OAK RIDGE NATIONAL LABORATORY

operated by

UNION CARBIDE CORPORATION

for the

U.S. ATOMIC ENERGY COMMISSION

ORNL-RSIC-38
(Vol. I)

TECHNIQUES for EFFICIENT MONTE CARLO SIMULATION

Volume I

SELECTING PROBABILITY DISTRIBUTIONS

E. J. McGrath
S. L. Basin
R. W. Burton
D. C. Irving
S. C. Jaquette
W. R. Ketler
C. A. Smith

RADIATION SHIELDING INFORMATION CENTER

ORNL-RSIC-38 (Vol. I)

Contract No. W-7405-eng-26

(SAI-72-590-LJ)

TECHNIQUES FOR EFFICIENT MONTE CARLO SIMULATION

Volume 1

SELECTING PROBABILITY DISTRIBUTIONS

E. J. McGrath
S. L. Basin
R. W. Burton
D. C. Irving
S. C. Jaquette
W. R. Ketler
C. A. Smith

APRIL 1975

Reprinted December 1974

Prepared for the

Office of Naval Research (Code 462)
Department of the Navy
Arlington, Virginia 22217

by Science Applications, Incorporated

NOTE:

This work partially supported by
DEFENSE NUCLEAR AGENCY

Reprinted by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37830
operated by
UNION CARBIDE CORPORATION
for the
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

SAI-72-590-LJ

TECHNIQUES FOR EFFICIENT MONTE CARLO SIMULATION

VOLUME I

SELECTING PROBABILITY DISTRIBUTIONS

Scientific Officer, Office of Naval Research (Code 462)

J. R. Simpson

Project Principal Investigator

E. J. McGrath

Co-Authors

S. L. Basin
R. W. Burton
D. C. Irving
S. C. Jaquette
W. R. Ketler
C. A. Smith

FOREWORD

One of the objectives of an information analysis center is to make available or call attention to particularly valuable information which otherwise might be overlooked in the great mass of published literature. To further that objective, we are reprinting this series of reports on the Monte Carlo method. The series was originally written to bring together for one technical community Monte Carlo information that was widely scattered. It was pointed out that techniques developed for neutron transport calculations could be used for quite different application. Here, we point out the reverse: a comprehensive review of Monte Carlo for defense applications is useful to the radiation transport community. In thinking of Monte Carlo as a simulation of the transport process, we sometimes forget that it is a powerful mathematical tool for solving multidimensional integral equations arising in many other situations.

We are grateful to the Office of Naval Research for granting us permission to reprint these reports. We feel this work will be of much greater usefulness as a result.

D. K. Trubey

Radiation Shielding Information Center
November 1974

ABSTRACT

This document is the first of three volumes which present techniques and methods for developing efficient Monte Carlo simulation: Each volume presents techniques for reducing computational effort in one of the following areas: Vol. I - Selecting Probability Distributions, Vol. II - Random Number Generation For Selected Probability Distributions, and Vol. III - Variance Reduction.

This volume provides a straightforward approach and associated techniques for selecting the most appropriate probability distributions for use in Monte Carlo simulations. Part I, BASIC CONSIDERATIONS, presents the underlying concepts and principles for selecting probability distributions. Part II, SELECTION OF DISTRIBUTIONS, gives the mathematical models representing stochastic processes and presents step-by-step procedures for identification and selection of the appropriate probability distributions based upon the degree of knowledge and available data for the random variable under study.

CONTENTS

EXECUTIVE SUMMARY

Monte Carlo simulation is one of the most powerful and commonly used techniques for analyzing complex physical problems. Applications can be found in many diverse areas from radiation transport to river basin modeling. Important Navy applications include: analysis of antisubmarine warfare exercises and operations, prediction of aircraft or sensor performance, tactical analysis, and matrix game solutions where random processes are considered to be of particular importance. The range of applications has been broadening and the size, complexity, and computational effort required have been increasing. However, such developments are expected and desirable since increased realism is concomitant with more complex and extensive problem descriptions.

In recognition of such trends, the requirements for improved simulation techniques are becoming more pressing. Unfortunately, methods for achieving greater efficiency are frequently overlooked in developing simulations. This can generally be attributed to one or more of the following reasons:

- Analysts usually seek advanced computer systems to perform more complex simulation studies by exploiting increased speed and/or storage capabilities. This is often achieved at a considerably increased expense.

- Many efficient simulation methods have evolved for specialized applications. For example, some of the most impressive Monte Carlo techniques have been developed in radiation transport, a discipline that does not overlap into areas where even a small number of simulation analysts are working.

- Known techniques are not developed to the point where they can be easily understood or applied by even a small fraction of the analysts who are performing simulation studies or developing simulation models.

In addition to the above reasons, comprehensive references describing efficient methodologies to improve Monte Carlo simulation are not available. It is the intent of these volumes to help alleviate the above shortcomings in Monte Carlo simulation.

This document is the first of three volumes which present techniques and methods for developing efficient Monte Carlo simulations. Each volume is essentially a self-contained discussion of useful techniques which can be applied in reducing computational effort in one of the following three major aspects of Monte Carlo simulation:

- Selecting Probability Distributions - Volume I
- Random Number Generation for Selected Probability Distributions - Volume II
- Variance Reduction - Volume III

The purpose of these volumes is to provide guidance in developing Monte Carlo simulations that accurately reflect the behavior of various characteristics of the system being simulated and are most efficient in terms of computational effort. The basic intent is to provide understanding of the concepts and methods for reducing analysis and computational effort as well as to serve as a practical guide for their application. They have been prepared primarily for the systems analyst and computer programmer who have a basic background and experinece in simulation and elementary statistics. Thus, the material is presented so as to preclude extensive knowledge of statistical techniques or of extensive literature search. However, it is assumed the reader has a grasp of the fundamentals of Monte Carlo methods, simulation modeling, and elementary statistics.

1. INTRODUCTION

The starting point in developing any Monte Carlo simulation is the construction of mathematical models which describe the stochastic behavior of the variables in the process under study. When the underlying processes are well understood and the functional forms of the variables are known, development of a model is straightforward. However, in many applications the exact functional form of the variable is not known, thus requiring selection from among a myriad of possible distributions to find the one that will best represent the process. This volume provides a straightforward approach and associated techniques for selecting the most appropriate probability distributions for use in Monte Carlo simulations.

Part I of this volume, BASIC CONSIDERATIONS, presents the underlying concepts and principles to be used in the selection of probability distributions. This background information provides the reader with an understanding of the important considerations, tasks, and methods and procedures involved in dealing with simulation events characterized by random variables.

Following Part I, the reader will find in Part II, SELECTION OF DISTRIBUTIONS, the mathematical models which will represent the stochastic behavior of the process as accurately as the data and understanding of the processes will allow. Part II presents step-by-step procedures for the identification and selection of appropriate probability distributions. Part II applies the rationale developed in Part I to the problems of developing distributions based on varying amounts of data and depth of understanding of the processes being simulated.

This volume also includes additional information useful in the selection of probability distributions. Appendix A contains background information of the complex parametric families of distributions which will be useful for the reader who has not encountered these distributions before. Appendix B contains tables which are needed in making computations involving distribution fitting and testing. Appendix C is an abstracted bibliography of publications relating to the subjects of probability distribution identification and selection.

PART I

BASIC CONSIDERATIONS

2. FUNDAMENTALS OF DISTRIBUTION SELECTION

Selection of an appropriate probability distribution for a given random variable in a simulation requires gathering and evaluating all the available facts, data, and knowledge concerning each variable. It is also important to know how the particular process which any given variable represents relates to the entire simulation model. For Monte Carlo applications this includes careful investigation of:

- Each individual process or event
- Underlying theory of the process
- Data representing the variability of the process
- Sensitivity of the process being simulated to probable values of the variable
- Simulation programming considerations

When the variable under consideration is just one among many variables which affect the overall problem or system, the simulation is often not very sensitive to the choice of the distribution. This can be likened to the phenomenon of summing a series of random variables, none of which dominates the sum. In this case the total tends to have a normal distribution irrespective of the individual distributions (see Refs. 7, 27). In other cases, the selection of a distribution is more critical to effective simulation. For example, when only a few variables dominate the process or the process is greatly influenced by rare occurrences (e.g., failure of a critical high reliability component) the selection of probability distributions becomes of paramount importance.[7, 27]

Choosing the form of probability distributions is often a trade-off between theoretical justification and empirical evidence. Typically, some form of parametric distribution can be justified, such as the normal, uniform, binomial, or Bernoulli distribution. Available data can then be used to estimate its parameters. In the absence of empirical data, one is forced to choose distributions on either theoretical or intuitive grounds, or often to use several distributions and conduct sensitivity or worst-case analyses. At the other extreme, where empirical data is abundant, either the histogram can be used or more elaborate parametric models can be employed.

The final choice of a particular distribution type is, of course, also dependent on ease of implementation. Computer storage space, computation time, and ease of programming are key considerations in most simulations. Generating random variables from a parametric distribution requires taking an inverse of the cumulative distribution function or using other random number generation techniques (see Volume II). For some distributions, such as the exponential or uniform, the inverse operation is a simple computation. For others, such as the normal, relatively simple techniques are available. Histograms are also fairly easy to use in computer simulations. Here, only a list of numbers must be stored (the more variable and detailed the histogram, of course, the longer the list). For many distributions, however, inverse algorithms for generation of random numbers do not exist, and other methods require lengthy computation. In this case, a compromise must be made between ease of computation and simulation accuracy. Making an estimate of how sensitive the total simulation will be to individual probability distribution assumptions is important in determining this compromise.

2.1 BASIS FOR MAKING SELECTIONS

Before proceeding to the techniques of distribution selection and their application in simulation development, it is necessary to understand the underlying concepts for making selections. Basically, the selection process described in Part II depends on two factors: the extent of knowledge of the process under study (qualitative) and the amount of data available (quantitative). Knowledge of the process refers to the level of understanding of its behavior and characteristics. For example, it is possible in some cases to be quite certain that the frequency distribution of a random variable is normal based on familiarity with the process. At the other extreme, little or nothing may be known. Similarly, the amount of data describing a particular variable may range from extensive to none. Each combination of the state of knowledge and amount of data poses particular problems in selecting the most appropriate distribution.

2.2 QUALITATIVE BASIS FOR SELECTION

Developing an understanding of some random process involves analysis to characterize the process. In general, such efforts attempt to identify the process on the basis of:

- Similarity to some other process whose behavior is known
- Underlying theory
- Certain qualitative aspects.

Often a process can be likened to some other, the behavior of which is known. In such circumstances, it can be reasonably justified that this known distribution might apply to the one under study. For example, consider the simulation of a process involving the human performance of some manual task. Even though the task may bear no particular resemblance to one in which the distribution is known, an assumption of similarity is reasonable. The frequency distribution of time of performance is likely to be from the same family of distributions even though the actual process might be quite different.

Many activities for which stochastic models must be developed can, at least generally, be identified by some applicable theory. Consider the case in which some repetitive human activity is involved such as in maintenance. Maintainability theory would indicate a strong likelihood that the frequency distribution of time to perform would have a log normal or a gamma distribution. Similarly, if the failure of electronic parts were to be modeled, it could be assumed that an exponential or possibly a Weibull might be applicable (53). Such reasoning is a fundamental part of the task of distribution selection.

There are, of course, many situations in which a theoretical basis for a particular distribution can be established. Consider the shots fired at a target or the velocity of a molecule in a stable solution. Under fairly weak conditions the velocity of the molecule or the deviation of shots (in three-dimensional space) from the bull's eye can be shown to have a Maxwell distribution (27). The component of velocity in any direction or the projection of shots onto any axis through the bull's eye follows the normal distribution. In two dimensions the resulting distribution is the Rayleigh. If the process being modeled involves reliability, the exponential distribution reflects the behavior of an item with a constant failure rate. If the process involves waiting or queueing phenomena, the exponential can be used to depict random arrival and service times. The gamma distribution also has wide application since it is related to the exponential distribution. The number of occurences up to a given point in time has a gamma distribution if the time between occurrences follows an exponential distribution.

In some cases, it will not be possible to relate the process being examined to anything which is known. This may be either because little understanding of the process exists or it simply bears no relation to any process whose behavior can be described on a theoretical basis.

However, there still may be some clues which are useful in identifying an applicable distribution, particularly where some data exist. A number of qualitative aspects of the process can be helpful. These include, for example, consideration of whether the variable is discrete or continuous, bounded, symmetric, or can be described in some other similar ways. Such clues, although probably not sufficient for positive identification above, are useful in making a rational selection of a distribution.

2.3 QUANTITATIVE BASIS FOR SELECTION

One of the most common problems in simulation is not having, or not being able to obtain, the data necessary to describe a particular variable. Collecting it may be too time consuming or expensive. In some cases it is simply not possible. Consequently, the amount of data available is one of the major considerations in the selection of probability distributions.

Where sufficient data are available, an empirical approach can be used. This means essentially using the data to derive a model. Combined with the state of knowledge of the process being modeled, graphical and analytical techniques can be employed to select the distribution most representative of the data.

In those cases where acquisition of the data is difficult, the application of the methodology of Part II can be useful in determining whether such effort is warranted. If a distribution can, in fact, be selected with little data, there may be no justification for collecting more. If, on the other hand, a distribution cannot be identified and the simulation results are sensitive to that particular variable, additional data may be essential for developing a valid model.

3. TECHNIQUES USED IN DISTRIBUTION SELECTION

Specific techniques for selecting a particular stochastic model depend on the information and the amount of data available. The situation can range from having practically nothing to work with to almost certain specification of the model based on sound theoretical and empirical evidence. The development of the theoretical evidence is entirely qualitative. Development of the empirical evidence, though, requires the use of a number of quantitative methods. These include:

- Sensitivity analysis
- Graphical analysis
- Parameter estimation
- Goodness-of-fit-testing.

Each of these is introduced briefly in the following sections.

3.1 SENSITIVITY ANALYSIS

The purpose of sensitivity analysis is to determine the extent to which the outcome of an analysis is dependent upon a particular variable or assumption. It is particularly applicable in simulation where little or no data is available to characterize some random variables. In such a situation, sensitivity analysis can indicate whether or not the behavior of the variable must be more accurately known. If, for instance, the outcome of the simulation is not sensitive to the variable, no further effort to characterize it is necessary. However, if it does prove sensitive, an attempt to develop an accurate distribution model is warranted.

The only practical way to perform the sensitivity analysis is to perform a simulation varying the values or assumptions concerning the variable in question. Comparison of the results using standard statistical tests can reveal whether significant differences are produced (see Sections 3.4 and 9.). This is not so formidable a task as it might at first appear. If the simulation is to have any real validity in the first place, the behavior of most of the variables must be known. If only a few variables can be accurately described, a simulation merely produces a precise but inaccurate result.

3.2 GRAPHICAL ANALYSIS

One of the topics in elementary applied statistics is the construction of frequency histograms and cumulative frequency polygons. These procedures provide one means for identifying appropriate distribution models under the proper circumstances. Where such techniques are applicable they do offer the advantage of relative simplicity. They are most useful when there is some knowledge of the process and at least minimal data available.

The histogram is constructed from data concerning the variable. It carries with it all the present empirical information available on the variable, nothing more. It does not try to estimate probable behavior. If rare events have not been observed, for instance, it will assign zero probability to their occurrence. Since it uses all data, it also perpetuates the mistakes of erroneous observations and may describe a model that is not valid.

The most common graphical procedure is the construction of the frequency histogram. This is simply a plot of the frequency with which each of various values occurs in the sample data. The histogram is useful in two ways. It provides visual evidence of the shape of the distribution which can be useful in selecting a distribution. It may also be used directly in the simulation as the model of the process.

When data is abundant the use of the histogram is often adequate for many Monte Carlo applications. In using the histogram, care must always be exercised to remove obvious errors and to consider low probability events. When only limited data is available the histogram approach suffers from sampling peculiarities and from lack of observations in any tails of the distribution. In this case more effective distributions can be developed by taking into consideration other information about the behavior of the variable or by obtaining additional information from the data, e.g., by estimating higher moments. This information can range from an understanding of the theoretical nature of the variable to intuition. It might be assumed, for example, that the underlying real distribution is continuous; then smoothing procedures can be applied to the histogram to obtain a continuous curve.

Another graphical procedure useful in the selection of probability distributions involves the use of probability paper. As with the histogram, there is a large element of subjectivity in this procedure. It involves selection of an appropriate probability paper from those available and plotting the sample distribution function. Judgment is required in deciding whether the plot sufficiently approximates a straight line.

The use of graphical procedures in simulation development is described in Section 6, Part II.

3.3 PARAMETER ESTIMATION

A parametric distribution is defined to be a functional or analytical representation for a probability distribution which depends on one or more parameters. Although use of such distributions requires that the parameter(s) be estimated, there are a number of reasons for using a parametric distribution function rather than a histogram in developing a mathematical model. In particular, a parametric distribution:

- Provides a convenient means for inclusion of additional information about the variable (such as known upper and lower limits on the data).
- Allows meaningful extrapolation into the tail(s) of the distribution and into regions where no data was available.
- Allows incorporation of the additional information inherent in the shape of the distribution if there is a theoretical justification.
- Provides for a reproducible means of representing the data since freehand "fit" to the same data will vary from person to person.
- Provides important summary information about the variable in the form of estimated parameters of the fitted distribution.
- Provides a more compact representation of the random variable usually resulting in less data storage requirements.
- Allows construction of reasonable and convenient models in cases of no data or very limited data.
- Provides for efficient and convenient random number generation in most cases.
- Facilitates analytic (rather than simulation) studies of portions of the process.
- Permits a convenient means whereby analysis of the sensitivity to the shape of the distribution can be accomplished.

To facilitate the presentation of parametric distributions, the individual parametric families have been classified as being either of a simple or of a complex nature. The difference between these two classifications is mainly the number of parameters necessary to describe the distribution. The simple distributions are characterized by no more than two parameters, the complex by more than two.

The other distinguishing feature is that simple distributions are those which are commonly encountered, relatively easy to recognize, and have some theoretical basis for their functional form and application. Thus, simple parametric families of distributions can often be derived from assumptions about the process generating the random variable or from graphical evidence based on the data.

The complex parametric families generally do not have a "nice" physical interpretation or a simple functional form. They can be viewed more as abstract inventions which admit enough shapes to insure a reasonable fit to any set of observations. They also provide greater flexibility than simple distributions in projecting events of the process that would appear in the tails of the distribution.

3.3.1 Simple Parametric Distributions

The simple distributions include, but are not limited to, the normal, gamma, binomial, exponential, and other distributions which can be defined by at most two parameters. For the purposes of selecting an appropriate probability model, a simple distribution will be indicated by the underlying theory of the process or by preliminary selection using graphical procedures referred to previously.

One of the most common and useful of the simple continuous probability functions is the normal distribution. Much of the appeal of this distribution is based on a the central limit theorem. In essence, this states that the sum of independent variables tends to be normally distributed.[27] This assumes, of course, that none of the individual elements of the sum dominates its behavior. Since many variables which are modeled in Monte Carlo simulations are in reality derived from several variables, the assumption of a normal distribution can often be justified.

Since simple parametric distributions are discussed in detail in most elementary textbooks on probability, they are not discussed in detail here. However, a summary of the more common simple parametric distributions is given in Section 4.3.

3.3.2 Complex Parametric Distributions

As used in this volume, complex parametric distributions are defined as the Weibull, Johnson, and Pearson distribution families. The functional form of these distributions is somewhat complicated, and three to five parameters are often required to define the specific distribution. Reverting to the analytic procedures to generate these distributions is most necessary when a simple distribution cannot be justified and the simulation results are dependent upon rare events. Rare events are usually related to the tails of the distribution. For certain events or processes to be simulated sufficient observations to accurately define the tail regions may not exist. In such cases, one usually employs smoothing techniques utilizing parametric functions to extend or infer the behavior of the tail regions from available data.

