, under the assumption of the first-order autoregressive process as a time dependence model, by using the mean run-length $\bar{N}^*(q)$, as a function of the probability, q, of the truncation level, for four annual runoff series: (XXV) Mississippi River at St. Louis, Missouri; (XXVI) St. Lawrence River at Ogdensburg, New York; (XXVII) Mississippi River at Keokuk, Iowa, and (XXVIII) Gota River at Sjotorp, Vanersburg; and for each station: (1) the expected mean run-length of independent series, $E\bar{N}^*(q) = 1/q$; (2) the observed $\bar{N}^*(q)$; (3) the 95% tolerance limits for given k as function of q, and (4) $k^* = (k^+ + k^-)/2$

Chapter VII

EXAMPLES OF COMPUTATION OF PROBABILITIES OF RUN-LENGTHS

7.1 Introduction. Present empirical techniques for determining probabilities of drought or low-flow durations, or probabilities of water surplus durations, for a well defined drought or water surplus level use sample data to derive the necessary information. The empirical procedure is as follows. Drought or surplus level is first defined; then a series of the hydrologic variable is plotted, with this level as the truncation level. Next, all durations as run-lengths equal to or greater than the truncation level are counted, and the relative frequency of run-lengths that are greater than the critical duration are computed. These frequencies are estimates of probabilities. Alternatively, the longest drought or water surplus duration is selected as the design drought or the design surplus. It is often difficult in practice to assign meaningful probabilities to a drought or a surplus because of large sampling fluctuations of these relative frequencies. Much confusion is often unavoidable in assigning probabilities to results empirically obtained.

It is not surprising then that the estimates of probabilities of historical droughts in some river basins or regions sometimes vary from 1:100 (one in a hundred years) to 1:3,000 (one in three thousand years) by different empirical approaches.

The method of using the run-length properties of the independent or of the first-order linear autoregressive series helps solve these important practical problems and also helps to avoid some confusion. Two aspects of these probabilities currently are of interest, probabilities of a given duration (say, probability of a 5-year drought), and probabilities of all events equal to or greater than a given duration (say, probability of all droughts of 3-years or more). The purpose of the techniques studied and developed in this paper are designed to solve these types of problems. In this study, however, only two cases are analytically approached: independent stationary time series, and dependent stationary time series of the first-order linear autoregressive model. For more complex models, such as the second-order, third-order, or higher-order linear autoregressive models, or for models with periodicity present in some parameters of a time series, the Monte Carlo simulation technique seems best suited at present as the alternative to the analytical method.

7.2 Determination of Run-Length Probabilities of Stationary and Independent Series. The annual series of precipitation and runoff of the following four gauging stations, found to be independent time series in Chapter VI, are used as examples for determining probabilities of run-lengths equal to or greater than a given length, for a given truncation level and its probability, q :

A. Precipitation Stations

1. Ord, Nebraska
2. Ravenna, Nebraska
3. Antioch F. Mills, California

B. Runoff Station

1. The Rhine River at Basle, Switzerland.

The probabilities of run-lengths being not less than a given value j are obtained as follows from Equation (2.21),

$$P(N^+ = k) = qp^{k-1} \quad .$$

Hence,

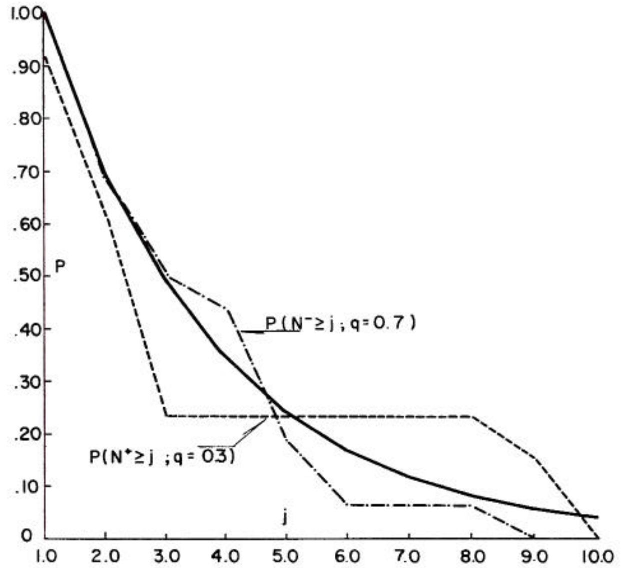
$$P(N^+ \geq j) = 1 - \sum_{k=1}^{j-1} P(N^+=k) = 1 - \sum_{k=1}^{j-1} qp^{k-1}$$

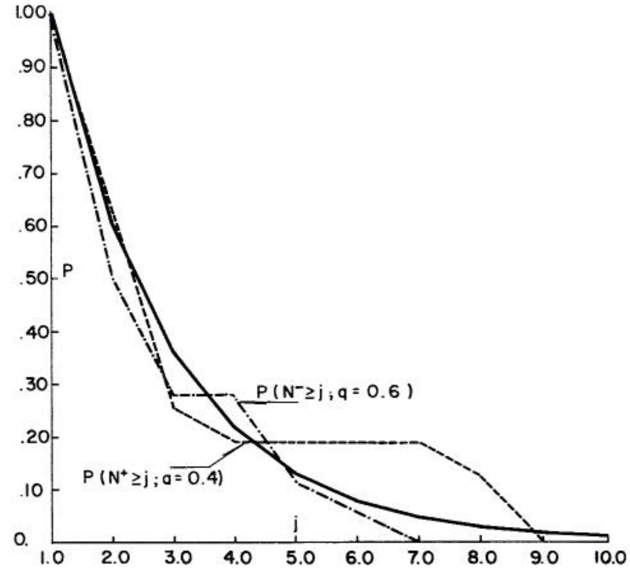
$$= 1 - q\frac{(1-p^{j-1})}{1-p} = p^{j-1} \quad . \tag{7.1}$$

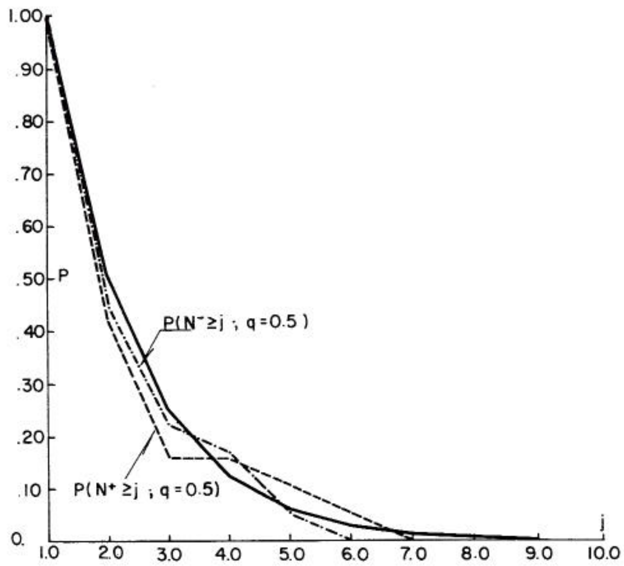
By using Equation (7.1) for probabilities of run-lengths of independent series, which are functions of q , or $p = 1 - q$, comparison is made between the relative frequencies of run-lengths, empirically determined, and the probabilities of the same run-lengths, analytically determined.

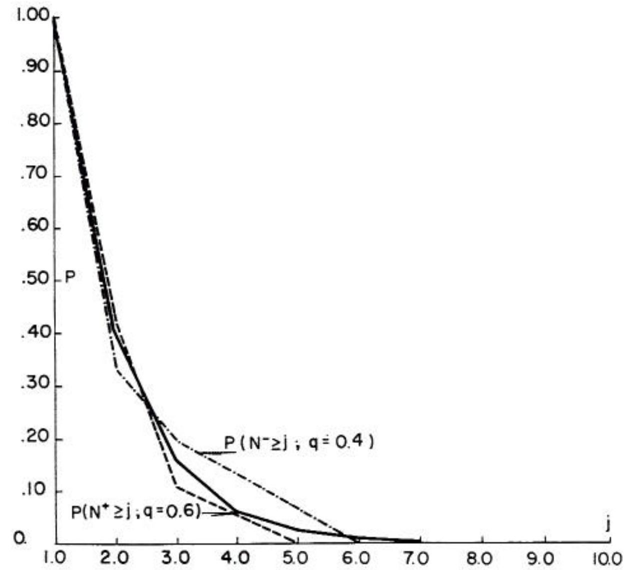
Figures 7.1 through 7.4 give these comparisons for the three annual precipitation series and the annual runoff series of the Rhine River, and for the run-lengths N_j , with the following j values: 2, 3, 4, 5, 6, 7, 8, 9, and 10, and in each case for five values of q , 0.3, 0.4, 0.5, 0.6, and 0.7.

As expected, probabilities of run-lengths $P(N \geq j)$ determined analytically by Equation (7.1) depart from the relative frequencies empirically determined from sample data. These deviations increase both with an increase of j and an increase of absolute departure of q values from $q = 0.50$. The unreliability of these relative frequencies of run-lengths $N \geq j$ empirically determined for extremes of q and for large values of j is the primary reason for the controversy between the various empirical methods of estimating probabilities of droughts or water surpluses of given durations for the prescribed levels of droughts or surpluses. These high sampling errors, associated with empirical methods currently used in practice, justify using the analytical method for computing probabilities $P(N \geq j)$ instead of using various empirical methods.









