PROBABILITIES OF OBSERVED DROUGHTS

by

Jaime Millan and Vujica Yevjevich

June 1971

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

50

PROBABILITIES OF OBSERVED DROUGHTS

by

Jaime Millan

and

Vujica Yevjevich

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80521

June 1971 No. 50

ACKNOWLEDGEMENTS

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

The National Center for Atmospheric Research, NCAR, sponsored by the National Science Foundation, NSF, is also acknowledged for some limited use of its CDC 6600 computer in the investigations leading to this paper. The Colorado State University CDC 6400 computer was used most of the time.

Acknowledgement also goes to Dr. D.C. Boes, Associate Professor of Statistics at Colorado State University, for his advice in connection with some of the statistical problems.

The cooperation of Mr. Pedro Guerrero, Ph.D. graduate student, in the computing phase of the study is appreciated.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF SYMBOLS

Symbol	Meaning
a, b, c, d, e	Coefficients of multiple regression equations
D	Negative run-sum or deficit corresponding to the longest run-length L_m
D_m	Largest negative run-sum in a sample of size N
D_{50}	D_m for truncation level q equal 0.5
E	Expectation
F(x)	Probability distribution function of x , also cumulative frequency distribution of x
F(x,y)	Joint cumulative frequency distribution of x and y
g	Sample skewness coefficient
L	Negative run-length corresponding to D_m
L_m	Longest negative run-length in a sample of size N
m	Number of squared normal deviates to add in order to get a gamma deviate
m_1, m_2, m_3, m_4	Number of cases in the simulation
m_r	Sample mean run
N	Sample size
N_r	Representative sample size
N(0,1)	Symbol for a standardized normal variable
0(1)	Of the order of 1
P_e	Probability obtained by the experimental method
P_i	Probability obtained by linear interpolation from figures of Appendix II
p	Probability of drowning an element of kind 0 in a binomial population, also $1 - F(x_o)$ or $1 - q$
q	Probability of drowning an element of kind 1 in a binomial population, also $F(x_o)$ or $1 - p$
$\hat{q}$	Sample q
R	Multiple correlation coefficient
R^2	Coefficient of determination
r	Sample autocorrelation coefficient of log 1
S_{50}	Largest positive run-sum at truncation level $q = 0.5$
t_j	Symbol for a standardized normal variable, also N(0,1)
u	Longest run length in a sample of size N
x_o	Truncation level
$\bar{x}$	Sample mean, also μ_x
γ	Skewness coefficient

LIST OF SYMBOLS (cont'd)

Symbol	Meaning
γ_x	γ for ε_i
γ_ε	γ for series x
ε_i	Independent random component with zero mean and unit variance
ΔF	Critical deviation for the Kolmogorov-Smirnov test
$(\Delta F)_{max}$	Maximum deviation for the Kolmogorov-Smirnov test
Δq	Difference between q and $\hat{q}$
$\Delta\gamma$	Difference between γ and g
$\Delta\rho$	Difference between ρ and r
η	Coefficient of variation
η_ℓ	η for the distribution of longest negative run-length
η_s	η for the distribution of largest negative run-sum
μ	Population mean
μ_n	Mean of the logarithms, parameter of the log-normal distribution
μ_x	Sample mean for series x, also $\bar{x}$
$(\mu_n)_\ell$	μ_n for the longest negative run-length
$(\mu_n)_s$	μ_n for the largest negative run-sum
ν	Degrees of freedom for the chi-squared distribution
ρ	Lag one autocorrelation coefficient (population value)
σ	Population standard deviation
σ_n	Standard deviation of the logarithms
$(\sigma_n)_\ell$	σ_n for the longest negative run-length
$(\sigma_n)_s$	σ_n for the largest negative run-sum

ABSTRACT

A method is presented for computing the probability or recurrence period of historical droughts by using the longest negative run-length and the largest negative run-sum as basic parameters of samples of a given size, and by using a given probability of the truncation level, a given autocorrelation coefficient, and a given skewness coefficient. The application of this method to selected annual runoff and precipitation series demonstrate its feasibility. The statistical experimental method in generating large numbers of samples is used to compute frequency distributions as the estimates of probability distributions of the longest negative run-length and of the largest negative run-sum in a sample of size N, as descriptors of the largest historical droughts, for normal and nonnormal independent and dependent stationary stochastic processes which follows the first-order linear autoregressive model. Experimentally obtained values are checked with theoretical results for the distribution of the longest negative run-length when the observations are independent. A set of graphs and a set of tables are presented to make the numerical values readily usable. Good approximations for practical computations are demonstrated by fitting lognormal probability distributions to the experimentally obtained frequencies.

PROBABILITIES OF OBSERVED DROUGHTS

by

Jaime Millan* and Vujica Yevjevich**

Chapter I

DEFINITION OF PROBLEMS INVESTIGATED

1.1 Objectives. The first objective of the present study is to determine the probabilities of historic hydrologic droughts. The second objective is to find the relations of these determined probabilities to statistics of corresponding hydrologic time series. This paper presents information in the form of equations, tables, and graphs that permit drought probabilities and the relations between drought magnitudes and the statistics estimated from time series to be quickly determined.

1.2 Significance. In the past, standard practice for designing reservoirs relied heavily on the "critical period," defined as that period in time when the historic record would have been most critical with respect to water demands required from a system. It is claimed [1] that design based on a critical period results in a reservoir storage capacity equal to the capacity obtained by using the total length of record. However, to determine this critical period accurately reliable knowledge of system performance, particularly demand patterns and operational rules and policies, is required. In the absence of this knowledge or because of complexity in obtaining this kind of information, the critical drought period is usually determined under simplified assumptions. Even though some current design practices take into account not only the critical drought period but also the total deficit of water supply by a reservoir under study, it still remains that this critical drought period represents, in most cases, the largest part of the deficit allowed by these design criteria.

W. Hall and A. T. Askew [1] found for 25 selected rivers across the continental United States that the dates of the critical periods agree with the dates of the major droughts in each region. Using this information, in general, a historic drought is considered that event for which most designs must perform satisfactorily. This is based on the assumptions that the most severe drought to be observed during the lifetime of a project will be about the same as a previously recorded maximum historic drought. The probability, however, that the critical drought period observed in the past will be the same as the critical drought period expected to be observed in the future is usually small. The probability is large that a very different drought will be observed.

The sample parameter describing the observed drought in a period of N years is a random variable like any sample statistic. Its distribution must be known before statistical inferences about the probability of exceedence or nonexceedence of the magnitude of that drought parameter can be made. The recurrence interval (return period) of a drought is often derived from the relative frequencies of historic drought records. This approach is unreliable because of the large sampling fluctuation of these frequencies. A much more accurate method is required for these estimations, one that is based on the properties of the structural mathematical model of a time series.

Because an unusually large or an unusually small value of a drought parameter may occur in the data of the historic sample by chance, the assessment of the probabilities of such droughts has a practical significance. This assessment requires the definition of a representative drought for a descriptive parameter of the drought and a given sample size. In this study the mean value of a drought parameter is defined as the representative drought.

1.3 Statistical Parameters Defining the Droughts Relevant to this Study. The definitions of drought parameters used in this investigation refer to statistics of samples of given sizes and not to the populations from which samples are derived. Therefore, interest is in finding the probabilities of exceedence, or nonexceedence, of a drought descriptor in a sample time series of size N, for a given type of hydrologic process. Logically, these probabilities of sample values of drought parameters are closely related to population characteristics.

The statistical definitions of droughts refer only to stationary stochastic processes, or with no trends, slippages (positive or negative jumps), or periodicities present in these processes. These conditions are met only by the homogeneous and consistent (without systematic errors) discrete time series of annual values of major hydrologic random variables. As soon as the discrete series refer to time units smaller than the year, periodicities in various parameters of these series complicate the analysis. This case of periodic-stochastic processes of hydrologic time series is outside the scope of this study, since the emphasis in this case is only on the annual time series.

Later in the text it is shown that the exact probabilities of some particular drought descriptors to be or not to be exceeded in a sample of size N may be derived only for independent processes. However, many hydrologic time series are rather dependent, either normal dependent or nonnormal independent or dependent processes, for which the available exact probability distributions are not applicable. Therefore, an experimental statistical (Monte Carlo) method is used to derive the properties of sample

*Ph.D. graduate student, 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, Fort Collins, Colorado.

drought parameters of dependent normal and independent or dependent nonnormal processes. To assess the reliability of this experimental statistical method, the known exact probability distributions of a simple descriptor are compared with the distributions experimentally determined.

1.4 Runs as an Objective Definition of Droughts. Runs are an objective definition of droughts [2]. Runs of the sequence of a stochastic variable (or a combination of stochastic and deterministic-periodic components constituting a composite sequence) also may be defined in various ways. Figure 1 represents a discrete and a continuous series of a variable x . By selecting an arbitrary value x_o the continuous series is truncated at many positive and negative discrete deviations. The parameter x_o , or the truncation level, can be any predeterminated level and is usually expressed as a function of the quantile, q , with $q = P[x \leq x_o]$. This level does not need to be a constant x_o , because it may be a deterministic, a stochastic, or a combined deterministic-stochastic process.

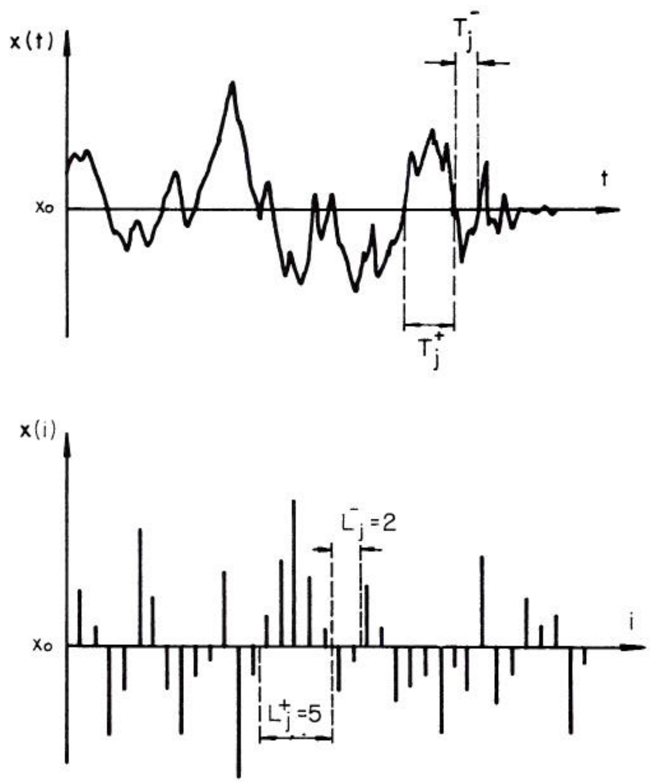


Fig. 1. Definitions of positive and negative runs for a truncation level x_o , with $P(x \leq x_o)=q$, for a discrete (lower graph) and a continuous (upper graph) time series.

Various definitions of runs have been used in hydrologic literature, with the two major run definitions, as descriptors of droughts, evident in Fig. 1. These runs may be directly associated with drought properties: (1) the distance between the successive downcross and upcross as the negative run-length for a continuous series, or the length of the uninterrupted sequence of negative discrete deviations, $x_i - x_o$; and (2) the integral of negative deviations between the successive downcross and upcross as the negative run-sum, for a continuous series, or the sum of negative discrete deviations, $x_i - x_o$, of an uninterrupted sequence for a discrete series. The negative run-length can be associated with the total length or duration of drought measured with respect to a given x_o , which does not need to be a constant but must be a function of time. Since the negative run integral or sum as the measure of the deficit is of more relevance for water resources problems, it is given special consideration in this study.

The critical drought period for a sample of length N , as it is used by many investigators in water resources computations, usually coincides with the largest deficit in record. As a consequence the largest deficit, as the negative run-integral or negative run-sum, and the longest negative run-length during a period of N years as the measure of the duration of critical drought period, are random variables whose probability distributions are of interest to hydrologists, mainly for determining the recurrence interval of the rare events.

1.5 Theoretical Background. The run-length, as the descriptor of drought durations, has been more widely studied because its treatment is simpler. Saldarriaga and Yevjevich [3] have presented a review of literature giving the exact properties of distributions of run-lengths for independent random variables, showing these probabilities to be independent of the underlying distributions. When observations follow the first-order autoregressive (Markov) linear model of dependence, the distributions in the power series expansion form are presented in reference [3]; then they are integrated in an approximate form, checked by the experimental statistical method, and finally presented as a series of graphs and tables.

The distribution of the run-sum (with the run-sum studied because only the discrete series are dealt with in the ensuing text) is more complex to obtain even for the independent normal process. For this case R. N. Downer, M. M. Siddiqui, and V. Yevjevich [4] obtained the exact properties of run-sums by using the cumulants of this process. The first few moments of the distribution of run-sums can also be obtained by using the crossing theory; however, there is no method, known to the writers of this paper, of obtaining the exact properties of run-sums for normal dependent or non-normal independent or dependent random variables.

The distribution of the length and size of the critical drought period (or the longest run-length and the largest run-sum for a given x_o in a period of N years) is more difficult to obtain and few references are available in the literature. Cramer [5] gives the asymptotic mean of the distribution of longest run-length in the sample of size N as

$$E(u) = - \frac{\log N}{\log (1-q)} + 0 \ (1) \quad , \qquad (1)$$

in which N is the sample size, q is $P[x \leq x_o]$, u is the longest run-length in the sample, and $0(1)$ is the error of the order of one.

The theory of recurrence events as applied to the computation of the probability distribution of the longest negative run-length in a sample of size

N for the independent Bernoulli trials is given by Feller [6, p. 322]. For q the probability of negative deviations $x_i - x_o$, with $x_i < x_o$, and $p = 1-q$ the probability of positive deviations $x_i - x_o$, with $x_i > x_o$, Feller gives the probability that the first negative run-length of size r occurs at the N-th trial, as an approximation,

$$f_N(r) = \frac{(y - 1)\ (1 - qy)}{[(r + 1) - ry]\ p} \cdot \frac{1}{y^{N+1}} \tag{2}$$

with

$$y = 1+pq^r + (r+1)(pq^r)^2+(r+1)^2(pq^r)^3+ \ldots \quad . \tag{3}$$

According to Feller the probability for no run of size $r+1$ or greater to occur in N trials is equivalent to the probability of the $(r+1)$-th run occurring for the first time in the sample sizes of $N+i$, $i = 1,2,\ldots,\omega$. This probability is then

$$P(u \leq r) = \sum_{i=1}^{\infty} f_{N+i}(r+1) = F_N(r) \quad . \tag{4}$$

By substituting Eq. (2) into Eq.(4), this probability becomes

$$F_N(r) = \frac{1-qy}{[r+2 - (r+1)y]\ q} \cdot \frac{1}{y^{N+1}} \quad . \tag{5}$$

Though Eqs. (2) through (5) are approximations, they may serve as the control of how the experimental statistical approach of the Monte Carlo method reproduces the probabilities of longest negative run-lengths of given size r to occur in a sample of size N , with $r < N$.

1.6 Experimental Approach for Computing Probabilities of Historic Droughts. As shown by Eqs. (1) through (5), even in the case of independent Bernoulli trials the probabilities of the longest negative run-length, to occur in the sample of size N , cannot be obtained exactly but can be only approximated. The accuracy of computed probabilities depends on the number of terms in the polynomial of Eq. (3) taken into computation. When the time series is dependent, the computation of probability of the longest run-length, r , to occur in the sample of size N are not available, to the writers' knowledge, in an explicit form of successive terms of various degrees of accuracy, as was done for the independent Bernoulli trials. The case of a largest run-sum not to exceed a given value in a sample of size N is still more complex to compute by an approximate method in the case of an independent normal process, and still more complex to determine in the case of series with various types of normal dependent and independent or dependent nonnormal processes. When faced with this difficulty the investigator must turn to the experimental method to obtain required results. In the following chapters techniques are presented for generating experimentally many equally likely realizations of the stochastic process and for computing distributions of the statistics of practical interest.

Chapter II

THEORETICAL BACKGROUND AND COMPUTATIONAL METHODS FOR DETERMINING DROUGHT PROBABILITIES

2.1 Selecting Parameters to Use for Investigating Droughts by the Data Generation Method. With runs accepted as the objective definition of drought descriptors it remains to choose which particular runs are of practical interest. As pointed out in Chapter 1, one objective is to determine the probability distributions of parameters of the largest drought in a sample. These drought parameters usually measure the drought duration (length) and its total deficit as the sum of negative uninterrupted deviations. These two parameters and their mutual relations are investigated subsequently. Since the total water deficit may be more critical in water resources problems than the drought duration, it was decided to investigate deficit as the primary parameter, while including both its relation to the duration parameter and the investigation of duration parameter itself.

A computer program was set up in such a way that the joint probability distribution of the longest run-length and its corresponding run-sum in the sample is first obtained from the generated samples as $F_N(L_{max}, D)$, with L_{max} the longest run-length for a given sample size N, and D the corresponding deficit of this run-length. Because L_{max} is a discrete variable while D is a continuous variable, class intervals equally spaced as states are used for D parameter, so that the problem is reduced to a joint distribution of two discrete variables. Just as for L_{max} and D, the joint probability distribution of the largest deficit and the corresponding duration are obtained from the generated samples as $F_N(D_{max}, L)$, with D_{max} the largest run-sum as the deficit for a given sample size N, and the corresponding run-length L as the duration. Again, the continuous variable D_{max} is divided into discrete states as equal class intervals, so that both D_{max} and L are considered as discrete random variables.

Both distributions, $F(L_{max}, D)$ and $F(D_{max}, L)$, which are subsequently designated as $F(L_m, D)$ and $F(D_m, L)$ respectively for this study, depend on the selection of four basic selective parameters: (1) q, the quantile probability (or the corresponding crossing level, x_o); (2) N, the sample size; (3) ρ, the population first serial correlation coefficient as the parameter of the assumed first-order autoregressive linear model, and (4) γ, the population skewness coefficient of a nonnormal independent or dependent stochastic process.

The two basic joint distributions, $F(L_m, D; q, N, \rho, \gamma)$ and $F(D_m, L; q, N, \rho, \gamma)$, permit, then, the computation of the following derived distributions:

1. $F(L_m; q, N, \rho, \gamma)$, as the marginal distribution of the longest run-length or drought duration;

2. $F(D_m; q, N, \rho, \gamma)$, as the marginal distribution of the largest run-sum or drought deficit;

3. $F(D|L_m; q, N, \rho, \gamma)$, as the conditional probability of drought deficit given the longest drought duration;

4. $F(D|L_m = \ell; q, N, \rho, \gamma)$, as the conditional probability of drought deficit given the longest run is equal to a given duration, ℓ;

5. $F(L|D_m; q, N, \rho, \gamma)$, as the conditional probability of the drought duration given the largest drought deficit;

6. $F(L|D_m = d; q, N, \rho, \gamma)$, as the conditional probability of the drought duration given the largest run-sum is equal to a given drought deficit;

7. $F(L_m = \ell|D = d; q, N, \rho, \gamma)$, the conditional probability of L_m being ℓ provided the drought deficit is a value d, and

8. $F(D_m = d|L = \ell; q, N, \rho, \gamma)$, as the conditional probability of the largest drought deficit being a given value d provided the drought duration is a value ℓ.

The problems at hand would decide which of these eight marginal and conditional variables and distributions would be used for a particular practical case of application.

2.2 Mathematical Models and Their Estimated Parameters of Time Series Structure. The generation of new hydrologic samples that preserve population characteristics (in the form of preserving mathematical models and their statistics of an available sample as the estimates of population models and parameters) is well known in hydrologic literature [7, 8, and others]. In every particular case, it is necessary to identify the models, and parameters in the form of sample statistics, that should be preserved to determine experimentally the approximate distributions of runs by the data generation method.

It has been shown [3] that the population run-length properties for stationary processes are independent of the mean and the standard deviation, but they are dependent on the truncation level q, the population dependence structure of a series, and the skewness of a population distribution. The same cannot be said for the population run-sum properties except that their magnitude is directly proportional to the standard deviation of the process. Once the run sum for the standardized variable $(\sigma = 1)$ is known, the run sum for any $\sigma \neq 1$ can be obtained by multiplying the run-sums of the standardized variable by σ of the nonstandardized variable. Therefore, the generated long samples with $\mu = 0$

and $\sigma = 1$, having a given truncation level (x_0) and measured by q, a given sample size N, a given time dependence model measured by ρ, and a given skewness coefficient γ, can be used for the analysis of probability distributions of runs, thus covering the situations most likely to occur under practical conditions.

Previous studies [9] support the thesis that the dependence structure of annual flows of most rivers in the world can be approximated by the first order autoregressive linear model. The general model is

$$x_i - \mu_x = \rho\,(x_{i-1} - \mu_x) + \sqrt{1-\rho^2}\,\sigma_x\,\varepsilon_i \quad , \qquad (6)$$

in which ε_i is an independent random component with zero mean and unit variance, and independent of x_{i-1}; μ_x is the population mean; σ_x is the population variance, and ρ is the first serial correlation coefficient of the x_i -series.

2.3 Obtaining the Independent Random Numbers by the Data Generation Method. When the skewness coefficient is zero and the kurtosis coefficient three, generating ε_i from a normal population probability distribution, $N(0,1)$, preserves the desired statistics.

When it is necessary to generate the independent random numbers with the skewness coefficient different from zero, it is convenient to use different population probability distributions, generally either Gamma or lognormal [8]. In choosing between a Gamma and a lognormal distribution, the Gamma distribution is used in this study because it was more convenient than the lognormal distribution. For values of skewness $0 < \gamma \leq 0.50$, the approximation used by Thomas and Fiering [10], as summarized in Appendix I, is used. This approximation is identical to the approximation given in the Handbook of Mathematical Functions [11].

If t_i are independent standard random numbers, $N(0,1)$, then they can be transformed into dependent random numbers ε_i, which follow a distribution with the skewness coefficient γ; this distribution is almost like the Gamma distribution with the first-order dependence measured by ρ. This Gamma transform is

$$\varepsilon_i = \frac{2}{\gamma_\varepsilon}\left(1 + \frac{\gamma_\varepsilon t_i}{6} - \frac{\gamma_\varepsilon}{36}\right)^3 - \frac{2}{\gamma_\varepsilon} \quad , \qquad (7)$$

where

$$\gamma_\varepsilon = \frac{(1-\rho^3)}{(1-\rho^2)^{3/2}} \cdot \gamma_x \quad , \qquad (8)$$

in which γ_x is the skewness to be preserved. For the proof of this transform see Appendix I. This approximation is good for $\gamma_x \leq 0.50$. For values $\gamma_x > 0.50$ it is necessary to use a much more time consuming but exact procedure used by Yevjevich [10]:

$$\varepsilon_i = \frac{1}{2}\sum_{j=1}^{m} t_j^2 \quad , \qquad (9)$$

and

$$x_i = \sqrt{\rho}\, x_{i-1} + \sqrt{1-\rho}\, \varepsilon_i \quad , \qquad (10)$$

in which ε_i is Gamma distributed with the mean $\mu = m/2$, with the variance $\sigma_\varepsilon^2 = m/2$, the first serial correlation coefficient ρ, and t_j are independent standard normal random numbers, $N(0,1)$.

To generate the standard Gamma dependent random numbers with the mean zero and the variance unity, the transformation is

$$y_i = \sqrt{\frac{2}{m}}\, x_i - \sqrt{\frac{m}{2}} \quad . \qquad (11)$$

2.4 Investigated Cases. For the four varying parameters, q, N, ρ, and γ, it was necessary to select several cases in the study for each of them. If the numbers of cases are m_1, m_2, m_3, and m_4 for each, respectively, then the total number of cases is $m_1 m_2 m_3 m_4$. Besides, the number of samples to be generated for each case out of $(m_1 m_2 m_3 m_4)$ -- cases must be determined.

In reference to the truncation level, x_o, the level used in the study of droughts is best expressed in the form of the quantiles of x_o, as the q -values, with $q = P(X \leq x_o)$. The selected number of q is $m_1 = 4$, with $q = 0.50$, 0.40, 0.30, and 0.20.

The selected number of sample sizes, N, is $m_2 = 5$, or with $N = 25$, 50, 100, 200, and 500 years. Although the values 25 and 50 cover the current samples in hydrology, the samples of sizes 100, 200, and 500 are used to determine experimentally the distributions of run parameters, so that the unrepresentatively large historical droughts may be referred to as samples of a larger size.

The selected number of the first serial correlation coefficient, ρ, is $m_3 = 5$, with $\rho = 0.00$, 0.10, 0.20, 0.30, and 0.70. The first four values are selected because they are the values most commonly found in practice for annual river flow series, and the last one was selected to study the effects of high dependence in series on properties of run parameters of samples.

The selected number of skewness coefficients, γ, is $m_4 = 4$, with $\gamma = 0.00$, 0.20, 0.50, and 1.00. It is assumed that annual series have the population skewness coefficients $\gamma \geq 0$ only. If a sample has a negative skewness coefficient, it is assumed that it does not differ significantly from $\gamma = 0$. This range of $0 \leq \gamma \leq 1.00$ covers most of the cases for annual time series of hydrologic variables of interest to drought analysis. In summary, the total number of cases selected for the study is $m_1 m_2 m_3 m_4 = 400$.

For a standardized series with $\mu = 0$ and $\sigma = 1$, for an approximate Gamma distribution in case $\gamma > 0.20$, and a normal distribution approximation in case $0 \leq \gamma \leq 0.20$, Table 1 gives the variable truncation values, t_o, for the four values of q and the four values of γ.

In using the generation of Gamma dependent random numbers by Eqs. (8) and (9), the value m in Eq. (8) is $m = 8$ for $\gamma = 1.00$ for a given ρ.

The results of the selected 400 cases studied are presented in a series of graphs. Besides covering the situations most often needed for practical conditions, the graphs permit an easy interpolation within the selected ranges of four parameters, as well as a limited range of extrapolation on one or both sides of these ranges.

TABLE 1

TRUNCATION VALUES, x_o, OF STANDARDIZED VARIABLE FOR FOUR VALUES OF q AND FOUR VALUES OF γ

γ \ q	0.20	0.30	0.40	0.50
0.00	-0.841	-0.524	-0.253	0.000
0.20*	-0.841	-0.524	-0.253	0.000
0.50	-0.8565	-0.5784	-0.3279	-0.0830
1.00	-0.8516	-0.6161	-0.3943	-0.1639

* For $\gamma = 0.20$, the truncation level, t_o, is taken the same as for $\gamma = 0.00$, with all differences being very small.

2.5 Selecting the Number of Samples to be Generated. The central limit theorem leads to the conclusion that the distribution of sample mean run, m_r, is asymptotically normal $N(\mu_r, \sigma_r/n)$, in which $\mu_r = E(m_r)$, σ_r is the variance of the sample run, and n is the number of samples from which runs and the mean run, m_r, are computed. Since one objective in this study is to determine the number n of samples of a given size N to be generated in experiments in such a way that the probability is at least 0.95 for the estimate m_r to be within the tolerance limits $\mu_r \pm \sigma_r/10$, then

$$P[(\mu_r - \frac{\sigma_r}{10}) \leq m_r \leq (\mu_r + \frac{\sigma_r}{10})] \geq 0.95$$

becomes

$$P[-\frac{\sqrt{n}}{10} \leq \frac{(m_r - \mu_r)\sqrt{n}}{\sigma_r} \leq \frac{\sqrt{n}}{10}] \geq 0.95 .$$

Putting $z = (m_r - \mu_r)\sqrt{n}/\sigma_r$, z is normally distributed, $N(0,1)$, so that at least

$$P(z \leq -\sqrt{n}/10) - P(z \geq \sqrt{n}/10) = 0.95 ,$$

or

$$P(z \leq -\sqrt{n}/10) = 0.025$$

is correct for $-\sqrt{n}/10 = -1.96$, or $n = 400$.

Since one value of a given run is obtained from a generated sample of size N, the accuracy in the determination of the mean run increases with an increase of the number n of samples. In this way generating a total of nN observations is necessary. This number nN is selected as a constant, $m = nN$, in this study and is $m = 95{,}000$. For the selected sizes, N, of samples, as $N = 25, 50, 100, 200$, and 500, the number of samples, n, becomes m/N, or

N	25	50	100	200	500
n	3800	1900	950	475	192

The reason for more samples of size N for $N = 25, 50, 100$, and 200, results from the need to generate at least $n = 400$ samples for $N = 200$, and about $n = 200$ for $N = 500$. Once the $m = nN$ random numbers are generated for $N = 500$, they are all used for the smaller values of N in order to increase the accuracy of estimating distributions of runs and of their general statistical parameters.

Chapter III

DISTRIBUTIONS OF DROUGHT PARAMETERS OBTAINED FROM GENERATED SAMPLES

The probabilities of run-lengths and run-sums, as defined in Chapter I, are obtained by using the experimental statistical (Monte Carlo) method in generating a multitude of samples for variables of given characteristics. The obtained probability frequency distributions are presented in this chapter and in Appendix II.

3.1 Verification that Distributions Obtained by the Experimental Method Converge to Exact Distributions. It is often convenient to verify how well the probability distributions of sample statistics are estimated by their cummulative frequency distributions obtained by generating a large number of samples of a given process by the experimental statistical method (in the subsequent text this method is called the experimental method).

In the numerical approximate integration of differential equations, a case is usually selected for which the exact solution is known so that the results of the approximate solution can be compared with the exact solution and thereby verifying the approximation. Similarly, the experimental method is an approximate method of estimating properties of sample statistics in the form of their approximate sampling frequency distributions. By selecting a known exact probability distribution of a sample statistic, and by experimentally determining its sampling cummulative frequency distribution, insight can be obtained on how well the experimental distribution approximates the exact distribution for a given number of generated samples of a process.

The distribution of the longest negative run-length to occur in a sample of N years for independent observations can be obtained by using the exact distributions given in Chapter I. Figure 2 shows the exact probability distribution of the longest run-length in a sample of size N for the independent standard normal variable, computed by Eq. (5) for the crossing level $q = 0.50$ and five values of N (25, 50, 100, 200, and 500), and indicated as solid lines in Figure 2. For the same five values, and indicated as dashed lines, the cummulative sampling frequency distributions of the longest run-length to occur in the sample of size N for $q = 0.50$ are also given in Figure 2 as the results of experimental method. The numbers n of generated samples for N = 25, 50, 100, 200, and 500, are respectively n = 3800, 1900, 950, 475, and 192. Visual inspection shows that the exact and experimentally determined distributions are essentially identical, through, as it should be expected, the deviations between these distribution curves increase as the number n of generated samples decreases with an increase of the sample size, N .

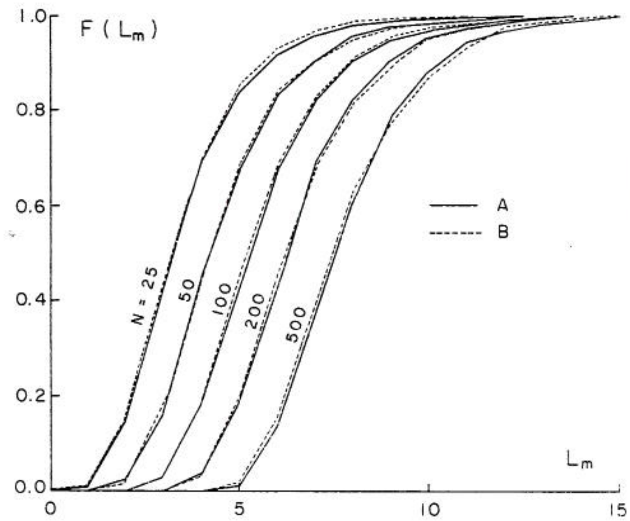


Fig. 2. Comparison of the exact sampling distributions (solid lines) and the experimentally determined frequency distribution (dashed lines) of the longest run-length, L_m , of a standard normal independent variable for the crossing level $q = 0.50$ and for five sample sizes N = 25, 50, 100, 200, and 500.

For an objective assessment of the proximity between the theoretical and experimental distributions, the means of these distributions are compared in Table 2.

TABLE 2

COMPARISON OF THE EXPECTED VALUE, E(x), OF EXACT DISTRIBUTIONS AND THE SAMPLE MEAN, $\overline{x}$, OF EXPERIMENTAL DISTRIBUTIONS FOR LONGEST NEGATIVE RUN-LENGTH OF INDEPENDENT PROCESSES FOR $q = 0.50$.

N	E [x]	$\overline{x}$	$\Delta\overline{x} = E(x) - \overline{x}$
25	3.99	3.97	0.02
50	4.99	5.00	-0.01
100	5.99	6.06	-0.07
200	6.98	7.03	-0.05
500	8.30	8.25	0.05

The differences, $\Delta\overline{x} = E(x) - \overline{x}$, are very small, ranging from 0.2-1.2 percent of $E(x)$, and they increase with an increase of N (or a decrease of n), as expected.

Kolmogorov-Smirnov tests were performed to determine the proximity of the theoretical and experimental distributions at the ten percent and one percent levels of significance. All tests give insignificant testing statistics, as summarized in Table 3 in which N is the sample size, $(\Delta F)_{max}$ is the maximum difference of probabilities of the two distribution curves, and ΔF's are the critical values of Kolmogorov-Smirnov statistics for given α -values.

TABLE 3

KOLMOGOROV-SMIRNOV TESTS OF PROXIMITY BETWEEN EXACT AND EXPERIMENTAL DISTRIBUTIONS OF LONGEST NEGATIVE RUN-LENGTH.

N	$(\Delta F)_{max}$	ΔF for $\alpha = 0.10$	ΔF for $\alpha = 1.01$
25	0.0113	0.0198	0.0265
50	0.0104	0.0280	0.0374
100	0.0368	0.0395	0.0550
200	0.0313	0.0555	0.0748
500	0.0213	0.0867	0.1171

In conclusion, the experimental method of generating samples of independent stochastic processes gives very precise results for exact and experimental longest negative run-lengths, for $q = 0.50$ and five different values of sample sizes.