Using a complex parametric distribution can be viewed as a convenient way of smoothing the raw data and expressing the smoothed data in functional form. These three families admit almost every type of probability distribution, one major exception being composite distributions made up of several distinct populations, e.g., multimodal distributions. In fact most of the simple parametric distributions are special cases of a Weibull, Johnson, or Pearson distribution.

If the reader is interested in a further discussion of these distributions, background information is contained in Appendix A. The material there is not, however, essential for understanding the principles discussed in Part I or the methods described in Part II.

3.4 GOODNESS-OF-FIT TESTS

After initial selections of a distribution for a Monte Carlo application and where sample data are available, it is usually worthwhile to try and validate or substantiate these choices. The validation step of the selection procedure is especially critical when it has been determined that the Monte Carlo result will be sensitive to distribution selection. More generally, developing confidence in the distributions used in any simulation adds to the confidence in the total simulation in addition to aiding in the overall understanding of the process.

One of the most useful methods used in validation is called goodness-of-fit-tests. These are statistical procedures for testing whether sample data can reasonably be expected to be representative of (drawn from) a particular probability distribution. Essentially, there are two such tests which have found wide application since they can be applied to any distribution. These are the Chi-Square test and the Kolmogorov-Smirnov test. A brief description of each of these two tests is presented below. In addition there are a number of specialized tests such as the W-test for a normal distribution and the WE-test for an exponential distribution which are useful. Specific details for applying these tests are contained in Part II, Section 9.

One word of caution should be noted in using these tests. The statistical inferences based on these tests rely on asymptotic properties. Thus a fair amount of data is required to obtain valid interpretations. Where limited data are available or many erroneous data points are believed to be in the sample, the usefulness of these tests may be questionable.

Chi-Square Test: This common goodness-of-fit-test is made by subdividing the data into groups or intervals and comparing the number of actual observations A_i in the i^{th} interval to the number expected E_i as computed from the assumed distribution. The statistic employed in this method is

$$\chi^2_{n-1} = \sum_{i=1}^{n} \frac{(A_i - E_i)^2}{E_i}$$

Under the null hypothesis (observations are from the assumed distribution) the distribution of this statistic asymptotically approaches a Chi-Square distribution with n-1 degrees of freedom.

The Chi-Square test has certain obvious shortcomings. In addition to being sensitive to sample size, this test is also sensitive to data grouping. Different investigators conducting this test will tend to get different results. One requirement in using the test is that each cell or subgroup should have a sufficient number of observations in it. Some authors (27) feel that a good test requires at least twenty observations per cell and that there should also be between five and twenty cells.

Kolmogorov-Smirnov Test: This goodness-of-fit test is made by computing the maximum difference between the sample cumulative distribution function and the assumed distribution function. This difference, under the null hypothesis, has a known asymptotic distribution which is available in table form (see Appendix B). The Kolmogorov-Smirnov is generally considered to be more sensitive than the Chi-Square test and also has the advantage that arbitrary data grouping decisions are not required. Its disadvantages are that it is usually more computationally difficult to apply, and if the hypothesis is rejected, the reason for the rejection is less clear.

PART II

SELECTION OF DISTRIBUTIONS

4. DISTRIBUTION SELECTION PROCEDURES

This section presents a systematic set of procedures for selecting the most representative model for a random variable in a simulation. The procedures selected depend on two types of knowledge of the random variable in question. These are:

1. Empirical Data (Quantitative Observations)
2. Understanding of the Random Process (Qualitative A Priori Knowledge).

Based on the degree of knowledge in each category, a set of procedures for selecting a distribution has been constructed. By following a particular procedure the most appropriate probability model can be easily selected.

The initial discussion in this section is devoted to a discussion of selecting the appropriate procedure to be used based on the degree of available knowledge of the random variable in question. Secondly, this section is devoted to presenting a brief guide to using the remaining sections of Part II. This section is concluded with a table listing all the candidate distributions considered here. This table also summarizes the characteristics of these distributions. The rest of Part II is concerned with how one performs the specific operations which lead to selection of the appropriate probability distribution model.

4.1 PROCEDURES FOR SELECTING DISTRIBUTIONS

The particular selection procedure for a probability model is determined by the extent of empirical data and knowledge of the random process in question. The extent of empirical data can, for convenience, be broken into three categories: none, some, and ample. This categorization is given in Table 4.1.

TABLE 4.1

Extent of Empirical Data (Observations)

Category	1	2	3
Description	none	some	ample
Number of Observations	0-5	5-20	over 20

The extent of knowledge of the random process is, for convenience, broken into four categories: no knowledge, qualitative knowledge, reasonably good ideas, and reasonable certainty. These categories are described further in Table 4.2. It should be clear that the more data and the greater the a priori qualitative knowledge available, the easier the selection process is and the greater the certainty of obtaining a good probability model.

TABLE 4.2

Extent of Qualitative Knowledge of the Random Process

Category	1	2	3	4
Description	None:	Qualitative:	Good ideas:	Reasonable certainty:
	No qualitative knowledge of the random process	Some knowledge of the random process, i.e. continuity, range, symmetry, shape of distribution, likely values, etc.	Reasonably based expectations that the random variable is one of a few known families	Good basis for expecting the distribution to be some known family

A concerted effort should be made to use all a priori knowledge. This means that all the qualitative characteristics listed under Category 2 in Table 4.2 should be written down, if known. This will also help in sketching a probability density or frequency curve. Table 4.3 should also be consulted to determine if Categories 3 or 4 are appropriate. Table 4.3 lists all of the probability distributions considered here. These are arranged in two groups, the simple parametric distributions and the complex parametric distributions. This table also summarizes the characteristics of these distributions. Table 4.3 is very useful as a reference in selecting a probability distribution since almost all of the information needed for selection is presented. To this end, therefore, the columns in Table 4.3 entitled Comments and Justification and Applications may give characteristics that fit the problem at hand. Any distributions that appear appropriate should be listed so that knowledge at a level of Category 3 or 4 can be used.

Once the categories for empirical data and knowledge of the random process have been established from Tables 4.1 and 4.2, a specific selection procedure can be identified from Table 4.4. Table 4.4 is simply a matrix indicating all possible combinations of data and knowledge categories. For each combination, a figure number is indicated. Each figure presents the details of the particular selection procedure that it represents.

A discussion of the selection procedures presented in Figure 4.1-4.12 and how that material is used is contained in the following section (4.2).

Simple Parametric Distributions:

TABLE 4.3
Summary of Probability Distributions

Distribution	Functional Form of f(x)	Recommended Parameter Estimators	Goodness-of-fit Test	Comments & Justification	Applications	References (See App. C)
Uniform	$\frac{1}{b-a}$ $a \le x \le b$	$b = \max[x_1,\dots,x_n]$ $a = \min[x_1,\dots,x_n]$	d-test(2) χ^2 - test	Equal probability in any interval	Wide use for events of equal probability	15, 24, 35
Exponential	$\lambda e^{-\lambda(x - \epsilon)}$ $x \ge \epsilon$ $\lambda > 0$	$\epsilon = \min[x_1,\dots,x_n]$ $\lambda = 1/(\bar{X} - \epsilon)$	WE - test or WE_0-test if ϵ is known	Wide applicability to any process with no 'memory' of the past and constant rate, e.g., a process where (probability of event)/time interval is constant and independent of time elapsed.	Queueing, quality control, reliability, etc.	6, 11, 15, 17, 24, 35
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ $\sigma^2 > 0$	$\mu = \bar{X}$ $\sigma^2 = S^2$	W - test	Any variable generated by the sum of many uniform random numbers. Wide applicability as it is often justified by the central limit theorem.	Physical measurements on living organisms, intelligence scores, product dimensions, bombing errors (1 dimensional), average temp., etc.	15, 24, 29, 35
Cauchy	$\frac{1}{\pi[1 + (x - \mu)^2]}$; $-\infty < x < \infty$	μ = sample median		The ratio of two independent normalized normal random variables. Distribution of tan θ if θ is uniform.	Ratio of standardized noise readings. Caution: Cauchy moments are infinite; behavior in a Monte Carlo program will be erratic.	15, 24, 29
Rayleigh	$\frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$ $x \ge 0$ $\sigma^2 > 0$	$\sigma^2 = \frac{1}{2}\overline{X^2}$	χ^2 - test d - test[2]	Sum of squares of two independent normalized random variables, i.e., radial error when x and y errors are independent and normal with same standard deviation.	Bomb sighting problems; amplitude of noise envelope for linear detector.	15, 24, 29
Gamma	$\frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}$ $x \ge 0$ $\lambda > 0$ $k > 0$	$\lambda = \frac{1}{S^2}\bar{X}$ $k = \lambda\bar{X}$	χ^2 - test d - test[2]	The time for exactly k independent events if events occur at constant rate λ.	Time between inventory restocking or recalibration, time to failure with standby, queueing time distributions.	5, 15, 17, 24, 29, 36

[1] $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ sample mean

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{X})^2$, sample variance

[2] d - test may not be strictly applicable but can still be used.

[3] References quoted can be found in the abstracted bibliography of Appendix

TABLE 4.3 (Continued)

Distribution	Functional Form of f(x)	Recommended Parameter Estimators	Goodness-of-fit Test	Comments & Justification	Applications	References (See App. C)
Beta	$\frac{1}{b-a}\,\frac{\Gamma(\gamma+\eta)}{\Gamma(\gamma)\Gamma(\eta)}\left(\frac{x-a}{b-a}\right)^{\gamma-1}\left[1-\frac{x-a}{b-a}\right]^{\eta-1}$ $a \le x \le b$ $\gamma > 0$ $\eta > 0$	$\eta = \frac{1}{S^2}\left[1-\frac{\overline{X}-a}{b-a}\right]\left[\frac{\overline{X}-a}{b-a}\left(1-\frac{\overline{X}-a}{b-a}\right)-S^2\right]$ $\gamma = \frac{\eta\left(\frac{\overline{X}-a}{b-a}\right)}{\left(1-\frac{\overline{X}-a}{b-a}\right)}$	χ^2 - test d - test[2]	A basic distribution of statistics for variables bounded at both sides.	Distribution of time to task completion, distribution of daily yield in manufacturing processes.	12, 15, 24, 29
Pareto	$\lambda\epsilon^{\lambda}\,x^{-\lambda-1}$ $\quad x \ge \epsilon$	$\epsilon = \min\,[x_1,\ldots,x_n]$ $\lambda = \left[\log\frac{\overline{X}}{\epsilon}\right]^{-1}$	χ^2 - test d - test[2]		Distribution of income and property values.	24, 26
Log-normal	$\frac{1}{\sigma(x-\epsilon)\sqrt{2\pi}}\cdot\exp\left[-\frac{1}{2\sigma^2}(\ln(x-\epsilon)-\mu)^2\right]$ $x \ge \epsilon$ $\sigma > 0$	$\epsilon = \min\,[x_1,\ldots,x_n]$ $\mu = \overline{\ln(x_i-\epsilon)}$ $\sigma^2 = \overline{\ln(x_i-\epsilon)^2}$	W - test	A random variable whose logarithm is normal. Model for a process arising from the multiplication of many uniform random numbers. Appropriate when the value of a variable is a random proportion of the previously observed value.	Size distribution from breakage, various economic distributions, various biological phenomena, life test data, reliability.	24, 37
Folded Normal	$\frac{1}{\sigma\sqrt{2\pi}}\left[e^{-(x-\mu)^2/2\sigma^2}+e^{-(x+\mu)^2/2\sigma^2}\right]$ $x > 0$	See Reference 9	χ^2 - test d - test[2]	Distribution of the absolute value of a normal variable.		9, 24
Kodlin's Distributions	$(\eta+\gamma x)e^{-(\eta x+1/2\,\gamma x^2)}$	See Reference 28	χ^2 - and d-test[2]		Suggested for a variety of survival time data.	28
Extreme Value Distributions	Maximum value: $\frac{1}{\sigma}\exp\left[-\frac{1}{\sigma}(x-\mu)-e^{-\frac{1}{\sigma}(x-\mu)}\right]$ Minimum value: $\frac{1}{\sigma}\exp\left[\frac{1}{\sigma}(x-\mu)-e^{\frac{1}{\sigma}(x-\mu)}\right]$	$\sigma = 1.283\,S$ $\mu = \overline{X} - .577\,\sigma$ Maximum $\mu = \overline{X} + .577\,\sigma$ Minimum	χ^2 - and d-test[2]	The asymptotic distribution for the maximum (or minimum) of a large number of values from such distributions as gamma, normal, log-normal, and exponential.	Maximum wind gust velocities, stock market maxima or minima, flood or drought distributions.	15, 34, 35

TABLE 4.3 (Continued)

Distribution	Functional Form of f(x)	Recommended Parameter Estimators	Goodness-of-fit Test	Comments & Justification	Applications	References (See App. C)
Binomial *	$\binom{n}{k} p^k (1-p)^{n-k}$	p = ratio of success to total trials	χ^2 - test	Describes probability of k successes in n independent trials	Quality control, reliability, sampling, etc.	15, 23, 29, 35
Multinomial *	$\binom{n}{k_1 k_2 \ldots k_m} p_1^{k_1} p_2^{k_2} p_3^{k_3} \ldots p_n^{k_m}$ $k_1 + k_2 + \ldots + k_m = n$	p_i = ratio of i^{th} outcome	χ^2 - test	Describes outcome of n independent trials where there are m alternatives for each trial.	Quality control, reliability, sampling, etc.	15, 23, 35
Poisson *	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ = mean value of k	χ^2 - test	Describes the number of occurrences in an interval when the rate of occurrence is constant.	Queueing, reliability, quality control, sampling.	15, 23, 29, 35
Hyper-geometric *	$\frac{\binom{}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	From formula $\frac{\text{mean value of } k}{n} = \frac{M}{N}$, either M or N can be estimated if the other is known.	χ^2 - test	Describes the probability of an event occurring k times in a sample of size n when it is known that M events will occur in the population of size N.	Reliability, quality control, sampling	15, 23, 35
Geometric *	$p(1-p)^{k-1}$	p = mean number of successes	χ^2 - test	Describes the number of trials to the first success in a sequence of Bernoulli trials.	Quality control, sampling, etc.	15, 23, 35
Pascal (also called negative binomial) *	$\binom{n+k-1}{k} (1-p)^n p^k$	p = mean number of successes in a series of trials.	χ^2 - test	Describes the probability of exactly k successes occurring before the n^{th} failure in a series of Bernoulli trials.	Quality control, sampling, etc.	15, 23, 35

*Discrete Distributions

Complex Parametric Distributions:

TABLE 4.3 (Continued)

Distribution	Functional Form of f(x)	Recommended Parameter Estimators	Goodness-of-fit Test	Random Number Generation (See Volume II)	Applications	References (See App. C)
Weibull	$\frac{\eta}{\lambda}(x-\epsilon)^{\eta-1} e^{-\frac{(x-\epsilon)^\eta}{\lambda}}$ $x \geq \epsilon$ $\eta, \lambda > 0$	ϵ can be estimated as outlined in Ref. 8. For η and λ, a simple technique is described in Ref. 38. A more complicated but more accurate method is the maximum likelihood estimate, as outlined in Ref. 51. For both techniques, the estimates should be multiplied by the unbiasing factors given in Ref. 51.	The WE_o test can be used on $\frac{(x-\epsilon)^\eta}{\lambda}$	An analytic selection technique requiring little data storage and using moderate amounts of computer time is available. With a little effort, very fast techniques requiring sizable data tables can be developed.	Data storage requirements are small. Random number selection is easy. All values in variable range, especially in the tail of the distribution, will be represented. Wide applicability. Has been used successfully for such diverse cases as yield strength of steel, size distribution of fly ash, fatigue life of steel, height of adult males, and width of beans.	1, 8, 15, 17, 21, 22, 34, 35, 36, 38, 51, 55
Johnson	S_L: $\frac{\eta}{\sqrt{2\pi}(x-\epsilon)}$ $\exp\left\{-\frac{\eta^2}{2}\left[\frac{\gamma}{\eta}+\ln(x-\epsilon)\right]^2\right\}$ $x \geq \epsilon$ S_B: $\frac{\eta}{\sqrt{2\pi}} \frac{\lambda}{(x-\epsilon)(\lambda-x+\epsilon)} \cdot$ $\exp\left\{-\frac{1}{2}\left[\gamma+\eta\ln\left(\frac{x-\epsilon}{\lambda-x+\epsilon}\right)\right]^2\right\}$ $\epsilon \leq x \leq \epsilon+\lambda$ S_U: $\frac{\eta}{\sqrt{2\pi}} \frac{1}{\sqrt{(x+\epsilon)^2+\lambda^2}} \cdot$ $\exp\left[-\frac{1}{2}\left(\gamma+\eta\ \ln\left\{\left(\frac{x-\epsilon}{\lambda}\right)+\left[\frac{(x-\epsilon)^2}{\lambda^2}+1\right]^{\frac{1}{2}}\right\}\right)^2\right]$ $-\infty < x < \infty$ $\eta, \lambda > 0$ $-\infty < \gamma, \epsilon < \infty$	Calculation of skewness and kurtosis for data determines which type (S_L, S_B, or S_U) of distribution to use. At some expense in computer time, maximum likelihood equations can be solved numerically. However, unbiasing factors are not yet available. Simpler, but less accurate, methods are the percentile technique (Ref.35) for S_L and S_B and the moments method for S_U (see Ref. 22).	S_L: Use W-test on $\ln(x_i-\epsilon)$ S_B: Use W-test on $\ln\left(\frac{x_i-\epsilon}{\lambda+\epsilon-x_i}\right)$ S_U: Use W-test on $\sinh^{-1}\left(\frac{x_i-\epsilon}{\lambda}\right)$	Fairly fast analytic selection techniques requiring little data storage are available. Very fast numerical techniques which would, however, require sizable data storage could be developed with a little effort.	Almost universal applicability. Data storage requirements are small. Random number selection is easy. All values in variable range, especially in the tail of the distribution, will be represented.	15, 21, 22, 35, 37

TABLE 4.3 (Continued)

Distribution	Functional Form of f(x)	Recommended Parameter Estimators	Goodness-of-fit Test	Random Number Generation (See Volume II)	Applications	References (See App. C)
Pearson	Twelve types. See Table 8.1.	The method of moments can be used to estimate parameters. This is described in detail for each type in Ref. 10. Alternatively the percentile technique, using tables and outline given in Ref. 25, may be employed. The maximum likelihood technique, which requires considerable computer time may also be used. This method is more powerful than the others, but its bias has been incompletely investigated (although Refs. 12 and 5 consider the bias for types 1 and 3).	χ^2-test d-test may also be tried although it is not strictly applicable.	Analytic selection techniques which are fast or which require small amount of computer storage are available only for integral values of exponents and only for a few types. In the general case, numerical techniques must be used which will involve: 1. Moderate to large numbers of storage cells for tabular data. 2. Considerable processing time spent generating the selection technique.	Almost universal applicability. All values in variable range, especially in the tail of the distribution, will be represented.	5, 10, 12, 15, 17, 18, 25 35, 40, 41, 42

TABLE 4.4

Sequence of Activity Selection (By Figure Number)

Knowledge of Random Process Category

Empirical Data Category	1	2	3	4
1	Figure 4.1	Figure 4.2	Figure 4.3	Figure 4.4
2	Figure 4.5	Figure 4.6	Figure 4.7	Figure 4.8
3	Figure 4.9	Figure 4.10	Figure 4.11	Figure 4.12

4.2 SELECTION TECHNIQUES

The following list provides a brief description of each selection technique used in the selection procedures and provides the location of further detailed discussion.

Sensitivity Analysis - (Section 5.) — Involves performing the simulation study using several different distributional assumptions or parameters to examine the effect it has on the final results.

Graphical Analysis - (Section 6.) — Involves plotting a histogram and/or using probability paper to judge what distributions appear likely. This analysis may reject some ideas as inappropriate or suggest several likely distributions. This analysis applies primarily to the simple or common distributions.