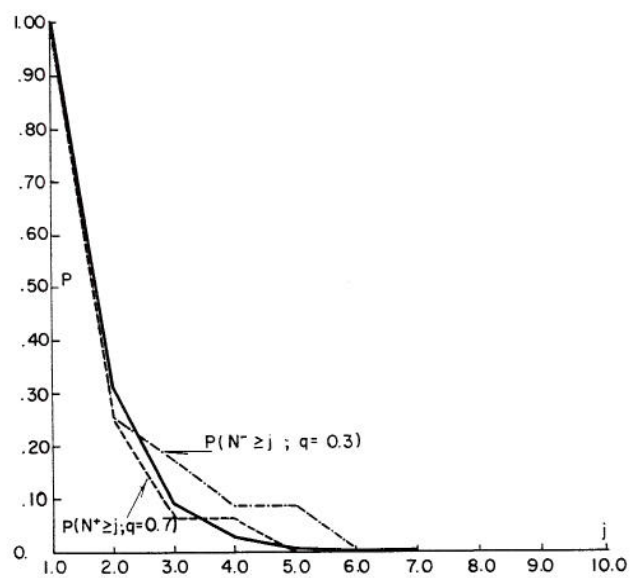
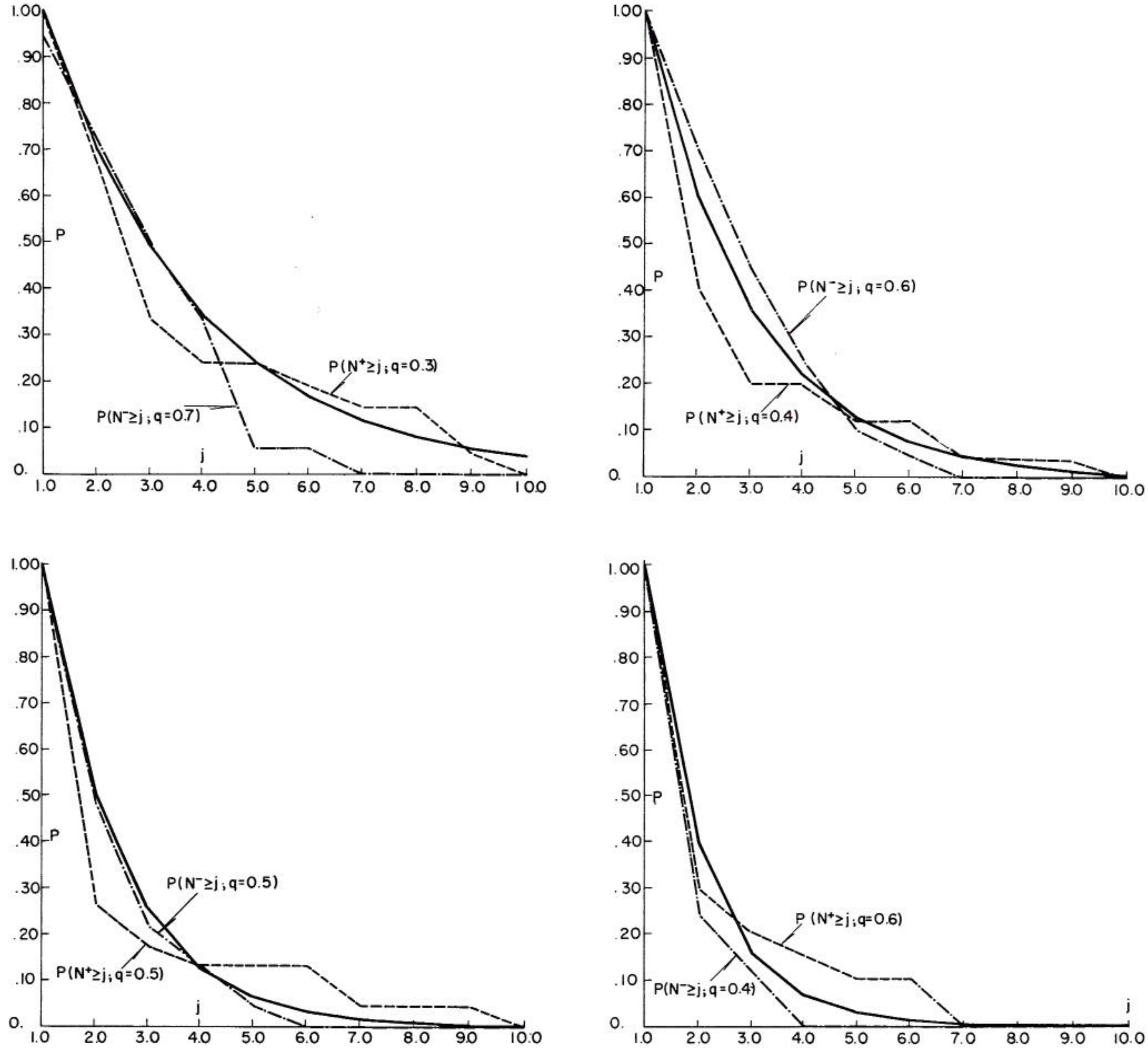


Fig. 7.1 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^+ \geq j \; ; \; q)$, and negative run-lengths, $P(N^- \geq j \; ; \; q)$, for the annual precipitation series at Ord, Nebraska (1896-1965) as compared with expected probabilities (solid lines) of positive and negative run-lengths of independent series for five truncation values: q = 0.3, 0.4, 0.5, 0.6, and 0.7.



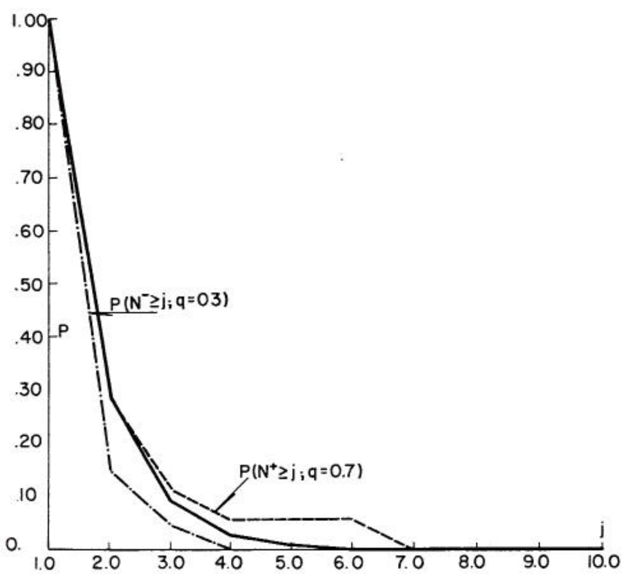
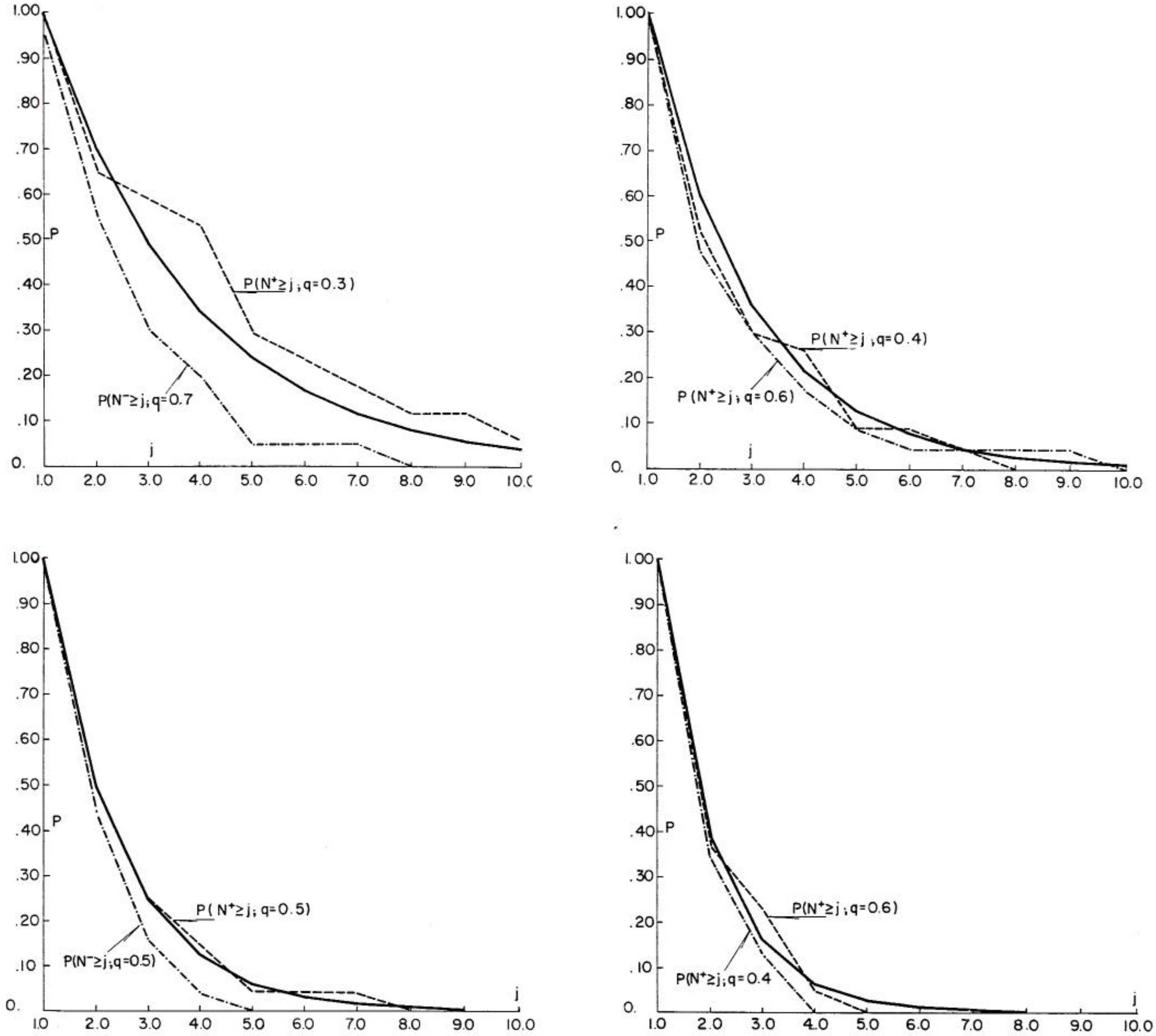


Fig. 7.2 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^+ \geq j\ ;\ q)$, and negative run-lengths, $P(N^- \geq j\ ;\ q)$, for the annual precipitation series at Ravenna, Nebraska (1878-1965) as compared with expected probabilities (solid lines) of positive and negative run-lengths of independent series for five truncation values: q = 0.3, 0.4, 0.5, 0.6, and 0.7.