3.2 Sampling Distributions of Drought Descriptors Obtained by the Experimental Method. As stated in Chapted II, several drought descriptors of practical relevance may be investigated by the experimental method. In particular, distributions of eight descriptors have been shown attractive for investigating droughts by this method. Although this investigation is attractive, it is not feasible to graphically present all information obtained for these eight descriptors, given the number of combinations of parameters q, N, ρ, and γ . This section shows the general form of the sampling distributions of descriptors as outlined in Chapter II. Appendix II presents graphically the sampling distributions of the two most relevant descriptors in drought investigations, the longest negative run-length and the largest negative run-sum.

Figure 3 shows in a comparative way the sampling distributions of the longest negative run-length, as solid lines, and the negative run-length which corresponds to the largest negative run-sum, as dashed lines, for two sample sizes $N = 25$ and $N = 100$. As it is expected, the negative run-length corresponding to the largest negative run-sum is always smaller than the longest negative run-length for a given probability. However, as the run-length increases the two distributions converge. For the short run-lengths the longest negative run-length is not necessarily the run-length with the largest deficit. However, for long run-lengths, the longest negative run-length will be in general, very close to the negative run-length of the largest negative run-sum.

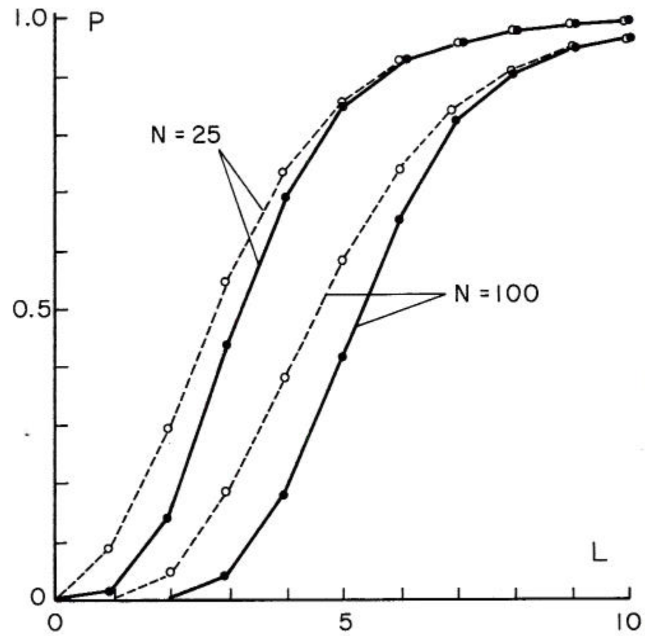


Fig. 3. Distributions of the longest negative run-length for $q = 0.50$, $\rho = 0.0$, and $\gamma = 00$, and two samples $N = 25$ and $N = 200$, (solid lines), and of the negative run-length for the largest negative run-sum (dashed lines).

Figure 4 shows also in a comparative way the sampling distributions of the largest negative run-sum, as solid lines, and the negative run-sum which corresponds to the longest negative run-length, as dashed lines, for two sample sizes $N = 25$ and $N = 100$. As expected, the negative run-sum corresponding to the longest negative run-length is always smaller than the largest negative run-sum for a given probability. Also in this case, as the probability increases the two distributions converge.

Figure 5 shows a presentation similar to that of Fig. 3 for the negative run-length except that the effect of two truncation levels, $q = 0.50$ and $q = 0.20$, is shown for the sample size of $N = 25$, instead of the effect of two sample sizes for a given q . Figure 6 shows a presentation similar to that of Fig. 4 for the negative run-sum except that the effect of two truncation levels, $q = 0.50$ and $q = 0.20$, is shown for the sample size of $N = 25$. The general conclusions of Figs. 5 and 6 are similar to those for Figs. 3 and 4. The mean and the variance of the negative run-length and the largest negative run-sum increase with an increase of the truncation level, q , and the distribution for the two parameters and the same q coverage with an increase of the probability value.

3.3 Correlation Between the Negative Run-Lengths and the Negative Run-Sums. To better represent the relations between the two parameters for given q, N, ρ , and γ , either between the longest negative run-length and the corresponding negative run-sum, or between the largest negative run-sum and the corresponding negative run-length, the correlation

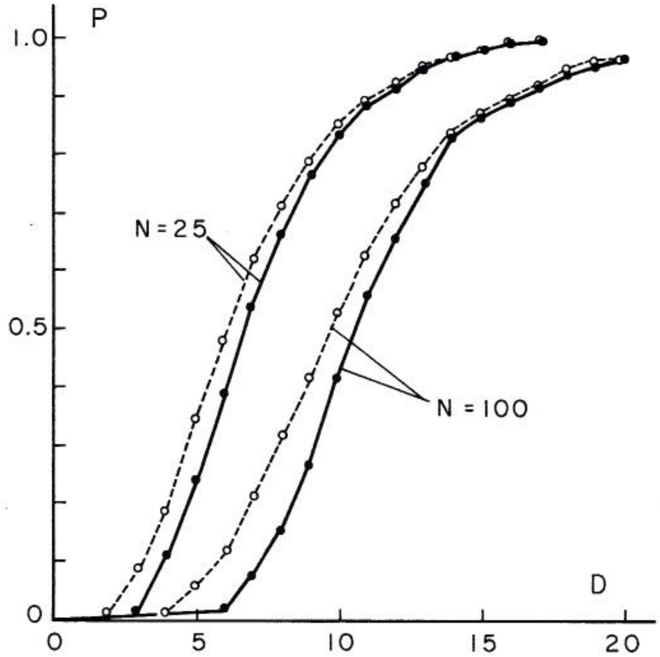


Fig. 4. Distributions of the largest negative run-sum for q = 0.50, ρ = 0.00, γ = 0.00, and two samples N = 25 and N = 100, (solid lines), and of the negative run-sum for the longest negative run-length (dashed lines).

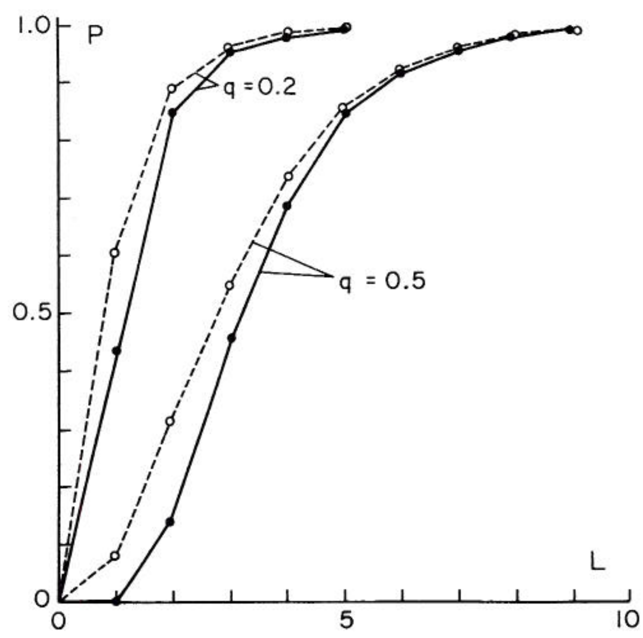


Fig. 5. Distributions of the same parameters as in Fig. 3 except for N = 25 , ρ = 0.00, γ = 0.00 , and two crossing levels, q = 0.50 and q = 0.20.

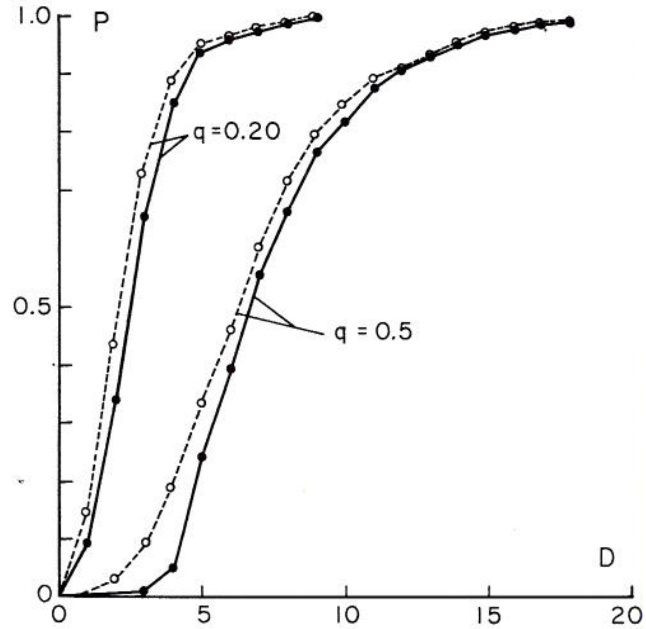


Fig. 6. Distributions of the same parameters as in Fig. 4 except for N = 25, ρ = 0.00, γ = 0.00 , and two crossing levels, q = 0.50 and q = 0.20.

coefficients are computed for these two statistics of all generated samples. Table 4 gives these correlation coefficients for both cases.

The general patterns for both correlated pairs of sample statistics is as follows. The correlation coefficients increase for small sample sizes with an increase of the first serial correlation coefficient, ρ , of the first-order autoregressive linear model. However, as the sample size N increases the difference between the correlation coefficients becomes less dependent on ρ . For N = 200 and N = 500 , the number of samples is not sufficient in the experimental method to clearly discern the patterns of the change of correlation coefficient with an increase of ρ . For all cases, however, the correlation coefficient decreases with an increase of the sample size. Simiarly, the general pattern is that this correlation coefficient decreases with a decrease of the truncation level, q , from q = 0.50 to q = 0.20 , for small N and small ρ . As ρ increases this pattern changes, and for ρ = 0.3 and ρ = 0.7 the pattern reverses, so that the correlation coefficients increase with a decrease of q from 0.50 to 0.20. However, for large N these two patterns are also valid, though the sampling variations for N = 500 do not clearly show the trend of how rapidly the correlation coefficient changes with a change of q for small values of ρ . The skewness coefficient (changing from γ = 0.00 , to γ = 0.20 , and to γ = 0.50) seems to little affect the correlation coefficient.

The general pattern of correlation between the above defined negative run-length and negative run-sum seems to indicate that an increase of sample size makes the correlation between the studied statistics less important.

TABLE 4. DEPENDENCE BETWEEN DROUGHT DESCRIPTORS, MEASURED BY THE CORRELATION COEFFICIENTS

	N=25				N=50				N=100				N=200				N=500			
q=0.5	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2
LARGEST NEGATIVE RUN-LENGTH AND CORRESPONDING RUN-SUM																				
Skewness = 0.0																				
ρ=0.0	.387	.365	.370	.304	.329	.289	.314	.248	.242	.232	.195	.181	.186	.165	.170	.126	.152	.158	.165	.159
=0.1	.409	.400	.418	.384	.354	.328	.360	.304	.253	.249	.253	.229	.202	.178	.166	.149	.143	.164	.190	.141
=0.2	.441	.429	.473	.452	.368	.351	.399	.360	.251	.265	.278	.278	.196	.200	.209	.176	.124	.152	.190	.146
=0.3	.469	.470	.516	.525	.383	.374	.424	.421	.256	.303	.287	.302	.187	.212	.253	.249	.107	.135	.197	.154
=0.7	.603	.642	.699	.707	.425	.526	.601	.664	.224	.345	.440	.520	.118	.216	.308	.427	.026	.073	.165	.264
Skewness = 0.2																				
=0.0	.393	.387	.388	.320	.327	.313	.335	.264	.221	.250	.229	.191	.198	.208	.145	.121	.172	.154	.047	.082
=0.1	.419	.427	.447	.416	.345	.334	.381	.360	.231	.277	.269	.240	.214	.221	.157	.170	.151	.149	.121	.069
=0.2	.445	.463	.487	.490	.355	.386	.404	.408	.255	.291	.301	.271	.218	.221	.189	.214	.137	.155	.169	.092
=0.3	.468	.507	.532	.544	.361	.409	.436	.468	.258	.303	.323	.335	.212	.236	.206	.234	.124	.144	.197	.135
=0.7	.590	.667	.711	.716	.417	.504	.600	.679	.244	.340	.455	.575	.125	.177	.315	.410	.033	.108	.156	.268
Skewness = 0.5																				
=0.0	.399	.400	.397	.348	.300	.315	.279	.271	.254	.245	.243	.204	.168	.188	.225	.122	.106	.155	.163	.106
=0.1	.416	.432	.448	.434	.319	.344	.317	.364	.259	.260	.276	.249	.183	.173	.226	.174	.141	.182	.181	.170
=0.2	.443	.461	.502	.507	.362	.366	.358	.419	.274	.279	.329	.308	.206	.234	.262	.248	.136	.184	.193	.186
=0.3	.473	.505	.546	.566	.403	.406	.404	.473	.278	.312	.356	.348	.224	.228	.261	.293	.134	.165	.193	.228
=0.7	.617	.681	.718	.713	.453	.557	.624	.678	.256	.375	.495	.583	.133	.235	.341	.446	.050	.111	.157	.301
LARGEST NEGATIVE RUN-SUM AND CORRESPONDING RUN-LENGTH																				
Skewness = 0.0																				
=0.0	.433	.423	.425	.380	.371	.360	.393	.358	.283	.282	.303	.281	.209	.224	.264	.216	.196	.199	.256	.251
=0.1	.453	.452	.468	.451	.393	.387	.428	.400	.287	.280	.333	.335	.221	.225	.251	.216	.177	.200	.243	.240
=0.2	.479	.483	.525	.507	.401	.405	.452	.441	.277	.285	.343	.366	.214	.229	.278	.264	.153	.176	.224	.216
=0.3	.506	.514	.568	.576	.410	.416	.471	.490	.279	.331	.340	.385	.194	.236	.302	.321	.126	.159	.222	.236
=0.7	.617	.661	.727	.736	.430	.538	.626	.698	.207	.346	.461	.552	.087	.199	.324	.457	.011	.052	.150	.279
Skewness = 0.2																				
=0.0	.434	.444	.461	.395	.358	.353	.399	.361	.257	.306	.307	.316	.236	.249	.237	.264	.173	.197	.127	.247
=0.1	.458	.481	.519	.481	.370	.373	.440	.451	.266	.322	.335	.341	.234	.259	.233	.273	.168	.198	.159	.217
=0.2	.482	.508	.551	.547	.381	.425	.461	.488	.281	.328	.346	.382	.240	.261	.226	.313	.130	.188	.200	.225
=0.3	.501	.549	.568	.600	.386	.451	.490	.542	.277	.342	.369	.425	.226	.261	.244	.328	.119	.167	.222	.197
=0.7	.603	.687	.733	.741	.415	.517	.625	.711	.230	.345	.475	.613	.099	.172	.321	.436	.013	.080	.166	.287
Skewness = 0.5																				
=0.0	.446	.464	.473	.429	.345	.379	.375	.393	.291	.307	.328	.320	.216	.252	.280	.275	.127	.196	.228	.231
=0.1	.454	.486	.511	.504	.363	.394	.395	.460	.290	.316	.335	.356	.214	.223	.280	.291	.170	.211	.236	.248
=0.2	.481	.506	.557	.562	.397	.410	.418	.493	.300	.328	.363	.407	.228	.263	.301	.308	.154	.212	.226	.244
=0.3	.507	.544	.596	.616	.427	.447	.462	.527	.304	.346	.394	.431	.248	.256	.307	.355	.142	.199	.213	.286
=0.7	.633	.702	.739	.739	.458	.571	.644	.708	.250	.384	.514	.616	.120	.231	.365	.478	.019	.093	.164	.319

3.4 Fitting Lognormal Probability Distributions to Frequency Distributions Obtained by the Experimental Method. The cummulative frequency distributions of the largest negative run-sum were plotted on lognormal graph paper, and the Kolmogorov-Smirnov tests for the goodness of the fit of the lognormal probability density function was performed. The exact probability distributions of the largest negative run-sum in samples of N years seem not to follow a simple probability function. However, practically for all cases investigated the Kolmogorov-Smirnov tests gave good results for fitting the lognormal function.

Appendix III presents the mean, the coefficient of variation, the mean of logarithms, and the standard deviation of logarithms for distributions of the longest negative run-length and the largest negative run-sum for all cases considered in this study.

Though the longest negative run-length is a discrete variable (only integers are random events of this statistic), the fit of a continuous lognormal variable is considered as an approximation to the discrete distribution, or, in other words, the probability densities at the inter variable values multiplied by the unit time interval represent the probability mass at the integer variable values.

The mean and the standard deviation of logarithms, as parameters of the lognormal distribution, are computed by

$$\mu_n = \frac{1}{2} \ln \frac{\mu^2}{\eta^2 + 1} , \quad (12)$$

and

$$\sigma_n = [\ln (\eta^2 + 1)]^{1/2} , \quad (13)$$

in which μ_n is the mean of logarithms, σ_n is the standard deviation of logarithms, μ is the mean, and η is the coefficient of variation of the variable.

Figures 7 and 8 give examples of the fitted lognormal probability distribution functions for the largest negative run-sum, respectively for $N = 25$ and $N = 50$. The five fitted curves are all for $q = 0.50$ (the median) and $\gamma = 0.0$ (normal variable), and each fitted curve for a different ρ ($\rho = 0.0$, 0.1, 0.2, 0.3, and 0.7). The general pattern is that for large values of ρ (the first serial correlation coefficient) the lognormal function starts to deviate at the extremes from the frequency distribution obtained by the experimental method.

The results of fitting the lognormal probability distributions to cummulative frequency distributions of the largest negative run-sum, obtained by the experimental method by using the Kolmogorov-Smirnov test are shown in Table 5. In all cases but $\rho = 7$ the fit of the lognormal distribution passes the Kolmogorov-Smirnov test even at the $\alpha = 0.10$ level.

TABLE 5

KOLMOGOROV-SMIRNOV TEST OF FITTING LOGNORMAL PROBABILITY DISTRIBUTIONS TO FREQUENCY DISTRIBUTIONS OF LARGEST RUN-SUM OBTAINED BY THE EXPERIMENTAL METHOD.

N	ρ	$(\Delta F)_{max}$	$(\Delta F)\alpha{=}0.10$	$(\Delta F)\alpha{=}0.01$
	ρ=0.0	0.0035	0.0198	0.0265
	ρ=0.1	0.0035		
25	ρ=0.2	0.0030		
	ρ=0.3	0.015		
	ρ=0.7	0.055		
	ρ=0.0	0.005	0.0280	0.0374
	ρ=0.1	0.020		
50	ρ=0.2	0.010		
	ρ=0.3	0.020		
	ρ=0.7	0.035		

3.5 Relations Between the Parameters of Fitted Lognormal Functions and the Basic Properties of Generated Samples. It is often useful to develop experimental relations between the parameters of a given process and the parameters of distributions of drought descriptors. The asymptotic value of the mean of the longest run-length for independent variables, as given by Cramer [5] and discussed in Chapter I, suggests a relationship of the form

$$u = f[\ln N, \ln q, \ln \gamma, \ln \rho] . \quad (14)$$

Stepwise regression analysis was used to perform a series of multiple regressions to the relations of the type of Eq. (14).

The independent variables used in this regression analysis are: q the truncation level, N the sample size, ρ the first serial correlation coefficient, and γ the skewness coefficient. The dependent variables are: μ_ℓ the mean of the longest negative run-length L_m, η_ℓ the coefficient of variation of L_m, μ_s the mean of the largest negative run-sum D_m, η_s the coefficient of variation of D_m, $(\mu_n)_\ell$ the mean of logarithms of L_m, $(\sigma_n)_\ell$ the standard deviation of logarithms of L_m, $(\mu_n)_s$ the mean of logarithms of D_m, and $(\sigma_n)_s$ the standard deviation of logarithms of D_m. The equations obtained are in the form

$$u = a + b \ln q + c \ln N + d \ln \rho + e \ln \gamma \quad (15)$$

in which u is a dependent variable. Table 6 gives the estimated regression coefficients. For all regression equations more than 90 percent of the variance of the dependent variable is explained by the four

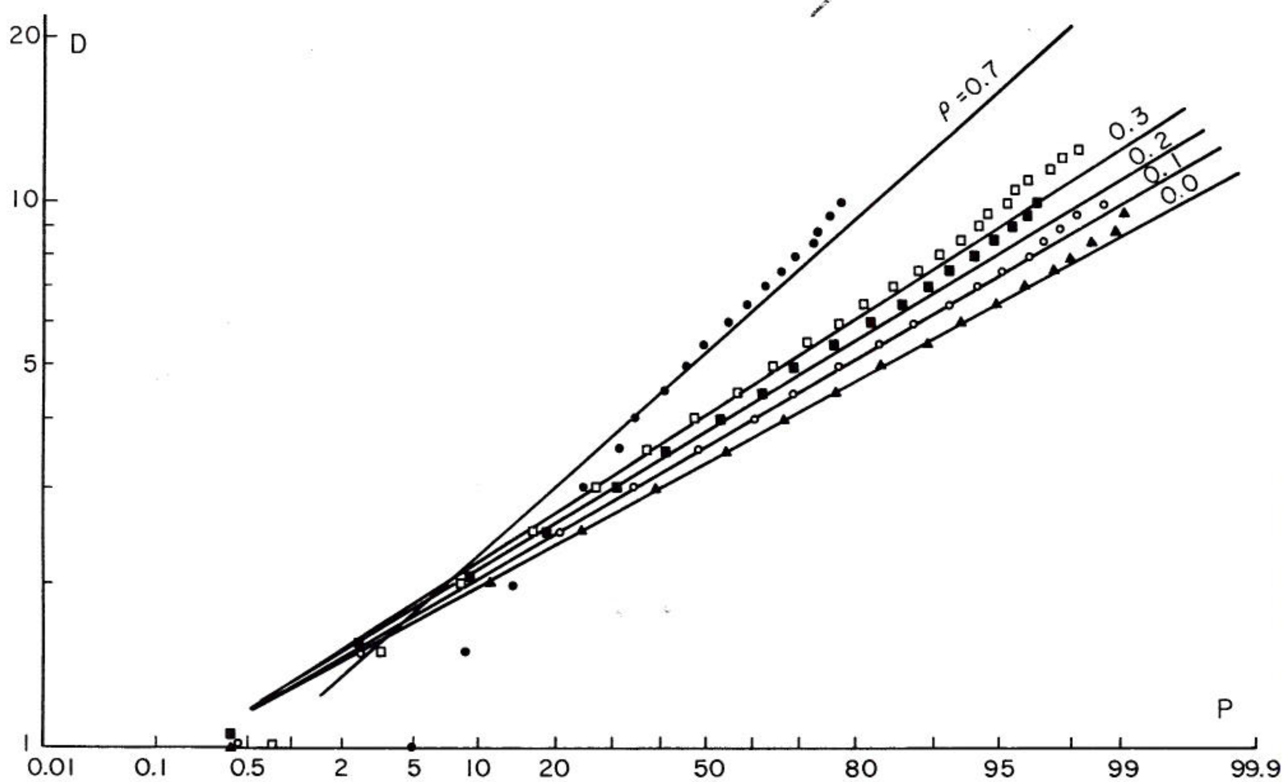


Fig. 7. Fitting the lognormal probability distribution for the largest negative run-sum for N = 25 , q = 0.50 , γ = 0.0 , and five cases of ρ (0.1, 0.2, 0.3, and 0.7).

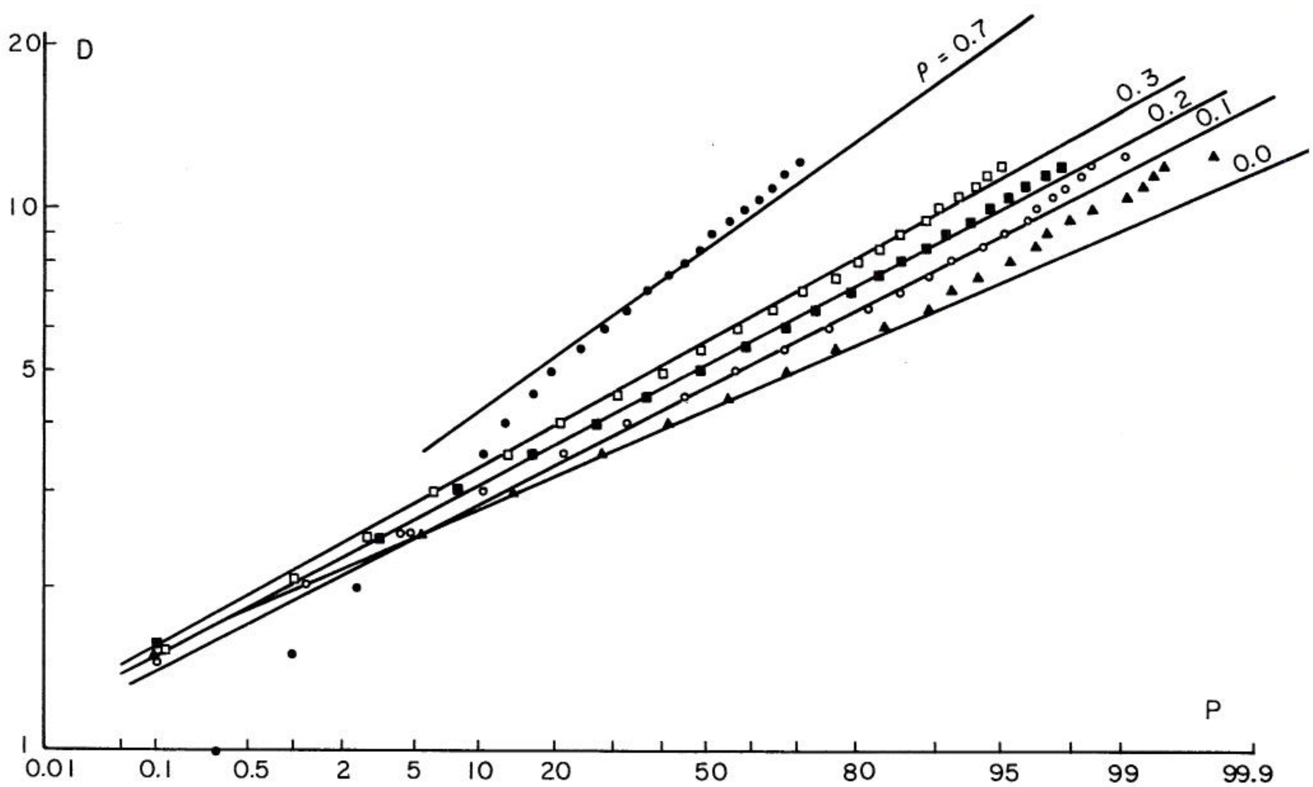


Fig. 8. Fitting the lognormal probability distribution functions for the largest negative run-sum for N = 50 , q = 0.50 , γ = 0.0 , and five values of ρ (0.0, 0.1, 0.2, 0.3, and 0.7).

parameters, q, N, ρ, and γ. Regression coefficients of Table 6 then represent a condensation of all information on the sampling distribution of two statistics, the largest negative run-sum and the longest negative run-length, provided Eq. (15) is used for the investigated ranges of four parameters, q, N, ρ and γ. Table 6, examining the multiple correlation coefficient, shows that the four parameters of logarithms $(\mu_n)_\ell$, $(\sigma_n)_\ell$, $(\mu_n)_s$, and $(\mu_n)_s$ have a larger explained variance in being predicted by the four independent variables q, N, ρ, and γ than the parameters $(\mu)_\ell$, $(\eta)_\ell$, $(\mu)_s$, and $(\eta)_s$.

A detailed analysis of regression coefficients and partial correlation coefficients reveals that the most significant independent variables are the truncation level q and the sample size N. Next in importance comes the first serial correlation coefficient ρ as the measure of dependence, while the skewness as a basic variable has the least effect. As expected, the skewness is more important for the largest negative run-sum than for the longest negative run-length. In fact, the longest run-length of an independent series should not depend on the skewness.

For practical application one may need to know the representative sample size, N_r, of given values q, ρ, and γ, for the longest negative run-length L_m, observed in a sample of size N. The representative sample size, N_r, is defined when L_m is put equal to the mean longest negative run-length, μ_ℓ, of N_r. The regression analysis then gives

$$\ln N_r = -2.1125 + 0.68649\,\mu_\ell - (2.82021 \ln q + 0.07962 \ln \rho + 0.00588 \ln \gamma) , \qquad (16)$$

with $\mu_\ell = L_m$, or the longest negative run-length of the available sample of size N is assumed equal to μ_ℓ of the representative sample of size N_r. Equation (16) has the multiple regression coefficient $R = 0.9046$, or $R^2 = 0.8183$.

Similarly for the largest negative run-sum the representative value, N_r, is

$$\ln N_r = -1.25358 + 0.65044\,\mu_s - (2.82957 \ln q + 0.08039 \ln \rho) - 0.03641 \ln \gamma \qquad (17)$$

with $\mu_s = D_m$, or the largest negative run-sum of the available sample of size N is assumed equal to μ_s of the representative sample of size N_r. Equation (17) has the multiple regression coefficient $R = 0.8330$ and $R^2 = 0.7023$.

Equations (16) and (17) show that the prediction of the representative sample size, N_r, of the sample longest negative run-length is somewhat better ($R^2 = 0.82$) than for the sample largest negative run-sum ($R^2 = 0.70$). In other words and for this latter case, the run-sum should have a larger variation than the run-length, or the linear multiple regression equation of logarithms is more appropriate for run-lengths than for run-sums.

3.6 Use of Distributions Presented as Curves in Appendix II. Probabilities of an observed longest negative run-length and an observed largest negative run-sum may be obtained from the graphs presented in Appendix II. It is necessary, however, to perform the interpolation between given curves when the parameters in a practical problem do not coincide with the parameters corresponding to the curves of the Appendix. Since there are four parameters for interpolation (q, N, ρ, γ) it seems at first glance that the interpolation procedure may be tedious and inaccurate. This is not necessarily so, and in most cases, the simple linear interpolations give the desired probabilities with an accuracy equal to the accuracy which is limited by both the sampling errors of these curves and the errors in reading the figure from the curves.

To illustrate the accuracy obtained in interpolating between curves of these graphs, the probability distribution of the longest negative run-length, L_m, and the largest negative run-sum, D_m, are produced both by using the experimental method and the linear interpolations in the curves of Appendix II for the following parameter values: $q = 0.45$, $N = 40$, $\rho = 0.45$, and $\gamma = 0.40$. The comparison between the two ways of obtaining probabilities P_e for the experimental method and P_i for the interpolation method is given in Table 7.

It should be noted that an interpolation technique more sophisticated than the linear interpolation would add little to the accuracy but would increase the computations. In summary, the probability distributions, estimated by the cummulative

TABLE 6
ESTIMATED REGRESSION COEFFICIENTS OF EQUATION (15)

U	a	b	c	d	e	R
μ_ℓ	4.3080	4.1082	1.1920	0.00856	0.1160	0.9562
η_ℓ	C.6200	-0.0400	-0.0702	0.00017	0.0023	0.9680
μ_s	4.0556	4.3502	1.0797	-0.05598	0.1236	0.9342
η_s	0.7203	-0.0618	-0.0861	-0.00045	0.0073	0.9421
$(\mu_n)_\ell$	1.275	0.9024	0.2703	0.00156	0.0237	0.9870
$(\sigma_n)_\ell$	0.5882	-0.0363	-0.0647	0.00015	0.0021	0.9706
$(\mu_n)_s$	1.1336	1.1876	0.3046	-0.01619	0.0273	0.9796
$(\sigma_n)_s$	0.6729	-0.0453	-0.0776	-0.00040	0.0065	0.9490

frequency distributions of the experimental method in generating samples, and presented as curves in Appendix II, may be used by linear interpolation to obtain the estimates of probability distributions of run-length and run-sum for any set of four parameters in the ranges of their studied variations.

TABLE 7

COMPARISONS OF PROBABILITIES OF RUN-LENGTHS AND RUN-SUMS OBTAINED BY THE EXPERIMENTAL METHOD (P_e) AND BY THE LINEAR INTERPOLATION OF CURVES IN APPENDIX II (P_i) .

Parameter	P_e	P_i	$\Delta P = P_e - P_i$
$L_m \leq 4$	0.479	0.490	-0.011
$L_m \leq 6$	0.824	0.830	-0.006
$L_m \leq 8$	0.946	0.940	+0.006
$D_m \leq 4$	0.570	0.580	-0.010
$D_m \leq 6$	0.863	0.850	+0.013
$D_m \leq 8$	0.963	0.950	+0.013

3.7 Relationship of the Difference Between the Population and Sample First Serial Correlation Coefficients to Other Sample Parameters. The parameters q , ρ , and γ in Eqs. (16) and (17) are population parameters, N is a selective parameter, whereas μ_ℓ and μ_s are the estimates of the $E(L_m)$ and $E(D_m)$, obtained as the means of L_m and D_m of a large number n of generated samples of size N , with the population values q , ρ , and γ . However, each of n generated samples has different values of $\hat{q}$ for given x_o , r , and g , with their means $\Sigma\hat{q}/n$, $\bar{r}$, and $\bar{g}$ approximating closely q , ρ , and γ .

In practice, only one series of size N is available from which the drought descriptors are obtained, with $\hat{q}$, r , and g only as estimates for q , ρ , and γ . By using the statistics $\hat{q}$, r , and g in Eqs. (16) and (17), instead of q , ρ , and γ , the computed N_r values have sampling error because of the sampling differences $q - \hat{q}$, $\rho - r$, and $\gamma - g$.

As it was shown earlier, the parameter γ does not significantly affect the values of μ_ℓ , μ_s , as estimates of L_m , and D_m . The variation of μ_ℓ and μ_s with a variation of q is more important, as it is also the effect of variation of ρ . Two ways can be considered for estimating or reducing the differences $q - \hat{q}$ and $\rho - r$.