Analytic Curve Fitting - (Section 7.) — Refers to fitting the data to one or more of the complex or uncommon distributions such as the Weibull, Johnson, and Pearson.

Parameter Estimation - (Section 8.) — Is the task of estimating the values of the parameters of a given distribution family to obtain the best fit with the data.

Goodness-of-Fit - (Section 9.)	Tests are used to determine if the candidate distribution is an adequate representation of the actual random process based on the data available.
Histogram - (Section 6.)	If all likely distributions fail the goodness-of-fit tests fail, a histogram should be used.

These techniques can best be applied by referring to the appropriate section. After application of any technique, refer to the appropriate figure to determine subsequent selection techniques to employ, if any.

Figure 4.1
No Data, No Knowledge

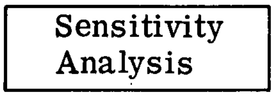


Figure 4.2
No Data, Qualitative Knowledge

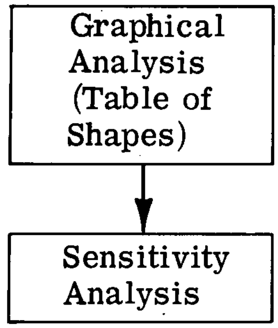


Figure 4.3
No Data, Good Knowledge

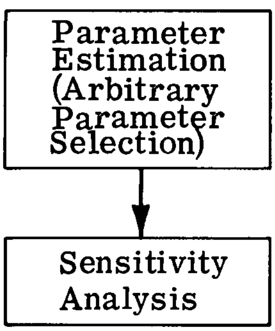


Figure 4.4
No Data, Certain Knowledge

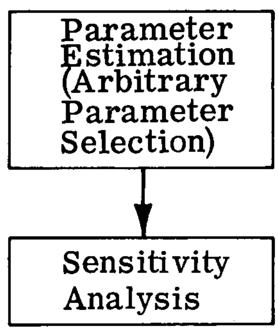


Figure 4.5
Some Data, No Knowledge

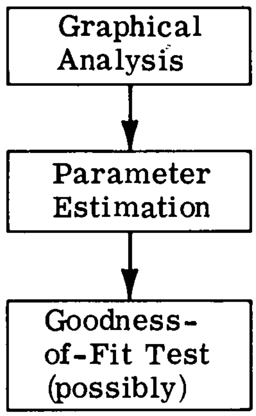


Figure 4.6
Some Data, Qualitative Knowledge

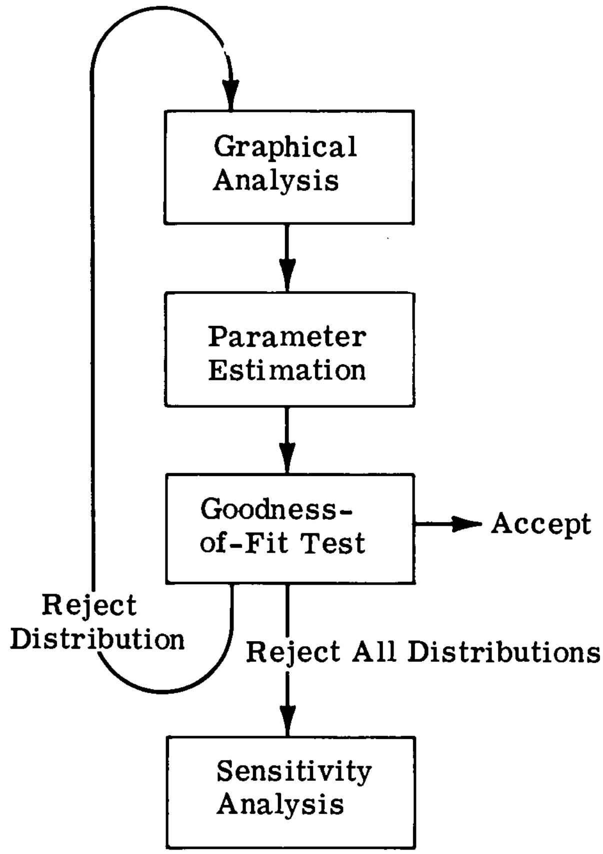


Figure 4.7
Some Data, Good Knowledge

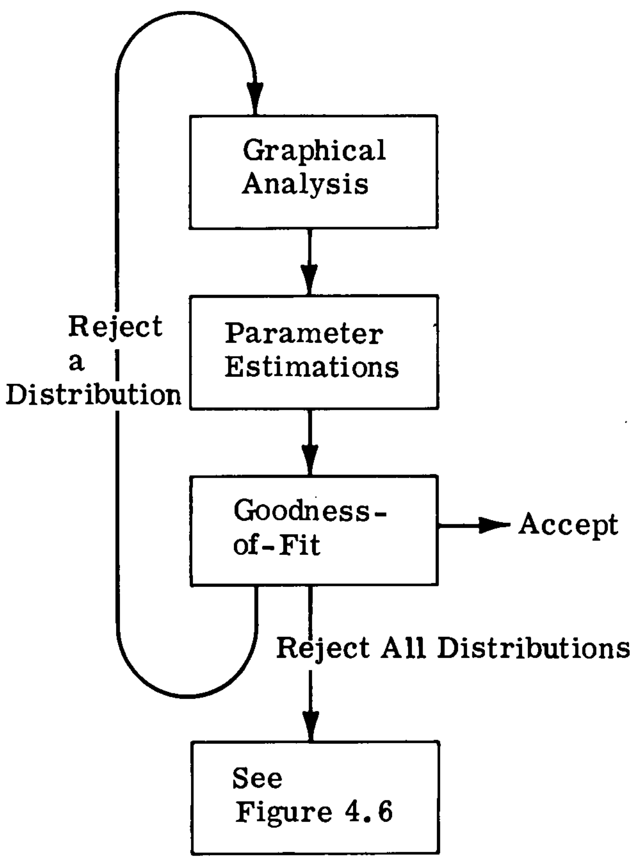


Figure 4.8
Some Data, Certain Knowledge

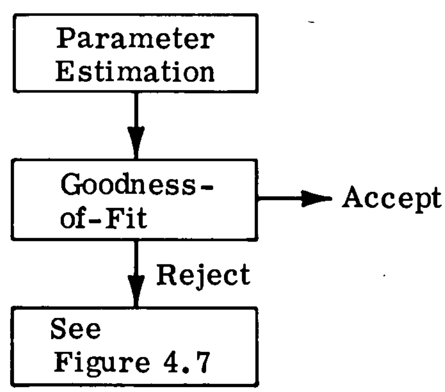


Figure 4.9
Ample Data, No Knowledge

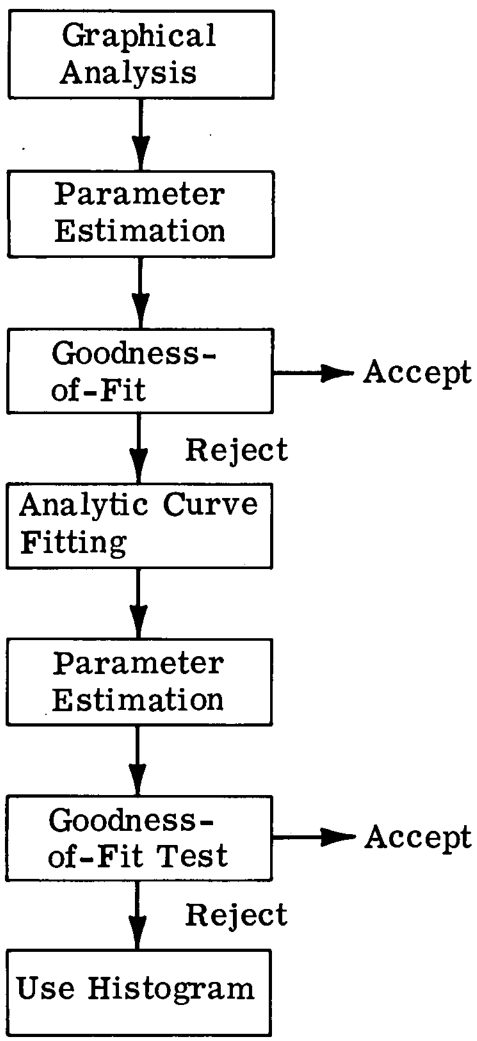


Figure 4.10
Ample Data, Qualitative Knowledge

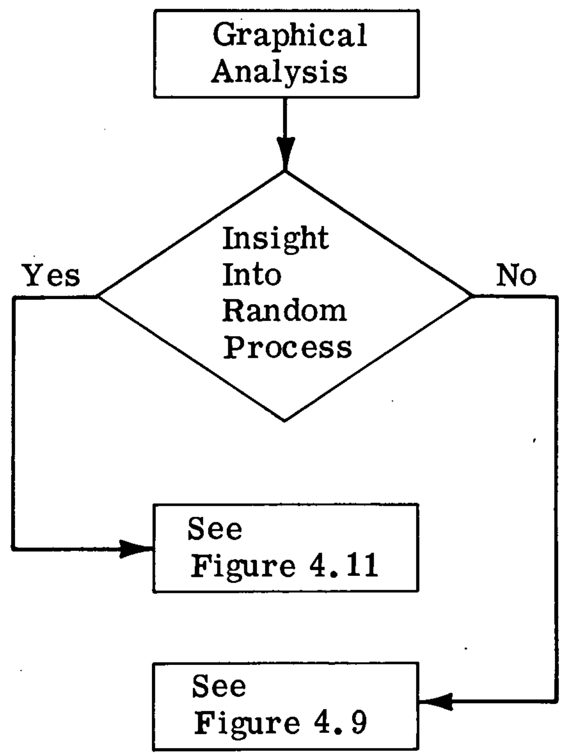


Figure 4.11
Ample Data, Good Knowledge

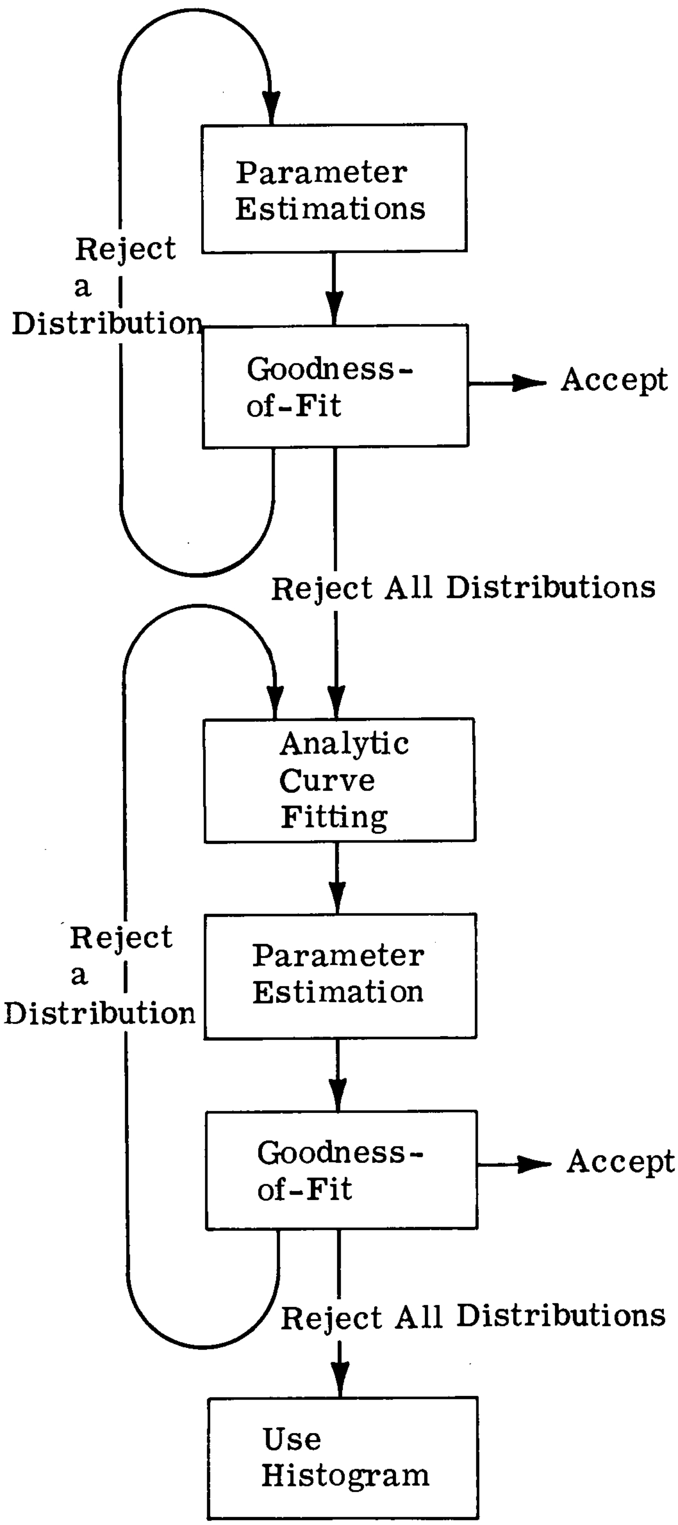
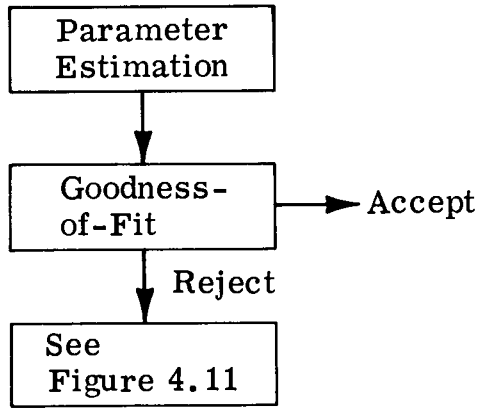


Figure 4.12
Ample Data, Certain Knowledge



5. SENSITIVITY ANALYSIS

The objective of sensitivity analysis is to determine the extent to which the final results of the simulation study are sensitive to a given probability distribution. To this end two general guidelines can be given.

The first is to attain a determination of sensitivity to the parameters of a distribution. It might be reasonable to vary the parameters to some extent in both directions. Suppose, for example, that a normal distribution with mean 100 and standard deviation 20 is postulated. Then five runs might be made to test sensitivity of the final simulation results to these parameters as follows [(mean, standard deviation)]: (100, 20), (110, 20), (90, 20), (100, 18), (100, 22).

A second sensitivity test that can be performed is one of shape of parametric family: it may be reasonable to make several simulations with different probability distributions, especially if unlikely events are important to the simulation results. In this case the shape of the tail of the distribution is important. Suppose, for example, that a gamma distribution has been chosen; then a lognormal or Weibull might also be tried, since these have similar shapes.

6. GRAPHICAL TECHNIQUES

There are two graphical techniques that are applicable here. The first deals with the empirical histogram and the second deals ing with the empirical cumulative distribution polygon. Both techniques can be quite useful in selecting a good functional fit to data. These graphical techniques are intended primarily for use in selecting one of the common or simple distributions. Although graphical techniques can be helpful in the selection of a complex distribution, this is discussed as analytical curve fitting in Section 7.

Graphical techniques can often suffice to determine a satisfactory probability model for a simulation variable. This is especially true if the simulation results are not sensitive to rare events of the several random variables. An example is given in Section 6. 3 to illustrate the histogram and cumulative distribution polygon methods.

6.1 USING THE EMPIRICAL HISTOGRAM

The empirical histogram can be used to determine what distributions are likely to fit a given set of data. This can best be accomplished by a visual comparison to find curves representing probability distributions that are similar to the data. The approach taken in this section is to find such visual fits by examining a series of figures representing the density function of most of the simple distributions.

The procedure is very straightforward. First plot the histogram from the data available. In some cases it may be helpful to sketch a smoothed version of the histogram, especially if the cells of the observation groupings are large or the data are few. Then examine the shapes given in Figure 6.1 and select those distributions whose densities are similar to the histogram. (Figure 6.1 does not include the Weibull, Johnson, or Pearson distributions. For these distributions, see Section 7.) It is also useful to rank the selections according to how good the fit is.

6.2 USING THE EMPIRICAL CUMULATIVE DISTRIBUTION POLYGON

An alternate technique is to use the cumulative distribution polygon in conjunction with probability paper. The horizontal axis of this paper represents the values of the variable under investigation; the vertical axis is a probability scale. The spacing on the vertical axis is constructed for a given probability family so that a cumulative distribution function belonging to that family will appear as a straight line on the paper.

The graphical method is quite general and can be applied to any known distribution; however, the probability paper which is commercially available is limited to the more commonly encountered distributions such as the normal (see Figure 6.2), lognormal, extreme value, chi-square, gamma, binomial, and Weibull.*

The procedure for using this graphical method is extremely simple although interpretation of the results is somewhat subjective. The sample cumulative distribution is plotted on the probability paper corresponding to the theoretical distribution of interest. If the points

*See, for example, TEAM Special Purpose Graph Papers, Box 25, Tamworth, N.H. 03886, also K+E papers.