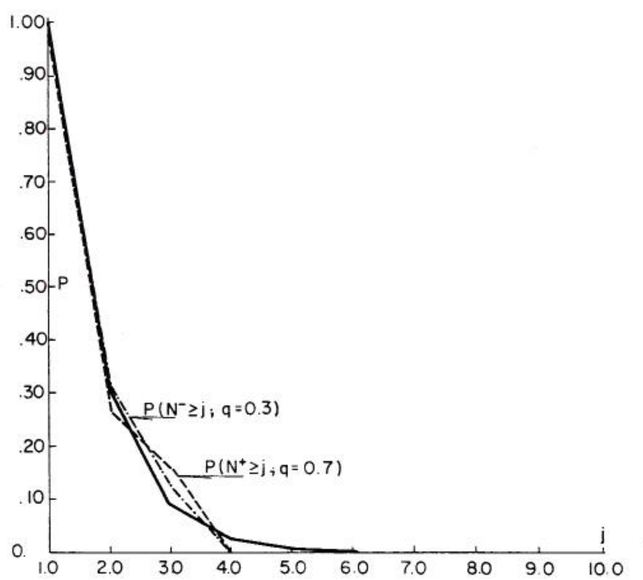
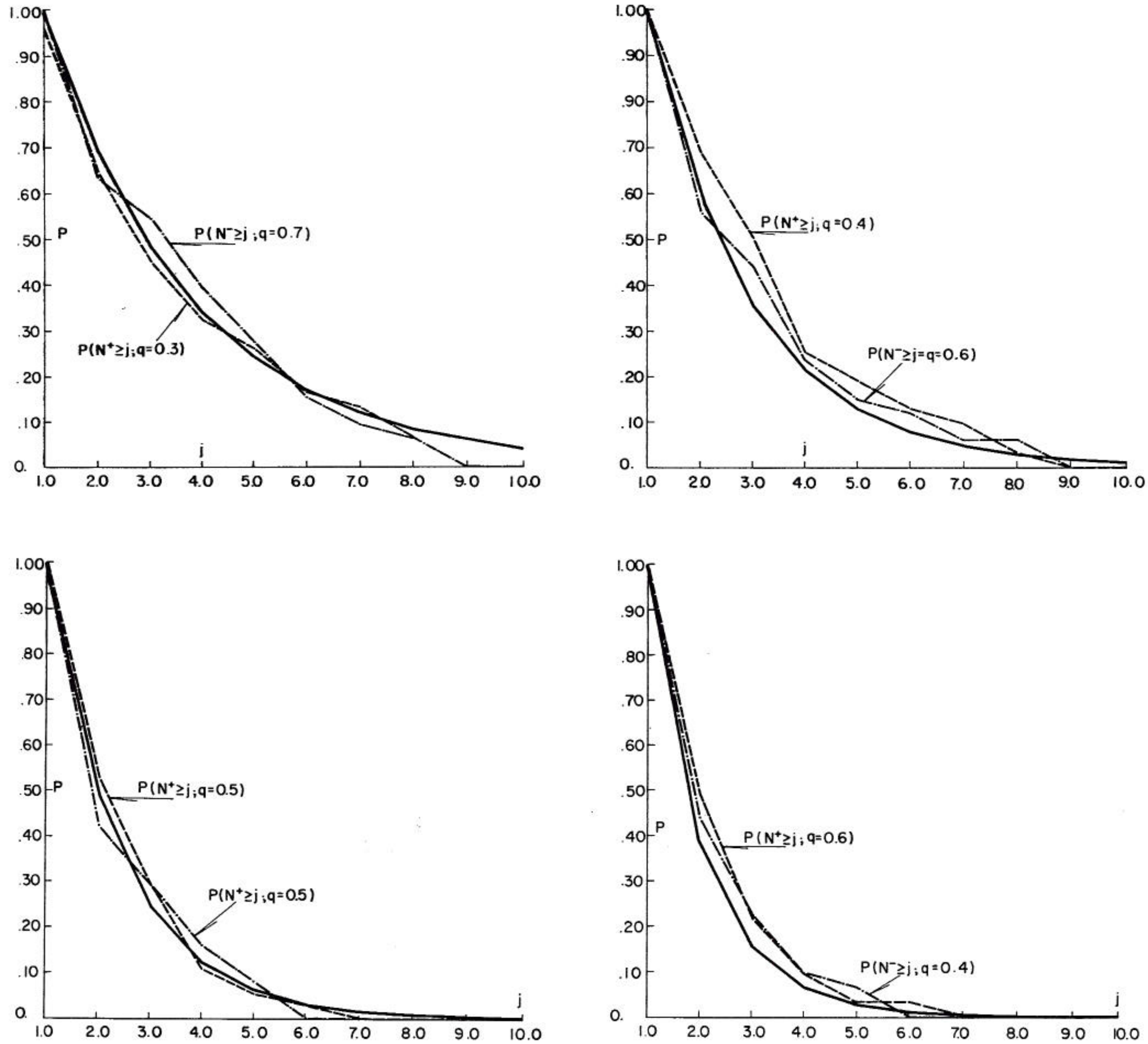


Fig. 7.3 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^+ \geq j\ ;\ q)$, and negative run-lengths, $P(N^- \geq j\ ;\ q)$, for the annual precipitation series at Antioch F. Mills, California (1879-1965) as compared with expected probabilities (solid lines) of positive and negative run-lengths of independent series for five truncation values: q = 0.3, 0.4, 0.5, 0.6, and 0.7.



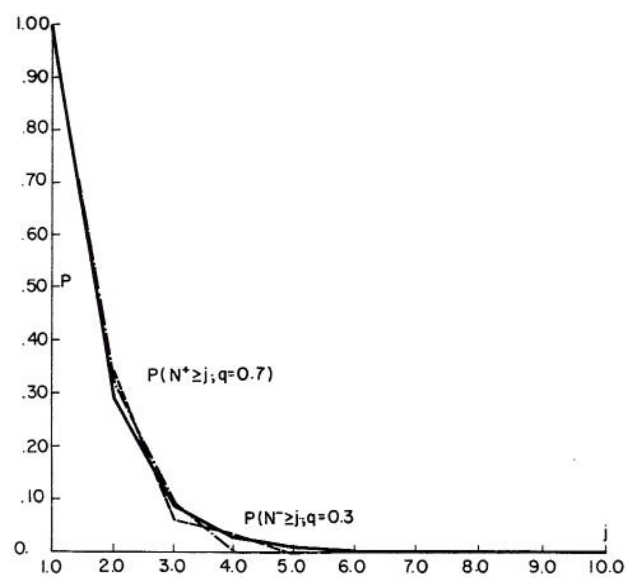


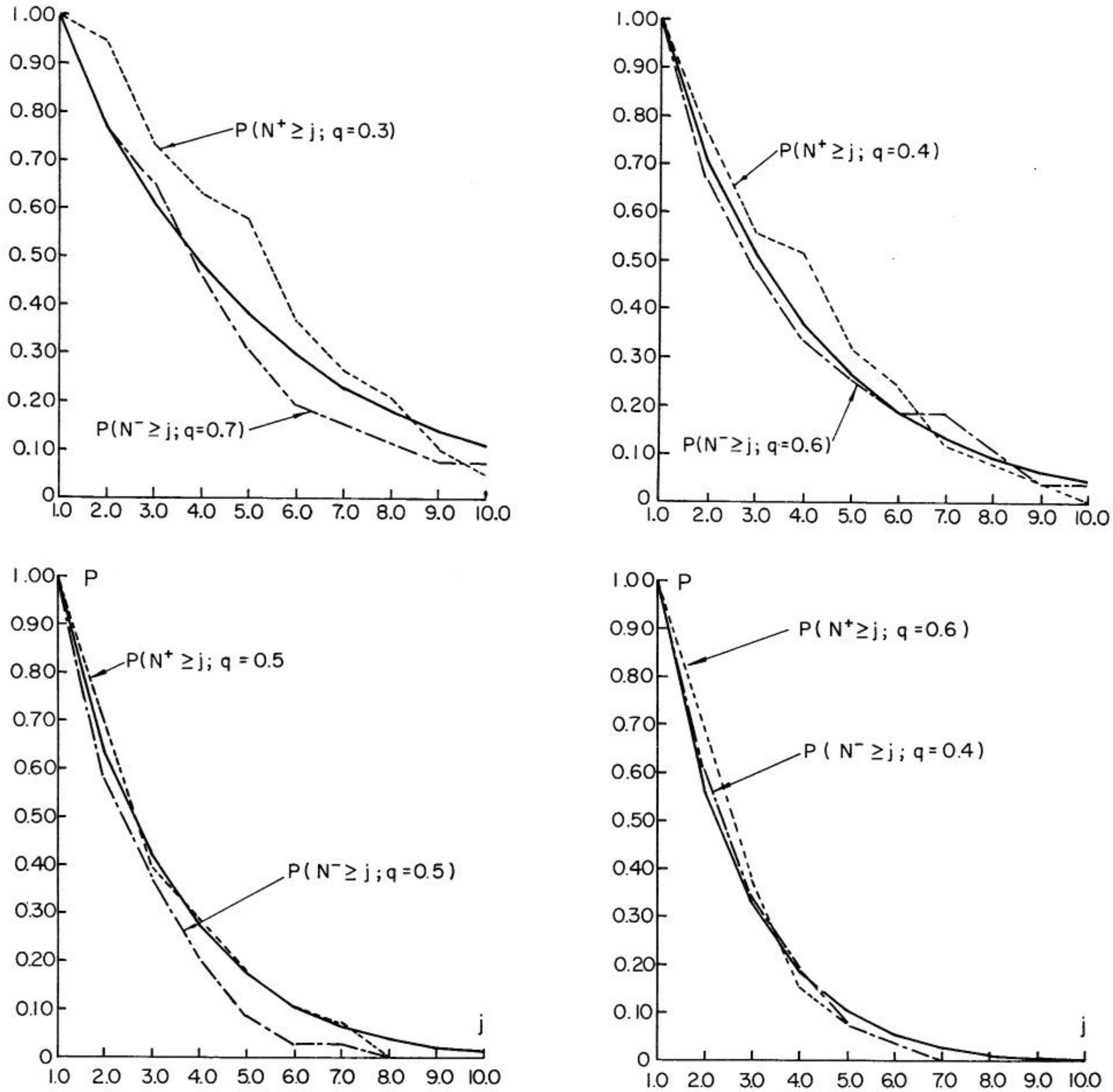
Fig. 7.4 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^{+} \geq j\ ;\ q)$, and negative run-lengths, $P(N^{-} \geq j\ ;\ q)$, for the annual runoff series of the Rhine River (1808-1957) as compared with expected probabilities (solid lines) of positive and negative run-lengths of independent series for five truncation values: $q = 0.3, 0.4, 0.5, 0.6$, and 0.7.

Because of large sampling fluctuations in series of limited size, either a drought of more than 3-4 years may not have been experienced in a period of 50-60 years, or a drought of nine years may have occurred, though no drought of 4-8 year duration was recorded. Many similar sampling biases are unavoidable in using current empirical methods in estimating probabilities of run-lengths.

7.3 Determination of Run-Length Probabilities of Stationary Dependent Series. The annual runoff series of the following three rivers, which are known to be dependent time series, are used as examples for determining probabilities of run-lengths equal to or greater than a given length, for a given truncation level and its probability, q :

5. The Göta River at Sjötörp, Vanersburg, Sweden, with $r_1 = 0.463$, where r_1 is the estimate of the first autocorrelation coefficient ρ_1 .
6. The Ashley Creek near Vernal, Utah, with $r_1 = 0.274$, and
7. The Trinity River at Lewiston, California, with $r_1 = 0.180$.

Probabilities of run-lengths being equal to or greater than a given value j are obtained by using the values $P(N^+ \geq j)$ of the Appendix, for values of $\rho_1 = 0.1, 0.2, 0.3, 0.4$, and 0.5. A linear interpolation is used for the r_1 values, which are between the ρ_1 values of the Appendix. These probabilities and the relative frequencies of run-lengths empirically determined are compared. Similarly, as in the case of independent series, probabilities of run-lengths $P(N \geq j)$ of dependent series depart from the relative frequencies empirically determined from the sample data. The Trinity River is analyzed in two ways: (1) by using the first serial sample correlation coefficient r_1 and the first-order autoregressive model for annual flows, and (2) by using $r_1^* = r_1 - r_1(P)$, where $r_1(P)$ is the first basin. This is done to remove the sampling fluctuation of $r_1(P)$ included in r_1 of runoff, because $E(r_1)$ for precipitation series should be close to zero. For the area of the Trinity River Basin, $r_1(P)$ has been calculated in a previous study [2] as $r_1(P) = 0.10$, so that the corrected value $r_1^* = 0.08$. Figures 7.5 through 7.8 show the results and comparisons of relative frequencies and expected probabilities for the corresponding cases.