First, there may be a dependence of these differences to sample statistics. If a regression equation of these differences, as dependent variables versus the statistics, as independent variables, may be established, with a relatively high explained variances for these differences, then the prediction equations for estimating $\Delta q = q - \hat{q}$, $\Delta\rho = \rho - r$, and $\Delta\gamma = \gamma - g$, may be considered as best estimates, so that $\hat{q} + \Delta g$, $r + \Delta\rho$, and $g + \Delta\gamma$ can be used in Eqs. (16) and (17) instead of $\hat{q}$, r , and g .

Second, the parameters q , ρ , and γ may be better estimated for a series if the regional analysis and/or the investigation of physical conditions produce more information, and thus better estimates, than by using the estimates $\hat{q}$, r , and g from only the data of a series. The procedure for obtaining better estimates of q , ρ , and γ for a series either by the regional information on a hydrologic variable or by studying the physical conditions is outside the scope of this paper.

As an example of the first approach, the study of the difference $\rho - r$ is presented. It is assumed for this analysis that the occurences of unusually prolonged droughts and prolonged high values, or of unusually short negative and positive runs, in a small sample affect all sample statistics or all differences $\alpha - \hat{\alpha}$ of a population parameter α estimated by the sample statistic $\hat{\alpha}$. Particularly these differences should affect the structure of autocovariances of this sample series in such a way that high autocorrelation will be obtained for unrepresentatively long run-lengths and sever large positive and negative run-sums, and low or even negative autocorrelation for unrepresentatively short run-lengths and mild positive and negative run-sums. More specifically, the occurence of very large or very small runs in a sample will be directly related to the sample autocorrelation coefficients. The hypothesis

$$\rho - r = f(N, g, D_{50}, S_{50}) \tag{18}$$

was tested as an example in which ρ is the population first serial correlation coefficient, r is the sample first serial correlation coefficient, N is the sample length, g is the sample skewness, D_{50} is the largest deficit in a sample using the median as the truncation level with $\hat{q} = 0.50$, and S_{50} is the largest surplus in a sample using the median as the truncation level, also with $\hat{q} = 0.50$.

The linear multiple regression of $\rho - r$ on on the other four parameters, N , g , D_{50} , and S_{50} gave a multiple coefficient of regression of $R = 0.49$, or 24.26 percent of the variation of $\rho - r$ was explained by the other four parameters. The replacement of N in Eq. 3.7 by $\ln N$ did not improve the correlation, because R^2 remained at 24.88 percent. For the case of $\rho = 0$ only, or for 500 values of of five parameters $\rho - r$, N , g , D_{50} , and S_{50} , the results of regression analysis were significantly improved, with the variance of $\rho - r$ explained by the remaining four parameters for 46.6 percent, or the multiple correlation coefficient $R = 0.68$. The regression analysis of $\rho - r$ for only D_{50} gave $R^2 = 0.33$, for only S_{50} gave $R^2 = 0.13$, and for both D_{50} and S_{50} gave $R^2 = 0.38$.

This example shows potential for improving the estimates of differences $q - \hat{q}$, $\rho - r$, and $\gamma - g$, provided a sufficient number of pertinant statistics from the sample are used in developing the prediction equations for these differences. The detailed analysis for development of these prediction equations is outside the scope of this paper, but any future development in this direction will increase the reliability of applying Eqs.(16) and (17) in determining the representative sample size , N_r , for estimating the return period, N_r , of an observed drought in a sample of size N.

The expectation is legitimate that a future combination of the first and second approach in improving the estimates $\Delta q = q - \hat{q}$, $\Delta\rho = \rho - r$ and $\Delta\gamma = \gamma - g$ and through them the parameters of Eqs. (16) and (17), as $q + \Delta q$, $r + \Delta\rho$, and $g + \Delta\gamma$ may significantly improve the estimates of probabilities of observed historic droughts in the already available samples.

Chapter IV

EXAMPLES OF COMPUTING PROBABILITIES OF OBSERVED DROUGHTS

This chapter presents some applications of the distributions of the longest negative run-length and the largest negative run-sum. For determining the recurrence intervals of an observed drought, it is frequently advocated that a particular drought has a recurrence period based solely on the length, N, of the historical data. The applications presented in this chapter make use of the principles of computations of such probabilities or recurrence periods, as developed so far in this paper.

4.1 Representative Drought of a Sample Size. The largest drought in a historical record of N years is only a sample statistic obtained from the populations of the largest droughts for the given sample size and the time series structure. If the probability of this event being or not being exceeded is 50 percent, this is by definition the median of the distribution of the largest drought. This median run, which is either the longest negative run-length or the largest run-sum, is defined as the "representative drought". In other words, when the historical drought is close to this representative drought, it is thought that the occurrence of the historical drought follows the structure of a particular stochastic process, and the length of the available time series can be used for computing the recurrence interval. If, on the contrary, the probability of exceedence of an observed drought in a sample of size N is either very small or very large, the largest historical drought is unrepresentative of the sample, or it does not behave according to the mean drought properties for the given structure and length of a series.

To illustrate this point, several graphs are presented that give relations among the representative drought, obtained for the largest negative run-sum, the sample length, and the structure of a stochastic process. These relations for observations belonging to normal variables, $\gamma = 0.0$, either independent or dependent following the first-order autoregressive linear model, are shown in Fig. 9. For a given first serial correlation coefficient, ρ, there is an approximate linear relation in the semi-log coordinates between the size of the representative drought, D_m (the median largest negative run-sum), and the sample size, N. The slope of the straight line fit increases with an increase in dependence, or with an increase of ρ. Figure 10 demonstrates the same relation for the skewness coefficient $\gamma = 0.20$. The patterns are the same as in Fig. 9 but with milder slopes of fitted lines. The same patterns can be observed in Fig. 11 for $\gamma = 0.50$ and in Fig. 12 for $\gamma = 1.00$. To summarize the results in Figs. 9 through 12, on the average, the size of the representative drought in a sample increases exponentially with the increase of the sample size. The size of the representative drought increases more rapidly for high dependences with ρ large, and more slowly for a large skewness coefficient.

The fit of straight lines in Figs. 9, 10, 11, and 12 through the computed points makes it possible to derive approximate relations of the intercept, a, and the slope, b, of these lines to the first serial correlation coefficient, ρ, for four values of γ (0.0, 0.2, 0.5 and 1.0). The intercept, a, is defined in this case as the values of these straight line fits at the intersection of a given N, for example $N = 20$. The relations of b to ρ are given in Figs. 9, 10, 11 and 12 as separate graphs, while the relations of a to ρ are not plotted.

The representative drought statistics, L_m or D_m, and the representative sample size, N_r, are two concepts, but they serve the same purpose. The representative droughts, as measured by the median negative run-length, L_m, and the median negative run-sum, are the droughts that would be exceeded or not exceeded 50 percent of the time if many samples of the given size N are generated by using the structural model of a stochastic process. The representative sample size, N_r, as defined at the end of Chapter III, is the size of the sample that should have the historical drought of a series equal to the mean of the longest negative run-length or the mean of the largest negative run-sum for a very large number of generates samples of this sample size. Both the representative drought and the representative sample size serve the purpose of studying the probabilities of historical droughts. It should be noticed that for $q = 0.50$ an approximate value of N_r can also be obtained from Figs. 9 through 12, for the historical drought used as the ordinate and N_r used along the N-axis.

4.2 Examples of Computing Probabilities of of Historical Droughts for Runoff Annual Series. Table 8 displays examples of annual series of ten river gauging stations, with the probabilities P_r given in the form of the return period ($P_r = 1/N_r$, with N_r the representative sample size) for the historical droughts. The historical droughts are given by both the longest negative run-length and the largest negative run-sum for the series available. First, general parameters are given: N, sample size, Q the mean river discharge, ρ the first serial correlation coefficient, and γ the skewness coefficient. Then the longest negative run-length and the largest negative run-sum of the standardized variables are obtained from the available samples. By using the graphs of Figs. 9 through 12, the representative sample size N_r is determined for the largest negative run-sum for $q = 0.50$. These N_r values are given in Table 8 as the median values of the sampling distributions of the largest negative run-sum.

The predicted representative values, N_r, for the longest negative run-length and the largest negative run-sum for four values of q (0.5, 0.4, 0.3, and 0.2) are computed by Eqs. (16) and (17), respectively. These representative drought values are the means of the sampling distributions of the run-length and run-sum.

Values of N_r obtained from Eqs. (16) and (17) that are greater than 500 must be extrapolated by these equations. This extrapolation outside the range of N values that was available for deriving these equations gives inaccurate results, so that all computed values of N_r greater than $N_r = 500$ are designated only as >500. For computing drought recurrence intervals greater than 500 years, simulation of samples of 1000, 2000, or so years must be

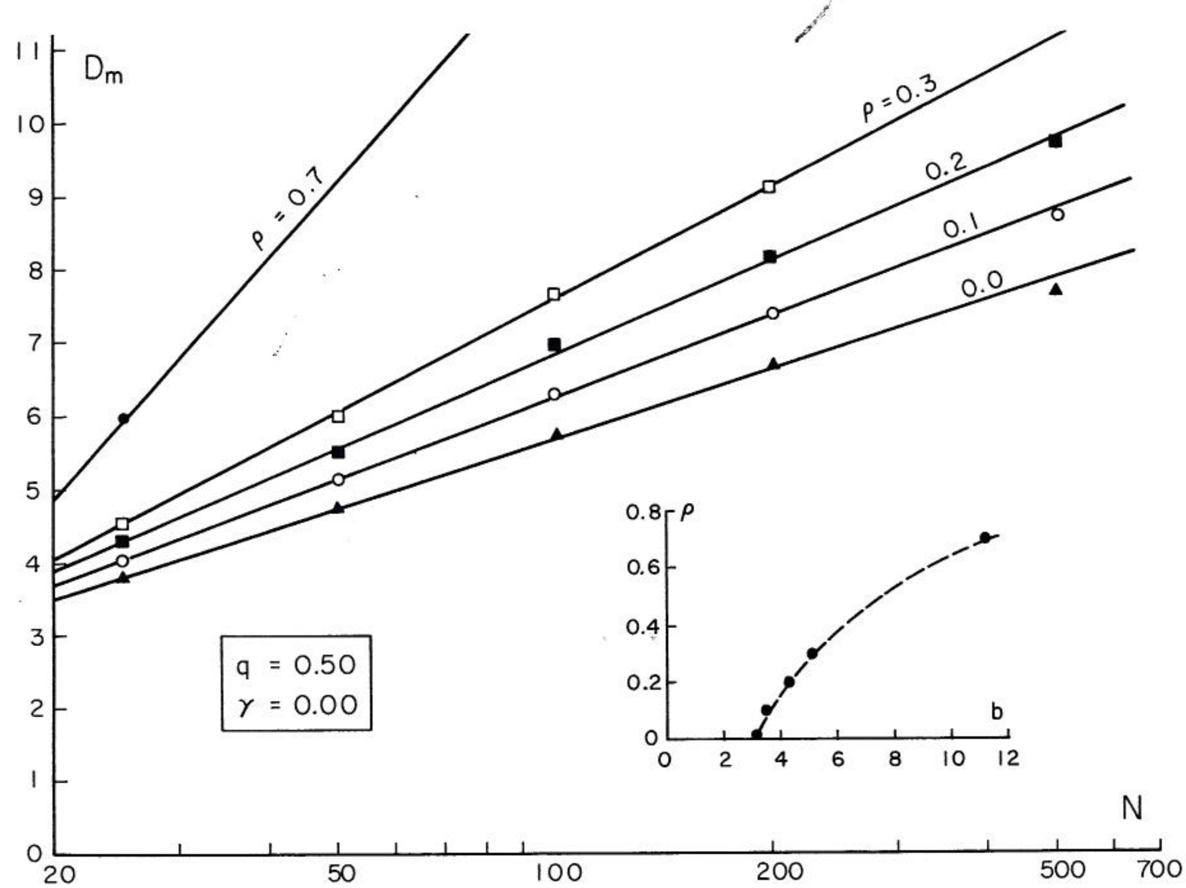


Fig. 9 Representative drought for normal variables

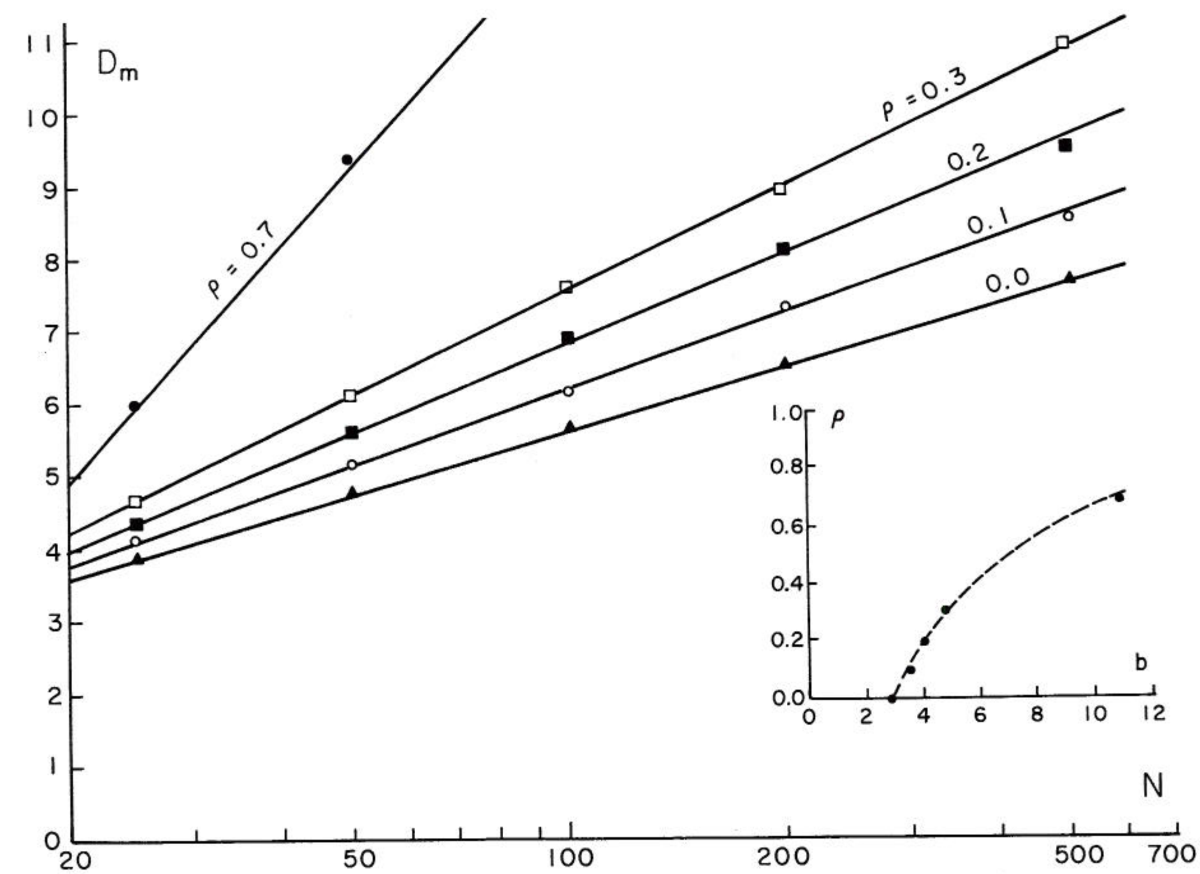


Fig. 10. Representative drought for skewness coefficient equal to 0.2

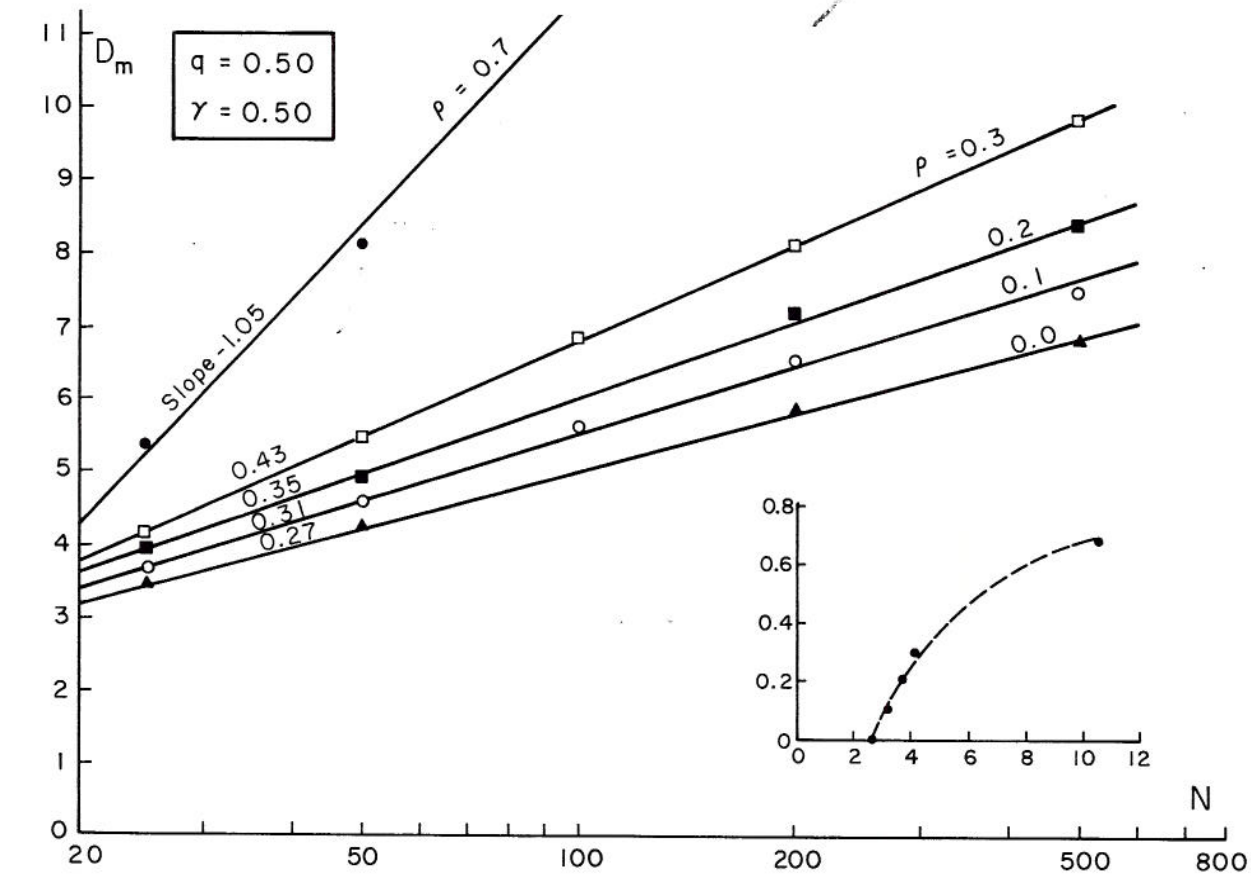


Fig. 11. Representative drought for skewness coefficient equal to 0.5

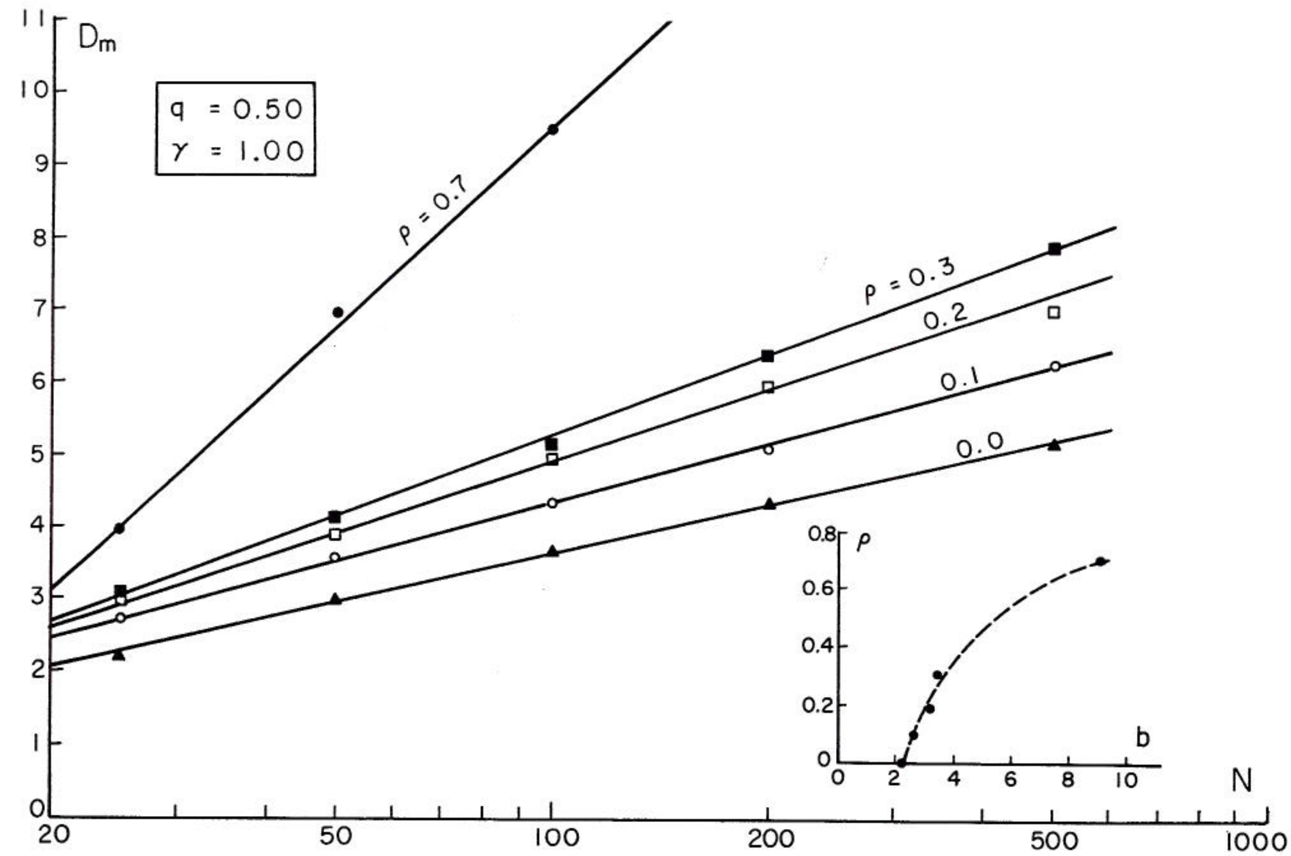


Fig. 12. Representative drought for skewness coefficient equal to 1.0

TABLE 8

EXAMPLES OF COMPUTING RETURN PERIODS, N_r, OF HISTORICAL DROUGHTS (LONGEST NEGATIVE RUN-LENGTH AND LARGEST NEGATIVE RUN-SUM) FOR ANNUAL FLOW SERIES OF TEN RIVERS

	q	Missouri River at Fort Benton, Montana, USA	Rio Grande River at El Paso, Texas, USA	Guadalupe River near Guadalupe, Colombia	Nasselle River near Naselle, Washington USA	Mekong River near Vientiane, Laos	Cherry Creek near Hetch Hetchy, California, USA	Rhine River at Basle, Switzerland	Danube River at Orshova, Rumania	Göta River at Vanersburg, Sweden	Nemunas River at Smolininkai Lithuania, USSR
N, the sample size		65	32	22	31	49	45	150	120	150	132
Q, the mean river discharge, c.f.s.		7635	585	710	430	162070	368	36253	189455	18921	19253
ρ, the first serial correlation coefficient		0.593	0.483	0.290	0.299	0.360	0.013	0.077	0.096	0.463	0.185
γ, the skewness coefficient		0.086	2.269	-1.076	0.019	-0.179	0.211	0.143	0.270	-0.058	0.465
L_m, the longest negative run-length with the truncation level q	0.5	13	14	3	7	8	4	5	6	7	5
	0.4	13	7	3	4	5	4	5	4	6	4
	0.3	12	7	3	2	3	3	3	3	6	3
	0.2	4	4	2	2	3	2	3	2	5	2
D_m, the largest negative run-sum with truncation level q, for standardized variables	0.5	15.237	11.168	2.848	3.124	6.653	4.032	4.155	5.104	7.494	4.039
	0.4	12.425	5.620	2.596	2.160	4.029	2.367	3.168	2.299	5.982	3.156
	0.3	7.933	3.391	1.799	1.495	3.197	1.55	2.538	1.683	4.530	2.158
	0.2	2.001	2.474	1.340	0.924	2.221	0.869	1.745	1.165	2.776	1.599
N_r, the representative sample size from graphs for D_m, the largest negative run-sum for q=0.5		>500	200	20	15	50	30	30	47	60	27
N_r, the representative sample size obtained by Eq. 16 for L_m, the longest negative run-length	0.5	>500	>500	8	117	237	19	33	64	117	30
	0.4	>500	206	15	28	57	36	62	30	110	29
	0.3	>500	465	33	16	32	40	35	34	249	32
	0.2	189	186	52	50	102	64	110	54	393	51
N_r, the representative sample size obtained by Eq. 17 for D_m, the largest negative run-sum	0.5	>500	>500	10	15	119	37	35	64	202	31
	0.4	>500	161	16	15	41	23	35	20	142	33
	0.3	>500	85	22	22	53	31	51	30	125	39
	0.2	87	148	51	47	89	63	97	67	125	85

added to results of this study. In general, the computed representative sample sizes, N_r, agree closely for the run-length and the run-sum for a given level of truncation, q . However, they change from one q to another. The Missouri and the Rio Grande rivers show historic droughts as having very rare occurrence, because the distributions of both the longest negative run-length and the largest negative run-sum show that the return periods of these historic droughts are much greater than $N_r = 500$ for $q = 0.50$.

The available annual series of the Mekong River has large runs at truncation levels $q = 0.5$ and $q = 0.2$ but just about the representative drought lengths at the truncation levels $q = 0.4$ and $q = 0.3$. It should be noticed that the Mekong River has negative skewness, so it is outside the range of the validity of the developed graphs and equations. The question is whether the value of $g = -0.179$ should be considered as significantly different from $\gamma = 0$; besides, the effect of the skewness is relatively small on the representative sample size, N_r .

The Guadalupe River in Colombia and Cherry Creek in California have run-lengths that just about produce the representative sample sizes for all truncation levels. For the Nasselle River the maximum differences among the available sample size and the computed representative sample sizes are less than 70 percent, and a little more than 100 percent for the Guadalupe River, both at the truncation level of $q = 0.2$. The drought probabilities of these two rivers can be adequately represented by the sample size and by using the parameters of their annual runoff series. The Rhine River, the Danube River, and the Nemunas River with the available sample sizes of more than 100 years have smaller representative sizes of historical droughts than the sizes of samples available while the Göta River has just about the same representative sample size as the sample size of the historical drought.

Hall, Askew and Yeh [1] considered from a study of 25 streams in the United States that the historical critical periods, as defined in Chapter I of this paper, were "significantly more severe than would be predicted by synthetically generated sequences of flows, using current standard methods." A study of the data in Table 8 shows which rivers support their statement and which do not support it by the results of this study, if the criterion used is that the most critical design period coincides with the historical drought. The Missouri and the Rio Grande rivers for all truncation levels and the Mekong River for the truncation level of $q = 0.5$ are within the category described [1]. However, all the other seven rivers taken as examples, and by using the first-order autoregressive linear model, not only would reproduce the historical droughts, but would also give, on the average, either longer or larger representative droughts for the sample available than the historical droughts. This statement is well supported by stressing that a value $N_r \leq N$ is equivalent to having a longer or larger historical drought in the available sample than the sample size would produce if the representative drought would occur.

Table 8 demonstrates a significant change of N_r with the truncation level q . In other words, because of large sampling variations in the lengths, sums, and shapes of the longest or largest negative runs, the variation of N_r with q should be expected. Therefore, it is evident that the selection of the truncation level for defining droughts represents an important decision in determining the probabilities (return periods) of historical droughts.

4.3 <u>Examples of Computing Probabilities of Historical Droughts for Annual Series of Precipitation.</u> Table 9 presents examples of 20 annual series of precipitation in the Upper Missouri River Basin the same analysis that was presented for the runoff series in the previous section.

Values of N_r for the precipitation series show in general the same pattern as observed for the runoff series. There are, however, some important points to emphasize. First, the 20 annual precipitation series may be divided in a rough manner into long and short series. The only instances for which the computed values of N_r exceed the length of the historical record by more than 500 percent is when the historical record is long. For small historical records the value of N_r never exceeds the historical sample by more than 100 percent. An explanation for this pattern is as follows. The short records, those with less than 40 years of data, have a relatively smaller mean than they would have were they longer, because for short records the drought of the 1930's represents a larger percentage of the total record than for the long records. This has a significant effect on the estimate of the truncation level, because a short record has a much lower truncation level, x_o , for a given q than a long record. The smaller the truncation level for a short record the smaller are the droughts.

The analysis of N_r values in Table 9 shows similar patterns for precipitation as for runoff, namely that for the same q the representative longest run-length and the representative largest run-sum have close values of N_r . However, the change of N_r is much more variable from one q to the next.

TABLE 9

EXAMPLES OF COMPUTING RETURN PERIODS, N_r, OF HISTORICAL DROUGHTS (LONGEST NEGATIVE RUN-LENGTH AND LARGEST NEGATIVE RUN-SUM) FOR ANNUAL PRECIPITATION SERIES OF TWENTY STATIONS IN THE UPPER MISSOURI BASIN

	q	Ashton Idaho (1)	Dubois Wyo. (2)	Salmon Idaho (3)	Augusta Montana (4)	Big Sandy Montana (5)	Bozeman Montana (6)	Butte Far Montana (7)	East Anaconda Montana (8)	Ennis Montana (9)	Fair Field Montana (10)
N, sample size years		59	39	50	56	37	66	66	55	38	34
$\bar{P}$, mean annual precipitation (inch)		16.41	10.90	9.05	14.52	12.02	17.55	12.89	13.58	10.75	11.89
ρ, the first serial correlation coefficient		-0.0440	-0.1030	-0.0300	0.1300	0.0210	0.2098	0.2962	0.0630	0.4340	-0.0351
γ, the skewness coefficient		0.117	0.670	0.340	0.340	0.350	0.099	0.205	-0.040	0.087	-0.021
L_m, the longest negative run-length with the truncation level q	0.5	6	5	6	10	5	9	8	8	6	4
	0.4	3	5	4	10	3	4	8	5	6	4
	0.3	3	5	4	3	3	4	5	5	6	3
	0.2	2	1	4	2	1	2	3	3	2	2
D_m, the largest negative run-sum with truncation level q, for standardized variables	0.5	4.43	1.67	3.80	8.93	3.06	6.10	7.22	6.13	6.00	4.10
	0.4	1.77	1.18	3.33	6.56	1.68	3.33	5.37	3.24	3.71	3.09
	0.3	1.26	1.09	2.81	1.67	1.11	2.27	3.27	2.24	1.16	2.34
	0.2	0.91	0.96	2.14	0.93	0.86	1.26	1.46	1.35	0.64	1.83
N_r, from Figs. 9 thru 12		40	<20	35	>500	<20	100	100	150	400	38
N_r, the representative sample size obtained by Eq. 16 for L_m, the longest negative run-length	0.5	110.	55.	110.	>500.	36.	473.	231.	273.	57.	29.
	0.4	26.	103.	52.	>500.	17.	29.	433.	65.	107.	55.
	0.3	60.	232.	118.	34.	39.	65.	124.	147.	240.	62.
	0.2	94.	46.	369.	53.	31.	51.	99.	117.	48.	98.
N_r, the representative sample size obtained by Eq. 17 for D_m, the largest negative run-sum	0.5	70.	13.	48.	500.	19.	112.	231.	98.	98.	44.
	0.4	16.	17.	67.	308.	15.	35.	130.	28.	41.	43.
	0.3	38.	36.	108.	29.	23.	39.	75.	33.	18.	59.
	0.2	95.	105.	220.	56.	62.	64.	73.	58.	40.	134.

TABLE 9 (cont'd)

	q	Great Falls Mont. (11)	Hebgen Dam Mont. (12)	Helena Mont. (13)	Holter Dam Mont. (14)	Lima Montana (15)	Morris Madison Mont. (16)	Virginia City Montana (17)	West Yellow-stone Montana (18)	Lamar Wyo. (19)	Yellow-Stone Park Wyo. (20)
N, sample size years		68	46	79	55	31	53	66	30	30	72
$\bar{P}$, mean annual precipitation (inch)		14.80	25.40	12.38	13.20	10.84	17.70	13.90	21.21	13.43	16.58
ρ, the first serial correlation coefficient		0.2430	-0.0740	0.2021	0.4380	-0.0375	0.2469	0.1600	0.0880	0.3142	0.3290
γ, the skewness coefficient		0.27	0.735	0.474	0.458	0.807	0.380	0.280	0.060	0.304	0.480
L_m, the longest negative run-length with the truncation level q	0.5	4	4	10	10	5	5	9	3	6	10
	0.4	3	3	3	8	3	4	5	3	4	7
	0.3	3	2	3	8	2	4	4	3	3	7
	0.2	3	1	2	3	2	3	3	2	1	5
D_m, the largest negative run-sum with truncation level q, for standardized variables	0.5	3.71	1.88	5.67	7.20	2.60	3.22	4.53	2.42	4.25	7.61
	0.4	3.42	1.47	2.23	5.24	1.23	2.16	3.24	1.86	2.80	4.88
	0.3	2.53	1.14	1.74	3.03	0.96	1.32	2.41	1.70	1.09	4.36
	0.2	1.92	0.96	1.15	1.16	0.70	0.73	1.50	0.59	0.99	2.72
N_r, from Figs. 9 thru 12		31	<20	90	60	20	<20	30	<20	25	150
N_r, the representative sample size obtained by Eq. 16 for L_m, the longest negative run-length	0.5	15.	28.	>500.	>500.	55.	30.	480.	8.	58.	>500.
	0.4	14.	26.	15.	417.	26.	28.	58.	15.	28.	215.
	0.3	31.	30.	32.	>500.	30.	63.	65.	35.	31.	483.
	0.2	99.	47.	51.	95.	93.	100.	103.	55.	24.	384.
N_r, the representative sample size obtained by Eq. 17 for D_m, the largest negative run-sum	0.5	24.	15.	90.	228.	23.	18.	43.	11.	34.	304.
	0.4	37.	21.	18.	120.	18.	17.	35.	14.	25.	97.
	0.3	47.	37.	29.	64.	33.	22.	46.	29.	18.	156.
	0.2	99.	105.	63.	60.	89.	47.	80.	44.	54.	169.