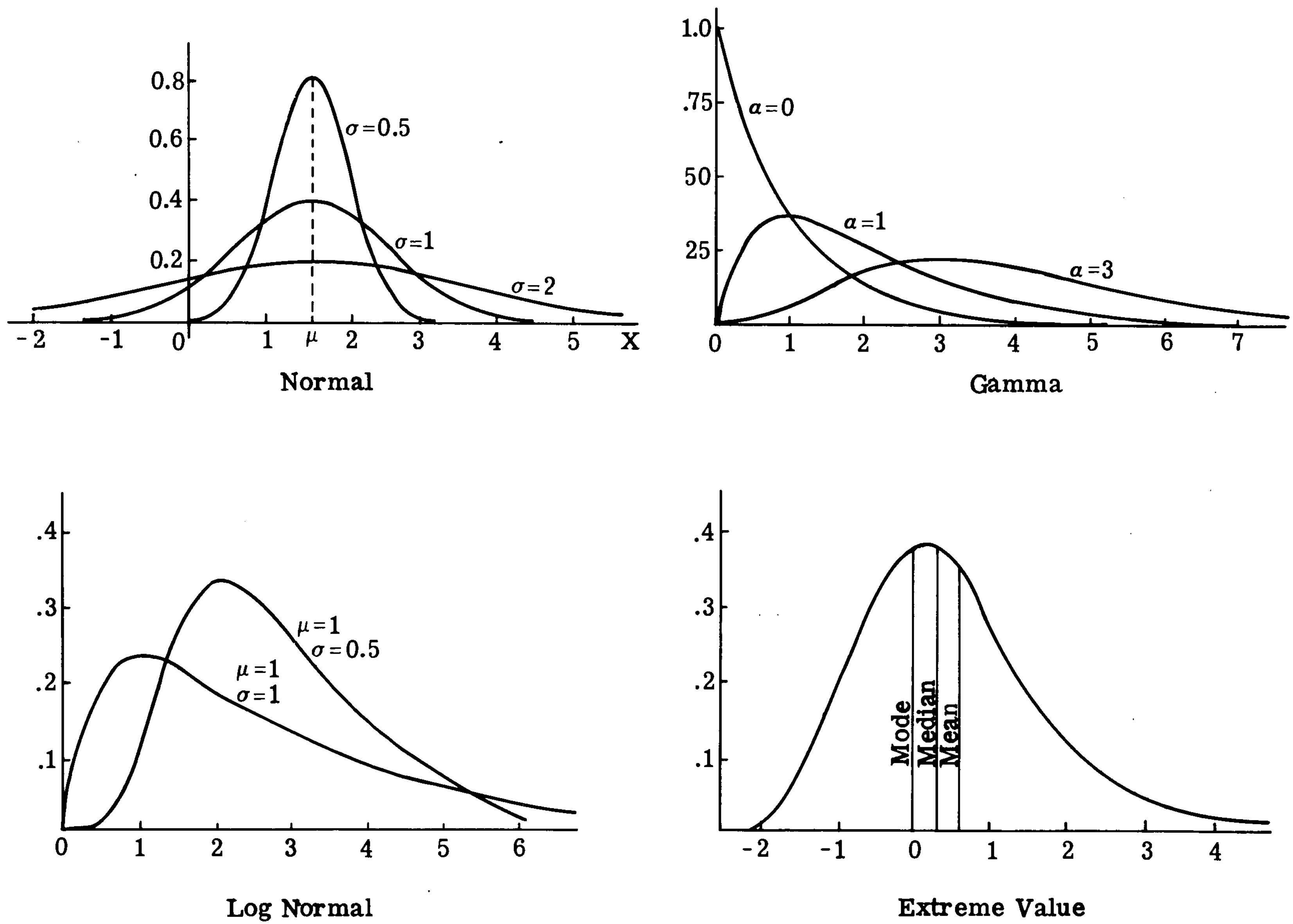


Fig. 6.1. Shapes of simple parametric distributions (Sheet 1 of 3)

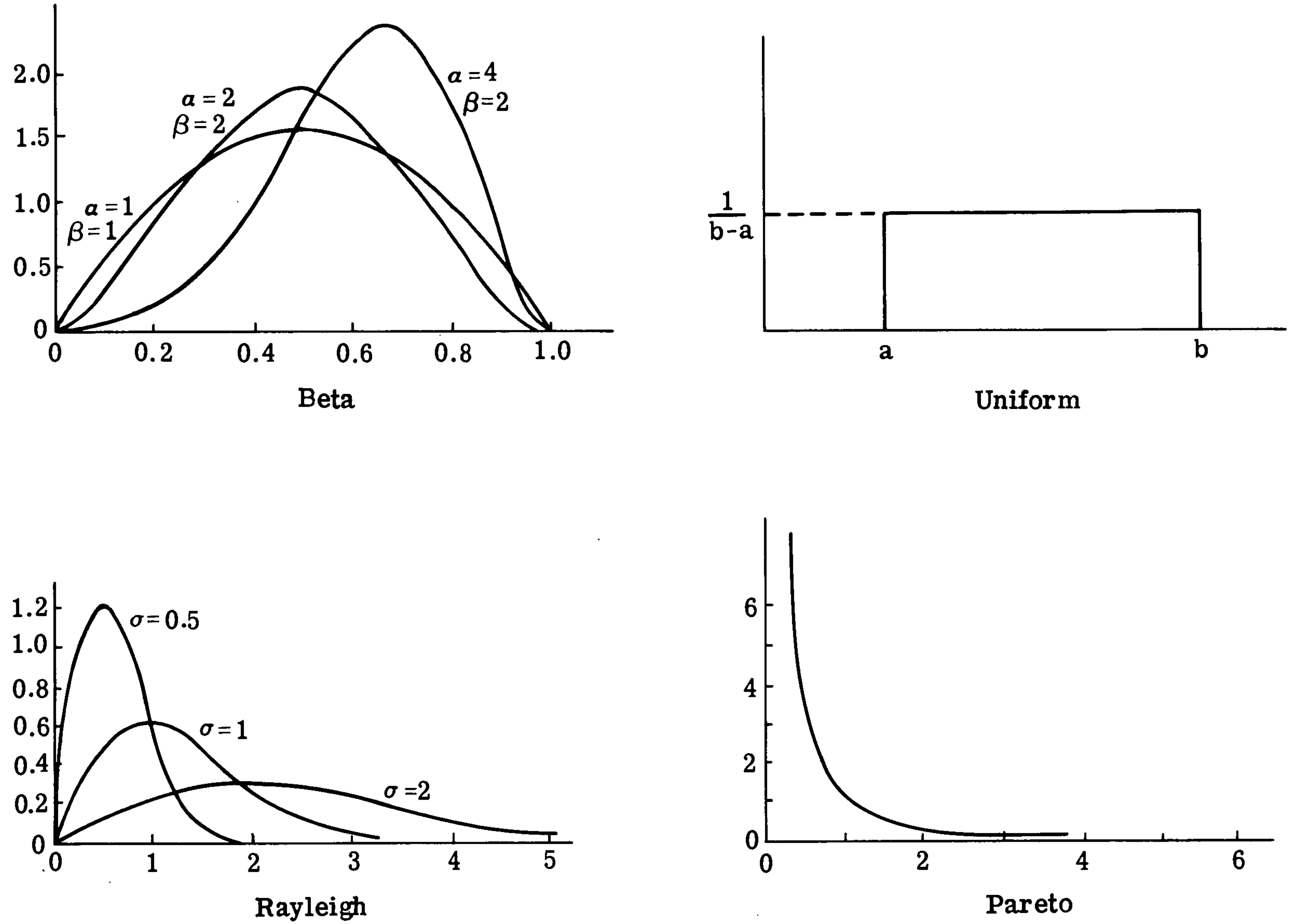


Fig. 6.1. Shapes of simple parametric distributions (Sheet 2 of 3)

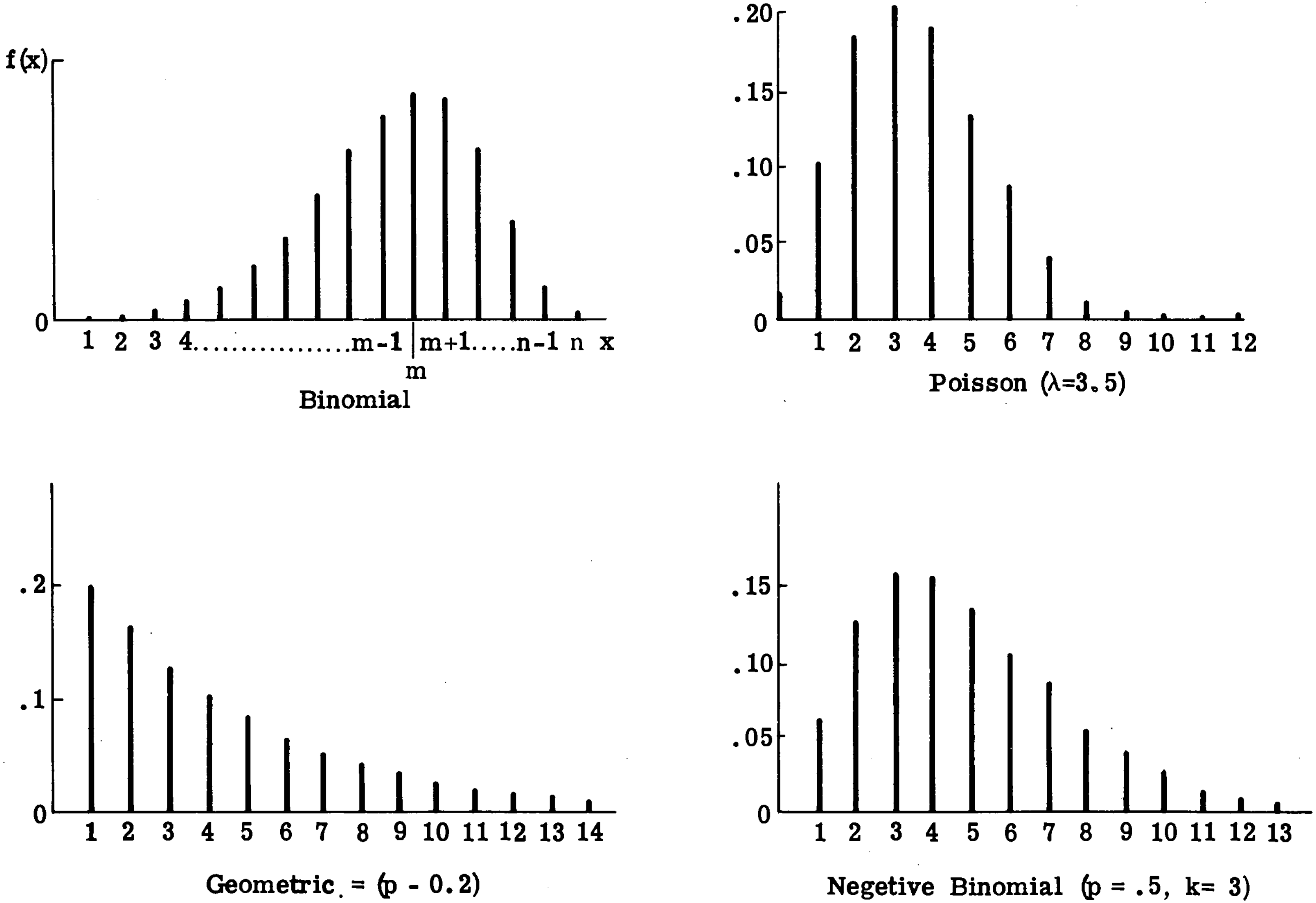


Fig. 6.1. Shapes of simple parametric distributions (Sheet 3 of 3)

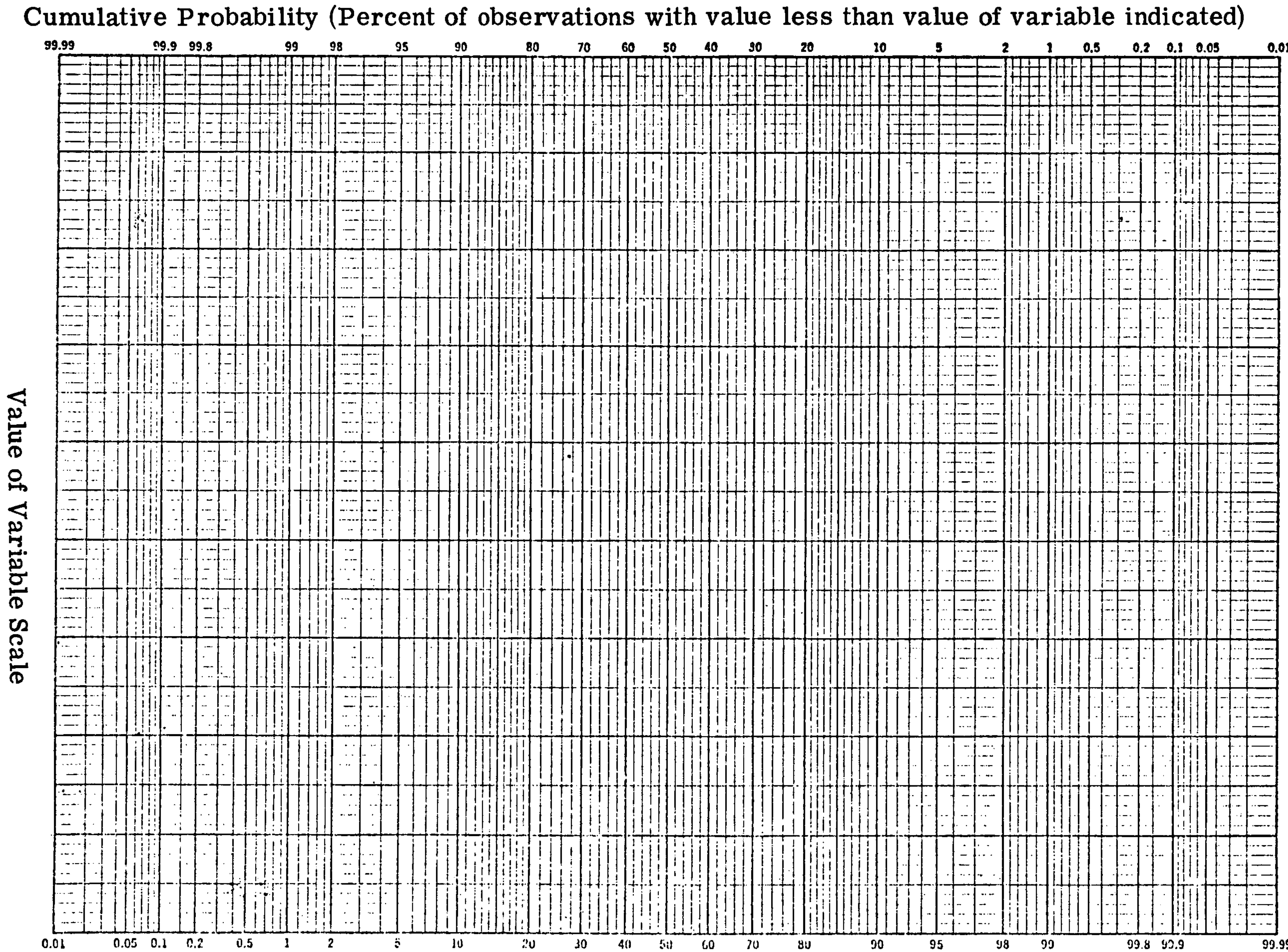


Fig. 6.2. Normal probability paper

fall on a straight line the theoretical distribution is accepted as representative of the data. If the line is badly curved, other distributions can be tried. The nature of the curve often suggests distributions which might be of better fit.

Another useful aspect of the graphical procedure is that estimates of the distribution's parameters can be read directly off the graph. For example, on normal probability paper, the difference in variable value between the .50 probability point and the .84 probability point on the fitted line corresponds to one standard deviation.

6.3 NUMERICAL EXAMPLE

An example will illustrate the use of these techniques. The data for the example is given in Table 6.1. Observations ranging from 66.75 to 75.25 have been divided into seventeen equal intervals or cells of 0.50 each. The frequency with which observations fall within each cell has been tabulated and summarized. This data was then plotted in Figure 6.3 to produce what is generally referred to as a histogram.

The histogram serves two purposes. First, it provides visual evidence on which to base preliminary selection of a distribution. Second, in the case of limited data, it may provide as good an estimate of the variability of the process as any other more elaborate approach.

On the basis of its symmetry and bell shape, the histogram of Figure 6.3 appears typical of data from a normal distribution. Making an assumption of normality, it is possible to proceed to the application of other quantitative methods to determine its validity.

TABLE 6.1

Sample Data

Cell Boundaries	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
66.75-67.25	2	0.005	2	0.005
67.25-67.75	2	0.005	4	0.011
67.75-68.25	5	0.014	9	0.025
68.25-68.75	6	0.016	15	0.041
68.75-69.25	7	0.019	22	0.060
69.25-69.75	24	0.066	46	0.126
69.75-70.25	36	0.099	82	0.225
70.25-70.75	48	0.132	130	0.357
70.75-71.25	64	0.176	194	0.533
71.25-71.75	51	0.140	245	0.673
71.75-72.25	41	0.113	286	0.786
72.25-72.75	32	0.088	318	0.874
72.75-73.25	24	0.066	342	0.940
73.25-73.75	12	0.033	354	0.973
73.75-74.25	5	0.014	359	0.986
74.25-74.75	4	0.011	363	0.997
74.75-75.25	1	0.003	364	1.000

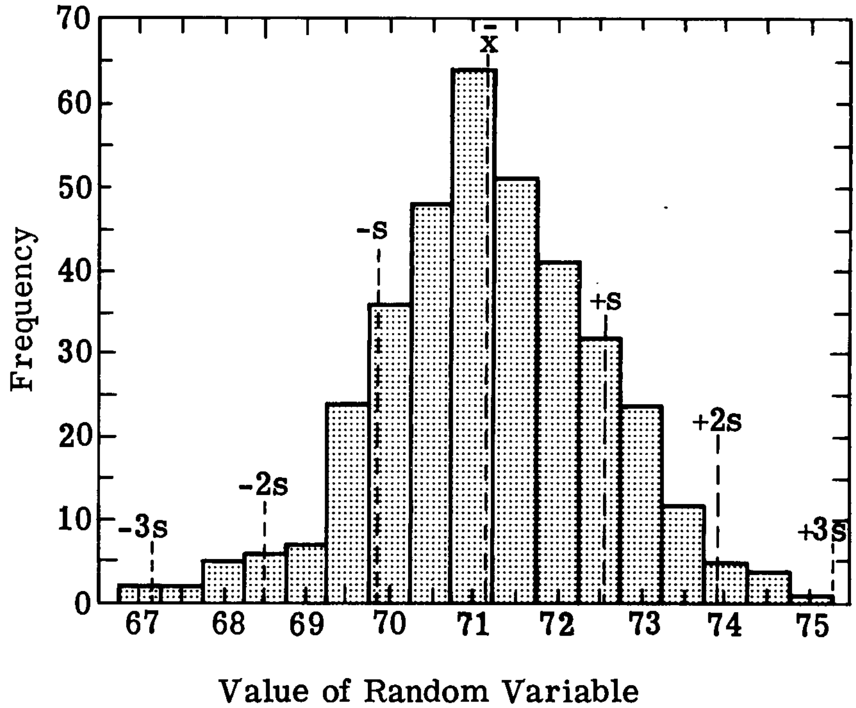


Fig. 6.3 Frequency Histogram for Data of Table 6.1

The data given in Table 6.1 can also be plotted on normal probability paper. This will verify the assumption of a normal distribution and also give the appropriate parameters for the distribution if the assumption of normality is accepted. The cumulative relative frequency (sample cumulative distribution function) when plotted on normal probability paper, shown in Fig. 6.4, turns out to be reasonably linear. Thus it can be concluded, at least tentatively, that the data in Table 6.1 has been drawn from a normal population. For many applications this will suffice to identify a satisfactory distribution. Note that the mean (μ) and the standard deviation (σ) can also be estimated from the graph.

Rather than go through the process of grouping the data into class intervals or cells as in Table 6.1 one can plot the data directly onto probability paper in the following way. The n observations $x_1, x_2, \ldots, x_n$ are placed in ascending order (ranked) such that:

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq --- \leq x_{(n-1)} \leq x_{(n)} .$$

To each $x_{(i)}$ associate the ordinate value $y_{(i)} = \frac{1}{n+1}$ and plot the ordered pairs $(x_{(i)}, y_{(i)})$ on the probability paper. This procedure is extremely fast, with the exception of having to rank the n observations. Therefore, it is probably most useful for sample sizes in the range 1-50, depending of course on how proficient one is at ranking observations. Many excellent examples of the use of probability paper for extreme value distributions may be found in Gumbel.[14]

This example is concluded with a visual verification of the selection of a normal distribution to fit the data in Table 6.1. Figure 6.5 gives the same information as Fig. 6.3 with the addition of the normal density curve scaled to the frequency polygon.