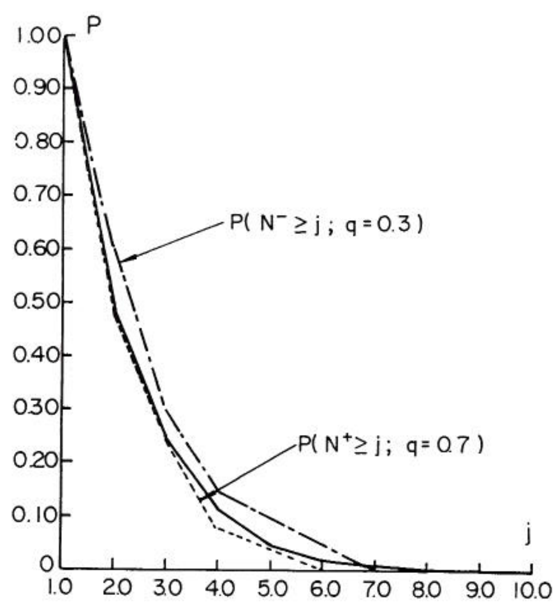
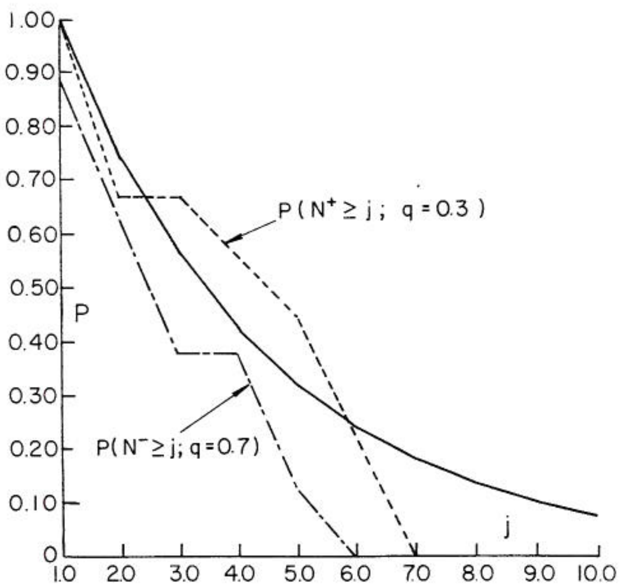
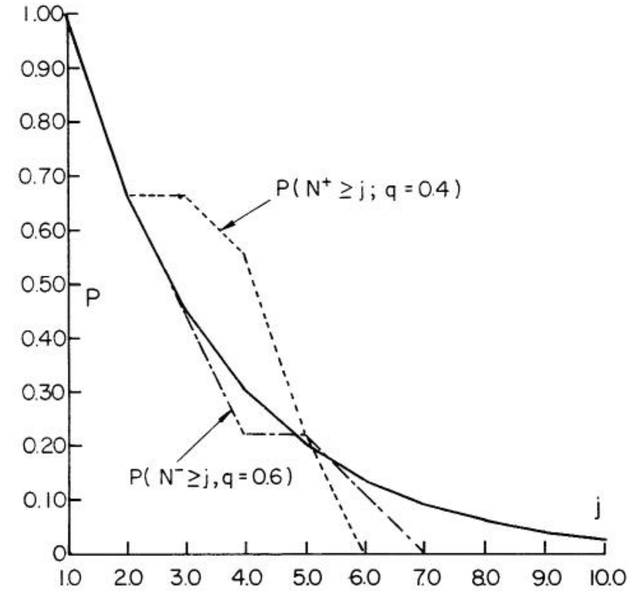
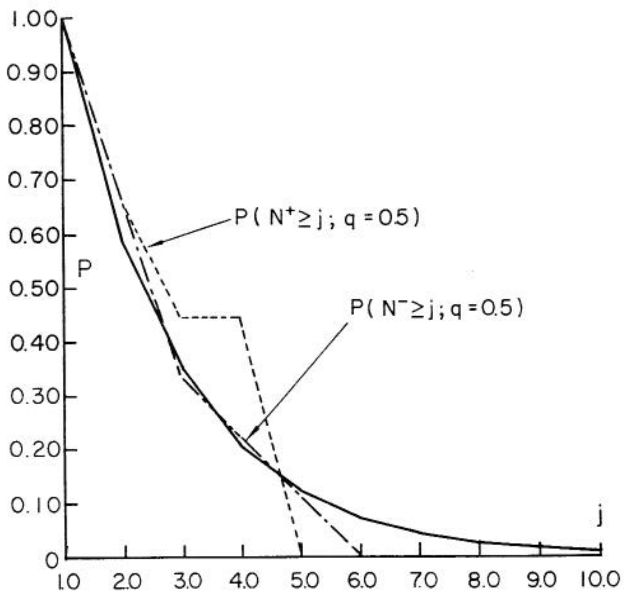
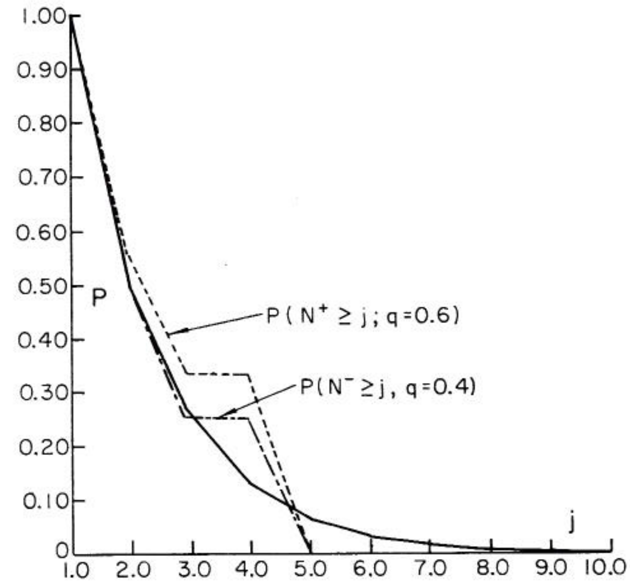


Fig. 7.5 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^+ \geq j\ ;\ q)$, and negative run-lengths, $P(N^- \geq j\ ;\ q)$, for the annual runoff series of the Göta River (1808-1957) as compared with expected probabilities (solid lines) of positive and negative run-lengths of a first-order autoregressive process with $\rho_1 = r_1 = 0.450$.









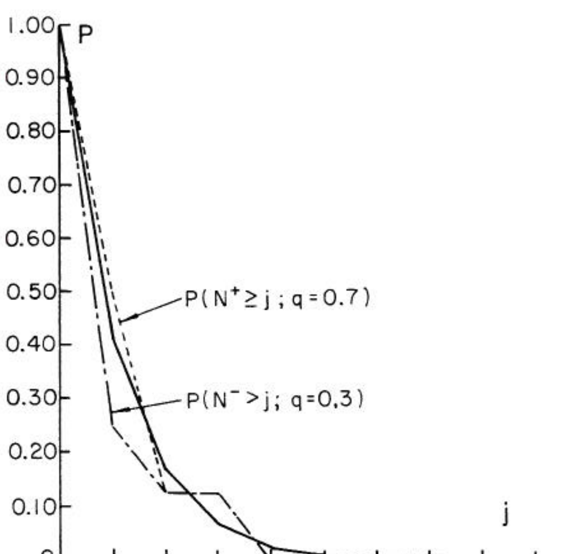
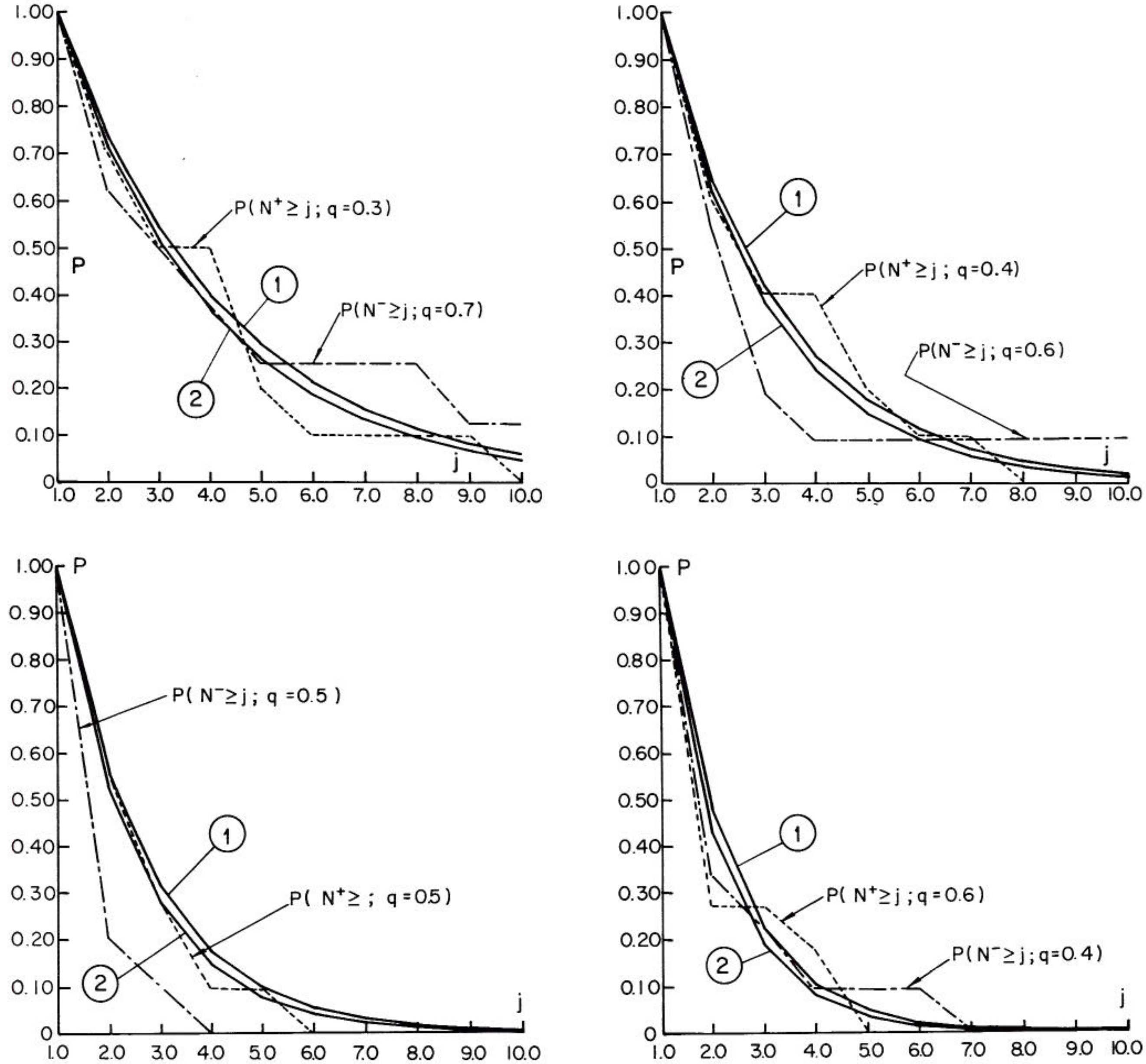


Fig. 7.6 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^+ \geq j \; ; \; q)$, and negative run-lengths, $P(N^- \geq j \; ; \; q)$, for the annual runoff series of Ashley Creek (1914-1957) as compared with expected probabilities (solid lines) of positive and negative run-lengths of a first-order autoregressive process with $\rho_1 = r_1 = 0.274$.