Chapter V

THE CONCLUSIONS

The developed methodology for determining the probabilities or the return periods of historical droughts (recurrence intervals given in years) is presented in this study. The basic statistical parameters used to describe the largest historical drought are the longest negative run-length and the largest negative run-sum in a sample of size N. The parameters describing the structure of a stochastic process for a given probability q of the truncation level of a series are the first serial correlation coefficient ρ measuring the time dependence and the skewness coefficient γ. This study leads to the following conclusions.

(1) The presented method can be used to determine when a historical or observed drought is unrepresentative of the sample size and the stochastic process for which the sample is observed.

(2) By determining the sample size N_r to which the historical or observed drought should belong to be representative, the return period or the recurrence interval of the drought can be indicated by the method.

(3) The frequency distributions as the estimates of sampling probability distributions of the longest negative run-length and the largest negative run-sum in the sample of size N are of practical relevance to water resources and hydrologic investigations. As such they are presented in Appendices II and III of this study for five sample sizes N(25, 50, 100, 200, 500), for four values of the probability 2 of the truncation level (0.5, 0.4, 0.3, 0.2), for five values of the first serial correlation coefficient ρ(0.0, 0.1, 0.2, 0.3, 0.7), and for four values of the skewness coefficient γ(0.0, 0.2, 0.5, 1.0).

(4) The sampling probability distribution of the longest negative run-length for independent stochastic processes can be obtained theoretically; however, for dependent stochastic processes following the first-order Markov linear model and for the sampling probability distribution of the largest negative run-sum it was necessary to use the statistical experimental (Monte Carlo) method in computing the frequency distributions from a large number of generated samples.

(5) To check the accuracy of the obtained results of the experimental method, the exact probability of the longest negative run-length was compared with the frequency distribution of this run-length of generated samples. Satisfactory agreements were obtained at the 0.10 level of significance by using the Kolmogorov-Smirnov test.

(6) The nonnormality of the underlying distribution as measured by the skewness coefficient only slightly effects the probability distribution of the longest negative run-length. For $\rho = 0.0$ it does not depend on the underlying distribution. For the largest negative run-sum, however, the nonnormality has a much greater effect than for the longest negative run-length.

(7) Although there are reasons to believe that the theoretical probability distributions for the longest negative run-length and the largest negative run-sum may be very complex mathematical experesions, the fit of the lognormal probability function with two parameters to the frequency distributions obtained by the experimental method is very good considering the expected complexity.

(8) Because the computation of probabilities (return periods, or recurrence intervals) of historical drought depends highly on the best estimates of q, ρ, and γ, for a given N, the use of all regional information for improving the accuracy of these estimates will produce much more reliable estimates of these probabilities.

(9) The differences between the population parameters (q, ρ, γ) and the sample estimates $(\hat{q}, r, g)$ are related not only to the population parameters but also to some other statistics of the available sample. This properly may be used to improve the estimates of the sample on which the probabilities of historical droughts depend.

(10) The application of the method developed in this study to 10 series of annual runoff and 20 series of annual precipitation indicates that for a given truncation level the computation of return periods for the longest negative run-length and the largest negative run-sum are approximately of the same order of magnitude. However, the representative drought, N_r, of a series is very sensitive to a change in the truncation level parametes, q. This sensitivity indicates the importance of searching for the most accurate estimation of the probability of a truncation level.

REFERENCES

[1] Hall W., and A.J. Askew, 1969, Use of critical period in reservoir design, Water Resources Research, Vol. 5, No. 6.

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

[3] Saldarriaga J., and V. Yevjevich, 1970, Applications of runs to hydrologic series, Hydrology Paper No. 40, Colorado State University.

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

[5] Cramer H., 1946, Mathematical methods of statistics, Princeton University Press, Princeton.

[6] Feller W., 1957, An introduction to probability theory and its applications, Vol. 1, Wiley and Sons, New York.

[7] Roesner L.A., and V. Yevjevich, 1966, Mathematical models for time series of monthly precipitation and monthly runoff, Hydrology Paper No. 15, Colorado State University.

[8] Matalas H.C., 1967, Mathematical assestment of synthetic hydrology, Water Resources Research, Vol. 3, No. 4.

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

[10] Thomas H. and M. Fiering, 1963, Statistical analysis of the reservoir storage-yield relation, Chap. 1, Operations research in water quality management, Report Harvard Water Resources Group.

[11] Handbook of Mathematical Functions, 1964, AMS 55, National Bureau of Standards, Washington, D.C.

[12] Yevjevich V., 1965, The applications of surplus, deficit and range in Hydrology, Hydrology Paper No. 10, Colorado State University.

APPENDIX I

PROOFS FOR EQUATIONS GIVEN IN CHAPTER II

Approximation to the chi-square distribution for large ν , where ν is the number of degrees of freedom is as follows.

For $\nu > 30$

$$x_i^2 = \nu\left[1 - \frac{2}{9\nu} + t_i\sqrt{\frac{2}{9\nu}}\right]^3 \tag{1.1}$$

in which t_i is the standard normal variable, $N(0,1)$, and x_i^2 is the corresponding chi-square variable.

The skewness is given by

$$\gamma = \sqrt{\frac{8}{\nu}} \quad , \tag{1.2}$$

so for values of $\gamma \leq \sqrt{\frac{8}{30}} = 0.51$, Eq. (1.1) becomes

$$x_i^2 = \frac{8}{\gamma^2}\left[1 - \frac{\gamma^2}{36} + \frac{t_i\gamma}{6}\right]^3 \tag{1.3}$$

which is chi-squared distributed with the mean ν and the variance 2ν . To obtain a standardized deviate x_i the mean is subtracted and divided by $\sqrt{2\nu}$ for Eq. (1.3), so that

$$x_i = \frac{2}{\gamma}\left[1 + \frac{\gamma}{6}t_i - \frac{\gamma^2}{36}\right]^3 - \frac{2}{\gamma} \quad . \tag{1.4}$$

For the dependence of the type of the first order Markov model a further transformation must be made. Referring to Eq. (6) in Chapter II, the standardized variable, y_i , of the variable following the first order Markov model, is

$$y_i = \frac{(x_i - \mu_x)}{\sigma_x} = \rho\, Y_{i-1} + (1 - \rho_1^2)^{1/2}\, \varepsilon_1 \quad . \tag{1.5}$$

By definition $E(y_i) = E(y_{i-1}) = 0$, $E(y_i^2) = 1$, and $E(y_i^3) = \gamma_y$; then Eq. (1.5) gives

$$y_i^3 = \rho^3 y_{i-1}^3 + 3\,\varepsilon_1\,(1 - \rho_1^2)^{1/2}\,\rho_1^2\, y_{i-1}^2 + 3\,\varepsilon_i^2(1 - \rho_1^2)\,\rho_1 y_{i-1} + \varepsilon_i^3\,(1 - \rho_1^2)^{3/2} \quad . \tag{1.6}$$

Taking the expectation of the terms in Eq. (1.6) then

$$\gamma_y = E(y_i^3) = \rho_1^3\, \gamma y + E(\varepsilon_1^3)(1 - \rho_1^2)^{3/2} \quad . \tag{1.7}$$

From Eq. (1.7) solved for $E(\varepsilon_i^3) = \gamma_\varepsilon$, then

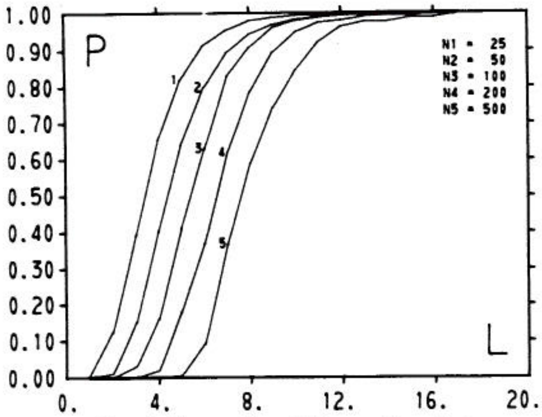
$$\gamma_\varepsilon = E(\ _{i+1}) = \frac{\gamma_y\,(1 - \rho_1^3)}{(1 - \rho_1^2)^{3/2}} \quad . \tag{1.8}$$

The skewness coefficient γ_y or γ_x is different from the skewness coefficient of γ_ε , because of the effect of serial correlation.

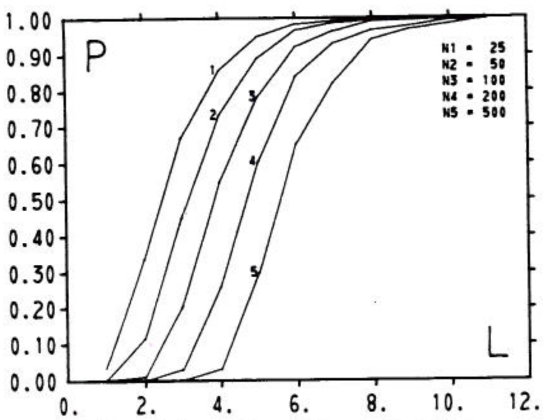
APPENDIX II

This appendix gives the plotted cumulative frequency distributions of the largest negative run-sum and the longest negative run-length in the sample of length N of the normal and nonnormal independent variables, with $\rho = 0.0$, and of the dependent variables following the first-order linear autoregressive model for four values with $\rho = 0.1, 0.2, 0.3, 0.7$. The nonnormality was accounted for by generating the one-parameter gamma variables that preserve the skewness coefficient, $\gamma = 0.2, 0.5,$ and 1.0. Four values $q = F(x_o)$, of the probabilities of truncation levels x_o, or $q = 0.50, 0.40, 0.30,$ and 0.20, are shown in the first five pages; pages 26 through 30 present the probabilities for all values of q, N, ρ and of $\gamma = 0.00$ and $\gamma = 0.20$ for the longest negative run-length, L_m, and pages 31 through 35 present the same probabilities for values of $\gamma = 0.50$ and $\gamma = 1.00$. Pages 36 through 45 give the probabilities for the largest negative run-sum, or the deficit D_m.

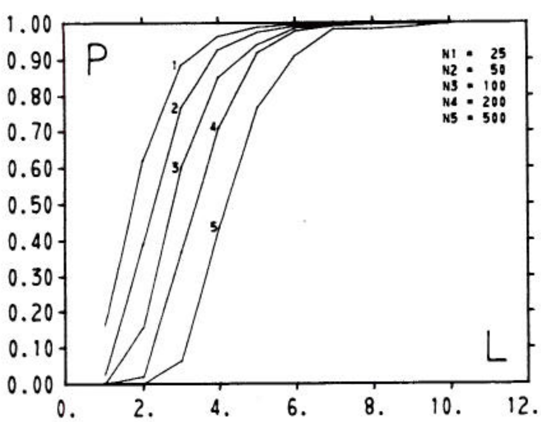
Each graph is identified by computer symbols Q, R, and G which correspond to the symbols used in this text, q, ρ, and γ. Five frequency distribution curves on each graph obtained by the experimental method, are given for five values of N (25, 50, 100, 200, 500).