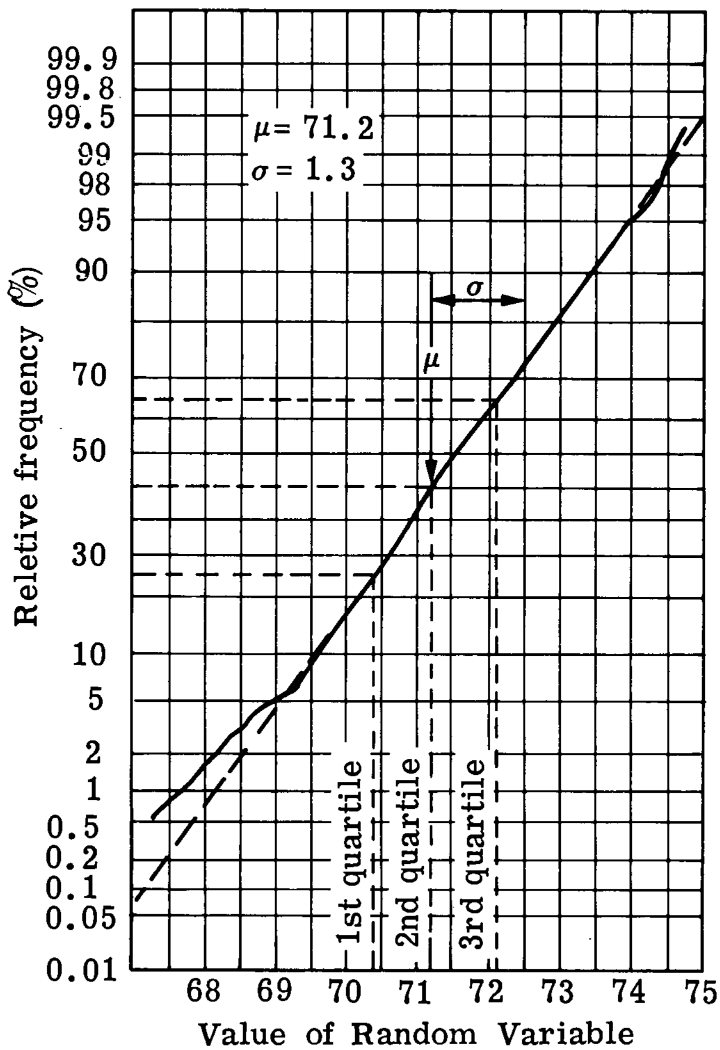


Fig. 6.4 Cumulative polygon on normal probability paper

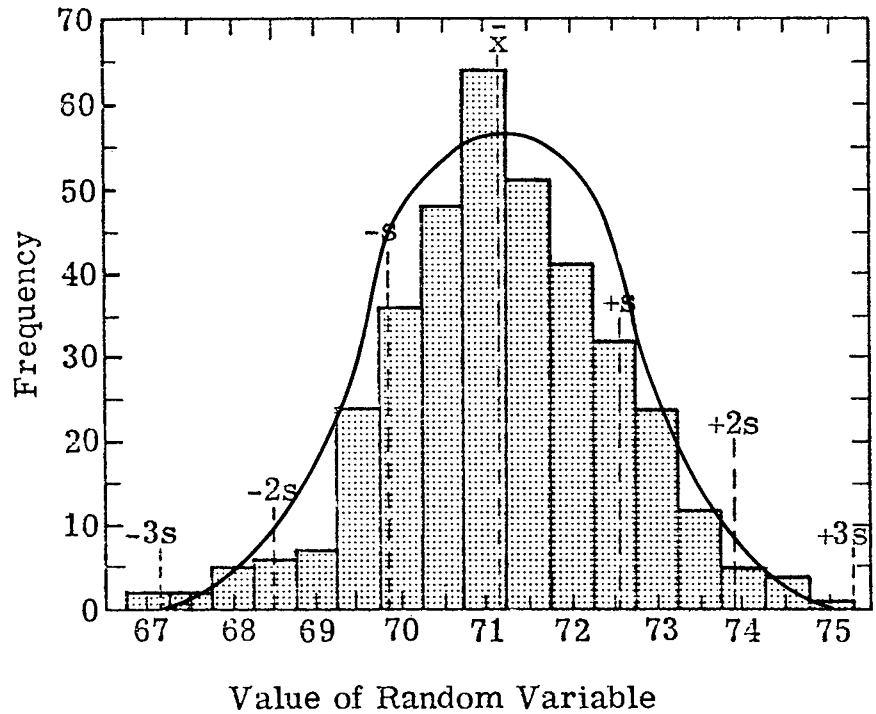


Fig. 6.5 Comparison of Histogram and Normal Distribution

7. ANALYTICAL CURVE FITTING

Analytical curve fitting encompasses a variety of techniques to smooth an empirical histogram for use. As discussed in Part I, the purpose of analytical curve fitting is to obtain a reasonable functional approximation of the empirical histogram to be used in a simulation.

For the purposes of Part II of this volume, analytical curve fitting will be restricted to the use of three families of probability distributions. These are the Weibull, Johnson, and Pearson distributions. The reader who is unfamiliar with these distributions may wish to refer to Appendix A to find a background discussion of these three distributions. The Weibull family is the easiest to work with and the Pearson family is the most difficult to work with. It is, therefore, recommended that analytical curve fitting be tried first with the Weibull, then if need be with the Johnson, and finally if necessary with the Pearson distributions.

The procedure for selecting one or more of these families is based on Table 7.1. The use of Table 7.1 is facilitated if qualitative information about the random processes and a sketch of the probability density are available. Once one or more families have been chosen, the selection procedure outlined in Section 4 should be followed.

Since using the Weibull, Johnson, or Pearson distribution is tantamount to using a smoothed histogram, some consideration should be given to using the histogram itself rather than a distribution. This is especially true if the histogram is drawn from an ample set of data, if the Weibull, Johnson, and Pearson curves do not give reasonably good fits, or if the histogram is multimodal. In the latter case the underlying population may actually be several distinct populations, and unless the user is prepared to separate that population by techniques not discussed here, using the histogram may be most expedient.

TABLE 7.1

Characteristics of Complex Probability Curves

Family Name	Number of Parameters	General Characteristics	Figures for Shapes of Densities
Weibull	3	Unimodal, finite left bound, tail to right	Figure 7.1
Johnson	4 (plus choice of three functions)	Bounded or unbounded, variety of shapes, mostly unimodal	Figures 7.2-7.3
Pearson	up to 4 (plus choice of twelve functions)	Great variety of curves	Figure 7.5

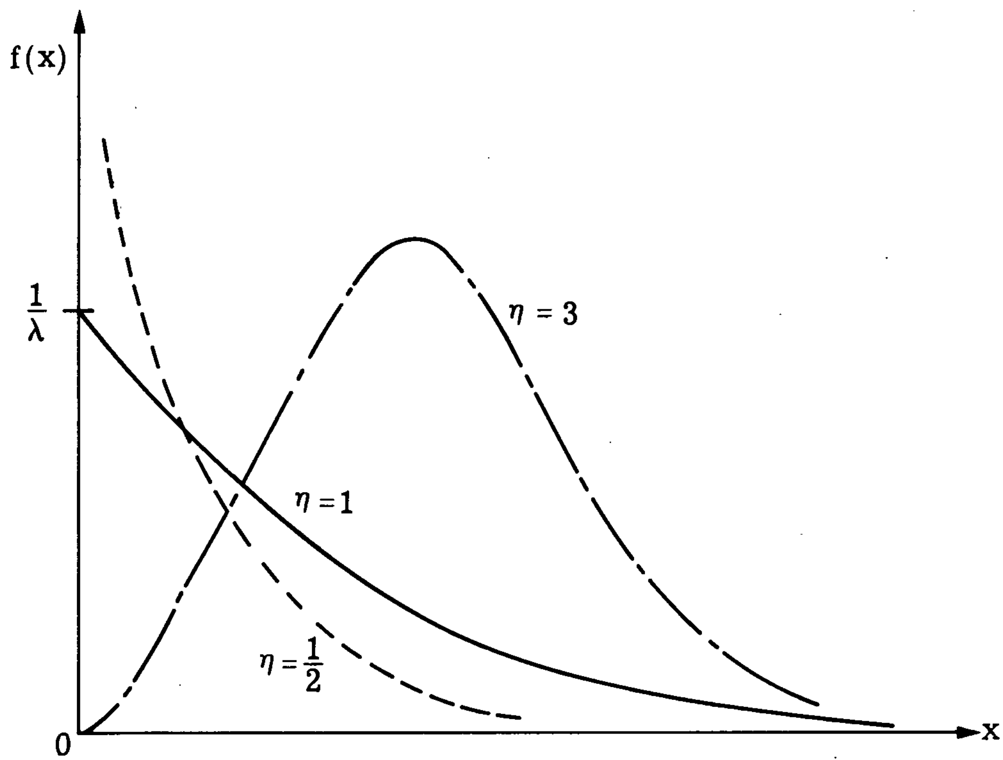


Fig. 7.1. Weibull Distribution for Various Values of Parameter η

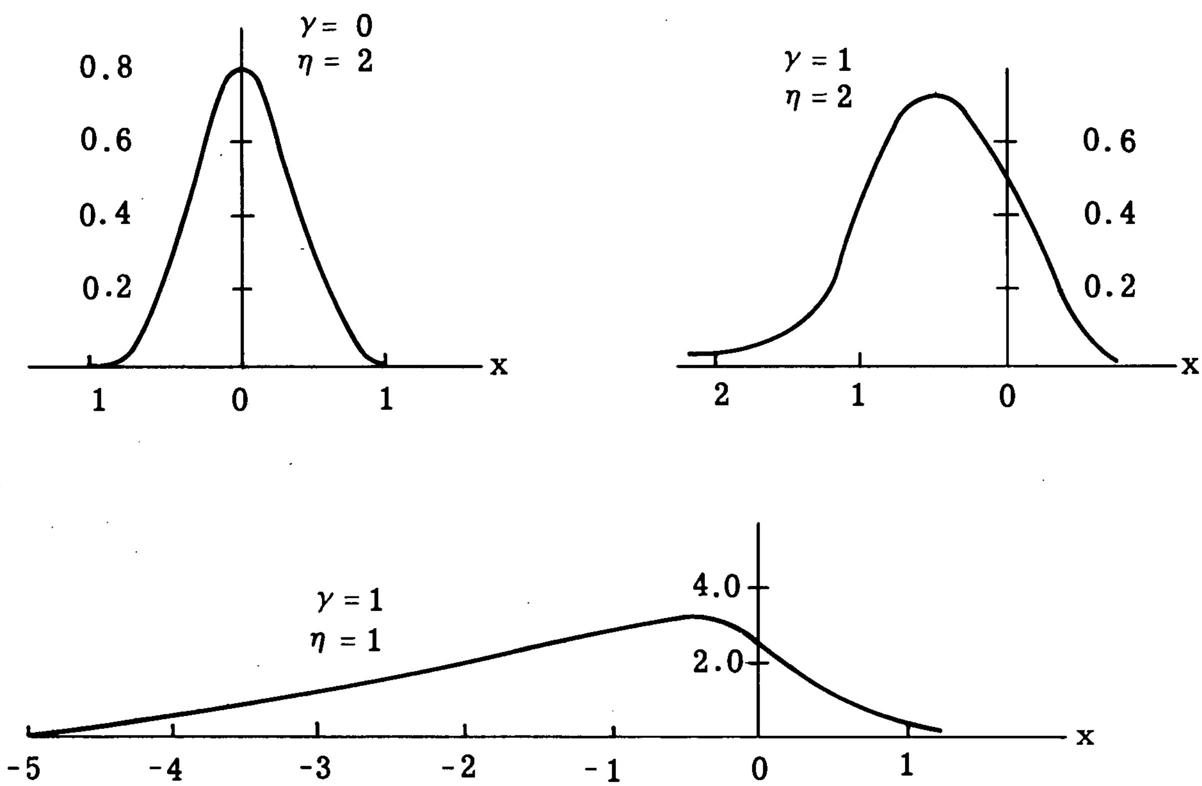


Fig. 7.2. Johnson Probability Density Functions for S_U

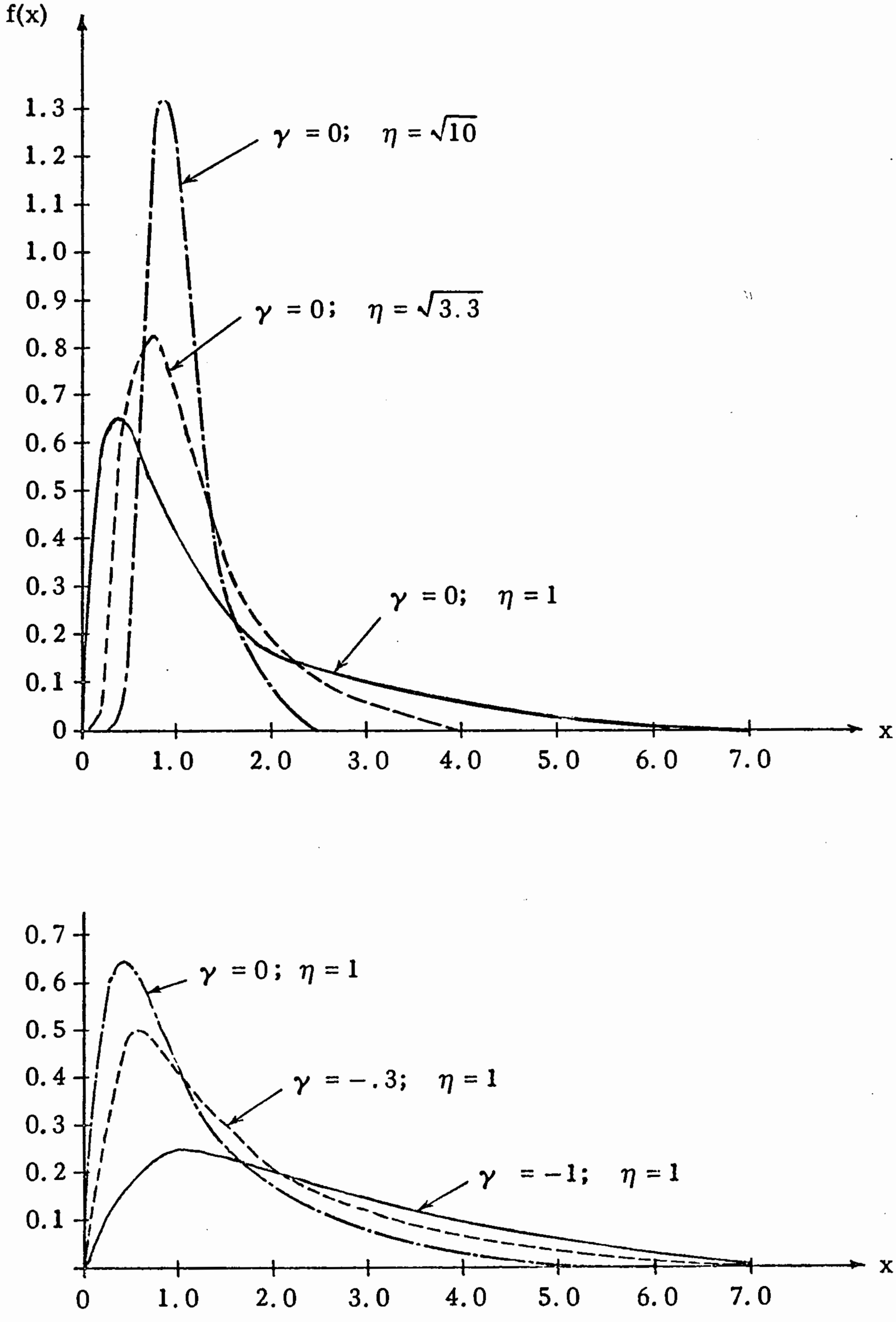


Fig. 7.3. Johnson Probability Density Functions for S_L ($\epsilon = 0$)

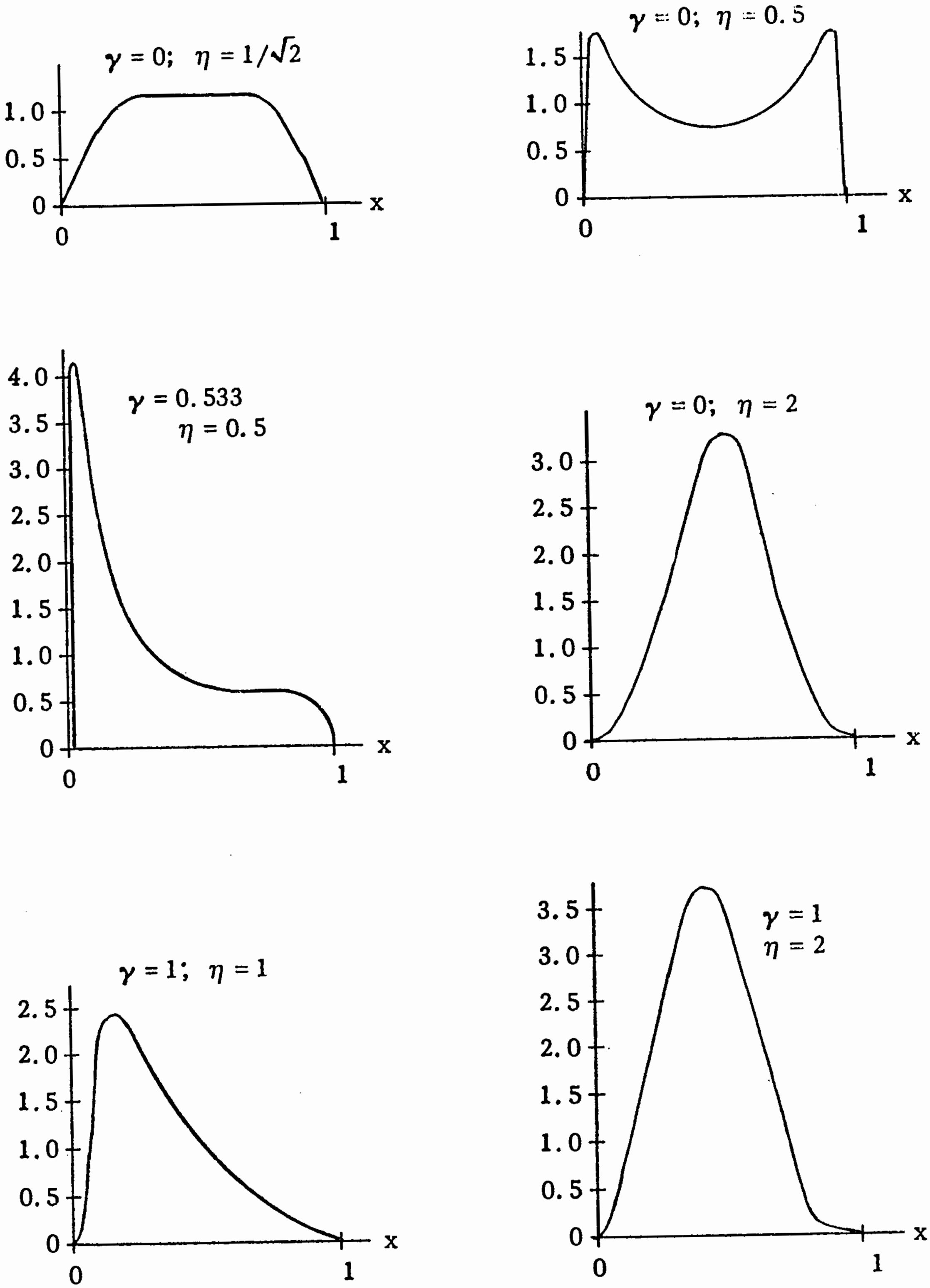
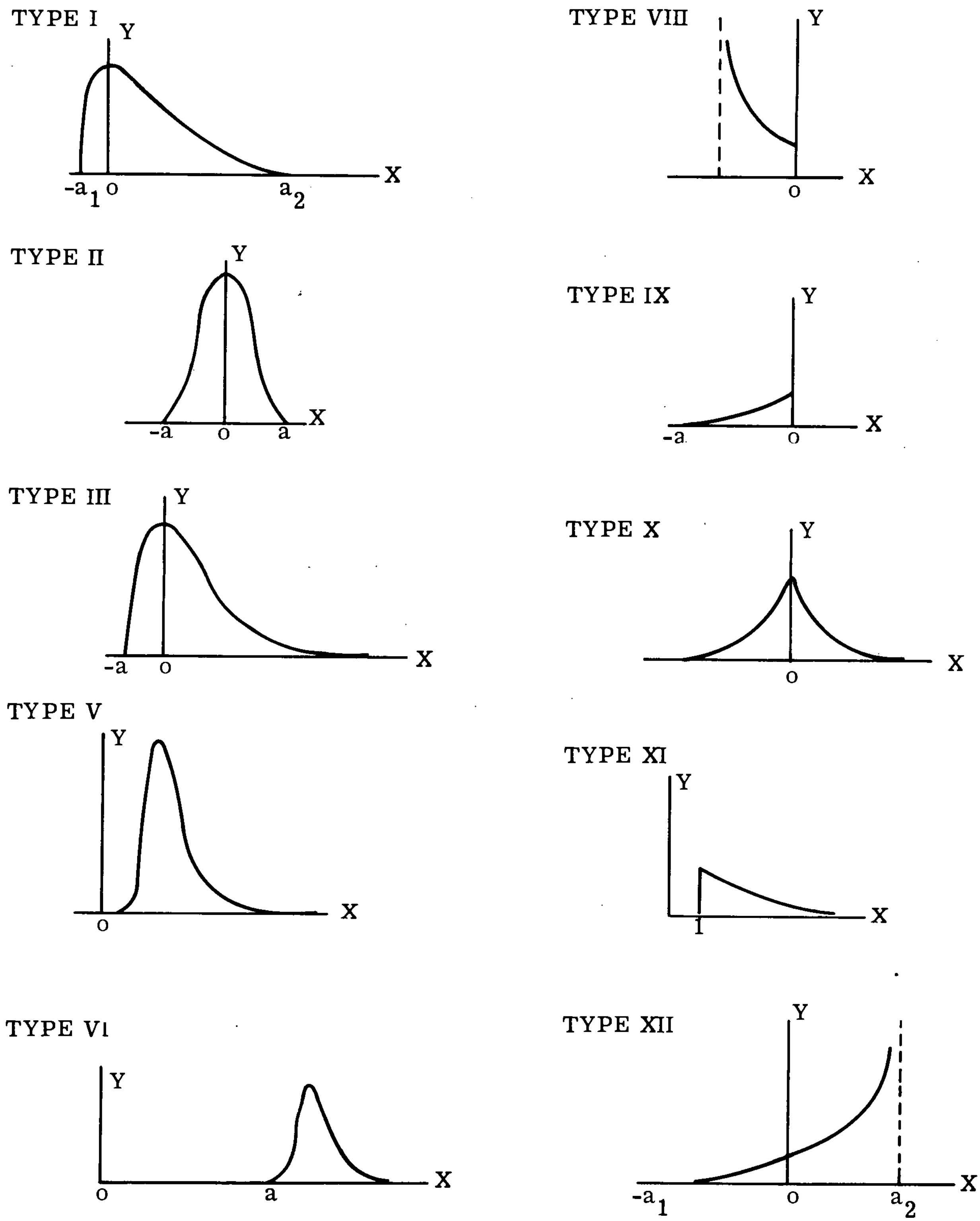


Fig. 7.4. Johnson Probability Density Functions for S_B ($\epsilon = 0$; $\lambda = 1$)



Note: Types IV and VII are similar to
Normal Distributions

Fig. 7.5. Typical Shapes of Pearson Distributions

8. PARAMETER ESTIMATION

Once a specific type from a family of probability distributions has been tentatively chosen to model a random variable, specific parameters for the distribution must be chosen. These parameters should be chosen so that the resulting specific distribution will best fit the data and knowledge available. This section is devoted to finding the specific parameter values based on the empirical data (observations) available.