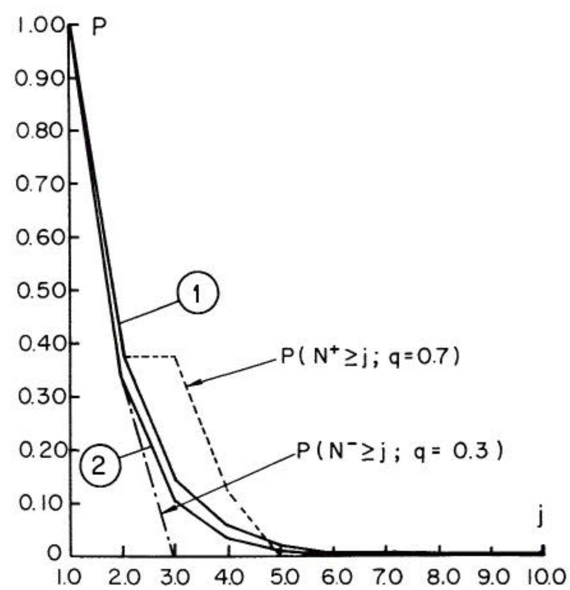


Fig. 7.7 Estimated probabilities by sample relative frequencies (dashed lines) of positive run-lengths, $P(N^+ \geq j \; ; \; q)$, and negative run-lengths, $P(N^- \geq j \; ; \; q)$, for the annual runoff series of Trinity River (1911-1956) as compared with expected probabilities (solid lines) of positive and negative of a first-order auto-regressive process with (1) $\rho_1 = r_1 = 0.180$, and (2) $\rho_1 = r_1^* = 0.080$.

Chapter VIII

CONCLUSIONS

A methodology is developed in this study for using the mean run-length as the parameter for investigating hydrologic series. The basic statistical parameters used are the mean total run-length and the mean positive or negative run-length, as they change with the probability of the truncation level of a series. This study leads to the following conclusions:

(1) The method is effectively used to investigate whether or not a time series of annual precipitation at a point is a sequence of stochastically independent variables with a common distribution.

(2) The method is also effectively used to investigate whether or not a time series of annual river flows is independent or first-order linear autoregressive dependence model.

(3) The method does not depend on the underlying distribution of the variables that are being investigated.

(4) Autoregressive linear models, widely used in hydrology, usually are referred to as stationary Gaussian processes, if their independent stochastic component is normally distributed. 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.

(5) An analytical approximation is developed for probabilities of a sequence of a given length of wet and dry years, when hydrologic time series are stationary, either independent, or first-order linear autoregressive process, and the truncation level is specified. Numerical values of these probabilities are obtained on a digital computer for the range of ρ_1 , the first autocorrelation coefficient between 0 and 0.50, with increments of 0.10, and the range of the probability of truncation level, q , between 0.30 and 0.70 with increments of 0.10, all for the first-order linear autoregressive model. These probabilities can be readily used for probability statements about the multiannual periods of water surplus or deficit, with respect to a specified truncation level that defines the surplus or the deficit.

(6) Probabilities of run-lengths of linearly dependent variables, with a common distribution, do not depend on the underlying, univariate distribution of the variable. They depend only on the probability q of the truncation level and the series dependence model.

(7) Examples of the investigation of stationary series by the run-length technique and examples of the computation of probabilities of lengths of surplus or deficit periods show the relevance of run theory for various applications in hydrologic and water resources investigations.

REFERENCES

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

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

3. Pinkayan, Subin, 1966, Conditional Probabilities of occurrence of wet and dry years over a large continental area. Hydrology Paper No. 12, Colorado State University.

4. Quimpo, Rafael G., 1967, Stochastic model of daily river flow sequences, Hydrology Paper No. 18, Colorado State University.

5. Quimpo, Rafael G., and Yevjevich, V., 1967, Stochastic description of daily river flows. Proceedings of the International Hydrology Symposium 1967, Fort Collins, Colorado

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

7. Saldarriaga, J., 1969, Investigation of wet and dry years by runs, unpublished Ph.D. dissertation, Colorado State University.

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

9. 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.

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

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

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

13. Mood, A.M., 1940, The distribution theory of runs, Annals of Mathematical Statistics, p. 367.

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

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

16. 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.

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

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

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

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

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

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

23. McFadden, J.A., 1955, Urn models of correlation and a comparison with the multivariate normal integral. Annals of Mathematical Statistics Vol. 26, pp. 478-489.

24. 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.

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

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

27. Yevjevich, V., and Jeng, E.I., 1969, Properties of non-homogeneous hydrologic series, Hydrology Paper No. 32, Colorado State University.

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

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

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

31. 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.

APPENDIX

This appendix gives the properties of run-lengths of the first-order linear autoregressive model for five values of the first auto-correlation coefficient, ρ_1, 0.1, 0.2, 0.3, 0.4, 0.5, and for five values of the probability, $q = F(x_0)$, of the truncation level, x_0, or q equal 0.3, 0.4, 0.5, 0.6, and 0.7. The first column gives these parameters: $q = F(x_0)$, $EN^+(q)$, and var $N^+(q)$; the second column gives the discrete value j of the run-length; the third column are probabilities, $P(j^+)$, of at least j consecutive values being above the truncation level, and subsequent ten columns give probabilities of at least k consecutive values being below the truncation level, x_0, followed by at least j consecutive values being above the truncation level, for k = 1,2, ..., 10, and j = 1,2, ..., 10. Finally, the last two columns give probabilities of run-lengths being greater than or equal to j, or being exactly equal to j, respectively. By using the values given in the subsequent five tables, it is feasible to make probability statements about durations of droughts of a river basin, with annual runoff series following the first-order linear auto-regressive process, of ρ_1 estimated by the sample r_1, and by finding $P(N^+ \geq j)$ values for a selected q in this appendix.

q / EN^+ / var N^+	j	$P(j^+)$	$P(1^-, j^+)$	$P(2^-, j^+)$	$P(3^-, j^+)$	$P(4^-, j^+)$	$P(5^-, j^+)$	$P(6^-, j^+)$	$P(7^-, j^+)$	$P(8^-, j^+)$	$P(9^-, j^+)$	$P(10^-, j^+)$	$P(N^+ \geq j^+)$	$P(N^+ = j)$
		$\rho_1 = 0.1$												
	1	.700000	.197744	.067781	.022847	.007705	.002601	.000877	.000295	.000099	.000033	.000011	1.000000	.283162
	2	.502256	.141493	.048273	.016403	.005569	.001886	.000637	.000214	.000072	.000024	.000008	.716838	.201899
0.3	3	.360763	.101655	.034698	.011785	.003998	.001353	.000457	.000154	.000051	.000017	.000006	.514939	.145080
	4	.259107	.073018	.024936	.008465	.002870	.000971	.000327	.000110	.000037	.000012	.000004	.369860	.104217
	5	.186089	.052448	.017918	.006079	.002060	.000696	.000234	.000079	.000026	.000009	.000003	.265643	.074863
3.45	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
7.77	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
	1	.600000	.225022	.098559	.043058	.018601	.008211	.003584	.001563	.000681	.000296	.000128	1.000000	.375750
	2	.374978	.140269	.061300	.026847	.011750	.005137	.002242	.000977	.000425	.000184	.000080	.624250	.233452
0.4	3	.234710	.087828	.038406	.016812	.007353	.003212	.001401	.000610	.000265	.000115	.000050	.390798	.146214
	4	.146881	.054972	.024055	.010524	.004599	.002007	.000874	.000380	.000165	.000071	.000031	.244584	.091529
	5	.091909	.034407	.015061	.006585	.002875	.001254	.000546	.000237	.000103	.000044	.000019	.153055	.057294
2.66	6	.057502	.021534	.009426	.004118	.001797	.000783	.000340	.000148	.000064	.000028	.000012	.095761	.035861
	7	.035968	.013475	.005897	.002575	.001122	.000488	.000212	.000092	.000040	.000017	.000007	.059900	.022442
	8	.022493	.008431	.003688	.001609	.000701	.000305	.000132	.000057	.000025	.000011	.000005	.037459	.014041
4.32	9	.014062	.005274	.002305	.001005	.000437	.000190	.000082	.000036	.000015	.000007	.000003	.023417	.008783
	10	.008788	.003298	.001440	.000627	.000273	.000118	.000051	.000022	.000010	.000004	.000002	.014635	.005492
	1	.500000	.234085	.124204	.066226	.035288	.018802	.010014	.005331	.002837	.001508	.000801	.999990	.468742
0.5	2	.265915	.124204	.065936	.035157	.018736	.009978	.005310	.002824	.001500	.000796	.000422	.531247	.248117
	3	.141711	.066226	.035157	.018736	.009978	.005310	.002824	.001500	.000796	.000422	.000224	.283131	.132318
	4	.075486	.035288	.018736	.009978	.005310	.002824	.001500	.000796	.000422	.000224	.000118	.150813	.070510
2.14	5	.040197	.018802	.009978	.005310	.002824	.001500	.000796	.000422	.000224	.000118	.000063	.080302	.037567
	6	.021395	.010014	.005310	.002824	.001500	.000796	.000422	.000224	.000118	.000063	.000033	.042735	.020008
	7	.011381	.005331	.002824	.001500	.000796	.000422	.000224	.000118	.000063	.000033	.000017	.022728	.010650
2.42	8	.006050	.002837	.001500	.000796	.000422	.000224	.000118	.000063	.000033	.000017	.000009	.012078	.005665
	9	.003213	.001508	.000796	.000422	.000224	.000118	.000063	.000033	.000017	.000009	.000005	.006413	.003011
	10	.001705	.000801	.000422	.000224	.000118	.000063	.000033	.000017	.000009	.000005	.000003	.003402	.001599
	1	.400000	.225022	.139948	.087762	.055010	.034477	.021604	.013534	.008476	.005307	.003321	.999796	.562698
	2	.174978	.098238	.061300	.038406	.024055	.015061	.009426	.005897	.003688	.002305	.001440	.437098	.245534
0.6	3	.076741	.043121	.026847	.016812	.010524	.006585	.004118	.002575	.001609	.001005	.000627	.191564	.107699
	4	.033619	.018904	.011750	.007353	.004599	.002875	.001797	.001122	.000701	.000437	.000273	.083865	.047186
	5	.014715	.008283	.005137	.003212	.002007	.001254	.000783	.000488	.000305	.000190	.000118	.036679	.020659
1.78	6	.006432	.003625	.002242	.001401	.000874	.000546	.000340	.000212	.000132	.000082	.000051	.016020	.009035
	7	.002807	.001584	.000977	.000610	.000380	.000237	.000148	.000092	.000057	.000036	.000022	.006985	.003945
	8	.001223	.000691	.000425	.000265	.000165	.000103	.000064	.000040	.000025	.000015	.000010	.003040	.001720
1.38	9	.000532	.000301	.000184	.000115	.000071	.000044	.000028	.000017	.000011	.000007	.000004	.001321	.000748
	10	.000231	.000131	.000080	.000050	.000031	.000019	.000012	.000007	.000005	.000003	.000002	.000572	.000325
	1	.300000	.197744	.141008	.101459	.072983	.052495	.037756	.027152	.019524	.014037	.010091	.997467	.657289
	2	.102256	.067297	.048273	.034698	.024936	.017918	.012873	.009247	.006641	.004769	.003424	.340178	.224278
0.7	3	.034959	.023040	.016403	.011785	.008465	.006079	.004365	.003134	.002249	.001614	.001158	.115900	.076493
	4	.011919	.007866	.005569	.003998	.002870	.002060	.001478	.001060	.000760	.000545	.000391	.039407	.026040
	5	.004053	.002680	.001886	.001353	.000971	.000696	.000499	.000358	.000256	.000184	.000132	.013367	.008847
1.51	6	.001374	.000910	.000637	.000457	.000327	.000234	.000168	.000120	.000086	.000062	.000044	.004519	.002997
	7	.000464	.000308	.000214	.000154	.000110	.000079	.000056	.000040	.000029	.000021	.000015	.001522	.001012
	8	.000156	.000104	.000072	.000051	.000037	.000026	.000019	.000013	.000010	.000007	.000005	.000511	.000340
0.79	9	.000052	.000035	.000024	.000017	.000012	.000009	.000006	.000004	.000003	.000002	.000002	.000171	.000114
	10	.000017	.000012	.000008	.000006	.000004	.000003	.000002	.000001	.000001	.000001	.000001	.000057	.000038