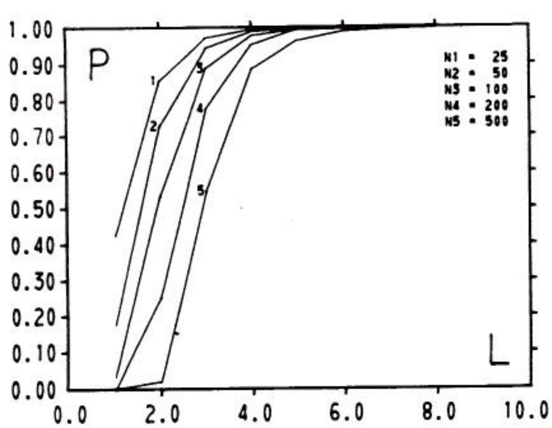
Q = 0.5 G = 0.2 R = 0.0



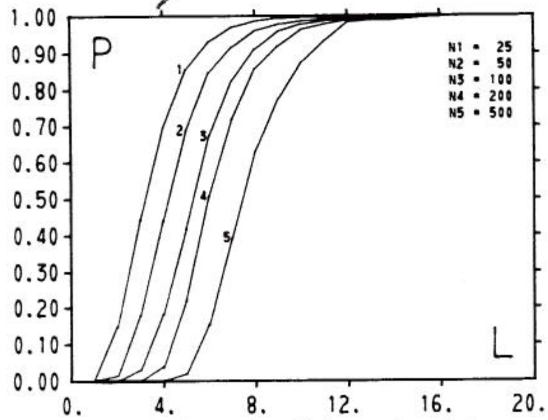
Q = 0.4 G = 0.0 R = 0.0



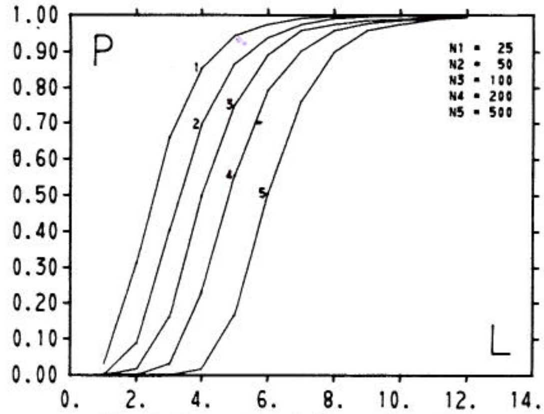
Q = 0.3 G = 0.0 R = 0.0



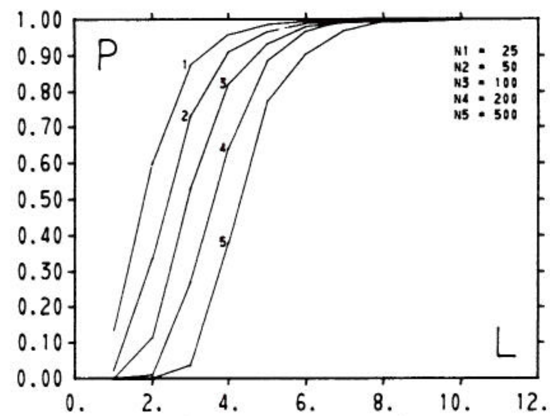
Q = 0.2 G = 0.0 R = 0.0



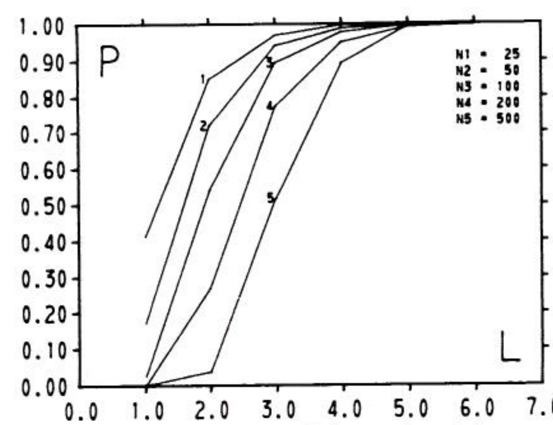
Q = 0.5 G = 0.0 R = 0.0



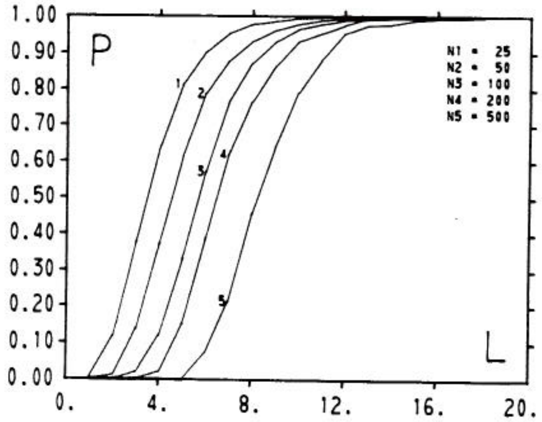
Q = 0.4 G = 0.2 R = 0.0



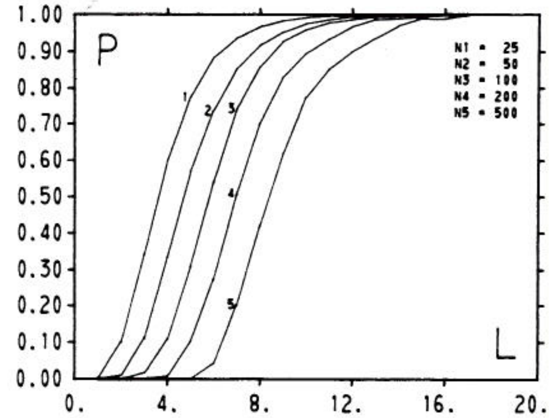
Q = 0.3 G = 0.2 R = 0.0



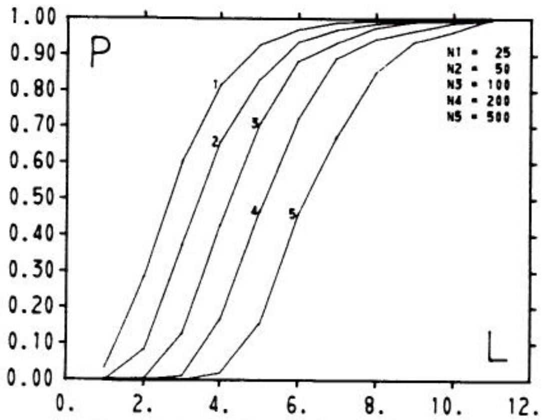
Q = 0.2 G = 0.2 R = 0.0



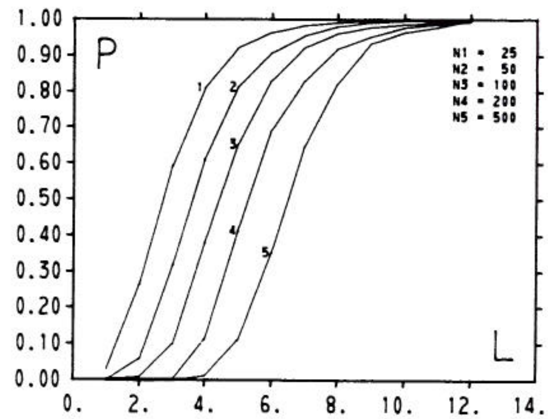
Q = 0.5 G = 0.0 R = 0.1



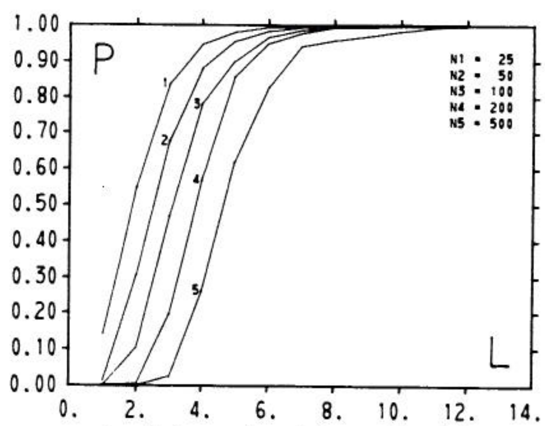
Q = 0.5 G = 0.2 R = 0.1



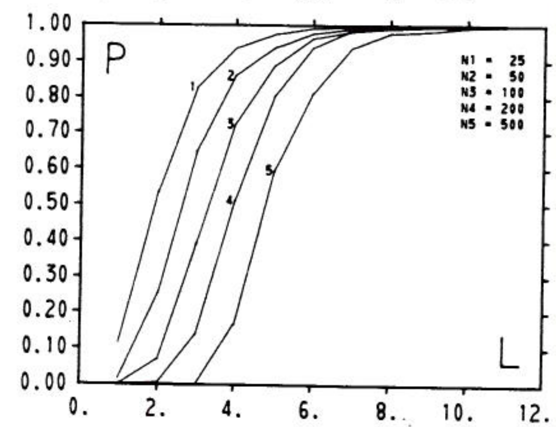
Q = 0.4 G = 0.0 R = 0.1



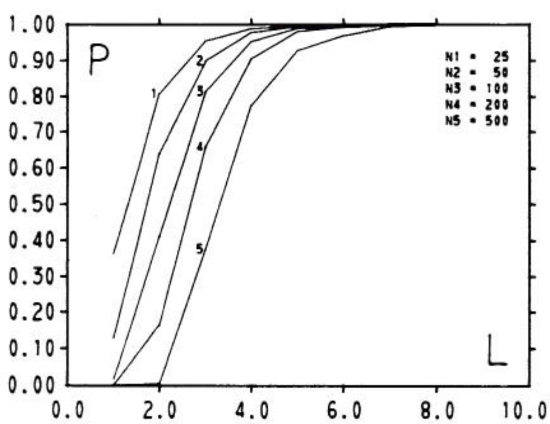
Q = 0.4 G = 0.2 R = 0.1



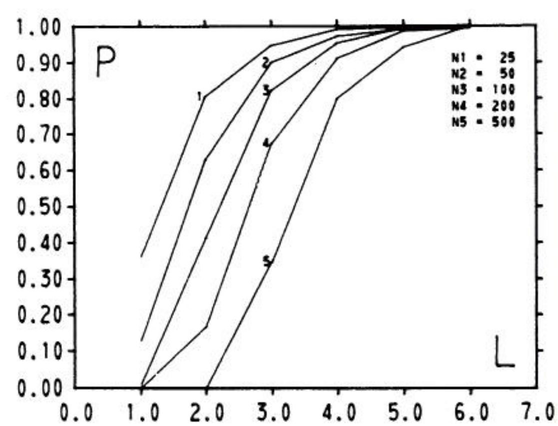
Q = 0.3 G = 0.0 R = 0.1



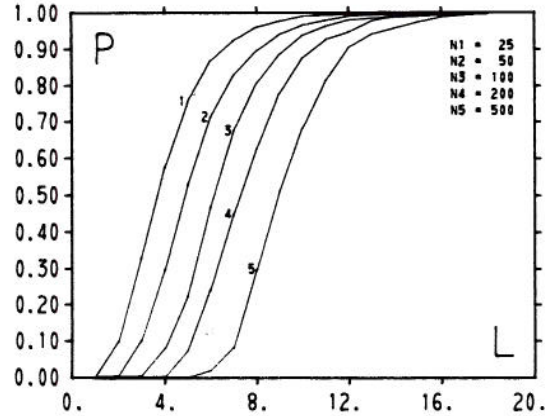
Q = 0.3 G = 0.2 R = 0.1



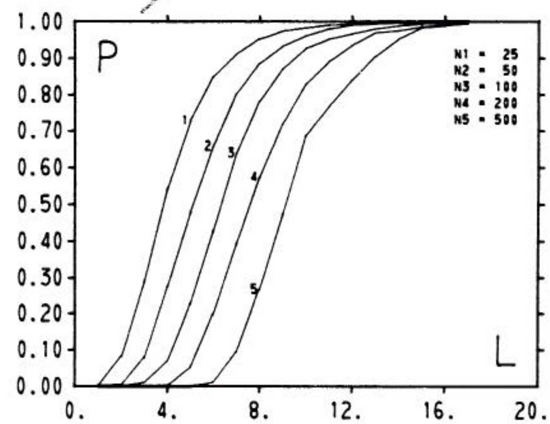
Q = 0.2 G = 0.0 R = 0.1



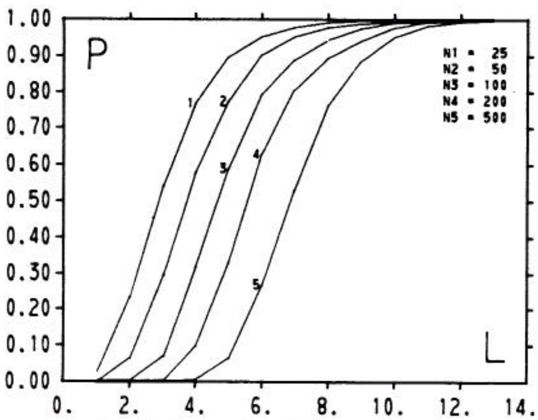
Q = 0.2 G = 0.2 R = 0.1



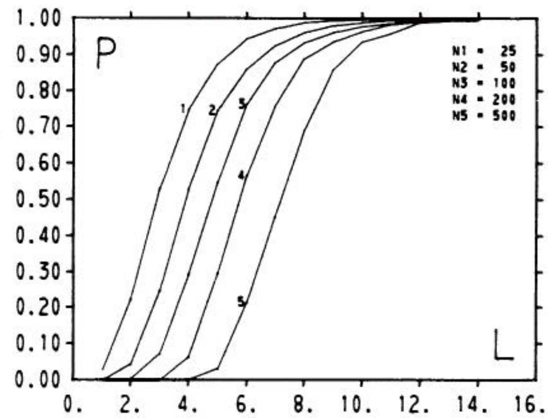
Q = 0.5 G = 0.0 R = 0.2



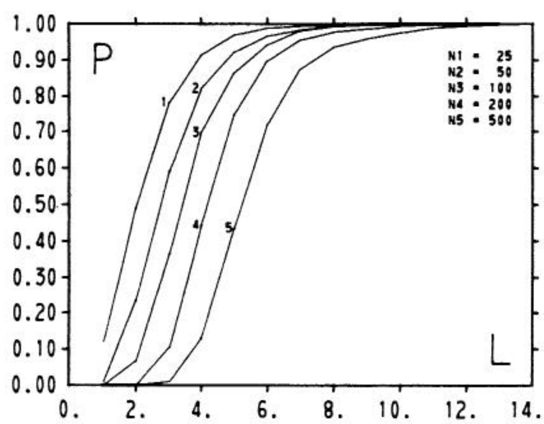
Q = 0.5 G = 0.2 R = 0.2



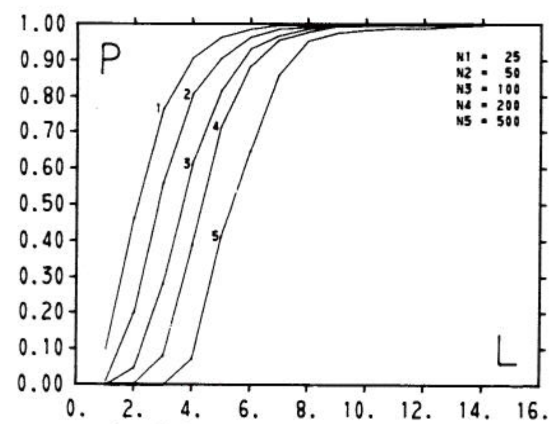
Q = 0.4 G = 0.0 R = 0.2



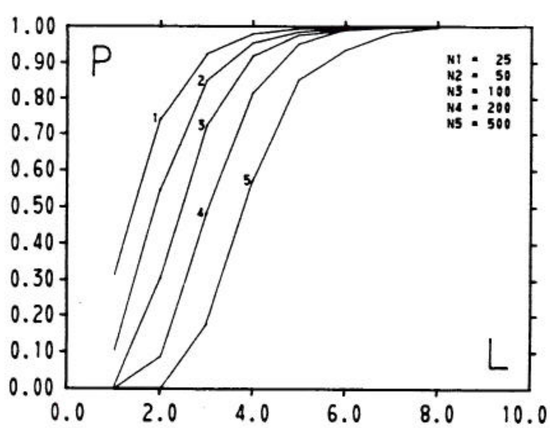
Q = 0.4 G = 0.2 R = 0.2



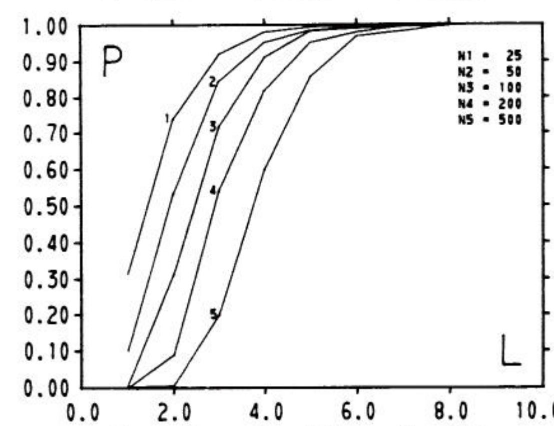
Q = 0.3 G = 0.0 R = 0.2



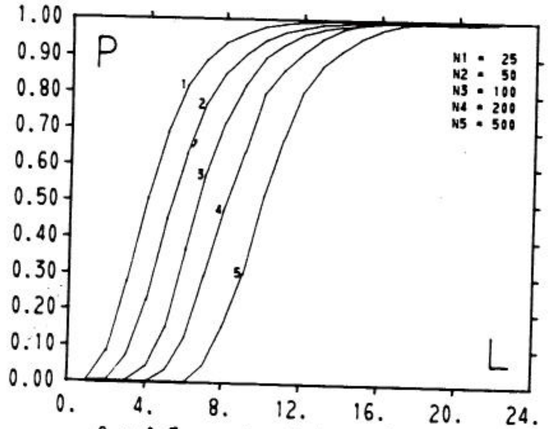
Q = 0.3 G = 0.2 R = 0.2



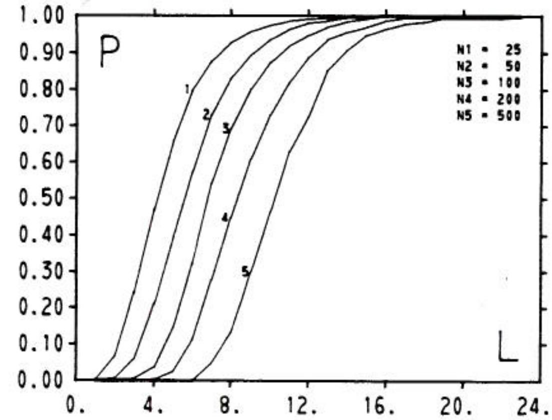
Q = 0.2 G = 0.0 R = 0.2



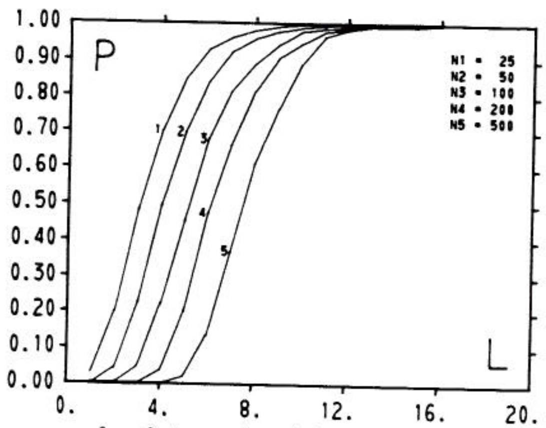
Q = 0.2 G = 0.2 R = 0.2



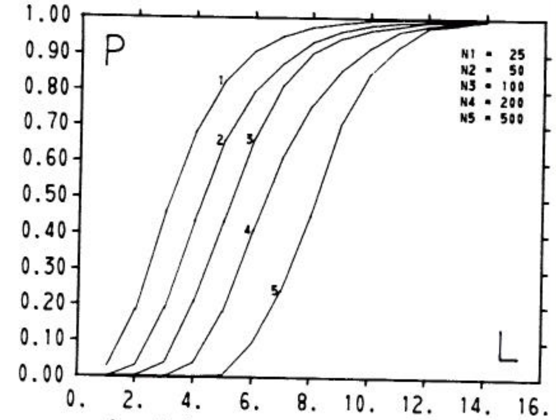
Q = 0.5 G = 0.0 R = 0.3



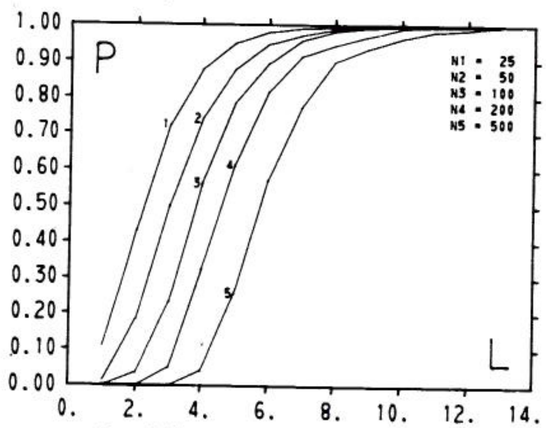
Q = 0.5 G = 0.2 R = 0.3



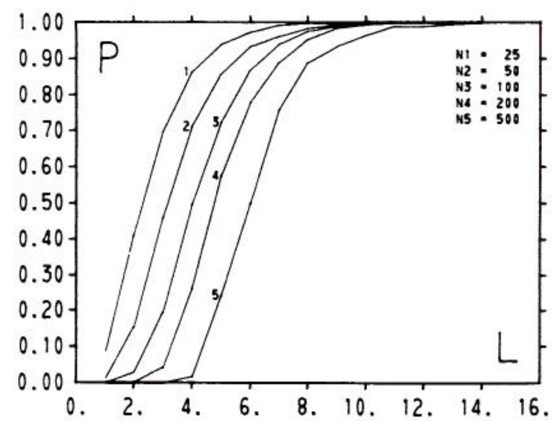
Q = 0.4 G = 0.0 R = 0.3



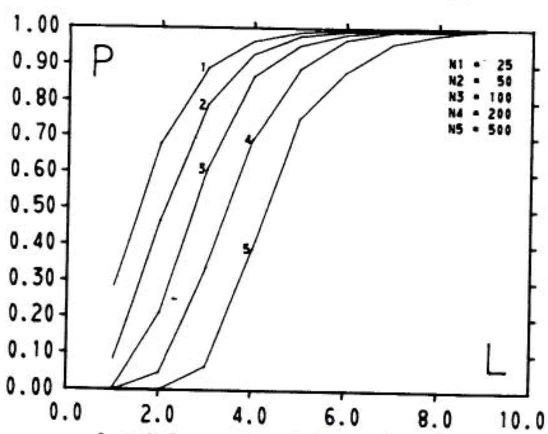
Q = 0.4 G = 0.2 R = 0.3



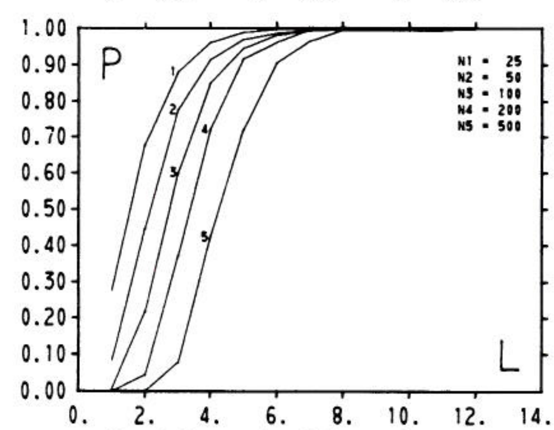
Q = 0.3 G = 0.0 R = 0.3



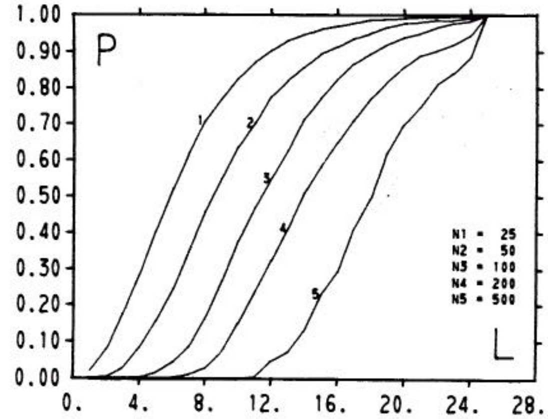
Q = 0.3 G = 0.2 R = 0.3



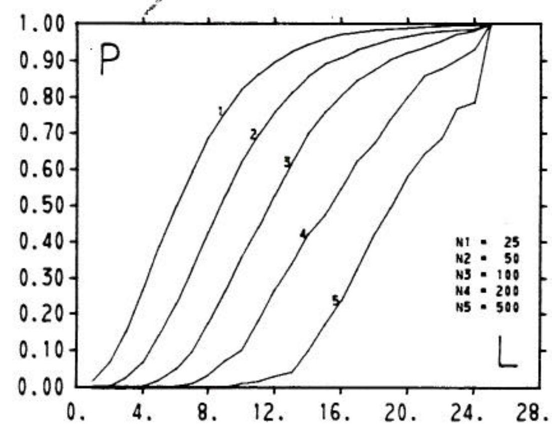
Q = 0.2 G = 0.0 R = 0.3



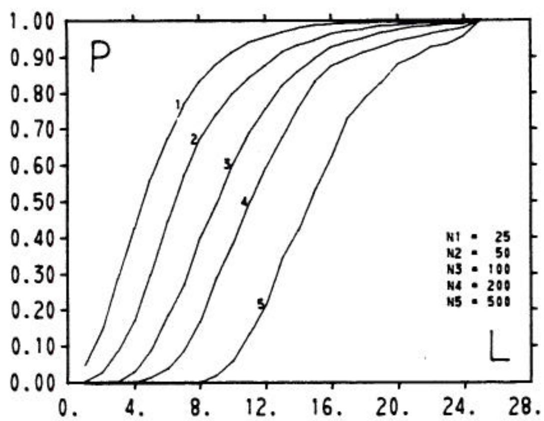
Q = 0.2 G = 0.2 R = 0.3



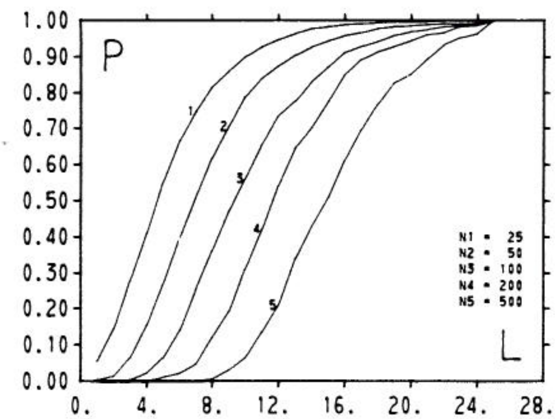
Q = 0.5 G = 0.0 R = 0.7



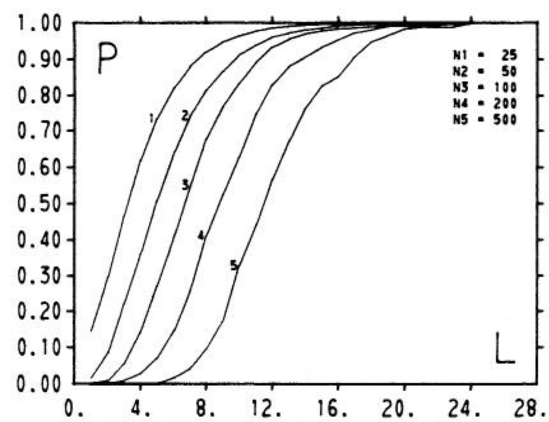
Q = 0.5 G = 0.2 R = 0.7



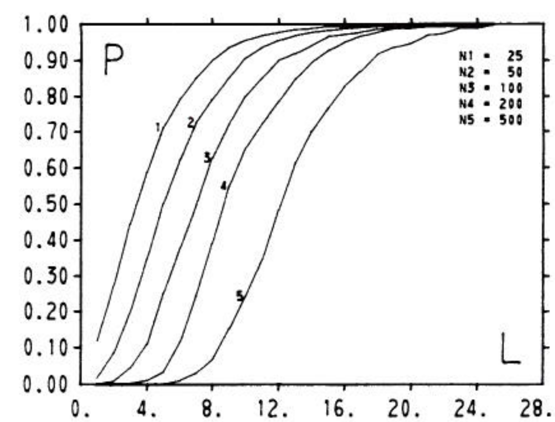
Q = 0.4 G = 0.0 R = 0.7



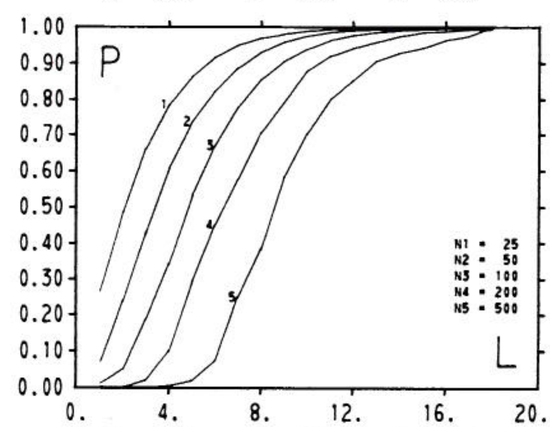
Q = 0.4 G = 0.2 R = 0.7



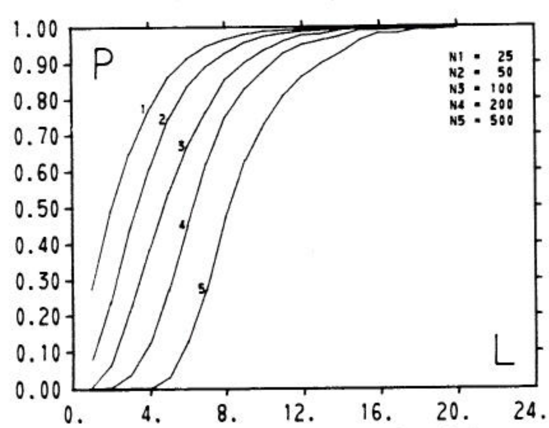
Q = 0.3 G = 0.0 R = 0.7



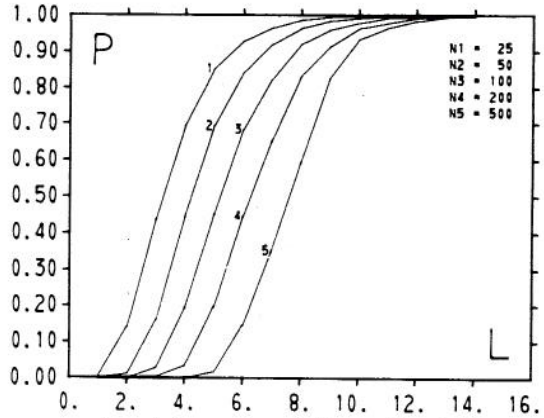
Q = 0.3 G = 0.2 R = 0.7



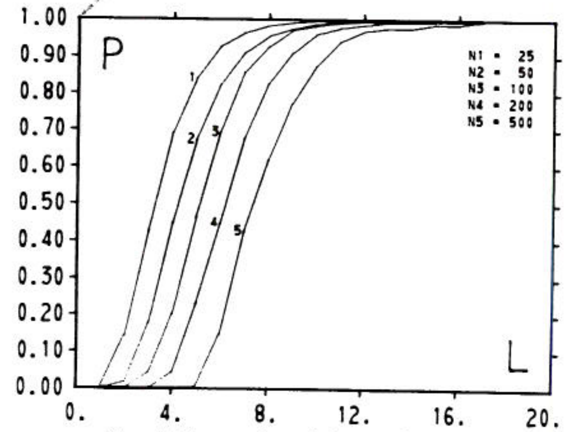
Q = 0.2 G = 0.0 R = 0.7



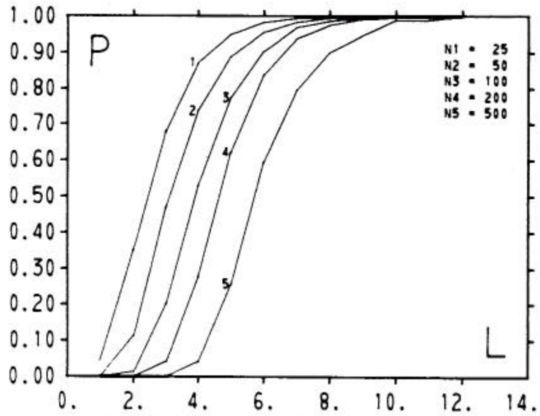
Q = 0.2 G = 0.2 R = 0.7



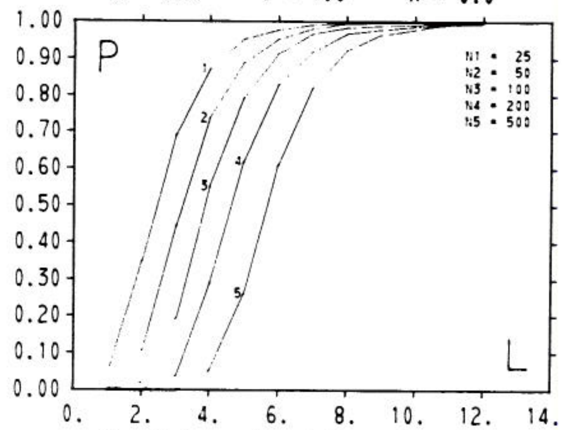
Q = 0.5 G = 0.5 R = 0.0



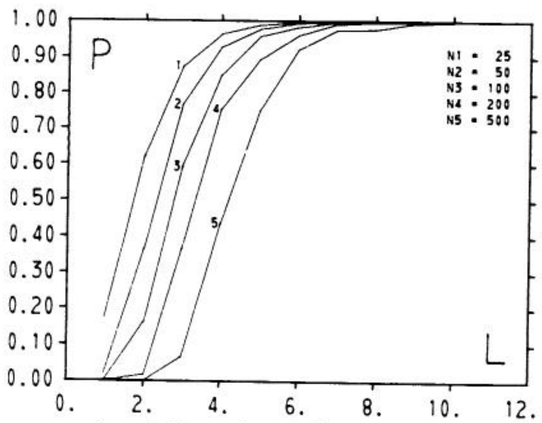
Q = 0.5 G = 1.0 R = 0.0



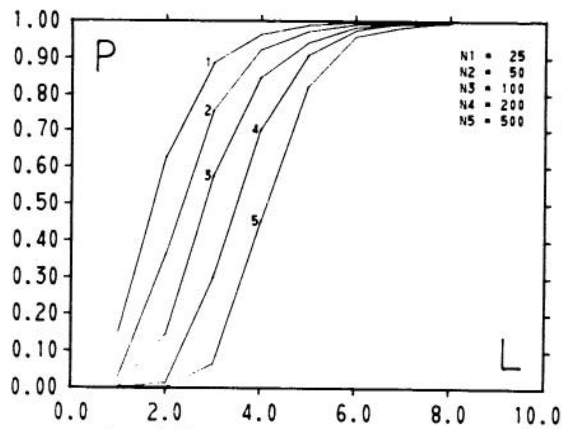
Q = 0.4 G = 0.5 R = 0.0



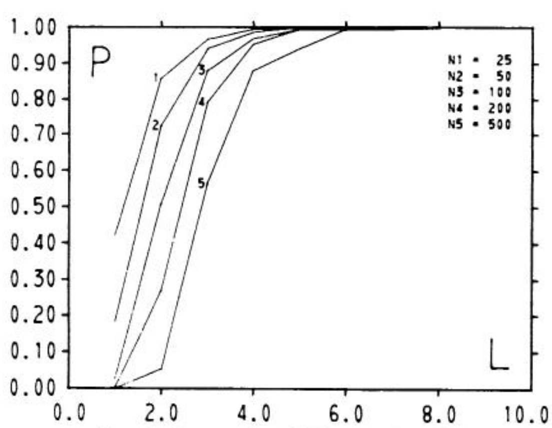
Q = 0.4 G = 1.0 R = 0.0



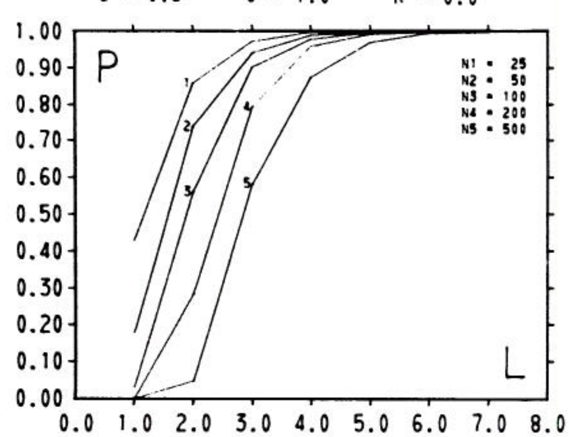
Q = 0.3 G = 0.5 R = 0.0



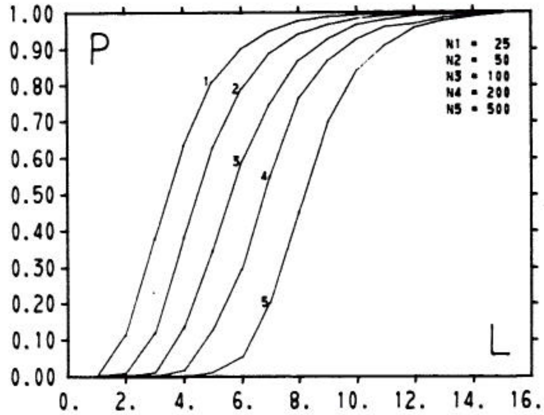
Q = 0.3 G = 1.0 R = 0.0



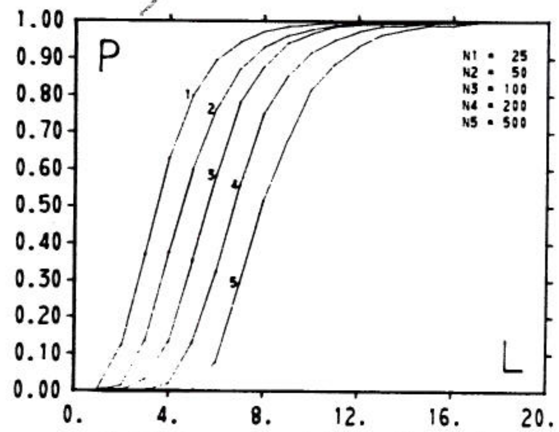
Q = 0.2 G = 0.5 R = 0.0



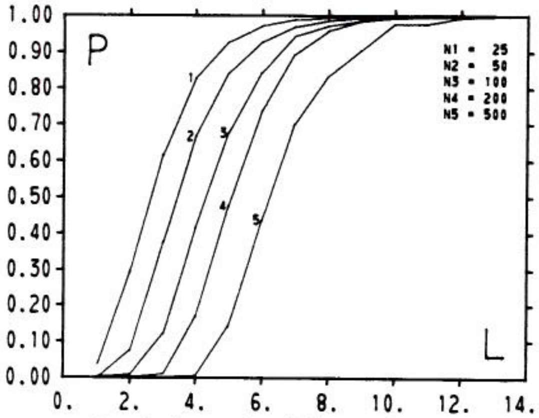
Q = 0.2 G = 1.0 R = 0.0



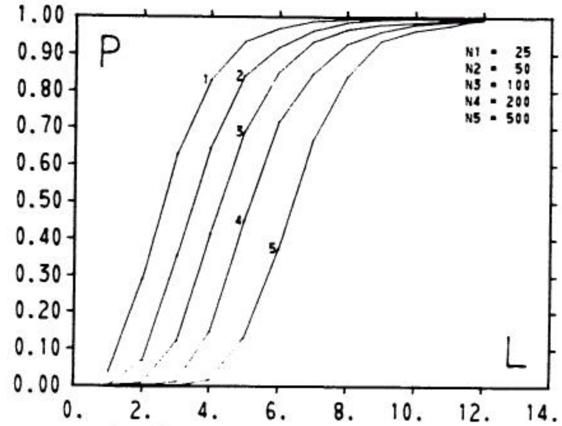
Q = 0.5 G = 0.5 R = 0.1



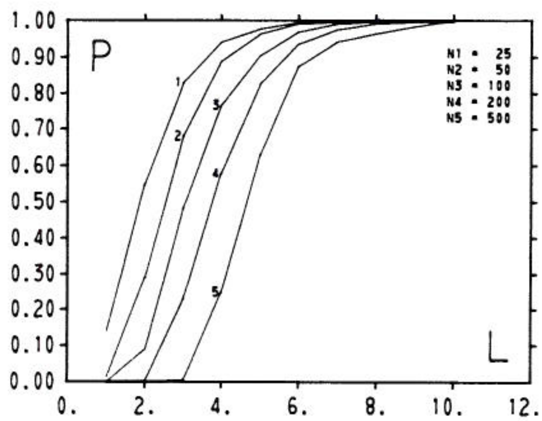
Q = 0.5 G = 1.0 R = 0.1



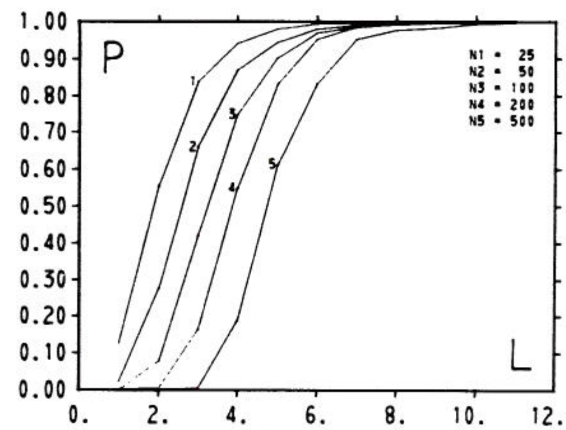
Q = 0.4 G = 0.5 R = 0.1



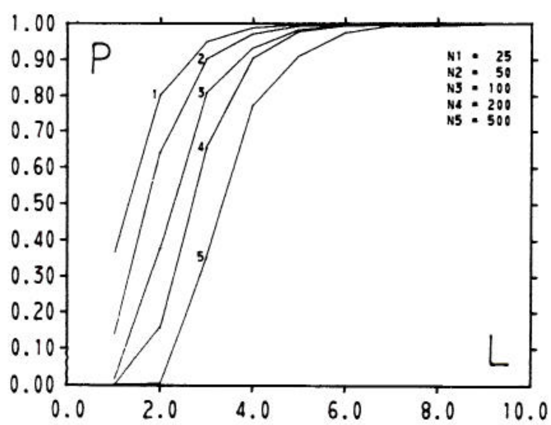