If no data is available, the parameters must be chosen arbitrarily. In this case no estimation procedure exists that is better than the analyst's intuition and judgment. If data is available, the parameters can be estimated based on the sample of data. Estimates, in this case, always begin with calculation of certain sample statistics which are given in Section 8.1. This section should be used in conjunction with the directions given in Section 8.2. This latter section gives formulas for estimating the specific parameters for all of the distributions considered. Since not all the sample statistics in Section 8.1 are needed for all the distributions and parameters in Section 8.2, Section 8.2 should be referred to before calculating sample statistics.

8.1 CALCULATING SAMPLE STATISTICS

The sample statistics given in this section include the sample mean, median, variance, skewness, kurtosis, 3rd moment, and 4th moment. To establish some standard notation, we define the following symbols:

n = number of data points

x_i = i^{th} data point (observation) for $i = 1, 2, \ldots, n$.

The sample statistics are calculated as follows:

Sample Mean (symbol $\bar{x}$)

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n \quad .$$

Sample Median

First rank the observations from smallest to largest. If n is odd, the median is given by the value of the $[(n+1)/2]^{th}$ observation. If n is even the median is given by the mean of the $[n/2]^{th}$ and $[(n/2)+1]^{th}$ observations.

Sample Variance (symbol s^2)

$$s^2 = \left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]/n$$

or, more conveniently

$$= \left(\sum_{i=1}^{n} x_i^2\right)/n - \bar{x}^2 \quad .$$

Sample m^{th} Centralized Moment (symbol μ_m)(only μ_3 and μ_4 needed)

$$\mu_m = \left[\sum_{i=1}^{n} (x_i - \bar{x})^m\right]/n$$

Sample Skewness (symbol β_1)

$$\beta_1 = \mu_3/s^3$$

Sample Kurtosis (symbol β_2)

$$\beta_2 = \mu_4/s^4$$

Interpretation of the last two estimators is usually in terms of how well the data fits the normal distribution. If the skewness is close to zero and the kurtosis is close to three the normal distribution should provide a good approximation to the distribution. Figure 8.1 gives an interpretation of the skewness value. Zero indicates a symmetric distribution, negative skewness means a long left tail, positive values a long right tail. Figure 8.2 illustrates the kurtosis measure. If the kurtosis is greater than three the distribution is more peaked than the normal (curve C). If it is less than three the curve is flatter than the normal (curve A).

8.2 CALCULATING PARAMETER ESTIMATES

This section is divided into two parts. Section 8.2.1 deals with the simple distributions. This section will be the one more commonly used. Section 8.2.2 is more complicated and deals with estimating parameters for the complex distributions.

8.2.1 Simple Parametric Distributions

Refer to Table 4.3 to obtain the recommended parameter estimates for the selected distribution. Use Section 8.1 to obtain the sample statistics required.

8.2.2 Complex Parametric Distributions

As can be seen in Table 4.3, estimating parameters for the Weibull, Johnson, and Pearson distributions is more involved than for the simple distributions. The reason for this is that the simple distributions generally have one or two parameters, whereas the complex distributions have 3 to 5 effective parameters. Background for the material which follows can be found in Appendix A.

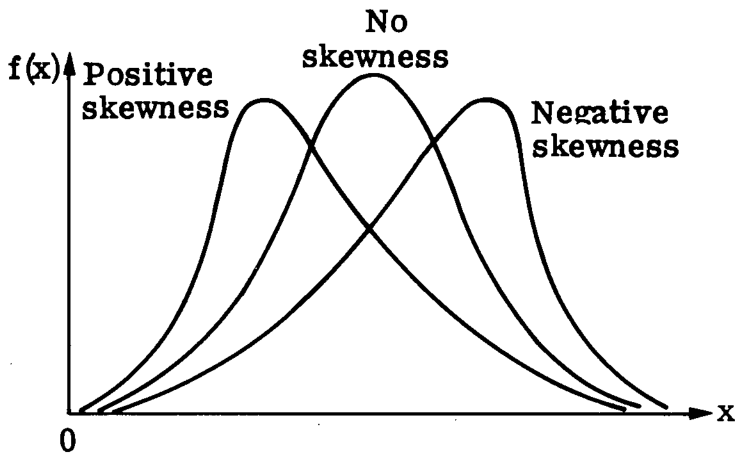


Fig. 8.1. Skewed distributions

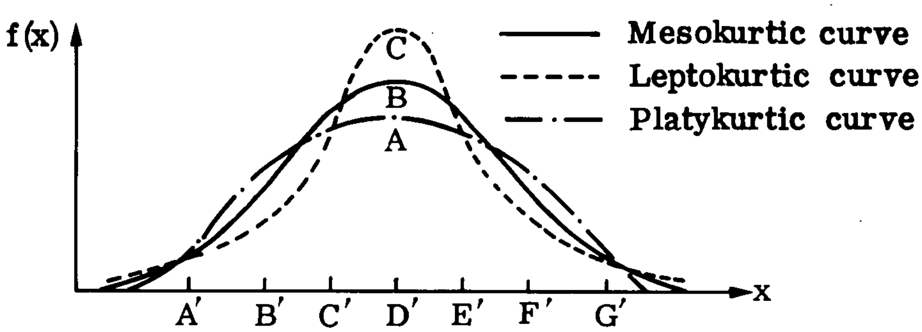


Fig. 8.2. Three frequency curves with different degrees of kurtosis

8.2.2.1 Weibull

The basic three-parameter Weibull distribution has a density given by:

$$f(x) = \frac{\eta}{\lambda}\left(\frac{x - \epsilon}{x}\right)^{\eta-1} \exp\left[-\left(\frac{x - \epsilon}{\lambda}\right)^{\eta}\right] , \qquad x \geq \epsilon$$

$$= 0 \qquad x \leq \epsilon$$

where:

$f(x)$ = Weibull probability distribution

ϵ = location parameter

λ = scale parameter

η = shape parameter

In most applications the location parameter, ϵ, is known. In cases where it is not known, it can be estimated from the observations:

$$\epsilon = \min[x_i] \ .$$

Better estimates of ϵ can be obtained using techniques developed by Dubey;[8] however, the improvement is not usually sufficient to warrant the extra effort involved.

The maximum likelihood estimators for the three-parameter Weibull distribution result in a set of equations that can be solved by iterative methods which are very tedious to perform. If the location parameter is known or estimated, the maximum likelihood equations for λ and η can be solved fairly easily[51] and are given by:

$$\frac{n}{\hat{\eta}} - n \frac{\sum_{i=1}^{n} x_i^{\eta} \ln x_i}{\sum_{i=1}^{n} x_i^{\hat{\eta}}} \sum_{i=1}^{n} \ln x_i = 0 \qquad (8.1)$$

and

$$\hat{\lambda} = (\Sigma x_i^{\hat{\eta}}/n)^{1/\hat{\eta}} \qquad (8.2)$$

where:

$\hat{\eta}$ = Maximum likelihood estimator of η

$\hat{\lambda}$ = Maximum likelihood estimator of λ

Equation 8.1 can be solved by the Newton-Raphson iterative procedure.

$$\hat{\eta}_{k+1} = \hat{\eta}_k + \frac{\dfrac{1}{\hat{\eta}_k} + \dfrac{S_1}{n} - \dfrac{S_3^k}{S_2^k}}{\dfrac{1}{\hat{\eta}_k^2} + \dfrac{S_2^k S_4^k - (S_3^k)^2}{(S_2^k)^2}}$$

where:

$$S_1 = \sum_{i=1}^{n} \ln x_i$$

$$S_2^k = \sum_{i=1}^{n} x_i^{\hat{\eta}_k}$$

$$S_3^k = \sum_{i=1}^{n} (\ln x_i) x_i^{\hat{\eta}_k}$$

$$S_4^k = \sum_{i=1}^{n} (\ln x_i)^2 x_i^{\hat{\eta}_k}$$

The estimate $\hat{\eta}$ is biased and should be corrected using the unbiasing factors in Table B-1 of Appendix B. Then, the estimate for $\hat{\lambda}$ can be obtained directly from (8.2). Further improvement can be obtained by using Menon's estimators.[38]

8.2.2.2 Johnson Distributions

As indicated in Table 4.3, there are three Johnson distributions. These three are generally denoted S_L, S_B, and S_U because these distributions are related to the normal distribution through a logarithmic transformation (S_L), bounded transformation (S_B), and unbounded transformation (S_U). The problem of estimating parameters of the Johnson distribution thus becomes a two-step procedure. First determine which distribution to use, then estimate the appropriate parameters.

The probability density functions for the three Johnson distributions are:

$$S_L: \quad f_1(x) = \frac{\eta}{\sqrt{2\pi}\,(x-\epsilon)} \exp\left\{-\frac{\eta^2}{2}\left[\frac{\gamma}{\eta} + \ln(x-\epsilon)\right]^2\right\} \quad ; \qquad x \geq \epsilon$$

$$S_B: \quad f_2(x) = \frac{\eta}{\sqrt{2\pi}} \frac{\lambda}{(x-\epsilon)(\lambda-x+\epsilon)} \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\epsilon}{\lambda-x+\epsilon}\right)\right]^2\right\}$$

$$\epsilon \leq x \leq \epsilon + \lambda$$

$$S_U: \quad f_3(x) = \frac{\eta}{\sqrt{2\pi}} \frac{1}{\sqrt{(x-\epsilon)^2 + \lambda^2}}$$

$$\exp\left[-\frac{1}{2}\left(\gamma + \eta \ln\left\{\left(\frac{x-\epsilon}{\lambda}\right) + \left[\frac{(x-\epsilon)^2}{\lambda^2} + 1\right]^{1/2}\right\}\right)^2\right]$$

$$-\infty < x < \infty$$

In these distributions η and γ are shape parameters, λ is a scale parameter, and ϵ is a location parameter. These must satisfy:

$$\eta > 0, \quad \lambda > 0, \quad -\infty < \gamma, \ \epsilon < +\infty$$

In Section 8.1, expressions are given for the skewness, β_1, and kurtosis, β_2, of the sample data. These are used to determine which distribution, S_L, S_B, or S_U to use. This can be accomplished by plotting the sample β_1 and β_2 on Fig. 8.3. The location of the sample point (β_1, β_2) indicates the distribution to select. One warning must be given. however. Figure 8.3 is accurate for categorizing distributions given the true value of β_1 and β_2. The values for β_1 and β_2 derived from the sample (Section 8.1) are estimates of the true values. Thus if the sample point falls near the edge of a region in Fig. 8.3, i.e., near the S_L line, then it would be prudent to try all three Johnson distributions or to select one or more based on possible boundedness of the random variable in question. Examining the density functions given above will aid in this determination.

The parameter estimates for the Johnson distributions are given below. The estimates of the Johnson parameters are not maximum likelihood estimates, except for the S_L (ϵ known) case, however they are the most practical to use. The approach taken is to use percentile points from the data. Recall that a $100\,\alpha$ percentile point for the population, x_α, is that value of x for which $P[x \le x_\alpha] = \alpha$. We assume that the random sample $x_1, \ldots, x_n$ has been ordered to give the order statistics $W_1 < \ldots < W_n$. Then the kth order statistic will provide an estimate for the 100α percentile of the population, where:

$$\alpha = \frac{k-1}{n-1} \tag{8.3}$$

This will be required in subsequent application, S_L (ϵ known). In this case the estimators for η and γ are respectively,

$$\hat{\eta} = \left\{ \frac{1}{n} \sum_{i=1}^{n} [\ln(x_i - \epsilon)]^2 - \left[\frac{1}{n} \sum_{i=1}^{n} \ln(x_i - \epsilon) \right]^2 \right\}^{-1/2}$$

and (8.4)

$$\hat{\gamma} = -\frac{\hat{\eta}}{n} \sum_{i=1}^{n} \ln(x_i - \epsilon)$$

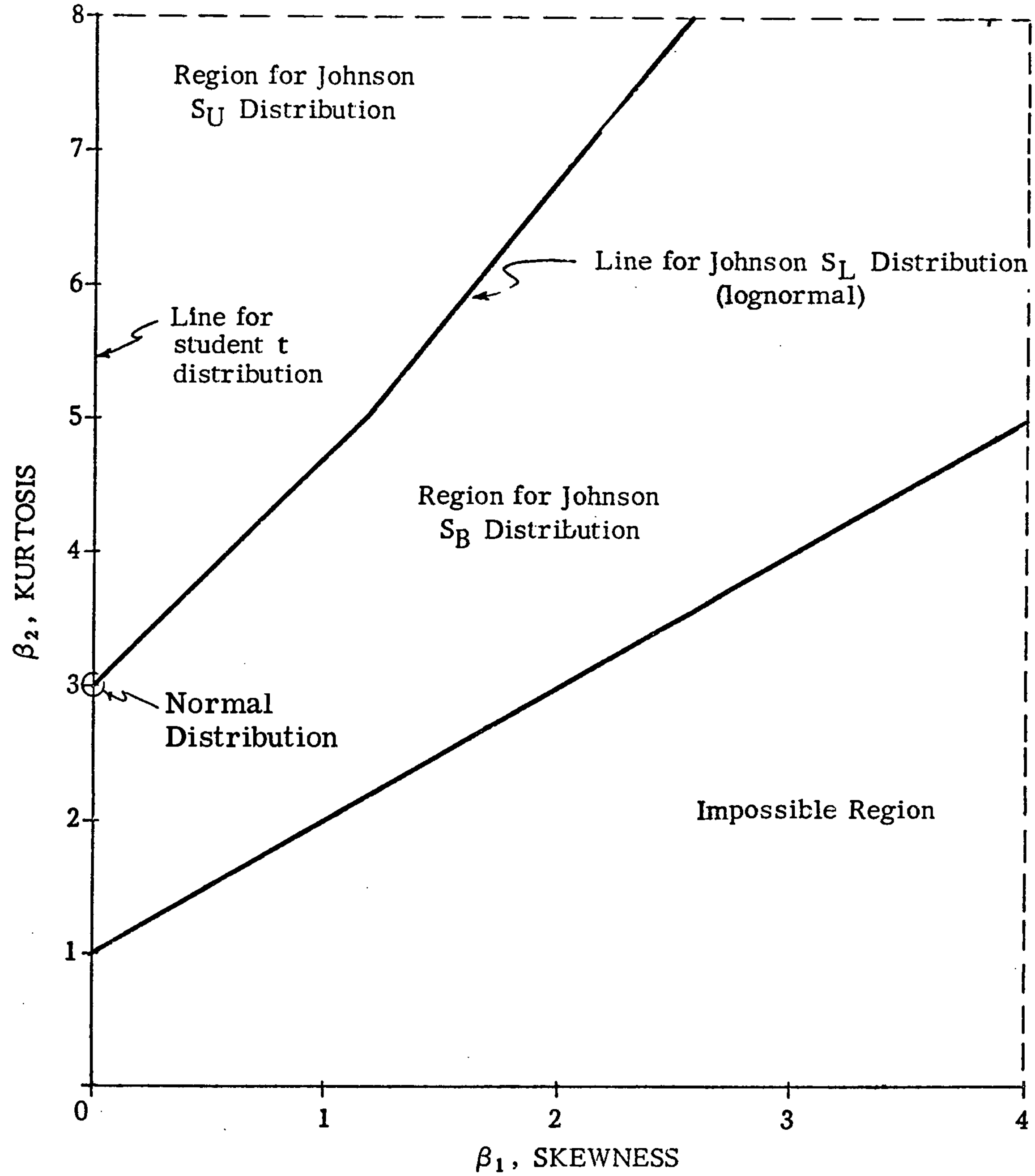


Fig. 8.3. Selecting a Johnson Distribution from Skewness and Kurtosis

Thus, from the sample $x_1, \ldots, x_n$ the parameters η and γ can be readily estimated with $\hat{\eta}$ and $\hat{\gamma}$, respectively.

S_L (ϵ unknown)

Again, the maximum likelihood estimators may be obtained but with some difficulty, and it is perhaps better to use the percentile approach. That is assume the percentile points x_{α_1}, x_{α_2}, and x_{α_3} have been estimated. These are required since there are three parameters ϵ, η, and γ to estimate. If z_α is defined as the value of the variable in the normal distribution function corresponding to the cumulative probability α, then.

$$z_{\alpha_1} = \gamma + \eta \ \ln(x_{\alpha_1} - \epsilon)$$

$$z_{\alpha_2} = \gamma + \eta \ \ln(x_{\alpha_2} - \epsilon)$$

$$z_{\alpha_3} = \gamma + \eta \ \ln(x_{\alpha_3} - \epsilon)$$

Explicit solutions cannot be obtained for ϵ, γ, and η from these equations although they can be determined iteratively. However, the following example will illustrate the use of one simplification.