q / EN^+ / var N^+	j	$P(j^+)$	$P(1^-, j^+)$	$P(2^-, j^+)$	$P(3^-, j^+)$	$P(4^-, j^+)$	$P(5^-, j^+)$	$P(6^-, j^+)$	$P(7^-, j^+)$	$P(8^- j^+)$	$P(9^-, j^+)$	$P(10^-, j^+)$	$P(N^+ \geq j^+)$	$P(N^+ = j)$
$\rho_1 = 0.2$														
	1	.700000	.185157	.072536	.026685	.009846	.003653	.001350	.000495	.000179	.000064	.000023	1.000000	.267114
	2	.514843	.134716	.051845	.019683	.007430	.002772	.001021	.000371	.000133	.000047	.000016	.732886	.191513
0.3	3	.380127	.099662	.038395	.014532	.005464	.002031	.000745	.000269	.000096	.000034	.000012	.541373	.141861
	4	.280464	.073585	.028402	.010716	.004014	.001486	.000543	.000196	.000070	.000025	.000009	.399513	.104777
	5	.206879	.054332	.020986	.007892	.002945	.001086	.000395	.000142	.000051	.000018	.000006	.294736	.077383
3.65	6	.152547	.040111	.015489	.005806	.002159	.000793	.000288	.000103	.000037	.000013	.000004	.217352	.057141
	7	.112435	.029605	.011420	.004267	.001581	.000579	.000209	.000075	.000026	.000009	.000003	.160212	.042181
	8	.082830	.021843	.008410	.003133	.001156	.000422	.000152	.000054	.000019	.000007	.000002	.118031	.031126
8.74	9	.060987	.016110	.006188	.002297	.000845	.000307	.000111	.000039	.000014	.000005	.000002	.086905	.022957
	10	.044877	.011875	.004548	.001683	.000617	.000224	.000080	.000028	.000010	.000003	.000001	.063948	.016923
	1	.600000	.209948	.100319	.047473	.022381	.010566	.004975	.002330	.001085	.000501	.000230	.999999	.352712
	2	.390052	.135102	.064040	.030625	.014575	.006888	.003230	.001502	.000693	.000318	.000144	.647287	.223851
0.4	3	.254950	.088551	.042060	.020044	.009498	.004468	.002086	.000966	.000444	.000203	.000092	.423436	.146985
	4	.166399	.057860	.027561	.013084	.006173	.002891	.001344	.000620	.000284	.000129	.000058	.276451	.096105
	5	.108539	.037811	.018017	.008520	.004003	.001867	.000864	.000397	.000181	.000082	.000037	.180346	.062828
2.84	6	.070728	.024697	.011750	.005534	.002589	.001202	.000554	.000254	.000115	.000052	.000023	.117518	.041046
	7	.046031	.016117	.007644	.003586	.001671	.000773	.000355	.000162	.000073	.000033	.000015	.076471	.026787
	8	.029914	.010505	.004962	.002318	.001076	.000496	.000227	.000103	.000047	.000021	.000009	.049684	.017458
5.09	9	.019408	.006837	.003213	.001495	.000691	.000317	.000145	.000066	.000030	.000013	.000006	.032227	.011360
	10	.012571	.004443	.002076	.000963	.000443	.000203	.000092	.000042	.000019	.000008	.000004	.020866	.007380
	1	.500000	.218169	.121817	.069445	.039396	.022338	.012642	.007133	.004011	.002246	.001253	.999978	.438640
0.5	2	.281831	.121817	.068288	.038919	.022100	.012497	.007035	.003943	.002200	.001222	.000676	.561337	.242472
	3	.160014	.069445	.038919	.022100	.012497	.007035	.003943	.002200	.001222	.000676	.000372	.318865	.138409
	4	.090570	.039396	.022100	.012497	.007035	.003943	.002200	.001222	.000676	.000372	.000204	.180456	.078551
2.29	5	.051174	.022338	.012497	.007035	.003943	.002200	.001222	.000676	.000372	.000204	.000112	.101905	.044531
	6	.028836	.012642	.007035	.003943	.002200	.001222	.000676	.000372	.000204	.000112	.000061	.057374	.025185
	7	.016194	.007133	.003943	.002200	.001222	.000676	.000372	.000204	.000112	.000061	.000033	.032189	.014198
2.94	8	.009061	.004011	.002200	.001222	.000676	.000372	.000204	.000112	.000061	.000033	.000018	.017991	.007974
	9	.005051	.002246	.001222	.000676	.000372	.000204	.000112	.000061	.000033	.000018	.000010	.010017	.004441
	10	.002804	.001253	.000676	.000372	.000204	.000112	.000061	.000033	.000018	.000010	.000005	.005555	.002486
	1	.400000	.209948	.133701	.088243	.058023	.038117	.025006	.016377	.010706	.006985	.004548	.999623	.525757
0.6	2	.190052	.098918	.064040	.042060	.027561	.018017	.011750	.007644	.004962	.003213	.002076	.473867	.247187
	3	.091134	.047734	.030625	.020044	.013084	.008520	.005534	.003586	.002318	.001495	.000963	.226679	.118978
	4	.043400	.022833	.014575	.009498	.006173	.004003	.002589	.001671	.001076	.000691	.000443	.107701	.056803
1.90	5	.020566	.010886	.006888	.004468	.002891	.001867	.001202	.000773	.000496	.000317	.000203	.050898	.027009
	6	.009680	.005158	.003230	.002086	.001344	.000864	.000554	.000355	.000227	.000145	.000092	.023888	.012760
	7	.004522	.002426	.001502	.000966	.000620	.000397	.000254	.000162	.000103	.000066	.000042	.011128	.005983
1.71	8	.002096	.001132	.000693	.000444	.000284	.000181	.000115	.000073	.000047	.000030	.000019	.005145	.002784
	9	.000964	.000524	.000318	.000203	.000129	.000082	.000052	.000033	.000021	.000013	.000008	.002361	.001285
	10	.000441	.000241	.000144	.000092	.000058	.000037	.000023	.000015	.000009	.000006	.000004	.001076	.000589
	1	.300000	.185157	.132602	.098767	.073418	.054536	.040479	.030020	.022244	.016467	.012179	.996068	.614041
0.7	2	.114843	.070421	.051845	.038395	.028402	.020986	.015489	.011420	.008410	.006188	.004548	.382026	.236021
	3	.044422	.027509	.019683	.014532	.010716	.007892	.005806	.004267	.003133	.002297	.001683	.146006	.090853
	4	.016912	.010552	.007430	.005464	.004014	.002945	.002159	.001581	.001156	.000845	.000617	.055153	.034558
1.61	5	.006360	.004003	.002772	.002031	.001486	.001086	.000793	.000579	.000422	.000307	.000224	.020594	.013009
	6	.002357	.001496	.001021	.000745	.000543	.000395	.000288	.000209	.000152	.000111	.000080	.007586	.004830
	7	.000861	.000551	.000371	.000269	.000196	.000142	.000103	.000075	.000054	.000039	.000028	.002756	.001768
0.99	8	.000310	.000200	.000133	.000096	.000070	.000051	.000037	.000026	.000019	.000014	.000010	.000988	.000638
	9	.000110	.000071	.000047	.000034	.000025	.000018	.000013	.000009	.000007	.000005	.000003	.000350	.000227
	10	.000039	.000025	.000016	.000012	.000009	.000006	.000004	.000003	.000002	.000002	.000001	.000123	.000080