Q = 0.4 G = 1.0 R = 0.1



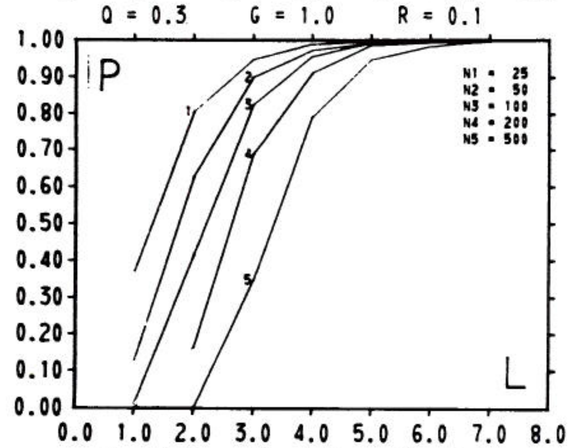
Q = 0.3 G = 0.5 R = 0.1



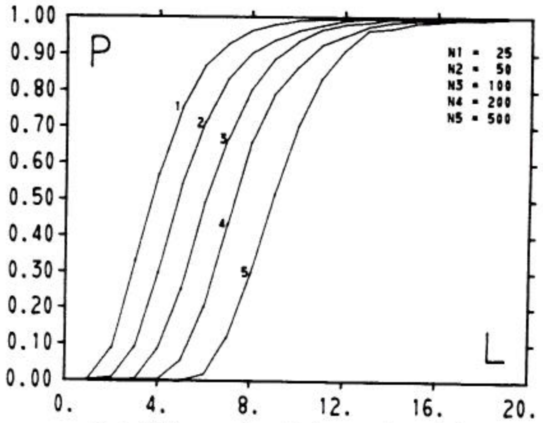
Q = 0.3 G = 1.0 R = 0.1



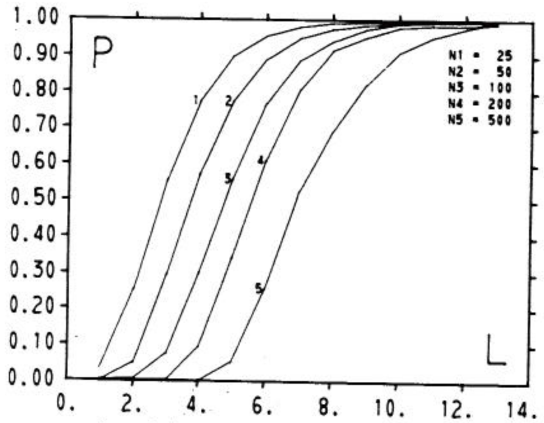
Q = 0.2 G = 0.5 R = 0.1



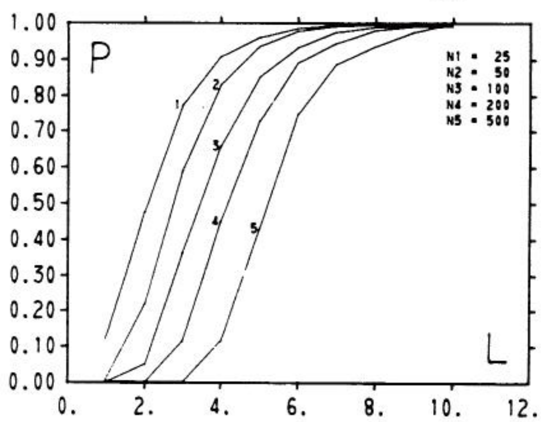
Q = 0.2 G = 1.0 R = 0.1



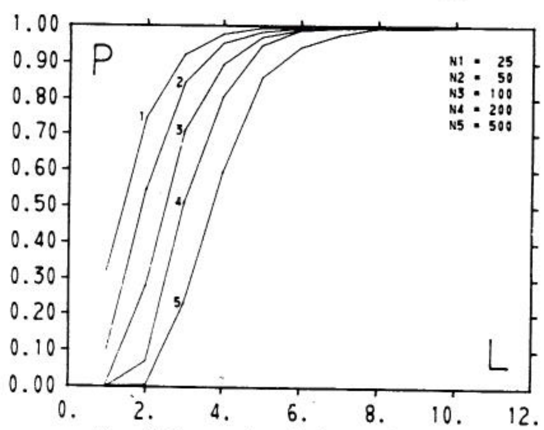
Q = 0.5 G = 0.5 R = 0.2



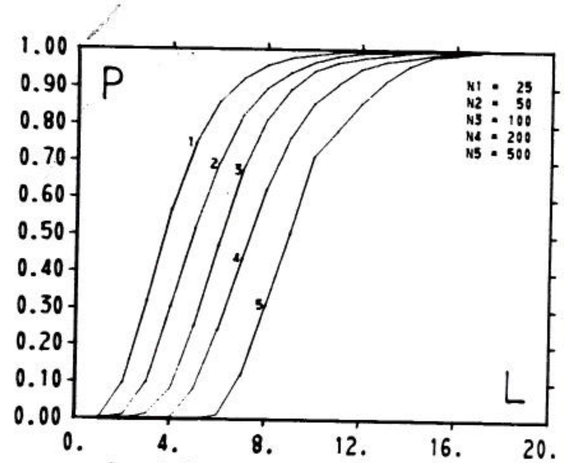
Q = 0.4 G = 0.5 R = 0.2



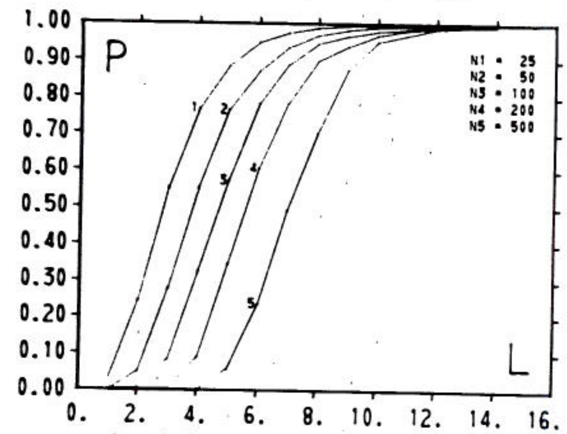
Q = 0.3 G = 0.5 R = 0.2



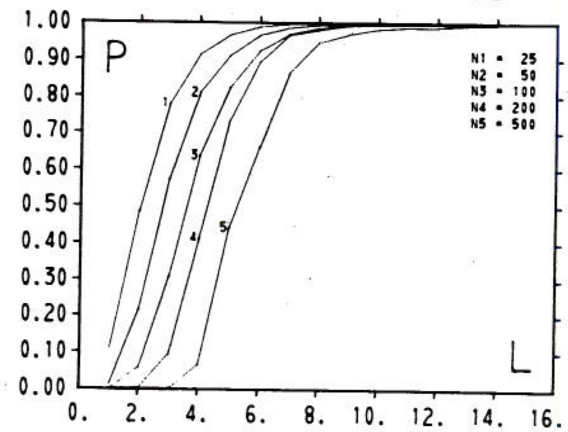
Q = 0.2 G = 0.5 R = 0.2



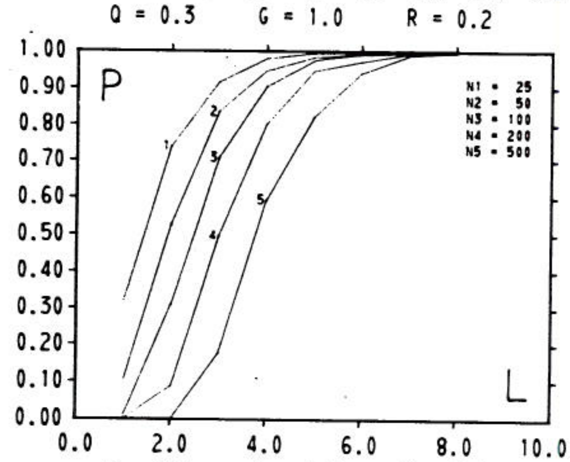
Q = 0.5 G = 1.0 R = 0.2



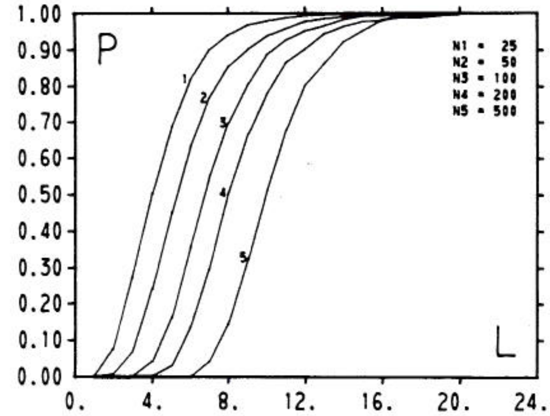
Q = 0.4 G = 1.0 R = 0.2



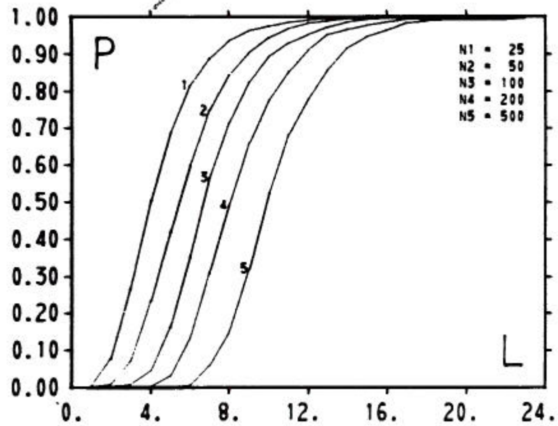
Q = 0.3 G = 1.0 R = 0.2



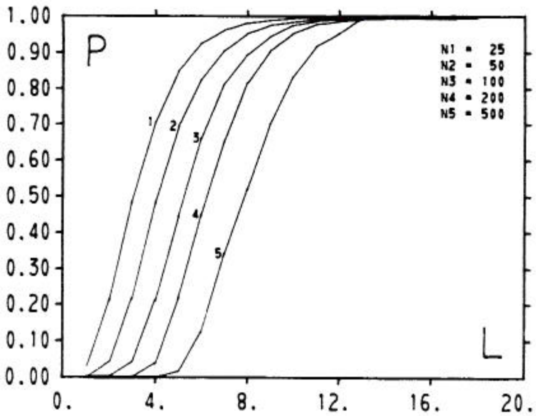
Q = 0.2 G = 1.0 R = 0.2



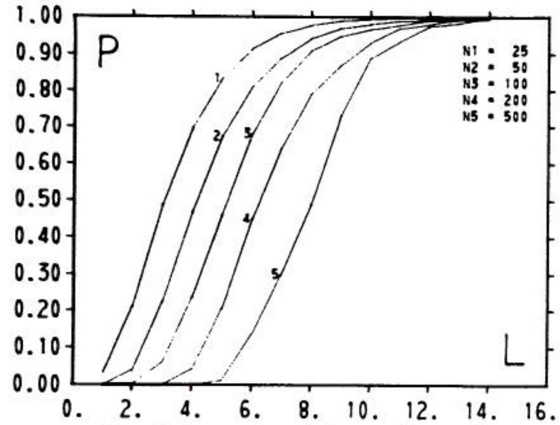
Q = 0.5 G = 0.5 R = 0.3



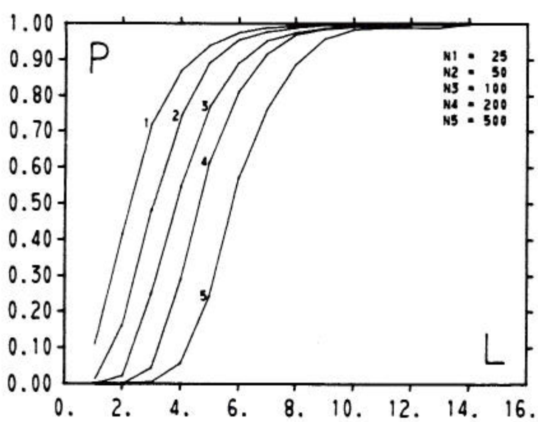
Q = 0.5 G = 1.0 R = 0.3



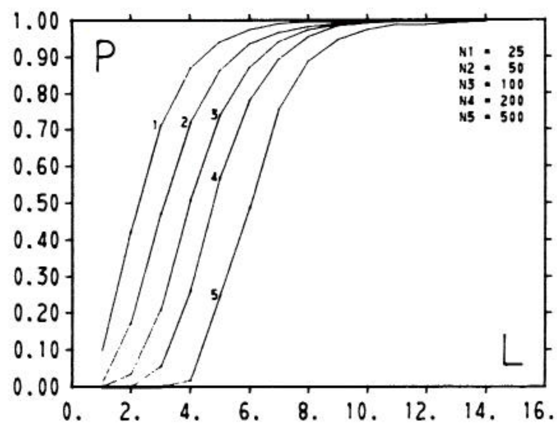
Q = 0.4 G = 0.5 R = 0.3



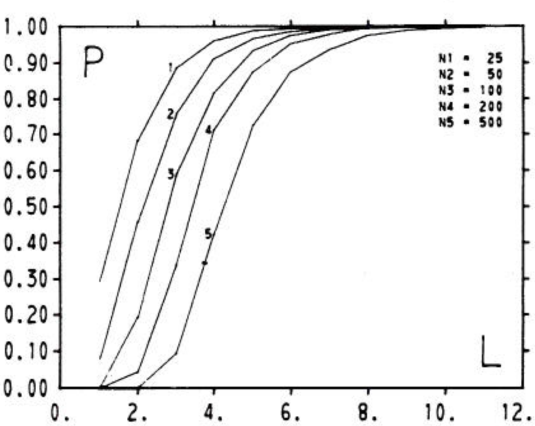
Q = 0.4 G = 1.0 R = 0.3



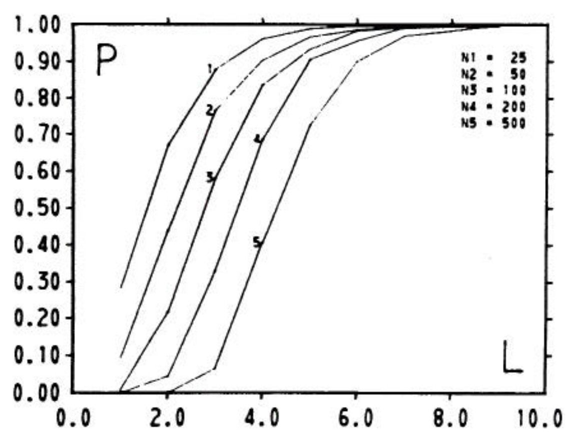
Q = 0.3 G = 0.5 R = 0.3



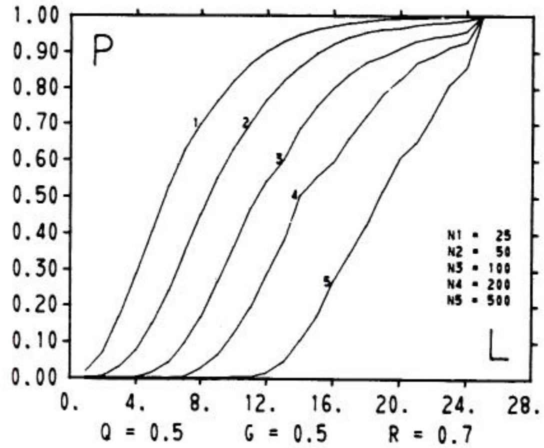
Q = 0.3 G = 1.0 R = 0.3

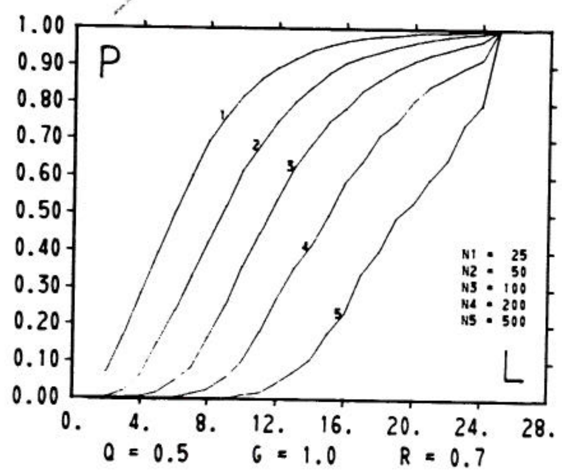


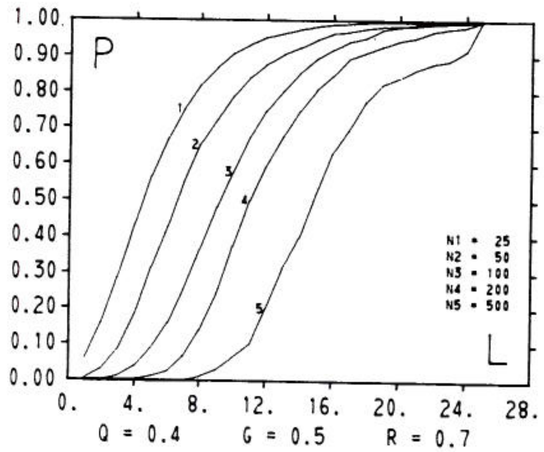
Q = 0.2 G = 0.5 R = 0.3

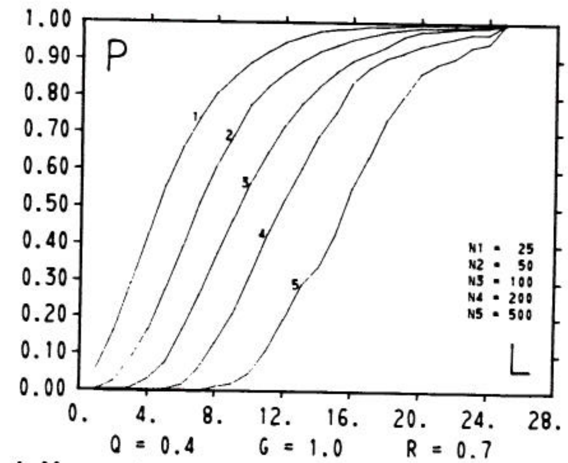


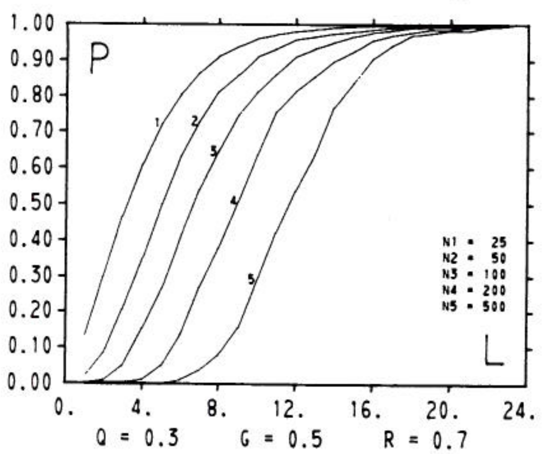
Q = 0.2 G = 1.0 R = 0.3

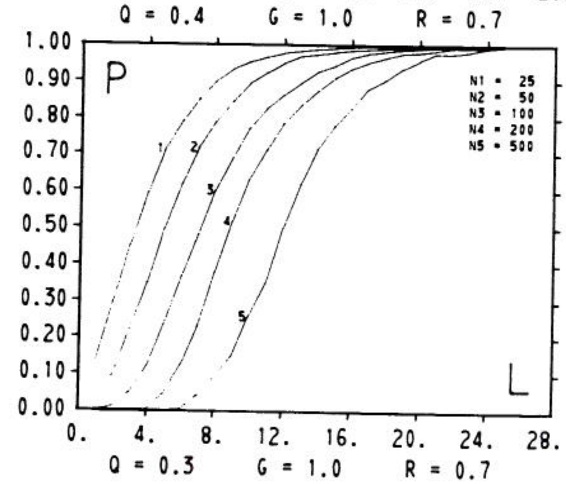

P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 0.5 R = 0.7

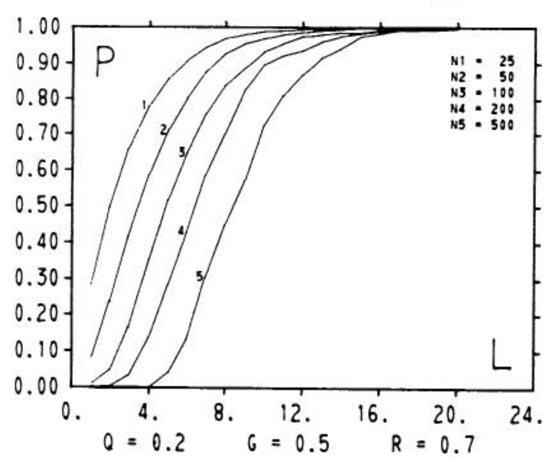

P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 1.0 R = 0.7

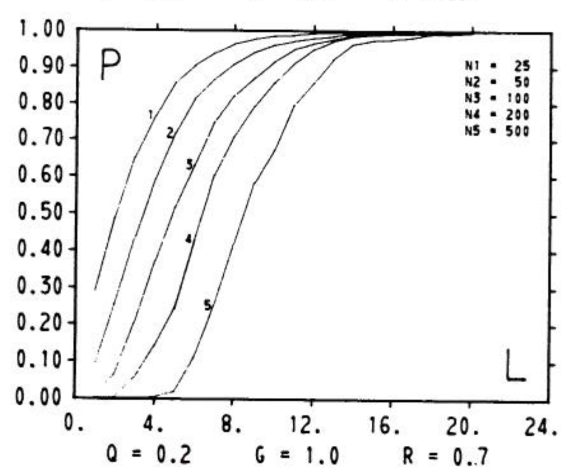

P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.4 G = 0.5 R = 0.7

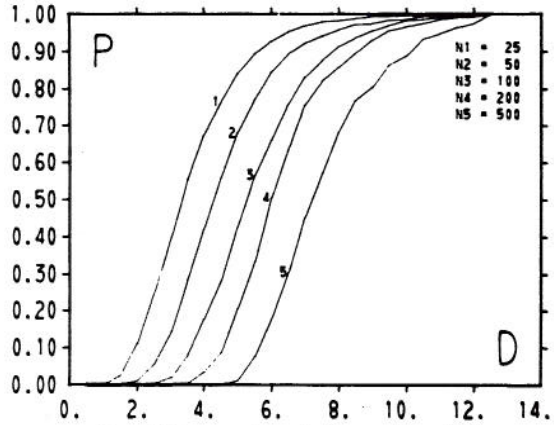

P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.4 G = 1.0 R = 0.7


P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.3 G = 0.5 R = 0.7

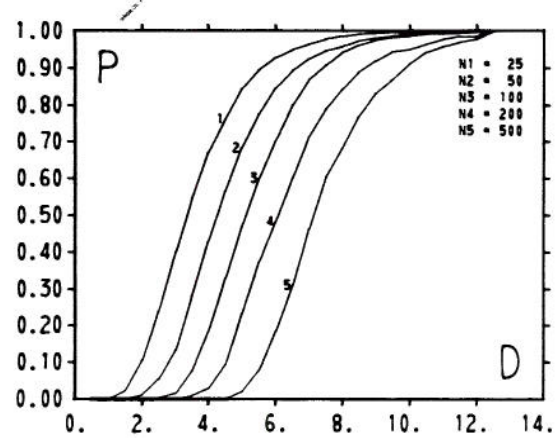

P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.3 G = 1.0 R = 0.7


P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.2 G = 0.5 R = 0.7

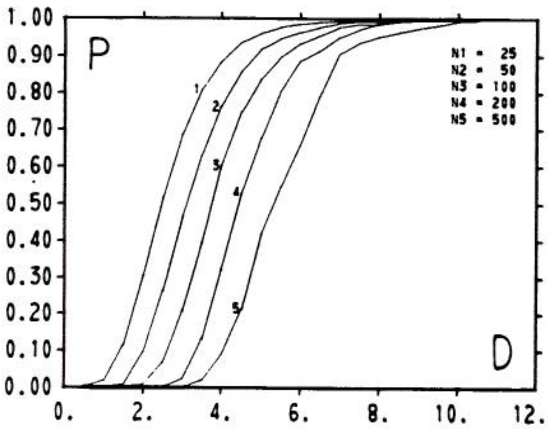

P
L
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.2 G = 1.0 R = 0.7



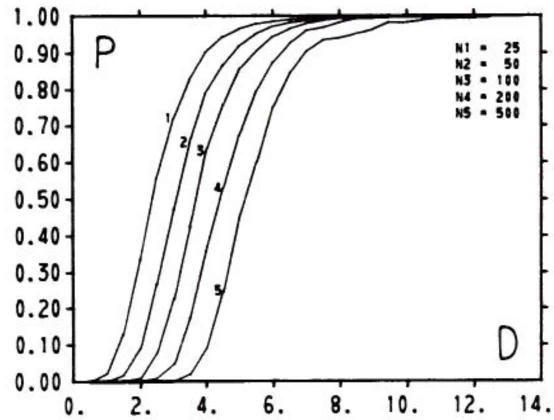
Q = 0.5 G = 0.0 R = 0.0



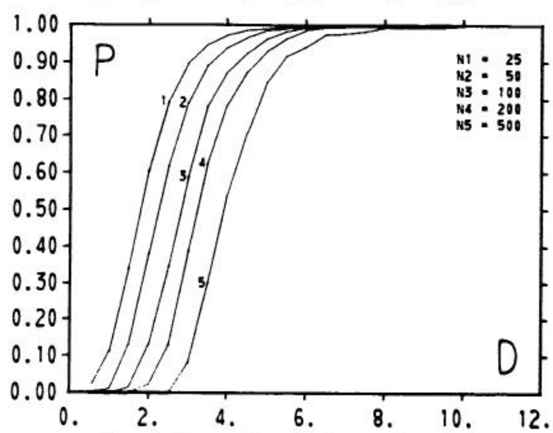
Q = 0.5 G = 0.2 R = 0.0



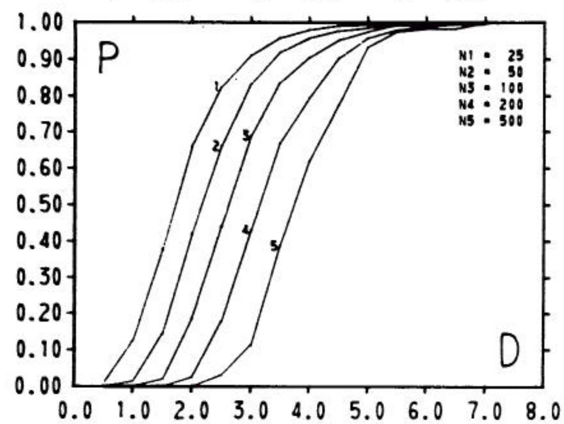
Q = 0.4 G = 0.0 R = 0.0



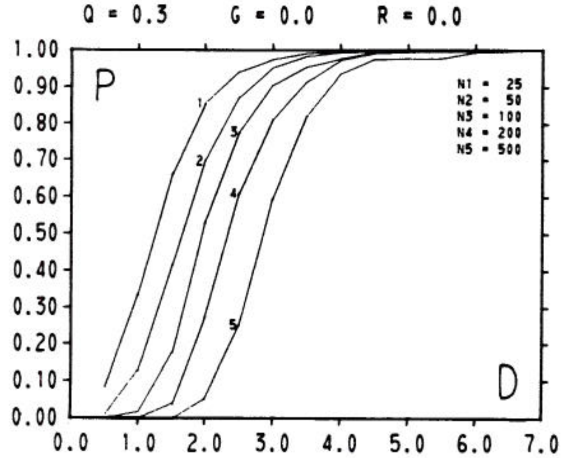
Q = 0.4 G = 0.2 R = 0.0



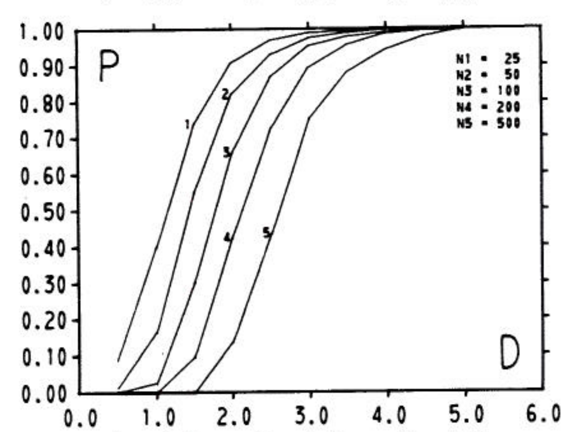
Q = 0.3 G = 0.0 R = 0.0



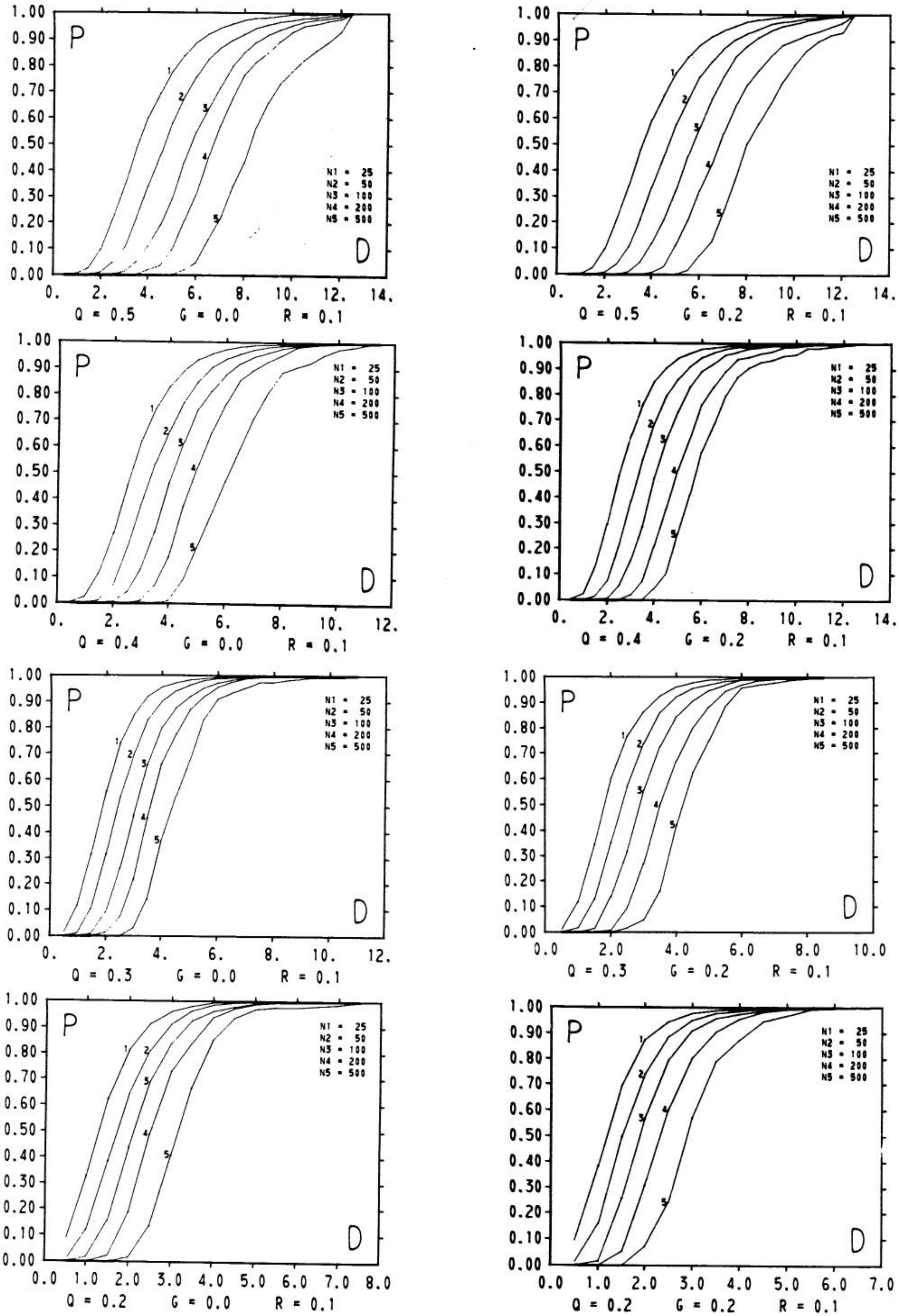
Q = 0.3 G = 0.2 R = 0.0

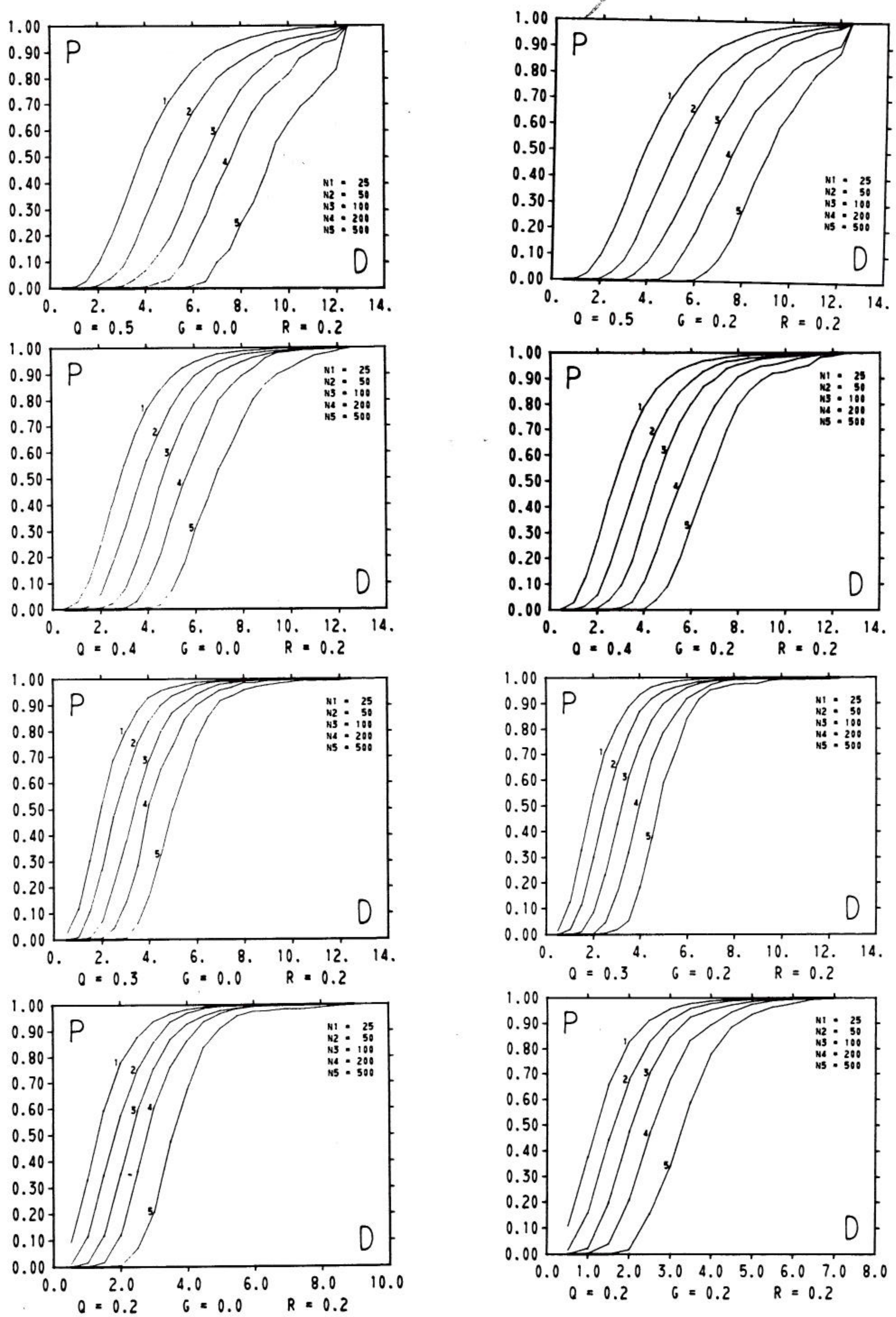


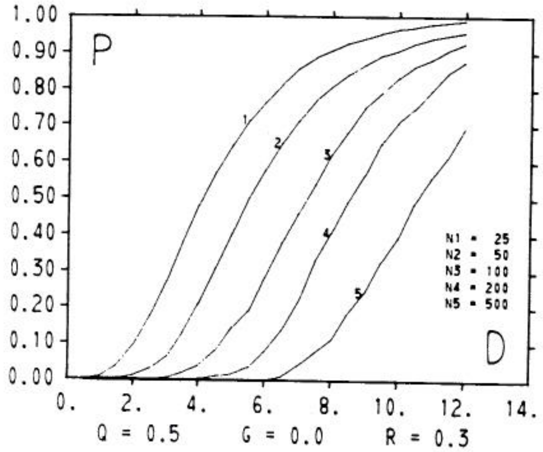
Q = 0.2 G = 0.0 R = 0.0

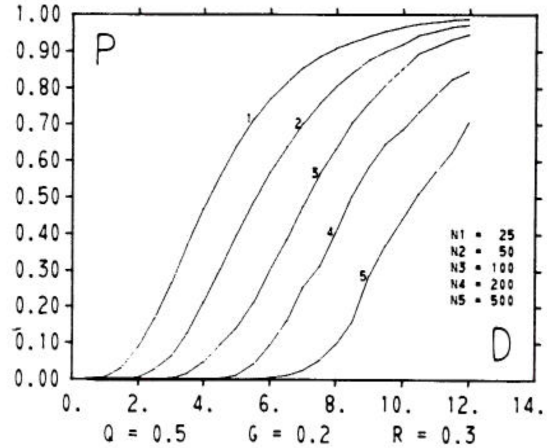


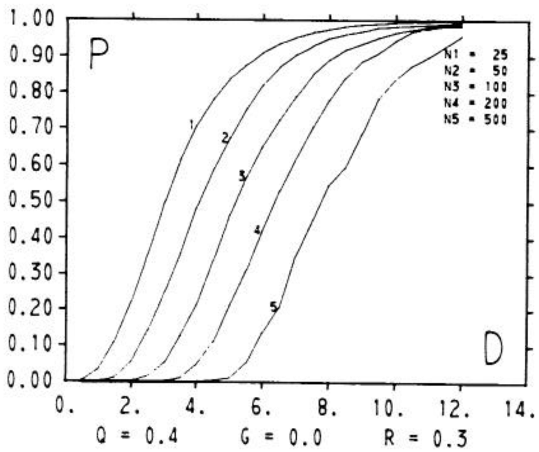
Q = 0.2 G = 0.2 R = 0.0

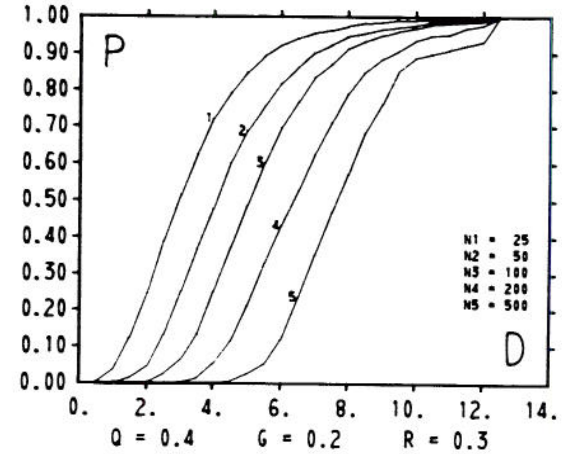

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 0.0 R = 0.1
Q = 0.5 G = 0.2 R = 0.1
Q = 0.4 G = 0.0 R = 0.1
Q = 0.4 G = 0.2 R = 0.1
Q = 0.3 G = 0.0 R = 0.1
Q = 0.3 G = 0.2 R = 0.1
Q = 0.2 G = 0.0 R = 0.1
Q = 0.2 G = 0.2 R = 0.1

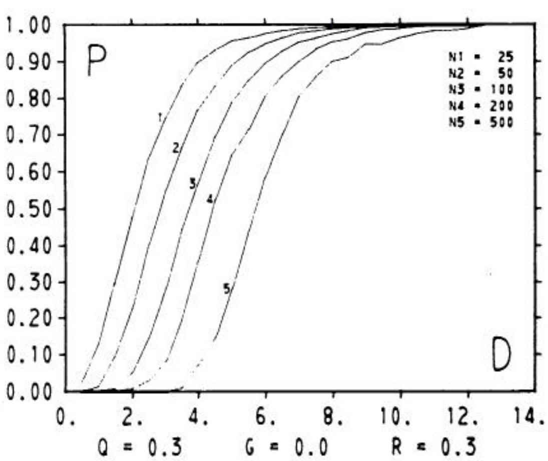

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 0.0 R = 0.2
Q = 0.5 G = 0.2 R = 0.2
Q = 0.4 G = 0.0 R = 0.2
Q = 0.4 G = 0.2 R = 0.2
Q = 0.3 G = 0.0 R = 0.2
Q = 0.3 G = 0.2 R = 0.2
Q = 0.2 G = 0.0 R = 0.2
Q = 0.2 G = 0.2 R = 0.2