Suppose a sample size of $n = 51$ has been obtained. The 6th, 26th, and 46th order statistic from $W_1 < W_2 < \ldots < W_{51}$ will be used to estimate the following percentiles:

$$x_{\alpha_1} = x_{.1} = W_6$$

$$x_{\alpha_2} = x_{.5} = W_{26}$$

$$x_{\alpha_3} = x_{.9} = W_{46}$$

where α_i $i = 1, 2, 3$, is obtained using Eq. 8.3. From Table B-2 in Appendix B:

$$z_{.1} = -1.28$$

$$z_{.5} = 0$$

$$z_{.9} = 1.28$$

From Eq. 8.4:

$$\hat{\eta} = 1.28 \left[\ell n \left(\frac{W_{46} - W_{26}}{W_{26} - W_6} \right) \right]^{-1}$$

$$\hat{\gamma} = \hat{\eta} \; \ell n \left[\frac{1 - e^{-1.28/\hat{\eta}}}{W_{26} - W_6} \right]$$

$$\hat{\epsilon} = W_{26} - e^{-\hat{\gamma}/\hat{\eta}}$$

The advantage of selecting $\alpha_1 = 1 - \alpha_3$ and $\alpha_2 = .5$ should be noted. The percentiles chosen are, of course, rather arbitrary and, therefore, many estimates could be obtained for ϵ, γ and η. In this case, comparisons of the relative goodness-of-fit for each selection may be appropriate.

$S_B(\epsilon, \lambda \text{ known})$

This case implies both end points of the distribution are known. Using the percentile method, estimators for γ and η may be obtained:

$$\hat{\eta} = \frac{z_{\alpha_2} - z_{\alpha_1}}{\ln \left[\frac{\left(x_{\alpha_2} - \epsilon \right)\left(\epsilon + \lambda - x_{\alpha_1} \right)}{\left(x_{\alpha_1} - \epsilon \right)\left(\epsilon + \lambda - x_{\alpha_2} \right)} \right]} \tag{8.5}$$

$$\hat{\gamma} = z_{\alpha_2} - \hat{\eta} \ln \left[\frac{x_{\alpha_2} - \epsilon}{\epsilon + \lambda - x_{\alpha_2}} \right]$$

S_B (General Case)

This case implies that none of the parameters are known and requires that the appropriate number of equations of the form

$$z_\alpha = \gamma + \eta \ln\left[\frac{x_\alpha - \epsilon}{\lambda + \epsilon - x_\alpha}\right]$$

be solved for the unknown parameters. Generally, this will lead to transcendental equations which can be solved numerically. There is one simplification in the case where ϵ is known and the percentiles are selected such that $\alpha = \alpha_1 = 1 - \alpha_3$ and $\alpha_2 = .5$ (only three equations of the type required for this case). The solution for $\hat{\lambda}$ for this case is

$$\hat{\lambda} = (x_{.5} - \epsilon)\left[\frac{(x_{.5} - \epsilon)(x_\alpha - \epsilon) + (x_{.5} - \epsilon)(x_{1-\alpha} - \epsilon) - 2(x_\alpha - \epsilon)(x_{1-\alpha} - \epsilon)}{(x_{.5} - \epsilon)^2 - (x_\alpha - \epsilon)(x_{1-\alpha} - \epsilon)}\right] \quad (8.6)$$

Equation 8.5 may then be used to generate estimates for η and γ since with 8.6 the problem reduces to one with both end points known.

S_U (General Case)

For general case of the S_U system, Johnson has generated tables that are useful for determining the parameters.[22] These are presented in Tables B-3 and B-4 of Appendix B. The tables were developed from solutions of equations defining the relationships of the first four moments to the parameters.

Use of the tables first requires that the mean, variance, skewness and kurtosis be calculated. The values for $\sqrt{\hat{\beta}_1}$ and $\hat{\beta}_2$ are then used to obtain the estimates for γ and η from Tables B-3 and B-4, respectively.

The $\hat{\lambda}$ and $\hat{\epsilon}$ estimators are calculated using:

$$\hat{\lambda} = \frac{s}{\left\{\frac{1}{2}(\omega-1)\left[\omega \cosh\left(\frac{2\hat{\gamma}}{\hat{\eta}}\right)+1\right]\right\}^{1/2}} \quad (8.7)$$

$$\hat{\epsilon} = x + \hat{\lambda}\sqrt{\omega}\,\sinh\left(\frac{\hat{\gamma}}{\hat{\eta}}\right) \quad (8.8)$$

where s is the sample standard deviation (see Section 8.1)

$$\omega = e^{1/\hat{\eta}^2} .$$

To illustrate use of the tables, assume a random sample gave $\sqrt{\hat{\beta}_1} = .5$ and $\hat{\beta}_2 = 6$. From Tables B-3, B-4

$$\hat{\gamma} = -.3278 \quad \text{and} \quad \hat{\eta} = 1.672$$

$\hat{\lambda}$ and $\hat{\epsilon}$ may now be calculated directly from Eqs. 8.7 and 8.8.

8.2.2.3 Pearson Distributions

There are twelve Pearson distributions. These are generally indicated by Roman numerals: Type I through Type XII. The problem of estimating Pearson parameters, like those of the Johnson, becomes a two-step problem. First determine which Pearson Type to use, then estimate the appropriate parameters.

To determine which Pearson distributions to use, the skewness, β_1, and kurtosis, β_2, of the sample data (see Section 8.1) are needed. The sample point $(\hat{\beta}_1, \hat{\beta}_2)$ should be plotted on Fig. 8.4. The location of the sample point indicates what distribution to use. A warning needs to be given on using this procedure. The point $(\hat{\beta}_1, \hat{\beta}_2)$ calculated from the data as in

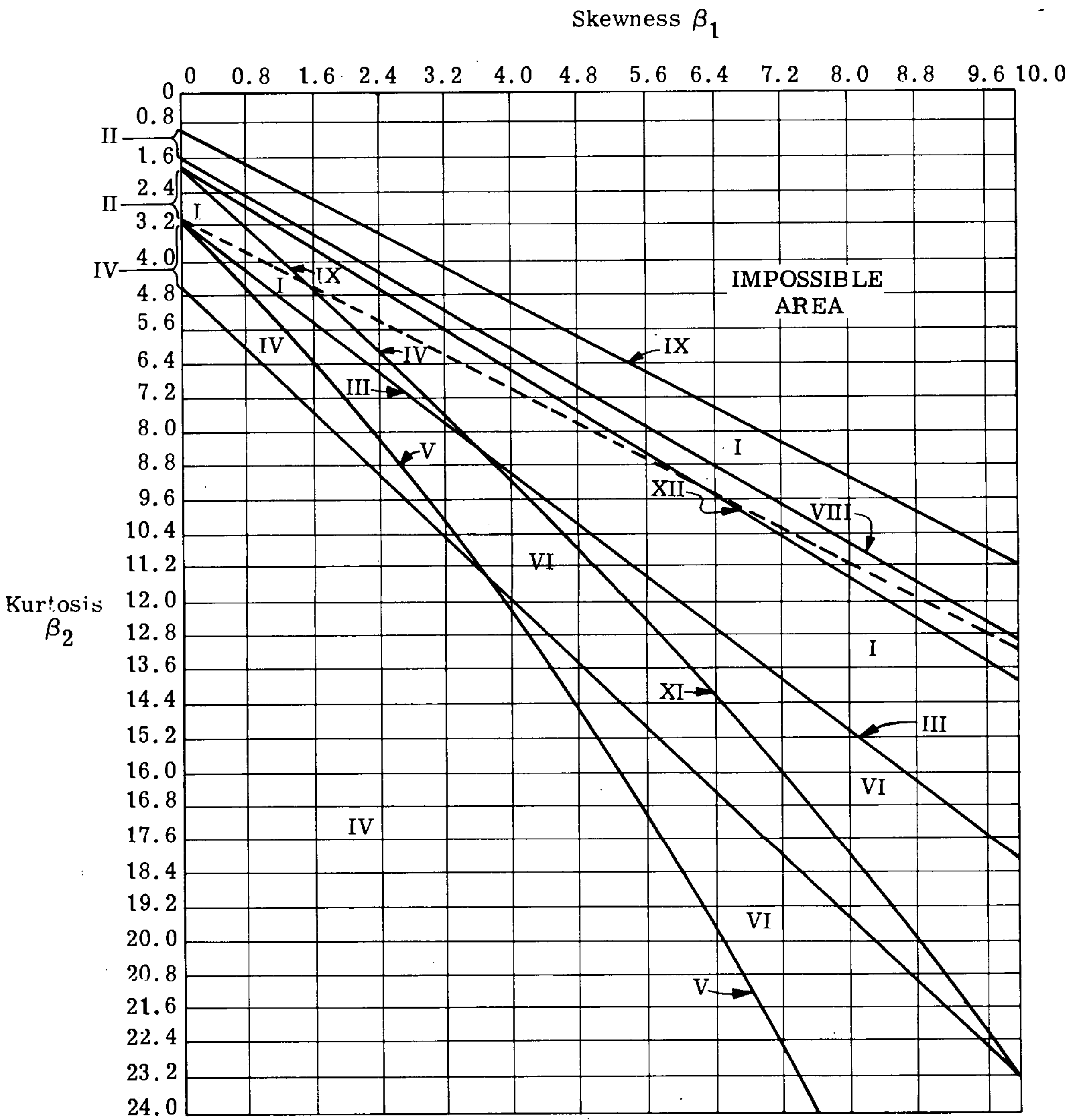


Fig. 8.4. Selection of Pearson Type from Skewness and Kurtosis

Section 8.1 is only an estimate of the true values. Thus if the sample point falls near a line separating two regions in Fig. 8.4, the Pearson Type in either region or in the line may fit the data. In this event more than one Pearson Type should be tried. It should also be clear from examination of Fig. 8.4 that only Types I, IV, and VI are indicated by regions; therefore, in practice only these types will be indicated by strict application of this selection procedure.

Selection of a Pearson Type may also be aided by examining the Remarks column of Table 8.1. This table lists all twelve Pearson Types and some information on each. The form of the density function should be obtained from Table 8.1.

The parameters for the density functions are given below.

Type I

$$f(x) = y_o \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2} \quad \text{where} \left(\frac{m_1}{a_1} = \frac{m_2}{a_2}\right) .$$

Calculate the quantities

$$r = 6(\beta_2 - \beta_1 - 1)/(6 + 3\beta_1 - 2\beta_2)$$

$$t = \frac{1}{2} s[\beta_1(r + 2)^2 + 16(r + 1)]^{1/2}$$

m_1 and m_2 are given by:

$$m = \frac{1}{2}\left\{r - 2 \pm r(r + 2)\left[\frac{\beta_1}{\beta_1(r + 2)^2 + 16(r + 1)}\right]^{\frac{1}{2}}\right\} .$$

If μ_3 is positive, take m_2 to be the positive root

$$a_1 = t/(m_2/m_1 + 1)$$

TABLE 8.1
Summary of Pearson Distributions

No. of type usually adopted (Pearson)	Equation to curve in form usually adopted (Pearson): Equation	Origin	Remarks
MAIN TYPES			
I	$y = y_o(1 + x/a_1)^{\nu a_1} (1 - x/a_2)^{\nu a_2}$	Mode (antimode)	Limited range ($-a_1$ to a_2); skew; usually bell-shaped, but may be U-shaped, J-shaped or twisted J-shaped
IV	$y = y_o(1 + x^2/a^2)^{-m} e^{-\nu \tan^{-1} x/a}$	$\nu a/(2m-2)$ after mean	Unlimited range; skew; bell-shaped
VI	$y = y_o(x - a)^{q_2} x^{q_1}$	a before start of curve	Unlimited range in one direction (a to ∞); skew; bell-shaped, but may be J-shaped
TRANSITION TYPES			
'Normal' curve	$y = y_o e^{-x^2/2\sigma^2}$	Mode (= mean)	Unlimited range; symmetrical; bell-shaped
II	$y = y_o(1-x^2/a^2)^m$	Mode (= mean)	Limited range (-a to a); symmetrical; usually bell-shaped, but U-shaped when $\beta_2 < \cdot 8$
VII	$y = y_o(1+x^2/a^2)^{-m}$	Mode (= mean)	Unlimited range; symmetrical; bell-shaped
III	$y = y_o(1+x/a)^{\gamma a} e^{-\gamma x}$	Mode	Unlimited range in one direction (-a to ∞); usually bell-shaped, but may be J-shaped
V	$y = y_o x^{-p} e^{-\gamma/x}$	Start of curve	Unlimited range in one direction (0 to ∞); bell-shaped
VIII	$y = y_o(1+x/a)^{-m}$	End of curve	Range from infinite ordinate at -a to finite ordinate at 0 (or from $-a(1-m)/(2-m)$ to $a/(2-m)$ with origin at mean)
IX	$y = y_o(1+x/a)^m$	End of curve	Range from x = -a where y = 0 to x = 0 where $y = y_o$ (or from $-a/m+1)/(m+2)$ $a/(m+2)$ with origin at mean)
X	$y = y_o e^{-x/\sigma}$	Start of curve	Exponential from finite ordinate at 0 (or $-\sigma$ with origin at mean) to infinitesimal ordinate at ∞; J-shaped
XI	$y = y_o x^{-m}$	b before start	J-shaped; starts at x=b (or $-b/(m-2)$ with origin at mean) where ordinate is finite
XII	$y = y_o\left(\frac{\sigma\{\sqrt{(3+\beta_1)}+\sqrt{\beta_1}\}+x}{\sigma\{\sqrt{(3+\beta_1)}-\sqrt{\beta_1}\}-x}\right)^{\sqrt{(\beta_1/(3+\beta_1))}}$	Mean	Twisted J-shaped; special case of Type I

$$a_2 = t/(m_1/m_2 + 1)$$

$$y_o = \frac{m_1^{\ m_1} m_2^{\ m_2}}{(m_1 + m_2)^{m_1 + m_2}} \cdot \frac{\Gamma(m_1 + m_2 + 2)}{\Gamma(m_1 + 1)\Gamma(m_2 + 1)}$$

Type II

$$f(x) = y_o \left(1 - \frac{x^2}{a^2}\right)^m$$

The function parameters are found as follows:

$$m = \frac{5\beta_2 - 9}{2(3 - \beta_2)}$$

$$a^2 = \frac{2s^2 \beta_2}{3 - \beta_2}$$

$$y_o = \frac{1}{a\sqrt{\pi}} \cdot \frac{\Gamma(m + 1.5)}{\Gamma(m + 1)}$$

Type III

$$f(x) = y_o (1 + x/a)^{\gamma a} e^{-\gamma x}$$

The function parameters are given by:

$$\gamma = \frac{2s^2}{\mu_3}$$

$$p = \gamma a = \frac{4}{\beta_1} - 1$$

$$a = \frac{2s^4}{\mu_3} - \frac{\mu_3}{s^2}$$

$$y_o = \frac{1}{a} \cdot \frac{p^{p+1}}{e^p \Gamma(p+1)}$$

Type IV

$$f(x) = y_o \left(1 + \frac{x^2}{a^2}\right)^{-m} \exp(-\nu \tan^{-1} x/a)$$

The function parameters are given by:

$$\gamma = 6(\beta_2 - \beta_1 - 1)/(2\beta_2 - 3\beta_1 - 6)$$

$$m = \frac{1}{2}(\gamma + 2)$$

$$\nu = -\gamma\ (\gamma - 2)\sqrt{\beta_1}\ [16(\gamma - 1) - \beta_1(\gamma - 2)^2]^{-1/2}$$

$$a = \left[\frac{s^2}{16}\ (16(\gamma - 1) - \beta_1(\gamma - 2)^2)\right]^{1/2}$$

$$y_o = 1/[aF(\gamma, \nu)]$$

where $F(\gamma, \nu)$ is given in Reference 42.

Type V

$$f(x) = y_o x^{-p} \exp(-\gamma/x) \qquad (x > 0)$$

The function parameters are given by:

$$p = 4 + \frac{8 + 4\sqrt{4 + \beta_1}}{\beta_1}$$

$$\gamma = s(p - 2)\sqrt{(p-3)}$$

$$y_o = \gamma^{p-1}/\Gamma(p - 1)$$

Type VI

$$f(x) = y_o(x - a)^{q_2} x^{q_1}$$

The function parameters are given by:

$$\gamma = 6(\beta_2 - \beta_1 - 1)/(6 + 3\beta_1 - 2\beta_2)$$

$$a = \frac{1}{2} [\beta_2 (\beta_1(\gamma + 2)^2 + 16(\gamma + 1))]^{1/2}$$

q_2 and $-q_1$ are given by:

$$q = \frac{\gamma - 2}{2} \pm \frac{\gamma(\gamma+2)}{2} [\beta_1/[\beta_1(\gamma + 2)^2 + 16(\gamma + 1)]]^{1/2}$$

$$y_o = \frac{a^{q_1 - q_2 - 1} \Gamma(q_1)}{\Gamma(q_1 - q_2 - 1) \Gamma(q_2 + 1)}$$

Type VII

$$f(x) = y_o \left(1 + \frac{x^2}{a^2}\right)^{-m}$$

The function parameters are given by:

$$m = \frac{5\beta_2 - 9}{2(\beta_2 - 3)}$$

$$a^2 = \frac{2s^2 \beta_2}{\beta_2 - 3}$$

$$y_o = \frac{1}{a\sqrt{\pi}} \frac{\Gamma(m)}{\Gamma[m - 0.5]}$$

Type VIII

$$f(x) = y_o\ (1 + x/a)^{-m}$$

The function parameters are given by:

$$a = \pm s\ (2 - m)\ \sqrt{(3 - m)/(1 - m)}$$

$$y_o = (1 - m)/a$$

where m is the solution of

$$m^3(4 - \beta_1) + m^2(9\beta_1 - 12) - 24\beta_1 m + 16\beta_1 = 0 \qquad 0 < m < 1$$

Type IX

$$f(x) = y_o(1 + x/a)^m$$

The function parameters are given by:

$$a = \pm s(m + 2)\ \sqrt{(m + 3)(m + 1)}$$

$$y_o = (m + 1)/a$$

where m is the solution of

$$m^3(\beta_1 - 4) + m^2(9\beta_1 - 12) + 24m\beta_1 + 16\beta_1 = 0 \qquad m > 0$$

Type X

$$f(x) = y_o\ \exp(-x/s)$$

The parameter is given by:

$$y_o = s$$

Type XI

$$f(x) = y_o x^{-m}$$

The function parameters are given by:

$$y_o = b^{m-1} (m - 1)$$

$$b = \pm s(m - 2) \sqrt{(m = 3)/(m - 1)}$$

where m is a solution of

$$m^3(4 - \beta_1) + m^2(9\beta_1 - 12) - 24\beta_1 m + 16\beta_1 = 0$$

Type XII

$$f(x) = y_o \left[\frac{s\,(\sqrt{3 + \beta_1} + \sqrt{\beta_1}) + x}{s\,(\sqrt{3 + \beta_1} - \sqrt{\beta_1}) - x}\right]^{\sqrt{\beta_1/(3+\beta_1)}}$$

y_o is given by

$$y_o = \Gamma(m + 1)\Gamma(1 - m)/b$$

where

$$m = \sqrt{\beta_1/(3 + \beta_1)}$$

$$b = 2s \sqrt{(3 + \beta_1)}$$

9. GOODNESS-OF-FIT TESTS

Goodness-of-fit tests are statistical tests for evaluating whether a group of data supports the assumption that the random variable from which the data are drawn has come from the assumed probability distribution. These tests are helpful in accepting or rejecting the conclusion that some random variable follows a tentatively selected probability distribution.

The technique of applying statistical tests of distributional assumptions follows three basic steps:

1. A number known as a test statistic is calculated from the observed data.

2. The probability of obtaining the calculated test statistic, assuming the selected distribution is correct, is determined. This can often be done by using precomputed tables of percentiles of the distribution of the test statistic.

3. If the probability of obtaining the calculated test statistic is low, the conclusion is that the assumed distribution does not provide an adequate representation. If the probability associated with the test statistic is not low, then the data provide no evidence that the assumed distribution is inadequate.

It should be clearly understood that although this procedure allows rejection of a distribution as inadequate, it never proves that the model is correct. In fact, the outcome of a statistical test depends highly upon the amount of data available - the more data there are, the better are the chances of rejecting an inadequate model. If too few data points are available, even a model that deviates grossly from the assumed model frequently cannot be established as inadequate.