$\rho_1 = 0.3$

q EN^+ $var\ N^+$	j	$P(j^+)$	$P(1^-, j^+)$	$P(2^-, j^+)$	$P(3^-, j^+)$	$P(4^-, j^+)$	$P(5^-, j^+)$	$P(6^-, j^+)$	$P(7^-, j^+)$	$P(8^-, j^+)$	$P(9^-, j^+)$	$P(10^-, j^+)$	$P(N^+ \geq j)$	$P(N^+ = j)$
	1	.700000	.172238	.077378	.030391	.012019	.004836	.001927	.000754	.000288	.000108	.000040	1.000000	.251747
	2	.527762	.126762	.054818	.023070	.009553	.003849	.001508	.000576	.000215	.000079	.000028	.748233	.178930
0.3	3	.401001	.097013	.041963	.017502	.007176	.002866	.001114	.000423	.000157	.000057	.000021	.569303	.137560
	4	.303988	.073698	.032006	.013234	.005376	.002129	.000822	.000310	.000115	.000042	.000015	.431743	.104587
	5	.230290	.056000	.024329	.009977	.004018	.001579	.000606	.000227	.000084	.000030	.000011	.327157	.079519
3.87	6	.174290	.042540	.018436	.007500	.002996	.001169	.000446	.000166	.000061	.000022	.000008	.247637	.060431
	7	.131750	.032291	.013930	.005624	.002229	.000864	.000327	.000122	.000044	.000016	.000006	.187206	.045886
	8	.099459	.024487	.010497	.004207	.001655	.000637	.000240	.000089	.000032	.000012	.000004	.141320	.034801
9.83	9	.074972	.018545	.007889	.003140	.001227	.000470	.000176	.000065	.000023	.000008	.000003	.106520	.026358
	10	.056427	.014025	.005916	.002339	.000908	.000345	.000129	.000047	.000017	.000006	.000002	.080162	.019933
	1	.600000	.194779	.101361	.051648	.026017	.013176	.006628	.003295	.001615	.000781	.000373	.999999	.330811
	2	.405221	.128577	.065818	.034373	.017690	.008941	.004438	.002166	.001041	.000494	.000231	.669188	.211363
0.4	3	.276644	.088616	.045523	.023518	.011964	.005982	.002940	.001423	.000679	.000320	.000149	.457885	.146490
	4	.188028	.060421	.031255	.015976	.008040	.003980	.001939	.000931	.000441	.000207	.000096	.311395	.100048
	5	.127607	.041230	.021310	.010782	.005372	.002635	.001273	.000607	.000286	.000133	.000061	.211347	.068326
3.05	6	.086377	.028100	.014435	.007233	.003570	.001736	.000833	.000394	.000185	.000086	.000039	.143022	.046576
	7	.058278	.019102	.009720	.004826	.002362	.001140	.000543	.000256	.000119	.000055	.000025	.096445	.031656
	8	.039175	.012942	.006509	.003204	.001555	.000745	.000353	.000165	.000076	.000035	.000016	.064789	.021437
5.98	9	.026233	.008735	.004337	.002117	.001020	.000485	.000228	.000106	.000049	.000022	.000010	.043352	.014458
	10	.017498	.005871	.002876	.001393	.000667	.000315	.000147	.000068	.000031	.000014	.000006	.028894	.009711
	1	.500000	.202254	.117838	.072157	.043495	.026197	.015694	.009330	.005499	.003213	.001861	.999959	.409691
	2	.297746	.117838	.069554	.042536	.025717	.015368	.009081	.005311	.003076	.001766	.001006	.590268	.233065
0.5	3	.179908	.072157	.042536	.025717	.015368	.009081	.005311	.003076	.001766	.001006	.000569	.357204	.143350
	4	.107751	.043495	.025717	.015368	.009081	.005311	.003076	.001766	.001006	.000569	.000320	.213854	.086508
	5	.064256	.026197	.015368	.009081	.005311	.003076	.001766	.001006	.000569	.000320	.000179	.127346	.052067
2.47	6	.038059	.015694	.009081	.005311	.003076	.001766	.001006	.000569	.000320	.000179	.000099	.075279	.031138
	7	.022366	.009330	.005311	.003076	.001766	.001006	.000569	.000320	.000179	.000099	.000055	.044140	.018471
	8	.013036	.005499	.003076	.001766	.001006	.000569	.000320	.000179	.000099	.000055	.000030	.025670	.010861
3.53	9	.007537	.003213	.001766	.001006	.000569	.000320	.000179	.000099	.000055	.000030	.000017	.014809	.006330
	10	.004325	.001861	.001006	.000569	.000320	.000179	.000099	.000055	.000030	.000017	.000009	.008479	.003659
	1	.400000	.194779	.125163	.087791	.060803	.041984	.028868	.019752	.013446	.009105	.006134	.999369	.489199
	2	.205221	.097947	.065818	.045523	.031255	.021310	.014435	.009720	.006509	.004337	.002876	.510170	.244750
0.6	3	.107275	.052237	.034373	.023518	.015976	.010782	.007233	.004826	.003204	.002117	.001393	.265420	.129842
	4	.055038	.027128	.017690	.011964	.008040	.005372	.003570	.002362	.001555	.001020	.000667	.135578	.067192
	5	.027910	.013971	.008941	.005982	.003980	.002635	.001736	.001140	.000745	.000485	.000315	.068386	.034413
2.05	6	.013939	.007086	.004438	.002940	.001939	.001273	.000833	.000543	.000353	.000228	.000147	.033973	.017354
	7	.006853	.003534	.002166	.001423	.000931	.000607	.000394	.000256	.000165	.000106	.000068	.016619	.008607
	8	.003319	.001733	.001041	.000679	.000441	.000286	.000185	.000119	.000076	.000049	.000031	.008013	.004200
2.08	9	.001585	.000837	.000494	.000320	.000207	.000133	.000086	.000055	.000035	.000022	.000014	.003812	.002020
	10	.000748	.000399	.000231	.000149	.000096	.000061	.000039	.000025	.000016	.000010	.000006	.001792	.000959
	1	.300000	.172238	.121607	.094701	.073237	.056488	.043448	.033324	.025487	.019440	.014787	.994092	.569185
	2	.127762	.072223	.054818	.041963	.032006	.024329	.018436	.013930	.010497	.007889	.005916	.424908	.244550
0.7	3	.055539	.032345	.023070	.017502	.013234	.009977	.007500	.005624	.004207	.003140	.002339	.180358	.106019
	4	.023194	.013763	.009553	.007176	.005376	.004018	.002996	.002229	.001655	.001227	.000908	.074339	.044453
	5	.009431	.005706	.003849	.002866	.002129	.001579	.001169	.000864	.000637	.000470	.000345	.029886	.018191
1.72	6	.003725	.002293	.001508	.001114	.000822	.000606	.000446	.000327	.000240	.000176	.000129	.011695	.007232
	7	.001432	.000894	.000576	.000423	.000310	.000227	.000166	.000122	.000089	.000065	.000047	.004464	.002797
	8	.000538	.000340	.000215	.000157	.000115	.000084	.000061	.000044	.000032	.000023	.000017	.001667	.001056
1.22	9	.000198	.000126	.000079	.000057	.000042	.000030	.000022	.000016	.000012	.000008	.000006	.000610	.000391
	10	.000072	.000046	.000028	.000021	.000015	.000011	.000008	.000006	.000004	.000003	.000002	.000220	.000142

q / EN^+ / var N^+	j	$P(j^+)$	$P(1^-, j^+)$	$P(2^-, j^+)$	$P(3^-, j^+)$	$P(4^-, j^+)$	$P(5^-, j^+)$	$P(6^-, j^+)$	$P(7^-, j^+)$	$P(8^-, j^+)$	$P(9^-, j^+)$	$P(10^-, j^+)$	$P(N^+ \geq j)$	$P(N^+ = j)$
							$\rho = 0.4$							
	1	.700000	.158986	.082409	.033964	.014149	.006120	.002605	.001074	.000429	.000166	.000063	1.000000	.237057
	2	.541014	.117691	.057189	.026564	.011937	.005115	.002098	.000829	.000318	.000119	.000044	.762943	.164159
0.3	3	.423323	.093824	.045400	.020696	.009134	.003858	.001565	.000613	.000234	.000087	.000032	.598784	.132432
	4	.329499	.073337	.035748	.016021	.006957	.002901	.001165	.000453	.000172	.000064	.000023	.466352	.103674
	5	.256162	.057385	.027949	.012333	.005278	.002174	.000865	.000334	.000126	.000046	.000017	.362679	.081208
4.29	6	.198777	.044886	.021715	.009447	.003989	.001626	.000641	.000246	.000092	.000034	.000012	.281470	.063561
	7	.153891	.035058	.016778	.007204	.003005	.001212	.000474	.000181	.000067	.000025	.000009	.217909	.049663
	8	.118833	.027323	.012899	.005472	.002257	.000902	.000350	.000133	.000049	.000018	.000006	.168246	.038711
12.78	9	.091510	.021239	.009873	.004141	.001690	.000670	.000258	.000097	.000036	.000013	.000005	.129535	.030090
	10	.070271	.016462	.007527	.003124	.001263	.000496	.000190	.000071	.000026	.000009	.000003	.099445	.023319
	1	.600000	.179514	.101761	.055605	.029622	.016000	.008539	.004463	.002281	.001141	.000560	1.000000	.309976
	2	.420486	.120765	.066636	.038091	.021095	.011295	.005866	.002968	.001469	.000713	.000341	.690024	.195884
0.4	3	.299721	.088088	.048795	.027236	.014752	.007752	.003965	.001981	.000970	.000467	.000221	.494140	.144940
	4	.211633	.062577	.035136	.019199	.010199	.005273	.002660	.001314	.000637	.000304	.000143	.349200	.103285
	5	.149056	.044579	.024939	.013370	.006983	.003558	.001774	.000867	.000417	.000198	.000093	.245914	.073670
3.32	6	.104477	.031686	.017483	.009215	.004741	.002385	.001176	.000570	.000272	.000128	.000060	.172244	.052365
	7	.072791	.022406	.012125	.006294	.003195	.001589	.000776	.000373	.000177	.000083	.000038	.119880	.037003
	8	.050384	.015740	.008330	.004265	.002139	.001053	.000510	.000243	.000114	.000053	.000025	.082877	.025964
7.34	9	.034645	.010976	.005676	.002870	.001424	.000694	.000333	.000158	.000074	.000034	.000016	.056913	.018082
	10	.023668	.007599	.003838	.001919	.000943	.000456	.000217	.000102	.000048	.000022	.000010	.038831	.012500
	1	.500000	.186338	.112268	.074363	.047512	.030339	.019171	.011941	.007325	.004429	.002642	.999994	.381926
	2	.313662	.112268	.069736	.046009	.029588	.018592	.011449	.006927	.004128	.002428	.001412	.618068	.219907
0.5	3	.201394	.074363	.046009	.029588	.018592	.011449	.006927	.004128	.002428	.001412	.000813	.398160	.147222
	4	.127032	.047512	.029588	.018592	.011449	.006927	.004128	.002428	.001412	.000813	.000464	.250939	.094271
	5	.079520	.030339	.018592	.011449	.006927	.004128	.002428	.001412	.000813	.000464	.000263	.156667	.060100
2.66	6	.049181	.019171	.011449	.006927	.004128	.002428	.001412	.000813	.000464	.000263	.000148	.096567	.037850
	7	.030010	.011941	.006927	.004128	.002428	.001412	.000813	.000464	.000263	.000148	.000082	.058717	.023483
	8	.018069	.007325	.004128	.002428	.001412	.000813	.000464	.000263	.000148	.000082	.000046	.035234	.014350
4.24	9	.010744	.004429	.002428	.001412	.000813	.000464	.000263	.000148	.000082	.000046	.000025	.020884	.008644
	10	.006315	.002642	.001412	.000813	.000464	.000263	.000148	.000082	.000046	.000025	.000014	.012240	.005139
	1	.400000	.179514	.114228	.086332	.063272	.046045	.033195	.023689	.016736	.011710	.008121	.999825	.453591
0.6	2	.220486	.095224	.066636	.048795	.035136	.024939	.017483	.012125	.008330	.005676	.003838	.546234	.238211
	3	.125262	.056614	.038091	.027236	.019199	.013370	.009215	.006294	.004265	.002870	.001919	.308023	.140357
	4	.068648	.031774	.021095	.014752	.010199	.006983	.004741	.003195	.002139	.001424	.000943	.167666	.078337
2.20	5	.036874	.017559	.011295	.007752	.005273	.003558	.002385	.001589	.001053	.000694	.000456	.089329	.042900
	6	.019314	.009443	.005866	.003965	.002660	.001774	.001176	.000776	.000510	.000333	.000217	.046428	.022862
	7	.009871	.004937	.002968	.001981	.001314	.000867	.000570	.000373	.000243	.000158	.000102	.023567	.011855
2.50	8	.004935	.002515	.001469	.000970	.000637	.000417	.000272	.000177	.000114	.000074	.000048	.011712	.005998
	9	.002419	.001253	.000713	.000467	.000304	.000198	.000128	.000083	.000053	.000034	.000022	.005713	.002971
	10	.001166	.000612	.000341	.000221	.000143	.000093	.000060	.000038	.000025	.000016	.000010	.002743	.001444
	1	.300000	.158986	.107823	.089087	.072330	.058297	.046656	.037089	.029299	.023010	.017973	.997330	.526590
	2	.141014	.072541	.057189	.045400	.035748	.027949	.021715	.016778	.012899	.009873	.007527	.470740	.250793
0.7	3	.068473	.037573	.026564	.020696	.016021	.012333	.009447	.007204	.005472	.004141	.003124	.219948	.122496
	4	.030900	.017540	.011937	.009134	.006957	.005278	.003989	.003005	.002257	.001690	.001263	.097452	.055961
	5	.013360	.007830	.005115	.003858	.002901	.002174	.001626	.001212	.000902	.000670	.000496	.041490	.024522
1.86	6	.005529	.003326	.002098	.001565	.001165	.000865	.000641	.000474	.000350	.000258	.000190	.016968	.010268
	7	.002204	.001353	.000829	.000613	.000453	.000334	.000246	.000181	.000133	.000097	.000071	.006700	.004131
	8	.000851	.000531	.000318	.000234	.000172	.000126	.000092	.000067	.000049	.000036	.000026	.002569	.001608
1.47	9	.000320	.000202	.000119	.000087	.000064	.000046	.000034	.000025	.000018	.000013	.000009	.000961	.000609
	10	.000118	.000075	.000044	.000032	.000023	.000017	.000012	.000009	.000006	.000005	.000003	.000352	.000225