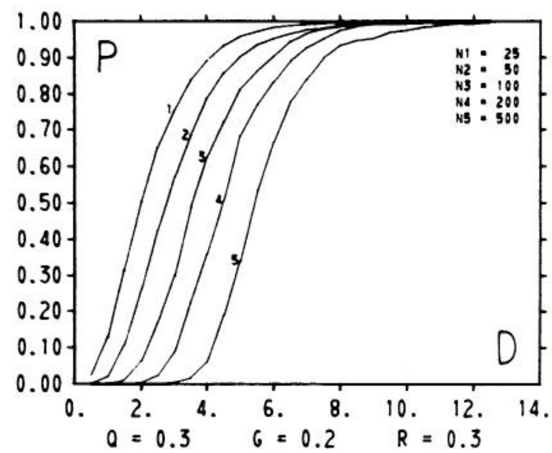

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 0.0 R = 0.3

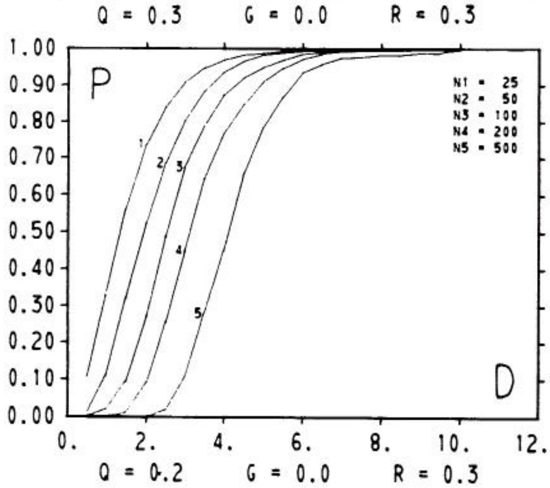

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 0.2 R = 0.3

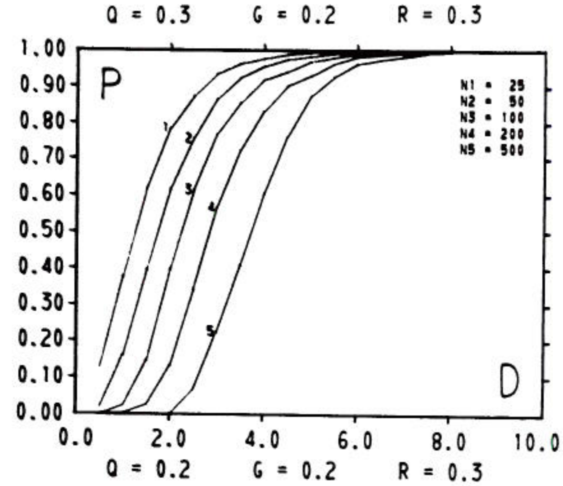

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.4 G = 0.0 R = 0.3

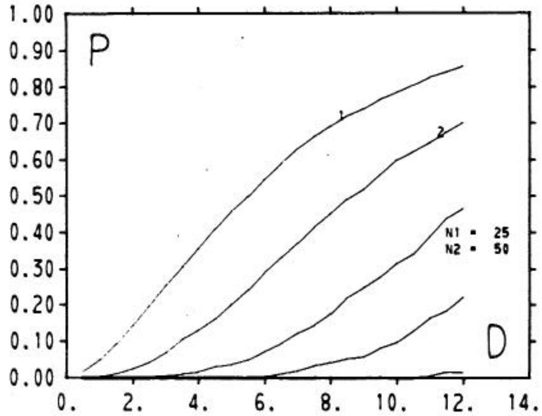

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.4 G = 0.2 R = 0.3


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.3 G = 0.0 R = 0.3

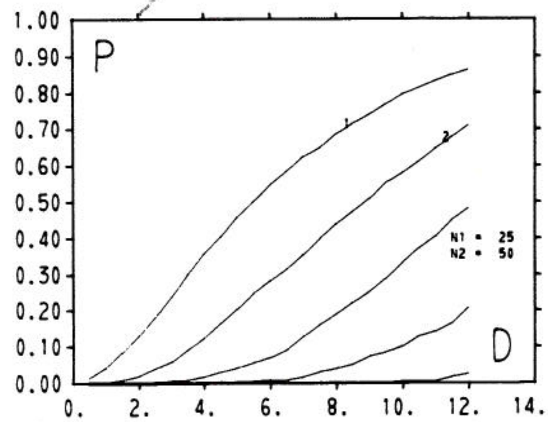

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.3 G = 0.2 R = 0.3


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.2 G = 0.0 R = 0.3

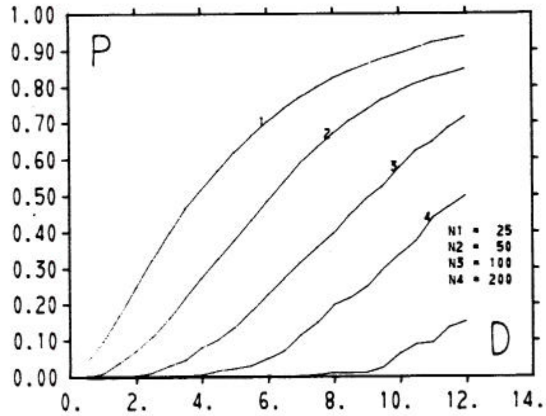

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.2 G = 0.2 R = 0.3



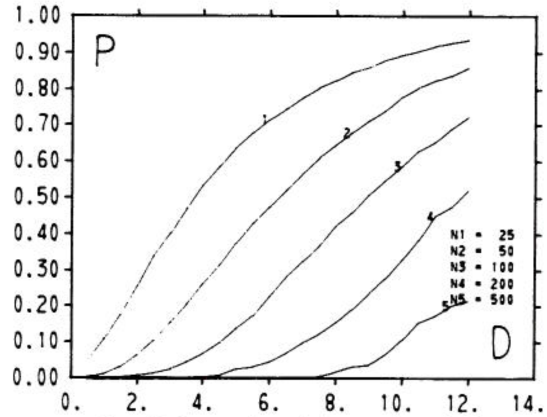
Q = 0.5 G = 0.0 R = 0.7



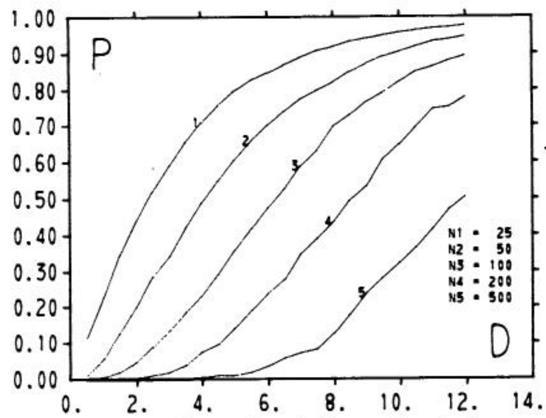
Q = 0.5 G = 0.2 R = 0.7



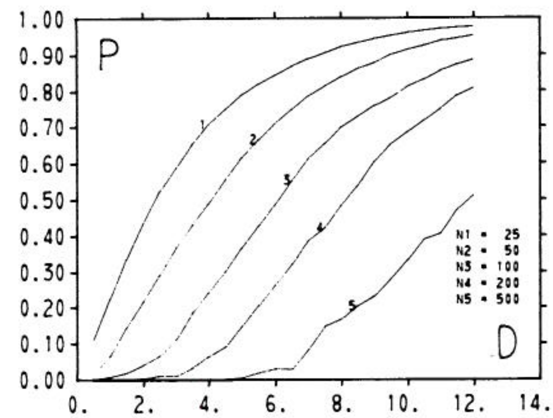
Q = 0.4 G = 0.0 R = 0.7



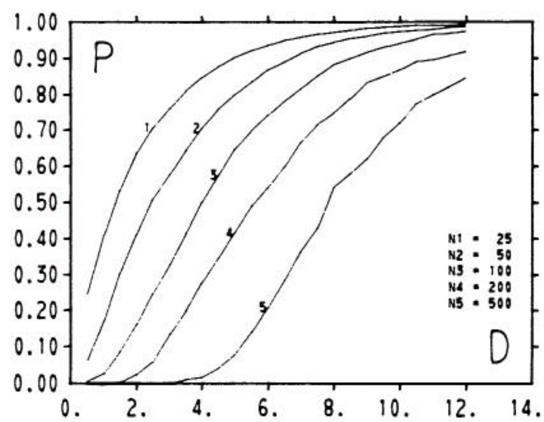
Q = 0.4 G = 0.2 R = 0.7



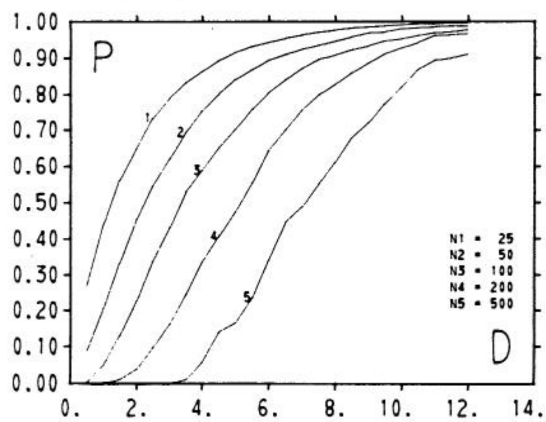
Q = 0.3 G = 0.0 R = 0.7



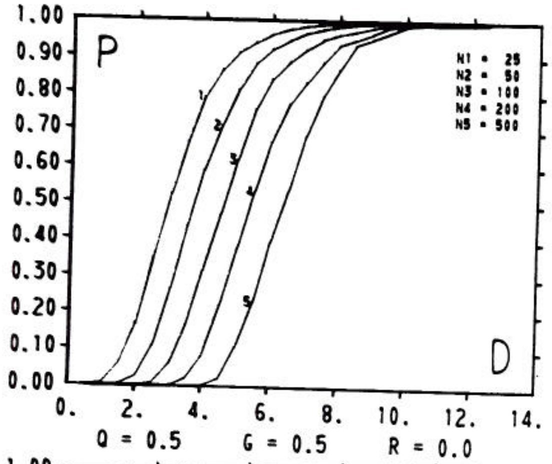
Q = 0.3 G = 0.2 R = 0.7

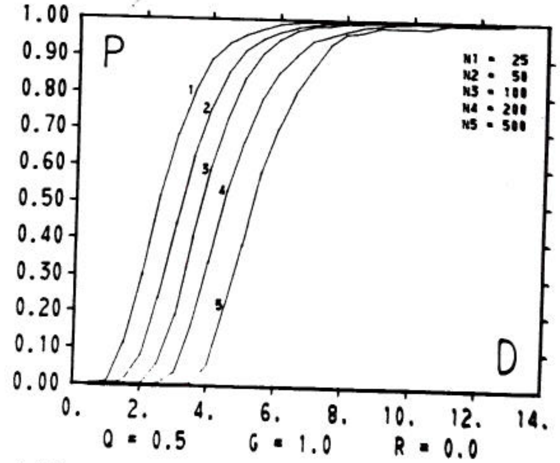


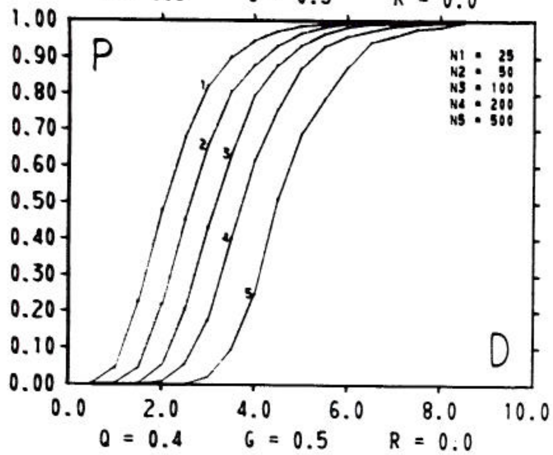
Q = 0.2 G = 0.0 R = 0.7

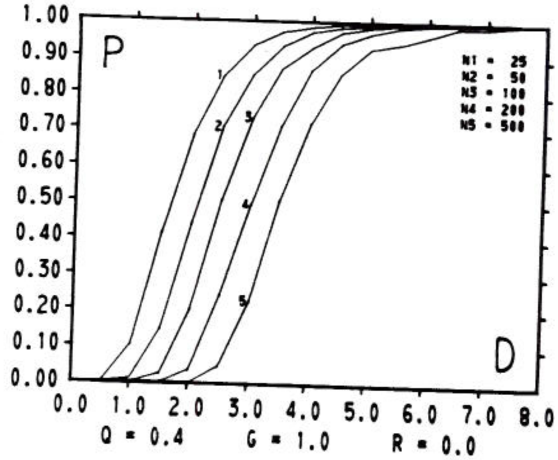


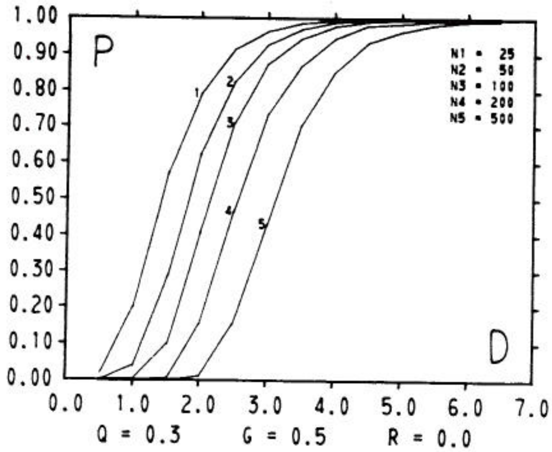
Q = 0.2 G = 0.2 R = 0.7

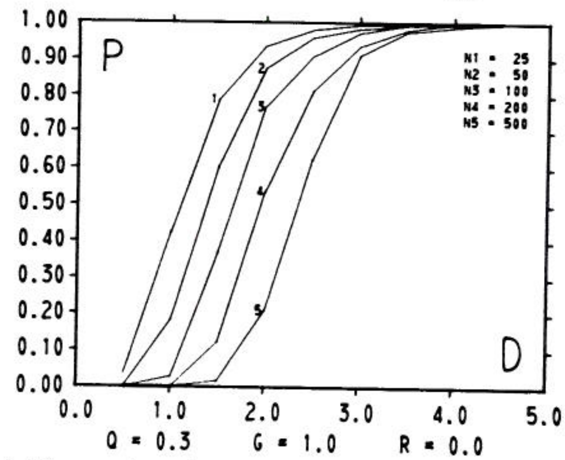

P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.5 G = 0.5 R = 0.0

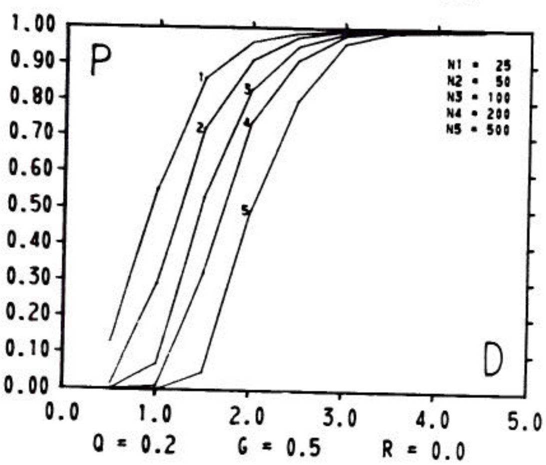

P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.5 G = 1.0 R = 0.0

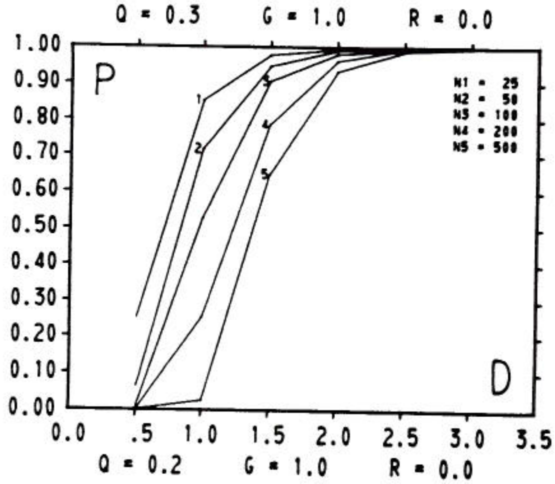

P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.4 G = 0.5 R = 0.0

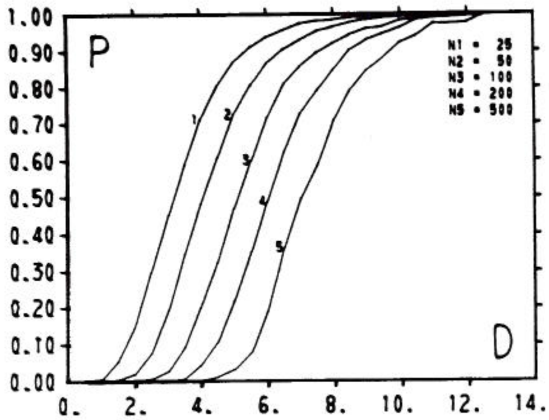

P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.4 G = 1.0 R = 0.0


P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.3 G = 0.5 R = 0.0

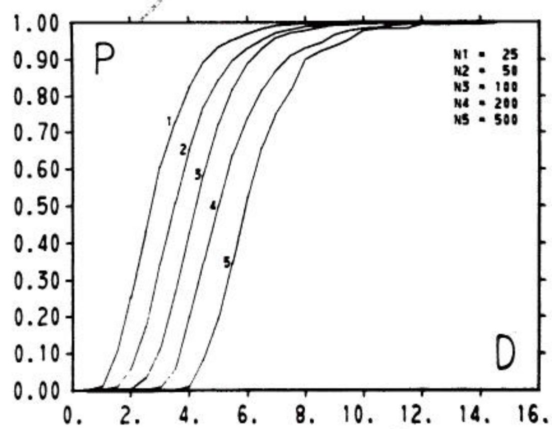

P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.3 G = 1.0 R = 0.0


P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.2 G = 0.5 R = 0.0

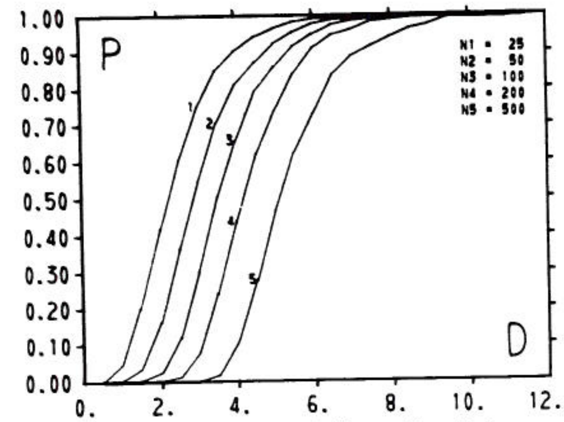

P
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
D
Q = 0.2 G = 1.0 R = 0.0



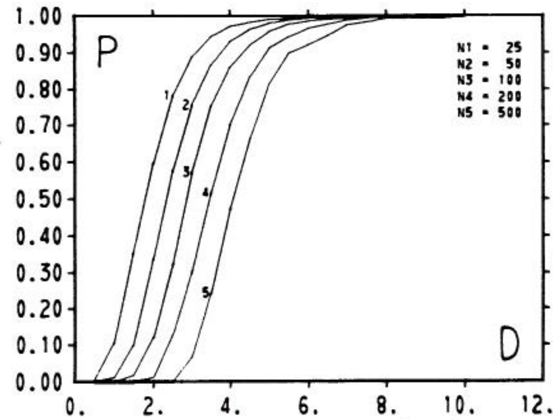
Q = 0.5 G = 0.5 R = 0.1



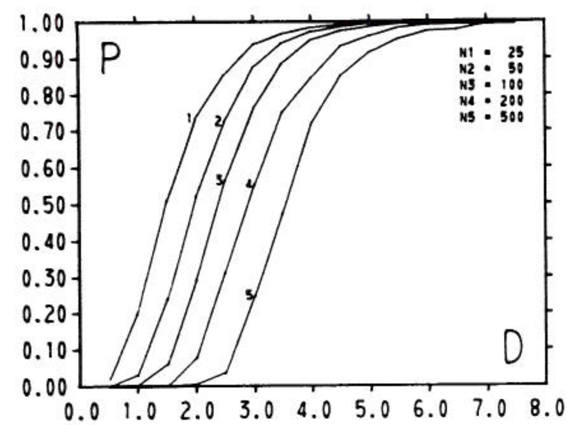
Q = 0.5 G = 1.0 R = 0.1



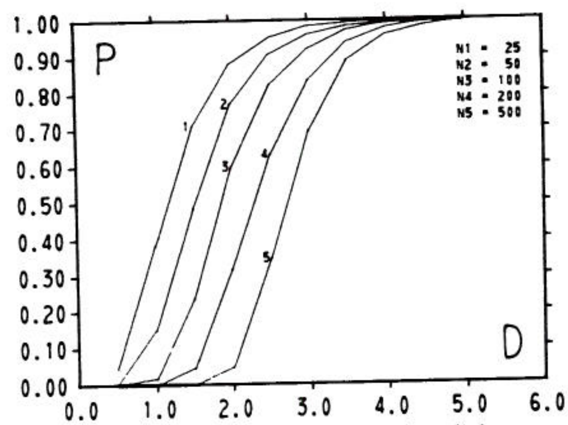
Q = 0.4 G = 0.5 R = 0.1



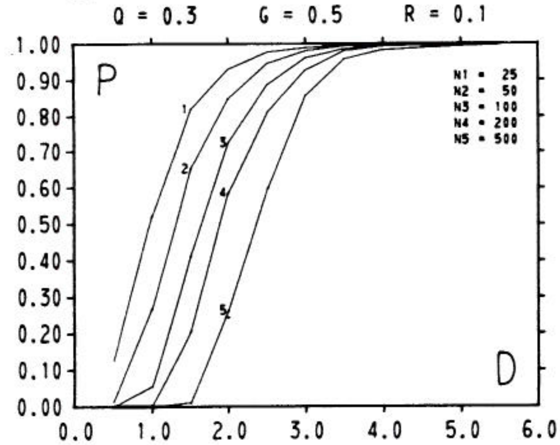
Q = 0.4 G = 1.0 R = 0.1



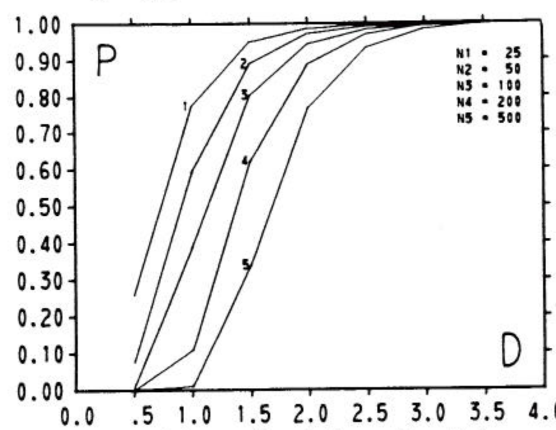
Q = 0.3 G = 0.5 R = 0.1



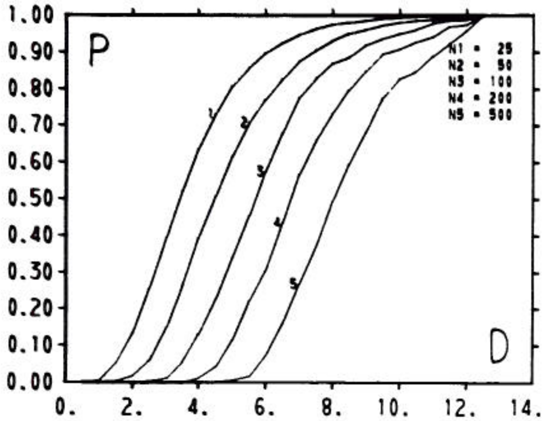
Q = 0.3 G = 1.0 R = 0.1



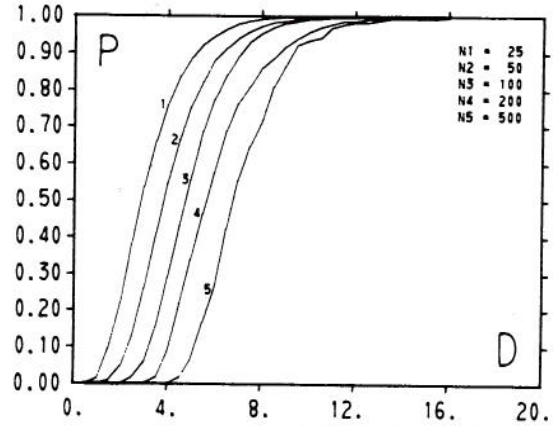
Q = 0.2 G = 0.5 R = 0.1



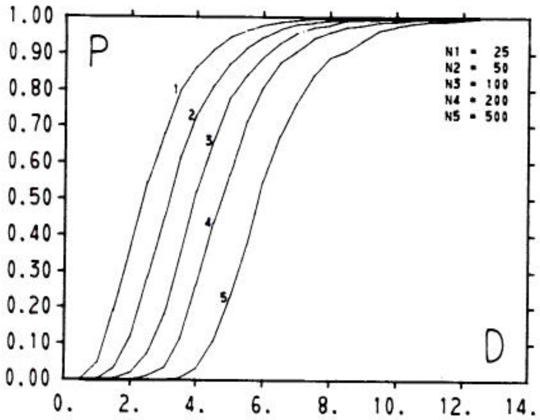
Q = 0.2 G = 1.0 R = 0.1



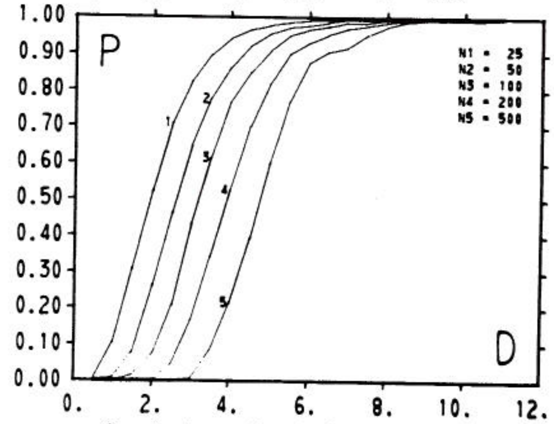
Q = 0.5 G = 0.5 R = 0.2



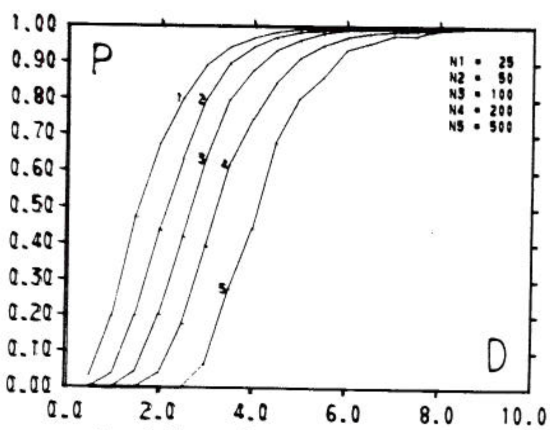
Q = 0.5 G = 1.0 R = 0.2



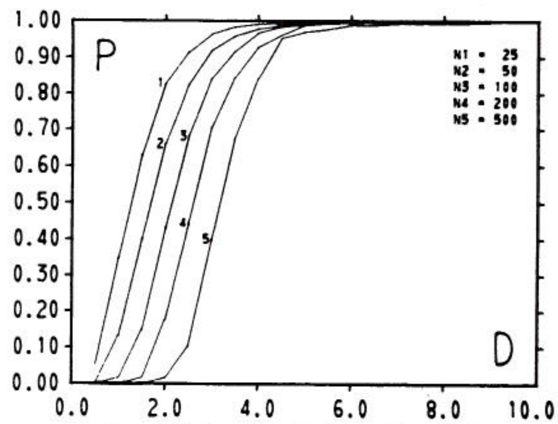
Q = 0.4 G = 0.5 R = 0.2



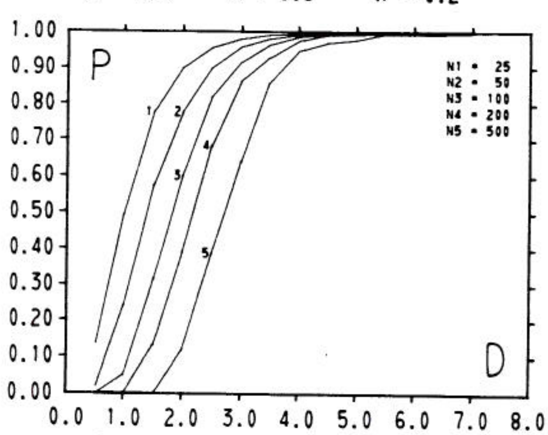
Q = 0.4 G = 1.0 R = 0.2



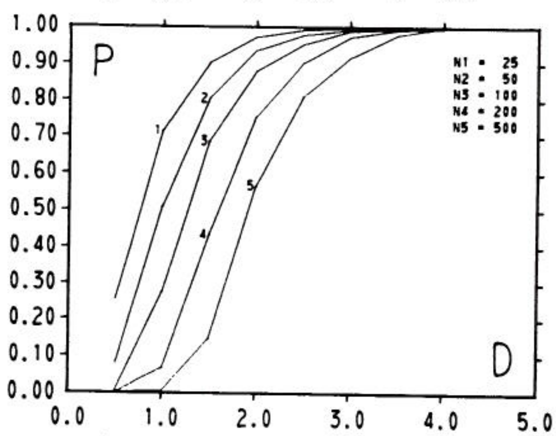
Q = 0.3 G = 0.5 R = 0.2



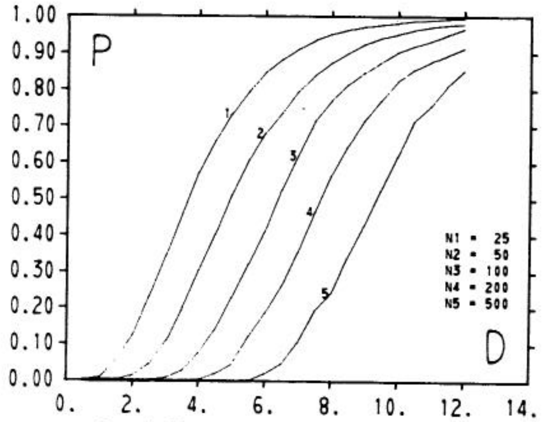
Q = 0.3 G = 1.0 R = 0.2



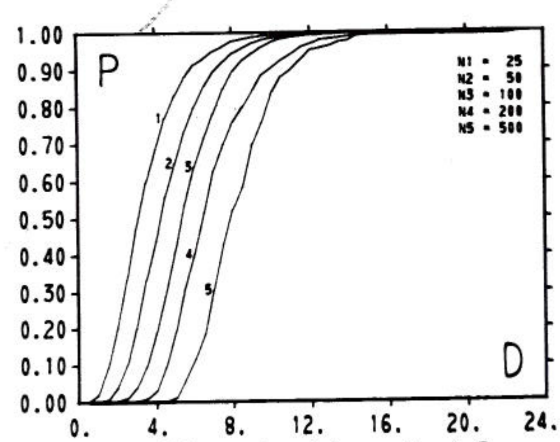
Q = 0.2 G = 0.5 R = 0.2



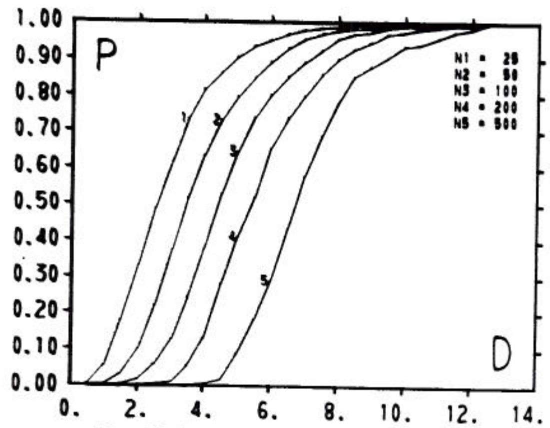
Q = 0.2 G = 1.0 R = 0.2



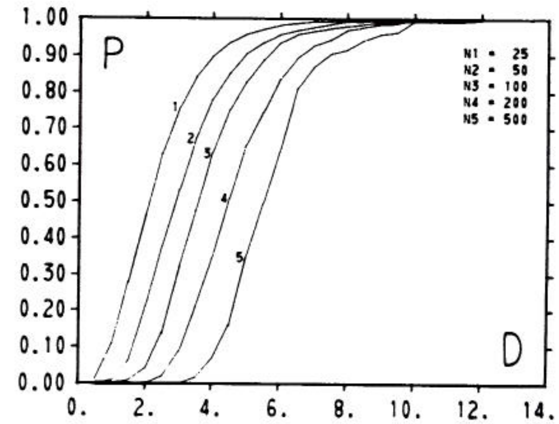
Q = 0.5 G = 0.5 R = 0.3



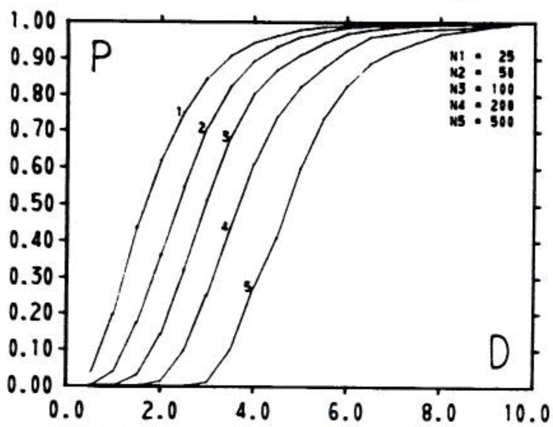
Q = 0.5 G = 1.0 R = 0.3



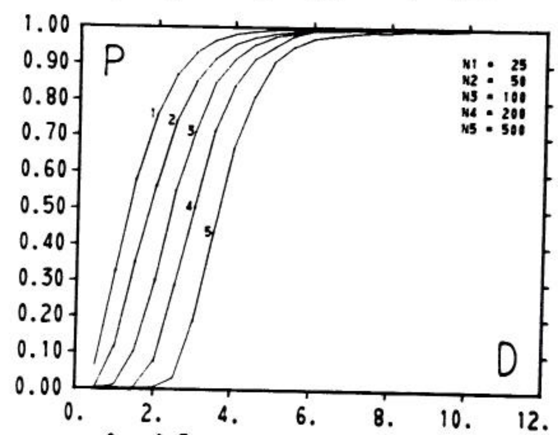
Q = 0.4 G = 0.5 R = 0.3



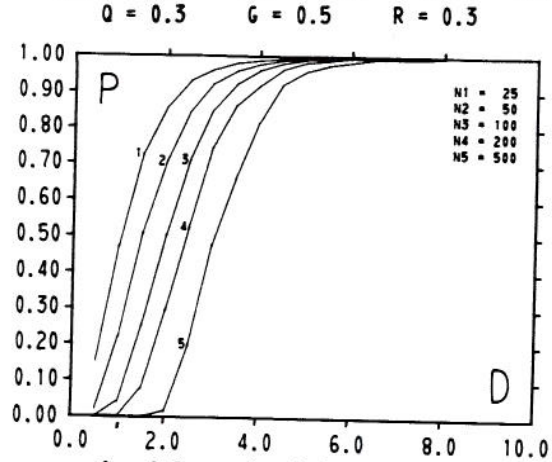
Q = 0.4 G = 1.0 R = 0.3



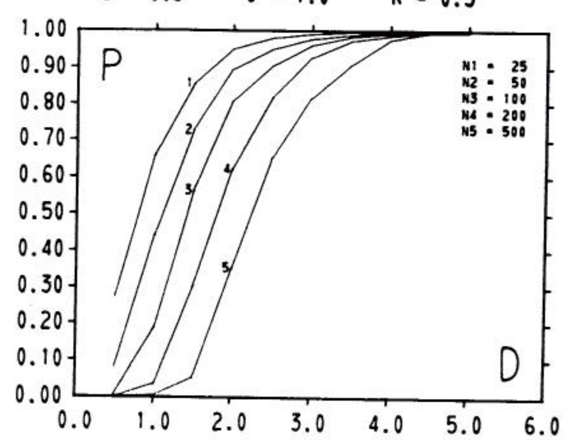
Q = 0.3 G = 0.5 R = 0.3



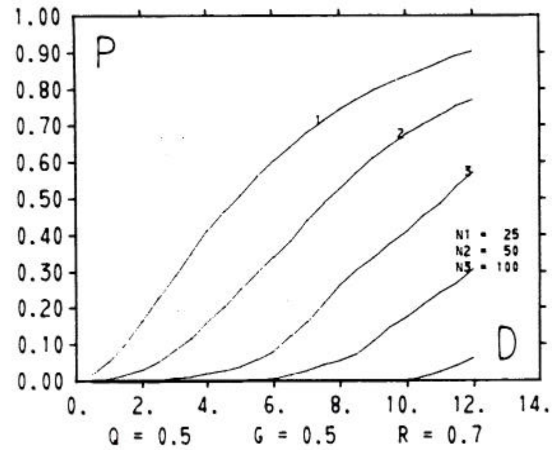
Q = 0.3 G = 1.0 R = 0.3

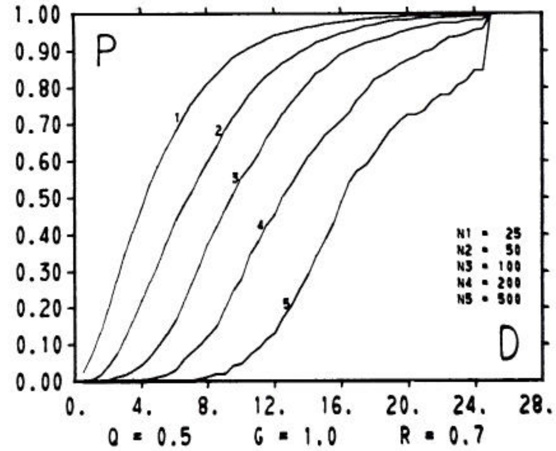


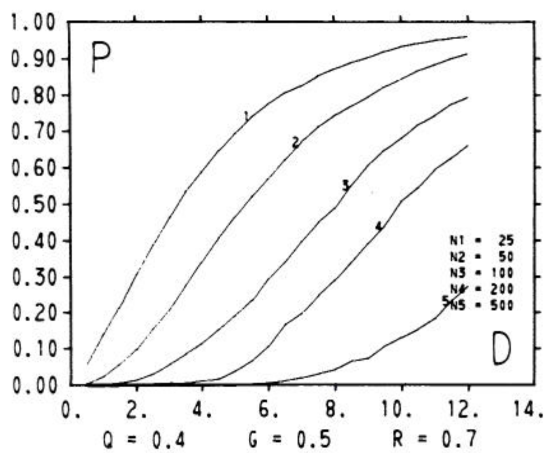
Q = 0.2 G = 0.5 R = 0.3

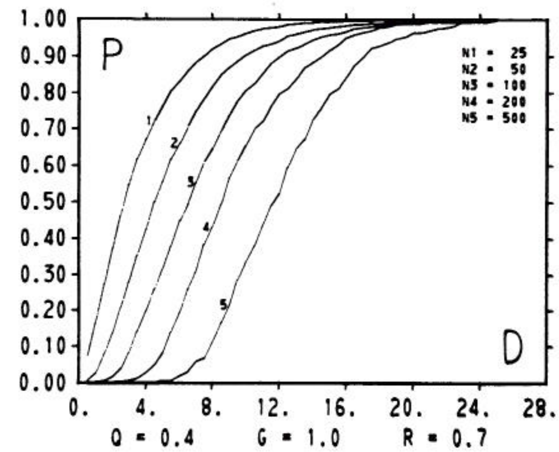


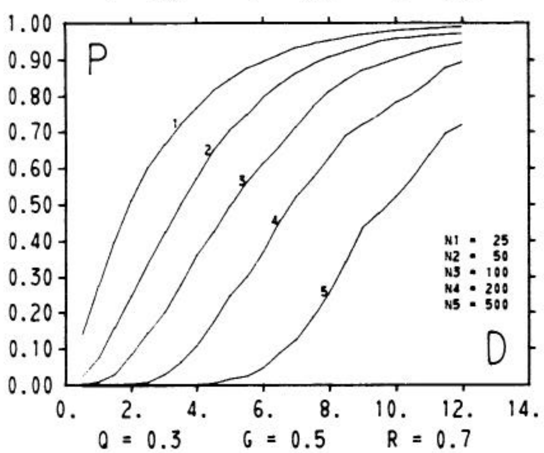
Q = 0.2 G = 1.0 R = 0.3

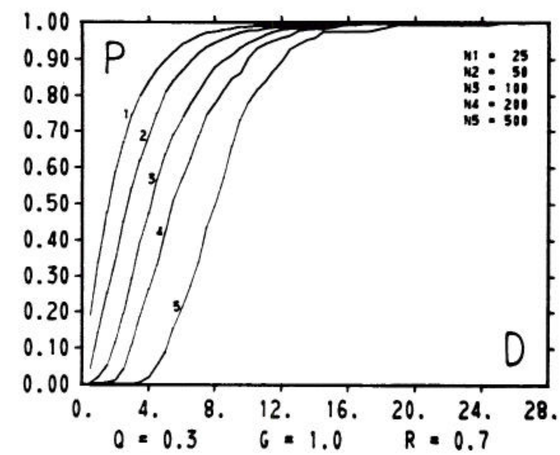

P
D
N1 = 25
N2 = 50
N3 = 100
Q = 0.5 G = 0.5 R = 0.7

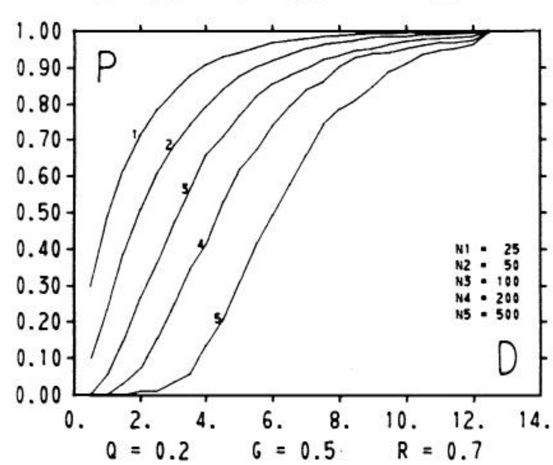

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.5 G = 1.0 R = 0.7

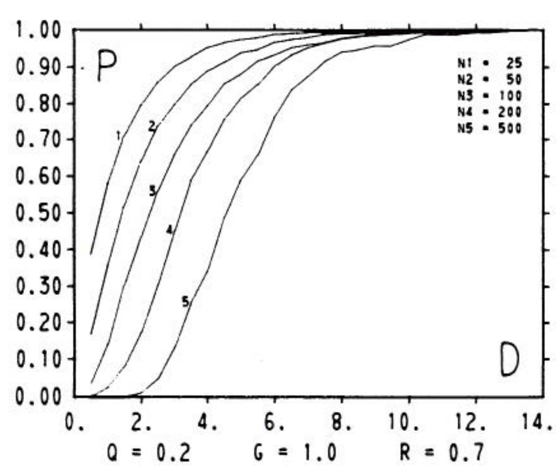

P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.4 G = 0.5 R = 0.7


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.4 G = 1.0 R = 0.7


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.3 G = 0.5 R = 0.7


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.3 G = 1.0 R = 0.7


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.2 G = 0.5 R = 0.7


P
D
N1 = 25
N2 = 50
N3 = 100
N4 = 200
N5 = 500
Q = 0.2 G = 1.0 R = 0.7

APPENDIX III

This appendix gives the values of parameters of frequency distributions presented in Appendix II. Table III-1 shows the mean and standard deviation of the distributions of the longest negative run-length for five values of N, (25, 50, 100, 200, 500) , five values of ρ, (0.0, 0.1, 0.2, 0.3, 0.5), four values of the truncation level probability $Q = F(x_o)$, (0.5, 0.4, 0.3, 0.2) , and three values of the skewness coefficient γ, (0.0, 0.2, 0.5) . Table III-2 gives the mean and the standard deviation for distributions of the largest negative run-sum for the same parameter values as in Table III-1. Tables III-3 and the standard deviation of the logarithms, μ_n , and the standard deviation of the logarithms, σ_n , as computed by Eqs. (12) and (13) in Chapter III, for the distributions of the longest negative run-length and the largest negative run-sum, respectively . These two parameters refer to the approximate fits of lognormal distributions to the computed frequency distributions.