q EN^+ $\text{var}\,N^+$	$\rho_1 = 0.5$													
	j	$P(j^+)$	$P(1^-, j^+)$	$P(2^-, j^+)$	$P(3^-, j^+)$	$P(4^-, j^+)$	$P(5^-, j^+)$	$P(6^-, j^+)$	$P(7^-, j^+)$	$P(8^-, j^+)$	$P(9^-, j^+)$	$P(10^-, j^+)$	$P(N^+ \geq j)$	$P(N^+ = j)$
	1	.700000	.145403	.087720	.037423	.016156	.007476	.003377	.001457	.000601	.000239	.000092	1.000000	.222952
	2	.554597	.107538	.058961	.030165	.014583	.006572	.002792	.001132	.000442	.000168	.000062	.777048	.147124
0.3	3	.447059	.090239	.048708	.024114	.011338	.005008	.002097	.000842	.000326	.000123	.000045	.629924	.126770
	4	.356820	.072494	.039629	.019077	.008757	.003800	.001571	.000624	.000240	.000090	.000033	.503154	.102085
	5	.284326	.058426	.031844	.014961	.006725	.002873	.001174	.000462	.000177	.000066	.000024	.401069	.082396
4.69	6	.225900	.047080	.025325	.011646	.005138	.002164	.000875	.000342	.000130	.000048	.000018	.318673	.066450
	7	.178820	.037847	.019964	.009007	.003909	.001625	.000650	.000252	.000095	.000035	.000013	.252223	.053437
	8	.140973	.030311	.015619	.006927	.002961	.001217	.000482	.000186	.000070	.000026	.000009	.198786	.042798
15.46	9	.110662	.024167	.012140	.005301	.002236	.000909	.000357	.000137	.000051	.000019	.000007	.155988	.034117
	10	.086495	.019175	.009381	.004039	.001682	.000677	.000264	.000100	.000037	.000014	.000005	.121871	.027060
	1	.600000	.164154	.101591	.059378	.033112	.018996	.010701	.005843	.003090	.001589	.000797	1.000000	.290142
	2	.435846	.111728	.066493	.041780	.024791	.013951	.007515	.003908	.001976	.000976	.000473	.709858	.177633
0.4	3	.324119	.087055	.051875	.031197	.017861	.009779	.005158	.002639	.001317	.000643	.000309	.532225	.142582
	4	.237063	.064254	.039204	.022752	.012651	.006770	.003507	.001769	.000872	.000422	.000201	.389643	.105759
	5	.172809	.047772	.028905	.016286	.008837	.004637	.002365	.001178	.000575	.000276	.000131	.283884	.078746
3.60	6	.125037	.035398	.020894	.011481	.006102	.003148	.001584	.000780	.000377	.000180	.000084	.205138	.058322
	7	.089639	.026002	.014859	.007991	.004171	.002121	.001054	.000514	.000247	.000117	.000054	.146816	.042777
	8	.063637	.018893	.010425	.005502	.002827	.001419	.000698	.000337	.000161	.000075	.000035	.104039	.031020
8.65	9	.044744	.013570	.007230	.003754	.001902	.000944	.000460	.000220	.000104	.000049	.000022	.073018	.022235
	10	.031174	.009639	.004964	.002540	.001272	.000624	.000301	.000143	.000067	.000031	.000014	.050784	.015741
	1	.500000	.170423	.105106	.076062	.051369	.034728	.023073	.014985	.009513	.005916	.003611	.999999	.355353
	2	.329577	.105106	.068833	.049337	.033712	.022169	.014137	.008793	.005357	.003209	.001894	.644696	.202982
0.5	3	.224472	.076062	.049337	.033712	.022169	.014137	.008793	.005357	.003209	.001894	.001104	.441714	.150092
	4	.148410	.051369	.033712	.022169	.014137	.008793	.005357	.003209	.001894	.001104	.000637	.291622	.101721
	5	.097041	.034728	.022169	.014137	.008793	.005357	.003209	.001894	.001104	.000637	.000364	.189901	.068552
2.88	6	.062314	.023073	.014137	.008793	.005357	.003209	.001894	.001104	.000637	.000364	.000206	.121349	.045298
	7	.039241	.014985	.008793	.005357	.003209	.001894	.001104	.000637	.000364	.000206	.000116	.076051	.029249
	8	.024256	.009513	.005357	.003209	.001894	.001104	.000637	.000364	.000206	.000116	.000065	.046802	.018468
4.98	9	.014743	.005916	.003209	.001894	.001104	.000637	.000364	.000206	.000116	.000065	.000036	.028333	.011428
	10	.008827	.003611	.001894	.001104	.000637	.000364	.000206	.000116	.000065	.000036	.000020	.016905	.006947
	1	.400000	.164154	.100784	.083772	.065349	.050267	.037998	.028224	.020623	.014847	.010548	.999929	.418424
0.6	2	.235846	.090647	.066493	.051875	.039204	.028905	.020894	.014859	.010425	.007230	.004964	.581505	.227150
	3	.145199	.060837	.041780	.031197	.022752	.016286	.011481	.007991	.005502	.003754	.002540	.354355	.150364
	4	.084362	.036760	.024791	.017861	.012651	.008837	.006102	.004171	.002827	.001902	.001272	.203990	.090138
2.38	5	.047601	.021681	.013951	.009779	.006770	.004637	.003148	.002121	.001419	.000944	.000624	.113852	.052466
	6	.025921	.012265	.007515	.005158	.003507	.002365	.001584	.001054	.000698	.000460	.000301	.061386	.029316
	7	.013655	.006664	.003908	.002639	.001769	.001178	.000780	.000514	.000337	.000220	.000143	.032069	.015745
2.90	8	.006991	.003497	.001976	.001317	.000872	.000575	.000377	.000247	.000161	.000104	.000067	.016304	.008204
	9	.003494	.001783	.000976	.000643	.000422	.000276	.000180	.000117	.000075	.000049	.000031	.008101	.004154
	10	.001711	.000887	.000473	.000309	.000201	.000131	.000084	.000054	.000035	.000022	.000014	.003947	.002055
	1	.300000	.145403	.091018	.081704	.070567	.059910	.050102	.041351	.033736	.027245	.021805	.998407	.481749
	2	.154597	.071200	.058961	.048708	.039629	.031844	.025325	.019964	.015619	.012140	.009381	.516637	.252704
0.7	3	.083397	.043205	.030165	.024114	.019077	.014961	.011646	.009007	.006927	.005301	.004039	.263933	.139615
	4	.040191	.021938	.014583	.011338	.008757	.006725	.005138	.003909	.002961	.002236	.001682	.124319	.068896
	5	.018253	.010425	.006572	.005008	.003800	.002873	.002164	.001625	.001217	.000909	.000677	.055422	.031982
2.00	6	.007828	.004626	.002792	.002097	.001571	.001174	.000875	.000650	.000482	.000357	.000264	.023440	.013951
	7	.003202	.001941	.001132	.000842	.000624	.000462	.000342	.000252	.000186	.000137	.000100	.009489	.005781
	8	.001261	.000779	.000442	.000326	.000240	.000177	.000130	.000095	.000070	.000051	.000037	.003707	.002300
1.72	9	.000482	.000302	.000168	.000123	.000090	.000066	.000048	.000035	.000026	.000019	.000014	.001408	.000886
	10	.000180	.000114	.000062	.000045	.000033	.000024	.000018	.000013	.000009	.000007	.000005	.000522	.000332

Key Words: Hydrology, Time Series, Theory of Runs, Run-Lengths, Stochastic Processes in Hydrology, Droughts, Precipitation, Run-Off

A method is developed for investigating time series structure by using the mean run-length parameter. This method is distribution-free. Applications to selected annual precipitation series and annual runoff series demonstrate the feasibility of this method. 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 is either an independent or a dependent stationary series of a variable which follows the first-order linear autoregressive model. Numerical values of probabilities of run-lengths are obtained by the digital computer integration of expansion equations for run-length probabilities of the first-order linear autoregressive model. A set of tables and a set of graphs are presented to make the numerical values readily useable. Probabilities of run-lengths of dependent variables with a common distribution are also distribution free. The significance of this investigation, and several applications in the text, are based on the premise that run-lengths, as statistical properties of time series, represent attractive parameters in studying droughts and surpluses.

Key Words: Hydrology, Time Series, Theory of Runs, Run-Lengths, Stochastic Processes in Hydrology, Droughts, Precipitation, Run-Off

A method is developed for investigating time series structure by using the mean run-length parameter. This method is distribution-free. Applications to selected annual precipitation series and annual runoff series demonstrate the feasibility of this method. 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 is either an independent or a dependent stationary series of a variable which follows the first-order linear autoregressive model. Numerical values of probabilities of run-lengths are obtained by the digital computer integration of expansion equations for run-length probabilities of the first-order linear autoregressive model. A set of tables and a set of graphs are presented to make the numerical values readily useable. Probabilities of run-lengths of dependent variables with a common distribution are also distribution free. The significance of this investigation, and several applications in the text, are based on the premise that run-lengths, as statistical properties of time series, represent attractive parameters in studying droughts and surpluses.

Key Words: Hydrology, Time Series, Theory of Runs, Run-Lengths, Stochastic Processes in Hydrology, Droughts, Precipitation, Run-Off

A method is developed for investigating time series structure by using the mean run-length parameter. This method is distribution-free. Applications to selected annual precipitation series and annual runoff series demonstrate the feasibility of this method. 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 is either an independent or a dependent stationary series of a variable which follows the first-order linear autoregressive model. Numerical values of probabilities of run-lengths are obtained by the digital computer integration of expansion equations for run-length probabilities of the first-order linear autoregressive model. A set of tables and a set of graphs are presented to make the numerical values readily useable. Probabilities of run-lengths of dependent variables with a common distribution are also distribution free. The significance of this investigation, and several applications in the text, are based on the premise that run-lengths, as statistical properties of time series, represent attractive parameters in studying droughts and surpluses.

Key Words: Hydrology, Time Series, Theory of Runs, Run-Lengths, Stochastic Processes in Hydrology, Droughts, Precipitation, Run-Off

A method is developed for investigating time series structure by using the mean run-length parameter. This method is distribution-free. Applications to selected annual precipitation series and annual runoff series demonstrate the feasibility of this method. 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 is either an independent or a dependent stationary series of a variable which follows the first-order linear autoregressive model. Numerical values of probabilities of run-lengths are obtained by the digital computer integration of expansion equations for run-length probabilities of the first-order linear autoregressive model. A set of tables and a set of graphs are presented to make the numerical values readily useable. Probabilities of run-lengths of dependent variables with a common distribution are also distribution free. The significance of this investigation, and several applications in the text, are based on the premise that run-lengths, as statistical properties of time series, represent attractive parameters in studying droughts and surpluses.