TABLE III -1

	N=25				N=50				N=100				N=200				N=500			
q=0.5	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2
LONGEST NEGATIVE RUN LENGTH MEANS																				
Skewness = 0.0																				
ρ=0.0	3.97	3.17	2.38	1.75	5.00	3.86	2.93	2.17	6.06	4.57	3.47	2.57	7.04	5.40	4.01	3.03	8.26	6.30	4.86	3.60
=0.1	4.23	3.38	2.55	1.88	5.35	4.17	3.20	2.36	6.48	4.95	3.80	2.82	7.34	5.84	4.45	3.29	9.06	6.94	5.42	3.96
=0.2	4.51	3.61	2.74	2.04	5.75	4.47	3.47	2.56	7.01	5.43	4.09	3.08	8.10	6.33	4.89	3.67	9.82	7.55	5.97	4.47
=0.3	4.87	3.88	2.94	2.19	6.21	4.85	3.79	2.77	7.60	6.00	4.55	3.38	8.94	7.00	5.38	4.08	10.81	8.28	6.58	4.97
=0.7	7.07	5.84	4.24	3.07	9.75	7.79	6.02	4.38	12.58	10.17	7.63	5.82	15.20	12.22	9.73	7.44	18.70	15.70	12.45	9.62
Skewness = 0.2																				
=0.0	4.17	3.23	2.45	1.77	5.22	4.04	3.06	2.18	6.11	4.78	3.63	2.57	7.24	5.58	4.23	3.01	8.53	6.73	4.95	3.56
=0.1	4.42	3.43	2.62	1.90	5.56	4.37	3.32	2.38	6.58	5.19	3.99	2.80	7.80	6.12	4.62	3.26	9.30	7.18	5.53	3.91
=0.2	4.69	3.70	2.82	2.04	5.96	4.73	3.58	2.60	7.15	5.59	4.36	3.07	8.49	6.56	5.00	3.62	10.02	7.87	6.13	4.38
=0.3	5.05	4.00	3.04	2.21	6.45	5.17	3.94	2.83	7.76	6.06	4.79	3.41	9.27	7.29	5.52	3.99	11.09	8.77	6.73	4.93
=0.7	7.26	5.81	4.51	3.08	9.97	8.14	6.16	4.34	12.79	10.61	8.07	5.76	16.13	12.78	9.96	7.24	19.68	15.77	13.16	9.30
Skewness = 0.5																				
=0.0	3.99	3.13	2.38	1.76	4.99	3.86	2.95	2.16	5.98	4.63	3.46	2.62	6.99	5.31	4.02	3.00	8.19	6.48	4.87	3.57
=0.1	4.24	3.35	2.57	1.89	5.30	4.17	3.17	2.35	6.43	5.03	3.79	2.89	7.54	5.77	4.46	3.31	8.90	7.03	5.35	3.99
=0.2	4.52	3.56	2.78	2.03	5.75	4.52	3.45	2.58	6.94	5.50	4.19	3.16	8.17	6.34	4.92	3.78	9.71	7.81	5.91	4.41
=0.3	4.86	3.87	2.99	2.19	6.24	4.95	3.80	2.85	7.67	6.03	4.62	3.51	8.95	7.02	5.38	4.10	10.79	8.62	6.56	4.99
=0.7	7.07	5.70	4.35	3.06	9.77	7.89	6.12	4.47	12.89	10.31	7.81	5.98	15.66	12.40	9.90	7.35	19.45	16.05	12.47	9.29
LONGEST NEGATIVE RUN LENGTH STANDARD DEVIATIONS																				
Skewness = 0.0																				
=0.0	1.58	1.28	1.02	.80	1.76	1.34	1.08	.82	1.85	1.33	1.11	.82	1.83	1.41	1.07	.81	1.88	1.36	1.18	.88
=0.1	1.72	1.40	1.12	.87	1.94	1.50	1.22	.92	2.03	1.44	1.25	.92	1.97	1.51	1.17	.91	2.05	1.55	1.48	1.02
=0.2	1.89	1.52	1.26	.98	2.10	1.62	1.37	1.02	2.19	1.66	1.32	1.02	2.14	1.65	1.36	.99	2.14	1.58	1.61	1.12
=0.3	2.12	1.72	1.40	1.10	2.35	1.80	1.53	1.15	2.41	1.96	1.48	1.16	2.31	1.88	1.57	1.18	2.43	1.76	1.66	1.23
=0.7	3.99	3.33	2.82	2.35	4.43	3.93	3.18	2.54	4.38	4.08	3.28	2.70	4.58	4.21	3.34	2.89	3.75	3.93	3.42	2.85
Skewness = 0.2																				
=0.0	1.68	1.31	1.04	.80	1.88	1.41	1.14	.83	1.75	1.54	1.16	.80	1.90	1.56	1.08	.83	2.02	1.43	1.12	.74
=0.1	1.83	1.44	1.15	.89	2.02	1.56	1.28	.95	1.95	1.67	1.28	.88	2.08	1.68	1.14	.88	2.21	1.54	1.24	.85
=0.2	1.97	1.60	1.26	.99	2.16	1.81	1.39	1.07	2.24	1.82	1.41	.99	2.43	1.77	1.32	1.05	2.24	1.66	1.52	1.08
=0.3	2.19	1.82	1.43	1.12	2.37	2.07	1.61	1.22	2.41	1.97	1.59	1.20	2.68	2.05	1.52	1.13	2.61	1.77	1.64	1.27
=0.7	4.00	3.49	3.04	2.39	4.51	3.99	3.28	2.57	4.61	4.24	3.52	2.76	4.74	3.99	3.44	2.79	3.98	3.98	3.72	2.85
Skewness = 0.5																				
=0.0	1.58	1.29	1.04	.81	1.65	1.35	1.02	.82	1.77	1.40	1.08	.84	1.76	1.32	1.12	.83	1.60	1.50	1.19	.95
=0.1	1.70	1.43	1.13	.89	1.82	1.49	1.11	.93	1.92	1.51	1.20	.97	1.96	1.39	1.27	.93	1.80	1.60	1.26	1.04
=0.2	1.86	1.54	1.31	1.00	2.10	1.66	1.25	1.05	2.11	1.69	1.41	1.10	2.16	1.76	1.45	1.05	2.09	1.83	1.37	1.23
=0.3	2.07	1.76	1.47	1.15	2.46	1.93	1.47	1.24	2.39	1.90	1.58	1.25	2.59	1.92	1.48	1.25	2.37	2.02	1.64	1.42
=0.7	3.92	3.48	2.96	2.40	4.47	4.06	3.30	2.62	4.81	4.00	3.42	2.88	4.66	4.15	3.43	2.88	3.79	4.27	3.12	2.77

TABLE III - 2

	N=25				N=50				N=100				N=200				N=500			
q=0.5	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2
LARGEST NEGATIVE RUN-SUM MEANS																				
Skewness = 0.0																				
ρ=0.0	3.60	2.66	1.92	1.33	4.55	3.32	2.40	1.73	5.55	3.95	2.94	2.80	6.41	4.64	3.42	2.44	7.56	5.54	4.11	2.95
=0.1	3.90	2.87	2.05	1.40	5.01	3.64	2.62	1.84	6.15	4.38	3.26	2.26	7.03	5.16	3.80	2.66	8.64	6.32	4.62	3.29
=0.2	4.22	3.11	2.19	1.47	5.50	4.00	2.86	1.98	6.85	4.90	3.59	2.48	7.95	5.82	4.28	2.95	9.73	7.12	5.22	3.73
=0.3	4.60	3.38	2.35	1.56	6.08	4.43	3.14	2.15	7.70	5.55	3.99	2.74	8.99	6.63	4.80	3.31	11.00	8.12	5.99	4.26
=0.7	6.74	4.90	3.27	2.07	9.99	7.30	5.00	3.22	13.57	10.01	6.94	4.67	17.38	12.84	9.21	6.41	22.34	17.53	12.53	8.81
Skewness = 0.2																				
=0.0	3.61	2.57	1.84	1.21	4.50	3.26	2.30	1.53	5.36	3.87	2.76	1.85	6.38	4.61	3.28	2.21	7.51	5.44	3.84	2.70
=0.1	3.91	2.78	1.97	1.27	5.00	3.59	2.50	1.65	5.95	4.32	3.06	2.02	7.20	5.21	3.69	2.44	8.51	6.08	4.36	3.02
=0.2	4.24	3.02	2.12	1.35	5.50	3.98	2.76	1.78	6.70	4.82	3.42	2.21	8.14	5.83	4.15	2.74	9.50	6.97	5.00	3.41
=0.3	4.62	3.30	2.29	1.43	6.06	4.44	3.05	1.94	7.47	5.37	3.84	2.45	9.18	6.65	4.64	3.06	10.98	7.99	5.73	3.87
=0.7	6.73	3.83	3.29	1.91	9.92	7.39	4.89	2.95	13.42	9.96	6.89	4.14	17.84	12.78	8.81	5.57	23.50	16.60	12.15	7.60
Skewness = 0.5																				
=0.0	3.15	2.22	1.53	1.00	3.95	2.78	1.92	1.30	4.81	3.31	2.25	1.57	5.64	3.88	2.69	1.78	6.52	4.71	3.23	2.11
=0.1	3.44	2.42	1.65	1.06	4.34	2.09	2.10	1.40	5.37	3.71	2.51	1.74	6.29	4.38	3.05	2.00	7.37	5.37	3.70	2.44
=0.2	3.72	2.62	1.78	1.13	4.83	3.41	2.31	1.53	5.99	4.19	2.83	1.93	7.59	4.98	3.46	2.26	8.35	6.15	4.26	2.80
=0.3	4.05	2.86	1.92	1.21	5.39	3.80	2.55	1.68	6.77	4.73	3.18	2.14	8.09	5.68	3.92	2.57	9.58	7.02	4.91	3.24
=0.7	5.94	4.19	2.70	1.61	8.79	6.28	4.12	2.59	12.50	8.73	5.70	3.74	15.69	10.87	7.62	4.88	19.05	15.34	10.31	6.47
LARGEST NEGATIVE RUN-SUM STANDARD DEVIATIONS																				
Skewness = 0.0																				
=0.0	1.52	1.13	.88	.68	1.66	1.20	.93	.69	1.75	1.21	.92	.70	1.57	1.15	.97	.67	1.75	1.30	1.06	.72
=0.1	1.76	1.31	1.02	.77	1.97	1.42	1.10	.80	2.04	1.39	1.12	.83	1.88	1.33	1.11	.77	2.13	1.54	1.23	.85
=0.2	2.05	1.53	1.20	.88	2.30	1.65	1.30	.94	2.34	1.63	1.32	.98	2.18	1.58	1.35	.91	2.33	1.77	1.40	1.01
=0.3	2.40	1.81	1.41	1.03	2.70	1.96	1.54	1.11	2.79	2.04	1.56	1.17	2.56	1.95	1.63	1.13	2.75	1.98	1.70	1.23
=0.7	5.21	4.03	3.13	2.28	6.18	5.05	3.75	2.70	6.49	5.17	4.01	3.15	7.10	5.64	4.23	3.69	6.90	6.27	4.76	3.68
Skewness = 0.2																				
=0.0	1.51	1.09	.82	.60	1.62	1.16	.87	.60	1.55	1.19	.87	.61	1.83	1.21	.87	.62	1.82	1.41	.81	.66
=0.1	1.75	1.29	.97	.56	1.89	1.35	1.06	.73	1.79	1.42	1.05	.73	2.16	1.46	1.01	.74	2.14	1.59	.99	.77
=0.2	2.03	1.52	1.14	.81	2.20	1.64	1.26	.88	2.14	1.66	1.24	.87	2.60	1.68	1.20	.91	2.34	1.78	1.27	.92
=0.3	2.37	1.81	1.34	.95	2.55	1.99	1.53	1.07	2.49	1.97	1.49	1.06	3.05	2.01	1.39	1.14	2.96	2.07	1.57	1.07
=0.7	5.02	4.03	3.13	2.09	5.91	5.10	3.68	2.55	6.52	5.31	3.96	2.94	7.16	5.16	3.91	2.83	8.92	5.52	4.50	2.89
Skewness = 0.5																				
=0.0	1.30	.96	.68	.49	1.37	1.00	.67	.49	1.47	1.03	.67	.49	1.42	1.04	.77	.49	1.32	1.07	.76	.47
=0.1	1.50	1.16	.81	.58	1.59	1.19	.81	.60	1.69	1.20	.82	.60	1.62	1.18	.93	.60	1.69	1.34	.92	.62
=0.2	1.74	1.31	.97	.68	1.93	1.40	.96	.73	1.99	1.43	1.00	.73	2.12	1.47	1.13	.73	1.95	1.53	1.10	.76
=0.3	2.04	1.57	1.15	.81	2.32	1.70	1.17	.89	2.36	1.74	1.22	.89	2.41	1.74	1.34	.90	2.24	1.77	1.30	.94
=0.7	4.40	3.54	2.55	1.75	5.27	4.30	3.01	2.20	6.49	4.40	3.36	2.53	6.45	4.56	3.51	2.63	6.22	5.63	3.29	3.55

TABLE III - 3

	N=25				N=50				N=100				N=200				N=500			
LARGEST NEGATIVE RUN-SUM - μ_n																				
Skewness = 0.0																				
ρ=0.0	1.20	.90	.56	.17	1.45	1.14	.81	.47	1.67	1.33	1.03	.68	1.83	1.50	1.19	.86	2.00	1.69	1.38	1.05
=0.1	1.27	.96	.61	.20	1.54	1.22	.88	.52	1.76	1.43	1.13	.75	1.92	1.61	1.29	.94	2.13	1.82	1.50	1.16
=0.2	1.33	1.03	.65	.23	1.62	1.31	.96	.58	1.87	1.54	1.21	.84	2.04	1.73	1.41	1.04	2.25	1.93	1.62	1.28
=0.3	1.41	1.09	.70	.26	1.72	1.40	1.04	.65	1.98	1.65	1.31	.92	2.16	1.85	1.51	1.14	2.37	2.07	1.75	1.41
Skewness = 0.2																				
=0.0	1.20	.86	.52	.08	1.45	1.12	.77	.35	1.64	1.31	.97	.56	1.81	1.50	1.15	.75	1.99	1.66	1.32	.96
=0.1	1.27	.93	.57	.15	1.54	1.21	.84	.41	1.74	1.41	1.06	.64	1.93	1.61	1.27	.85	2.11	1.77	1.45	1.07
=0.2	1.34	.99	.63	.15	1.63	1.30	.92	.47	1.85	1.52	1.17	.72	2.05	1.72	1.38	.96	2.22	1.91	1.58	1.19
=0.3	1.41	1.06	.68	.18	1.72	1.40	1.00	.53	1.96	1.62	1.28	.81	2.16	1.85	1.49	1.05	2.36	2.05	1.71	1.32
Skewness = 0.5																				
=0.0	1.07	.71	.33	-.11	1.32	.96	.59	.20	1.53	1.15	.77	.40	1.70	1.32	.95	.54	1.85	1.52	1.15	.72
=0.1	1.15	.78	.39	-.07	1.41	1.06	.67	.25	1.63	1.26	.87	.50	1.81	1.44	1.07	.65	1.97	1.65	1.28	.86
=0.2	1.22	.85	.45	-.03	1.50	1.15	.76	.32	1.74	1.38	.98	.59	1.99	1.56	1.19	.77	2.10	1.79	1.42	.99
=0.3	1.29	.92	.50	0.00	1.60	1.24	.84	.40	1.86	1.49	1.09	.68	2.05	1.69	1.31	.89	2.23	1.92	1.56	1.14
LARGEST NEGATIVE RUN-SUM - σ_n																				
Skewness = 0.0																				
=0.0	.41	.41	.44	.48	.35	.35	.37	.39	.31	.30	.31	.33	.24	.24	.28	.27	.23	.23	.25	.24
=0.1	.43	.44	.47	.52	.38	.38	.40	.42	.32	.31	.33	.36	.26	.25	.29	.29	.24	.24	.26	.25
=0.2	.46	.47	.51	,55	.40	.40	.43	.45	.33	.32	.36	.38	.27	.27	.31	.30	.24	.24	.26	.27
=0.3	.49	.50	.55	.60	.42	.42	.46	.49	.35	.36	.38	.41	.28	.29	.33	.33	.25	.24	.28	.28
Skewness = 0.2																				
=0.0	.40	.41	.42	.47	.35	.35	.37	.38	.28	.30	.31	.32	.28	.26	.26	.28	.24	.26	.21	.24
=0.1	.43	.44	.46	.42	.37	.37	.40	.42	.29	.31	.33	.35	.29	.28	.27	.30	.25	.26	.22	.25
=0.2	.45	.47	.50	.55	.39	.40	.44	.47	.31	.34	.35	.38	.31	.28	.28	.32	.24	.25	.25	.27
=0.3	.48	.51	.54	.60	.40	.43	.47	.51	.32	.36	.37	.41	.32	.29	.29	.36	.27	.26	.27	.27
Skewness = 0.5																				
=0.0	.40	.41	.43	.46	.34	.35	.34	.36	.30	.30	.29	.31	.25	.26	.28	.27	.20	.22	.13	.22
=0.1	.42	.45	.47	.51	.36	.37	.37	.41	.31	.31	.31	.33	.25	.27	.30	.29	.23	.25	.25	.25
=0.2	.44	.47	.51	.56	.38	.39	.40	.45	.31	.33	.35	.37	.27	.29	.31	.31	.23	.25	.26	.27
=0.3	.48	.51	.55	.61	.41	.43	.44	.50	.34	.36	.37	.40	.29	.30	.33	.34	.23	.25	.26	.28

TABLE III - 4

	N=25				N=50				N=100				N=200				N=500			
q=0.5	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2
LONGEST NEGATIVE RUN-LENGTH - μ_n																				
Skewness = 0.0																				
ρ=0.0	1.30	1.08	.78	.47	1.55	1.29	1.01	.71	1.76	1.48	1.20	.90	1.92	1.65	1.35	1.07	2.09	1.82	1.55	1.25
=0.1	1.37	1.14	.85	.53	1.62	1.37	1.10	.79	1.82	1.56	1.28	.99	1.96	1.73	1.46	1.15	2.18	1.91	1.65	1.34
=0.2	1.43	1.20	.91	.61	1.69	1.44	1.17	.87	1.90	1.65	1.36	1.07	2.06	1.81	1.55	1.27	2.26	2.00	1.75	1.47
=0.3	1.50	1.27	.98	.67	1.76	1.51	1.26	.94	1.98	1.74	1.46	1.16	2.16	1.91	1.64	1.37	2.36	2.09	1.85	1.57
Skewness = 0.2																				
=0.0	1.35	1.10	.81	.48	1.59	1.34	1.05	.71	1.77	1.51	1.24	.90	1.95	1.68	1.41	1.07	2.12	1.88	1.57	1.25
=0.1	1.41	1.15	.88	.54	1.65	1.42	1.13	.79	1.84	1.60	1.34	.98	2.02	1.78	1.50	1.15	2.20	1.95	1.69	1.34
=0.2	1.46	1.22	.95	.61	1.72	1.49	1.20	.88	1.92	1.67	1.42	1.07	2.10	1.85	1.58	1.25	2.28	2.04	1.78	1.45
=0.3	1.53	1.29	1.01	.68	1.80	1.57	1.29	.95	2.00	1.75	1.51	1.17	2.19	1.95	1.67	1.35	2.38	2.15	1.88	1.56
Skewness = 0.5																				
=0.0	1.31	1.06	.78	.47	1.56	1.29	1.03	.70	1.75	1.49	1.19	.91	1.91	1.64	1.35	1.06	2.08	1.84	1.55	1.24
=0.1	1.37	1.13	.86	.54	1.61	1.37	1.10	.78	1.82	1.57	1.28	1.01	1.99	1.72	1.46	1.16	2.17	1.92	1.65	1.35
=0.2	1.43	1.18	.92	.60	1.69	1.45	1.18	.87	1.89	1.66	1.38	1.09	2.07	1.81	1.55	1.29	2.25	2.03	1.75	1.45
=0.3	1.50	1.26	.99	.66	1.76	1.53	1.27	.96	1.99	1.75	1.48	1.20	2.15	1.91	1.65	1.37	2.35	2.13	1.85	1.57
LONGEST NEGATIVE RUN-LENGTH - σ_n																				
Skewness = 0.0																				
=0.0	.38	.39	.41	.43	.34	.34	.36	.37	.30	.29	.31	.31	.26	.26	.26	.26	.22	.21	.24	.24
=0.1	.39	.40	.42	.44	.35	.35	.37	.37	.31	.29	.32	.32	.26	.26	.26	.27	.22	.22	.27	.25
=0.2	.40	.40	.44	.45	.36	.35	.38	.38	.31	.30	.31	.32	.26	.26	.27	.27	.21	.21	.26	.25
=0.3	.42	.42	.45	.47	.37	.36	.39	.40	.31	.32	.32	.33	.26	.26	.29	.28	.22	.21	.25	.24
Skewness = 0.2																				
=0.0	.39	.39	.41	.43	.35	.34	.36	.37	.28	.31	.31	.30	.26	.28	.25	.27	.23	.21	.22	.21
=0.1	.40	.40	.42	.45	.35	.35	.37	.38	.29	.31	.31	.31	.26	.27	.24	.27	.24	.21	.22	.21
=0.2	.40	.41	.43	.46	.35	.37	.38	.40	.31	.32	.32	.31	.28	.27	.26	.28	.22	.21	.24	.24
=0.3	.41	.44	.45	.48	.36	.38	.39	.41	.30	.31	.32	.34	.28	.28	.27	.28	.23	.20	.24	.25
Skewness = 0.5																				
=0.0	.38	.39	.42	.44	.32	.34	.34	.37	.29	.30	.31	.31	.25	.25	.27	.27	.19	.23	.24	.26
=0.1	.38	.41	.42	.45	.33	.35	.34	.38	.29	.29	.31	.33	.26	.24	.28	.28	.20	.22	.23	.26
=0.2	.40	.41	.45	.47	.36	.36	.35	.39	.30	.30	.33	.34	.26	.27	.29	.27	.21	.23	.23	.27
=0.3	.41	.43	.46	.49	.38	.38	.37	.42	.30	.31	.33	.35	.28	.27	.27	.30	.22	.23	.25	.28

Key Words: Droughts, Drought Probability, Return Period, Critical Period, Runs, Longest Negative Run-Length, Largest Negative Run-Sum, Stochastic Processes in Hydrology, Precipitation, Run-Off.

Abstract: A method is presented for computing the probability or recurrence period of historical droughts by using the longest negative run-length and the largest negative run-sum as basic parameters of samples of a given size, and by using a given probability of the truncation level, a given auto-correlation coefficient, and a given skewness coefficient. The application of this method to selected annual runoff and precipitation series demonstrate its feasibility. The statistical experimental method in generating large numbers of samples is used to compute frequency distributions as the estimates of probability distributions of the longest negative run-length and of the largest negative run-sum in a sample of size N, as descriptors of the largest historical droughts, for normal and nonnormal independent and dependent stationary stochastic processes which follows the first-order linear autoregressive model. Experimentally obtained values are checked with theoretical results for the distribution of the longest negative run-length when the observations are independent. A set of graphs and a set of tables are presented to make the numerical values readily usable. Good approximations for practical computations are demonstrated by fitting lognormal probability distributions to the experimentally obtained frequencies.

Reference: Millan, Jaime and Vujica Yevjevich, Colorado State University, Hydrology Paper No. 50 (June 1971) "Probabilities of Observed Droughts"

Key Words: Droughts, Drought Probability, Return Period, Critical Period, Runs, Longest Negative Run-Length, Largest Negative Run-Sum, Stochastic Processes in Hydrology, Precipitation, Run-Off.

Abstract: A method is presented for computing the probability or recurrence period of historical droughts by using the longest negative run-length and the largest negative run-sum as basic parameters of samples of a given size, and by using a given probability of the truncation level, a given auto-correlation coefficient, and a given skewness coefficient. The application of this method to selected annual runoff and precipitation series demonstrate its feasibility. The statistical experimental method in generating large numbers of samples is used to compute frequency distributions as the estimates of probability distributions of the longest negative run-length and of the largest negative run-sum in a sample of size N, as descriptors of the largest historical droughts, for normal and nonnormal independent and dependent stationary stochastic processes which follows the first-order linear autoregressive model. Experimentally obtained values are checked with theoretical results for the distribution of the longest negative run-length when the observations are independent. A set of graphs and a set of tables are presented to make the numerical values readily usable. Good approximations for practical computations are demonstrated by fitting lognormal probability distributions to the experimentally obtained frequencies.

Reference: Millan, Jaime and Vujica Yevjevich, Colorado State University, Hydrology Paper No. 50 (June 1971) "Probabilities of Observed Droughts"

Key Words: Droughts, Drought Probability, Return Period, Critical Period, Runs, Longest Negative Run-Length, Largest Negative Run-Sum, Stochastic Processes in Hydrology, Precipitation, Run-Off.

Abstract: A method is presented for computing the probability or recurrence period of historical droughts by using the longest negative run-length and the largest negative run-sum as basic parameters of samples of a given size, and by using a given probability of the truncation level, a given auto-correlation coefficient, and a given skewness coefficient. The application of this method to selected annual runoff and precipitation series demonstrate its feasibility. The statistical experimental method in generating large numbers of samples is used to compute frequency distributions as the estimates of probability distributions of the longest negative run-length and of the largest negative run-sum in a sample of size N, as descriptors of the largest historical droughts, for normal and nonnormal independent and dependent stationary stochastic processes which follows the first-order linear autoregressive model. Experimentally obtained values are checked with theoretical results for the distribution of the longest negative run-length when the observations are independent. A set of graphs and a set of tables are presented to make the numerical values readily usable. Good approximations for practical computations are demonstrated by fitting lognormal probability distributions to the experimentally obtained frequencies.

Reference: Millan, Jaime and Vujica Yevjevich, Colorado State University, Hydrology Paper No. 50 (June 1971) "Probabilities of Observed Droughts"

Key Words: Droughts, Drought Probability, Return Period, Critical Period, Runs, Longest Negative Run-Length, Largest Negative Run-Sum, Stochastic Processes in Hydrology, Precipitation, Run-Off.

Abstract: A method is presented for computing the probability or recurrence period of historical droughts by using the longest negative run-length and the largest negative run-sum as basic parameters of samples of a given size, and by using a given probability of the truncation level, a given auto-correlation coefficient, and a given skewness coefficient. The application of this method to selected annual runoff and precipitation series demonstrate its feasibility. The statistical experimental method in generating large numbers of samples is used to compute frequency distributions as the estimates of probability distributions of the longest negative run-length and of the largest negative run-sum in a sample of size N, as descriptors of the largest historical droughts, for normal and nonnormal independent and dependent stationary stochastic processes which follows the first-order linear autoregressive model. Experimentally obtained values are checked with theoretical results for the distribution of the longest negative run-length when the observations are independent. A set of graphs and a set of tables are presented to make the numerical values readily usable. Good approximations for practical computations are demonstrated by fitting lognormal probability distributions to the experimentally obtained frequencies.

Reference: Millan, Jaime and Vujica Yevjevich, Colorado State University, Hydrology Paper No. 50 (June 1971) "Probabilities of Observed Droughts"

PREVIOUSLY PUBLISHED PAPERS

Colorado State University Hydrology Papers

No. 35 "Stochastic Process of Precipitation," by P. Todorovic and V. Yevjevich, September 1969.

No. 36 "Suitability of the Upper Colorado River Basin for Precipitation Management," by Hiroshi Nakamichi and Hubert J. Morel-Seytoux, October 1969.

No. 37 "Regional Discrimination of Change in Runoff," by Viboon Nimmannit and Hubert J. Morel-Seytoux, November 1969.

No. 38 "Evaluation of the Effect of Impoundment on Water Quality in Cheney Reservoir," by J.C. Ward and S. Karaki, March 1970.

No. 39 "The Kinematic Cascade as a Hydrologic Model," by David F. Kibler and David A. Woolhiser, February 1970.

No. 40 "Application of Run-Lengths to Hydrologic Series," by Jaime Saldarriaga and Vujica Yevjevich, April 1970.

No. 41 "Numerical Simulation of Dispersion in Groundwater Aquifers," by Donald Lee Reddell and Daniel K. Sunada, June 1970.

No. 42 "Theoretical Probability Distributions for Flood Peaks," by Emir Zelenhasic, November 1970.

No. 43 "Flood Routing Through Storm Drains, Part I, Solution of Problems of Unsteady Free Surface Flow in a Storm Drain," by V. Yevjevich and A.H. Barnes, November 1970.

No. 44 "Flood Routing Through Storm Drains, Part II, Physical Facilities and Experiments," by V. Yevjevich and A.H. Barnes, November 1970.

No. 45 "Flood Routing Through Storm Drains, Part III, Evaluation of Geometric and Hydraulic Parameters," by V. Yevjevich and A.H. Barnes, November 1970.

No. 46 "Flood Routing Through Storm Drains, Part IV, Numerical Computer Methods of Solution," by V. Yevjevich and A.H. Barnes, November 1970.

No. 47 "Mathematical Simulation of Infiltrating Watersheds," by Roger E. Smith and David A. Woolhiser, January 1971.

NO. 48 "Models for Subsurface Drainage," W.E. Hedstrom, A.T. Corey and H.R. Duke, February 1971.

No. 49 "Infiltration Affected by Flow of Air," David B. McWhorter, May 1971.