Table 9.1 provides summary information on goodness-of-fit tests and also indicates on which distributions the tests are applicable. After a test is selected from this table, instructions on how to perform the test can be found in the subsequent sections.

A comment on using goodness-of-fit tests on the complex distributions (Weibull, Johnson, and Pearson) may also be helpful. These distributions are designed to fit almost any set of data well. It is, therefore, unlikely that any of them will be rejected by a goodness-of-fit test. Using goodness of fit tests on any of these distributions will not generally give the analyst much further information on the form of the true distribution, and he may elect to accept one of these complex distributions without a goodness-of-fit test.

This brief background should suffice for practical use of goodness-of-fit test in simulation modeling. In the following section, a simple selection procedure is given to determine what goodness-of-fit test to use based upon the probability distribution tentatively selected to model the random variable in question. In the following sections these tests are described and instructions for performing the tests are given. Although there are numerous statistical tests, these are the most powerful available.

9.1 CHI-SQUARE GOODNESS-OF-FIT TEST

The Chi-square goodness-of-fit test is probably the most widely used and versatile technique for evaluating distributional assumptions. It can be applied to test any distributional assumption without having to know the values of the distribution parameters. Its major drawbacks are its lack of sensitivity in detecting inadequate models when few observations are available and the frequent need to arrange the data within arbitrary cells which can affect the outcome of the test.

TABLE 9.1
Goodness-of-Fit Tests

PROBABILITY DISTRIBUTION	APPROPRIATE TEST (SUBSECTION)	APPLICABILITY	EASE OF USE	TABLES OR DATA REQUIRED (APPENDIX B)	GENERAL COMMENTS	REFERENCES (SEE APPENDIX C)
Any	Chi-square test (9.1)	Test to evaluate a sample for any distributional assumption for any type of distribution. A non-parametric or distribution free test.	Test is relatively easy to apply. Requires placing the sample values into intervals and some minor computations.	Chi-square distribution table (1-2 pages).	A general and powerful statistical test that is widely used. However, it is not a good test for small samples.	4, 15, 20
Any continuous	"d" - test Kolmogorov-Smirnov test Kolmogorov test (9.2)	Test to evaluate the agreement between the distribution of a set of sample values and any completely specified continuous distribution. A non-parametric or distribution free test.	Test is easy to apply. However requires ordering of data which may be tedious for large samples.	d - distribution table (1-2 pages).	A powerful statistical test for continuous theoretical distributions. It is a good test for small samples and where it is applicable it is usually a better test than the Chi-square.	4, 33
Normal Log Normal Johnson (see Table 4.4)	"W" - test (9.3)	Test to evaluate the assumption that a sample comes from a normal or log-normal distribution.	Relatively easy to use.	Tables (2-3 pages) used with equations are required.	A test more powerful than the χ^2 for testing the normal distribution assumption.	15, 47
Exponential (origin unknown)	"WE" - test (9.4)	Test to evaluate the assumption that a sample comes from an exponential distribution with origin unknown.	Relatively easy to use.	Requires tables (2-3 pages) used with equations.	A test more powerful than the χ^2 for testing the exponential distribution assumption.	15, 47
Exponential (origin known)	"WE_o" - test (9.5)	Test to evaluate the assumption that a sample comes from an exponential distribution with origin known.	Relatively easy to use.	Requires tables (2-3 pages) used with with equations.	A test more powerful than the χ^2 for testing the exponential distribution assumption	15, 47

The Chi-square test is used as follows:

Step 1. Estimate each of the unknown parameters of the assumed distribution.

Step 2. Divide the data into k classes or cells and determine the probability of a random value from the assumed model falling within each class. Two methods for this are presented: the first method is applicable if the data are initially arranged in frequency classes, and the second applies when the data are not initially tabulated in classes.

Method a. The number of cells, k, will be the number of classes of the tabulated data subject to the restriction that the expected number of observations in each cell under the assumed model is at least 5. Let CL_i and CU_i denote the lower and upper bounds of the i^{th} frequency cell. The distribution of the assumed model (using the estimated parameters) is then used to estimate:

$$Pr(CL_i \leq x < CU_i) \quad , \quad i = 1,2,\ldots k \quad .$$

Method b. When the number of observations, n, is large (>200) a good rule is to take k as the integer closest to

$$k' = 4[0.75(n-1)^2]^{1/5} \quad .$$

For moderate values of n a good rule is to make k as large as possible but with the restriction $k < n/5$. The cell boundaries $x_1, x_2, \ldots x_3$ are determined from the cumulative distribution for the assumed model (using the estimated parameters) as the values such that:

$$Pr(x \leq x_1) = \frac{1}{k}, Pr(x \leq x_2) = \frac{2}{k}, \ldots \; Pr(x \leq x_{k-1}) = \frac{(k-1)}{k} \quad .$$

Step 3. Multiply each of the cell probabilities by the sample size n. This yields the expected number E_i of observations for each cell under the assumed model. For Method 2b:

$$E_i = n/k \quad , \quad i = 1,2,\ldots k \quad .$$

Step 4. If the data are not initially tabulated, count the number of observed values, m_i, in each cell. Otherwise, determine m_i directly.

Step 5. Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - E_i)^2}{E_i} .$$

For Method 2b this simplifies to

$$\chi^2 = \frac{k}{n}\left(\sum_{i=1}^{k} m_i^2\right) - n .$$

Step 6. Compare the computed value χ^2 with the tabulated percentiles for a chi-square variate (Table B-5) using k-r-1 degrees of freedom, where r is the number of parameters that were estimated in Step 1. High values of χ^2 signify that the observed data contradicts the assumed model. For example, if the above calculated value χ^2 exceeds the 0.95 tabulated value of Chi-square, the chances are less than one in twenty that the data could have come from the assumed distribution.

9.2 KOLMOGOROV-SMIRNOV TEST

This test is used to evalute the assumption that a sample belongs to a specified known continuous distribution. It is a distribution-free test and is a good test for small samples. In general, it is a more powerful test than the Chi-square where it is applicable. Although the test is designed for comparing a sample against a specified and known distribution, the test is robust enough that it may still be applied to distributions whose parameters are estimated from the sample data. The effect of estimating the parameters of the distribution from the sample is to reduce the critical level of the $d_\alpha(N)$ statistic, i.e., the level of significance is really higher than the α associated with the chosen $d_\alpha(N)$. Hence, if the chosen $d_\alpha(N)$ statistic value is exceeded in the test, it can be safely concluded that the discrepancy is significant. Grouping observations into intervals also tends to lower the value of d. For grouped data, therefore, the appropriate significance levels for testing should be chosen smaller than the significance levels used for a complete sample.

The test is used as follows:

Step 1. Rearrange the sample of size n to obtain the ordered sample $x_1, x_2, \ldots x_n$ where $x_1 \le x_2 \le \ldots \le x_n$.

Step 2. The sample cumulative distribution then takes on values of $1/n, 2/n, \ldots, n/n$ at the points $x_i, \ldots, x_n$.

Step 3. Calculate the cumulative frequency values for the assumed distribution at the sample values of $x_1, x_2, \ldots x_n$.

Step 4. Determine the maximum deviation, d, between the sample cumulative distribution and assumed cumulative distribution from Steps 2 and 3.

Step 5. Compare the calculated deviation d with the test statistic $d_\alpha(n)$ found from Table B-6 for the desired level of significance. If d exceeds the value $d_\alpha(n)$ then the assumption that the sample comes from the assumed distribution may be rejected at the $100\alpha\%$ significance level.

9.3 W-TEST

This test is used to evalute the assumption that a sample has a normal distribution. It can be used to test the assumption that a sample fits log-normal distribution by using the log of the sample values. The W-test has been shown to be an effective technique for evaluating the assumption of normality against a wide spectrum of non-normal alternatives, even if only a relatively small number of observations are available. It is generally more powerful than the χ^2, especially for small sample sizes.

The W-test is used as follows:

Step 1. Rearrange the sample to obtain the ordered sample $x_1, x_2, \ldots, x_n$, where $x_1 \le x_2 \le \ldots x_n$.

Step 2. Compute

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \quad ,$$

where $\bar{x}$ is the data mean.

Step 3. If n is even, set $k = n/2$; if n is odd, set $k = (n-1)/2$. Then compute.

$$b = \sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i) \quad ,$$

where the values of a_{n-1+1} for $i = 1, \ldots . k$ are given in Table B-7 for $n = 3, \ldots , 50$. Note that when n is odd x_{n+1} does not enter into this computation.

Setp 4. Compute the test statistic

$$W = b^2/S^2 \quad .$$

Step 5. Compare the calculated value of W with the percentiles of the distribution of this test statistic shown in Table B-8. This table gives the minimum values of W that we would obtain with 1, 2, 5, 10, and 50 percent probability as a function of n, if the data actually came from a normal distribution. If the percentile is lower than the selected level of significance, then the hypothesis of normality can be rejected and accepted otherwise.

9.4 WE-TEST

This test is used to evaluate the assumption that a sample has an exponential distribution with the origin unknown. Percentiles of the WE distribution have not yet been tabulated for sample sizes other than 7 to 35. The comments on the W-test are also applicable here.

The WE-test is used as follows:

Step 1. Calculate the test statistic:

$$WE = \frac{(\bar{x}-x_i)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\left(\sum_{i=1}^{n} x_i/n - x_1\right)^2}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ n}$$

where x_i, $i = 1, \ldots n$, are the n observed values, x_1 is the smallest value, and $\bar{x}$ is the data mean.

Step 2. Compare the computed value WE with the 95 percent and 90 percent ranges given in Table B-9. This test is two-sided in that too-low or too-high values indicate non-exponentiality. Thus, if the computed WE value falls outside the 95 or 90 percent range, then the chances are less than one in 20 or one in 10, respectively, that the observed sample was drawn from an exponential distribution.

9.5 WE_o-TEST

This test is used to evaluate the assumption that a sample has an exponential distribution with the origin ϵ known. However, percentiles of the distribution WE_o have not been tabulated for sample sizes other than 7 to 35. The comments on the W-test are also applicable here.

The WE_o-test is used as follows:

Step 1. Subtract the known location ϵ from each of the sample values x_i.

Step 2. Calculate the test statistic

$$WE_o = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n} x_i\right)^2} \quad ,$$

where x_i, $i = 1, \ldots, n$, are the n sample values and $\bar{x}$ is the sample mean.

Step 3. Determine whether the computed value WE_0 lies outside the tabulated 95 percent and 90 percent ranges shown in Table B-10 as a function of n. This test is two-sided in that too-low or too-high values indicate non-exponentiality. Thus, if the computed value of WE_0 falls outside the 95 percent range, the chances are less than one in twenty that the observed sample was drawn from an exponential distribution with the assumed origin.

APPENDIX A

COMPLEX PARAMETRIC DISTRIBUTIONS

A.1 INTRODUCTION

Although the reader probably has a good general knowledge of the simple parametric distributions, he is likely to be unfamiliar with the complex parametric distributions. The main text of this volume indicates when and how to use these distributions, but all without requiring a thorough understanding of the complex distributions. This appendix is intended to give the reader some background information on the complex distributions so that he will be better able to understand and use the related material in the main text.

A. 2 WEIBULL DISTRIBUTION

The Weibull distribution is best known for its application to reliability analysis where it is known to fit a large class of life (time to failure) distributions (53). The basic distribution suggested by Weibull is to define $\varphi(x)$, where the cumulative distribution function F is given by

$$F(x) = P[X \leq x] = 1 - e^{-\varphi(x)} = \int_{-\infty}^{x} f(x)dx \quad .$$

One of the simplest forms for $\varphi(x)$ is

$$\varphi(x) = \frac{(x-\epsilon)^{\eta}}{\lambda} \qquad x \geq \epsilon$$

$$= 0 \qquad x \leq \epsilon$$

in which case

$$F(x) = 1 - e^{-\frac{(x-\epsilon)^{\eta}}{\lambda}} \qquad x \geq \epsilon$$

$$= 0 \qquad x \leq \epsilon$$

and

$$f(x) = \eta/\lambda \, (x-\epsilon)^{\eta - 1} e^{-\frac{(x-\epsilon)^{\eta}}{\lambda}} \qquad x \geq \epsilon$$

$$= 0 \qquad x \leq \epsilon$$

The parameter ϵ is called the location parameter in the sense that it defines the lower limit for the random variable x. For the special case where $\epsilon = 0$,

$$f(x) = \eta/\lambda \, x^{\eta-1} e^{-x^{\eta/\lambda}}$$

and

$$F(x) = 1 - e^{-x^{\eta}/\lambda} \quad .$$

The values of η and λ may be selected to provide a large number of shapes some of which are sketched below in Fig. A.1. For this reason η is called a shape parameter and λ is called a scale parameter since it scales the value of x.

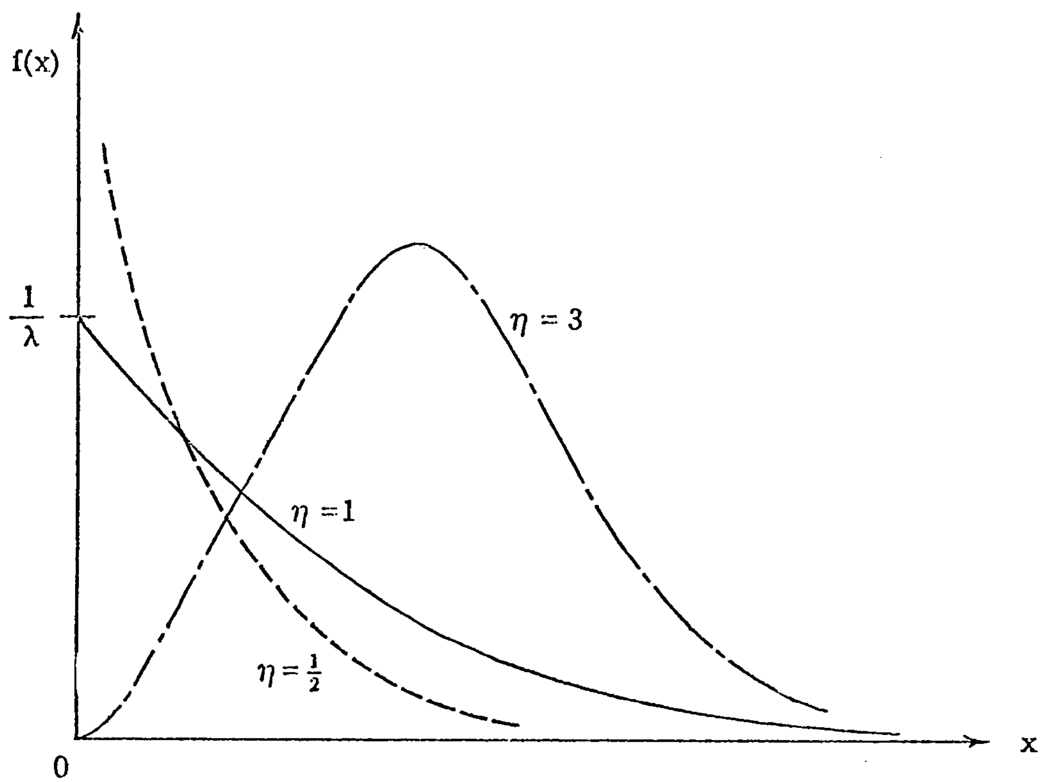


Fig. A.1. Weibull Distribution for Various Values of Parameter η

It should be noted from Fig. A.1 that $\eta_1 \cong 1/2$ might represent the shape parameter for the early failure region and $\eta_3 = 3$ the shape parameter for the wear-out region in a typical reliability application.

A. 3 JOHNSON DISTRIBUTIONS

These distributions were proposed by Johnson who used transformations of the normalized normal random variable to generate empirical distributions(21, 22). The main advantages of this approach are that percentiles of the empirical distribution may be obtained using a table of the normal probability distribution, as will become apparent later, and that the approach encompasses a broad class of problems.

To introduce the Johnson distributions, assume that it is desired to obtain a probability density function for the random variable X about which little or no information is available. Then, a general transformation from X to Z is postulated, where Z is a normalized normal random variable, as follows

$$Z = \gamma + \eta\, T(X) \quad ,$$

where γ and η and parameters to be determined.

In most situations, the transformation T(X) will be unknown. However, Johnson proposed three families of distributions, referred to as the S_L, S_B, and S_U systems, respectively, defined as follows

$$S_L \text{(Log-normal)}\; T(x) = \ell n\,(x-\epsilon) \quad ; \qquad x > \epsilon$$

$$S_B \text{(Bounded)}\; T(x) = \ell n\left(\frac{x-\epsilon}{\lambda+\epsilon-x}\right) \quad ; \qquad \epsilon < x < \epsilon + \lambda$$

$$S_U \text{(Unbounded)}\; T(x) = \sinh^{-1}\left(\frac{x-\epsilon}{\lambda}\right) \quad ; \qquad -\infty < x < \infty$$

The undefined regions for x above imply T(x) = 0.

Similar to the Weibull distribution, η and γ are shape parameters, λ is a scale parameter, and ϵ plays the role of location parameter which shifts the region of relevancy for x. These parameters are subject to the following constraints:

$$\eta > 0 \quad ; \quad \lambda > 0$$

$$-\infty < \gamma < \infty$$

$$-\infty < \epsilon < \infty$$

In some cases, these parameters may be identified from a basic understanding of the process. For example, if the random variable x must be non-negative, then $\epsilon = 0$ and the S_L, or lognormal distribution, might be appropriate. If x is restricted to a finite region, $\epsilon \le x \le \epsilon + \lambda$, then S_B (bounded distribution) may be appropriate. An infinite range for x would suggest the S_U (unbounded) distribution.

The probability density function for the distributions are as follows:

$$S_L: \quad f_1(x) = \frac{\eta}{\sqrt{2\pi}\,(x-\epsilon)} \exp\left\{-\frac{\eta^2}{2}\left[\frac{\gamma}{\eta} + \ln(x-\epsilon)\right]^2\right\} \quad ; \qquad x \geq \epsilon$$

$$S_B: \quad f_2(x) = \frac{\eta}{\sqrt{2\pi}} \frac{\lambda}{(x-\epsilon)(\lambda-x+\epsilon)} \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\epsilon}{\lambda-x+\epsilon}\right)\right]^2\right\}$$

$$\epsilon \le x \le \epsilon + \lambda$$

$$S_U: \quad f_3(x) = \frac{\eta}{\sqrt{2\pi}} \frac{1}{\sqrt{(x-\epsilon)^2 + \lambda^2}}$$

$$\exp\left[-\frac{1}{2}\left(\gamma + \eta \ln\left\{\left(\frac{x-\epsilon}{\lambda}\right) + \left[\frac{(x-\epsilon)^2}{\lambda^2} + 1\right]^{1/2}\right\}\right)^2\right]$$

$$-\infty < x < \infty$$

The density function for the S_L system is a three-parameter distribution commonly called the log-normal distribution. This is known to describe many familiar events such as amount of inheritance, income, particle size from breakage, etc.

As previously mentioned, the class of situations encompassed within these distributions is large. An indication of the flexibility in defining a large number of shapes is evidenced in Figs. A.2 to A.4 which illustrate several forms of the S_L, S_B and S_U density functions.

The difference between the three types of Johnson distributions can be characterized by the relationship between the distribution skewness and distribution kurtosis. Section 8 of this volume contains a discussion of skewness and kurtosis; however, a summary definition is that

$$\beta_1 = \mu_3/s^3 \quad \text{(skewness)}$$

and

$$\beta_2 = \mu_4/s^4 \quad \text{(kurtosis)}$$

To help in the definition of the relative variation in β_1 and β_2, Johnson prepared the results as shown in Fig. A.5. Note that the log-normal distribution is defined by a line given by:

$$\beta_1 = (\omega-1)(\omega+2)^2 \quad ; \quad > 0$$

$$\beta_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3 \quad ; \quad > 0$$

where

$$\omega = e^{1/\eta^2}$$

is the shape parameter for S_L.

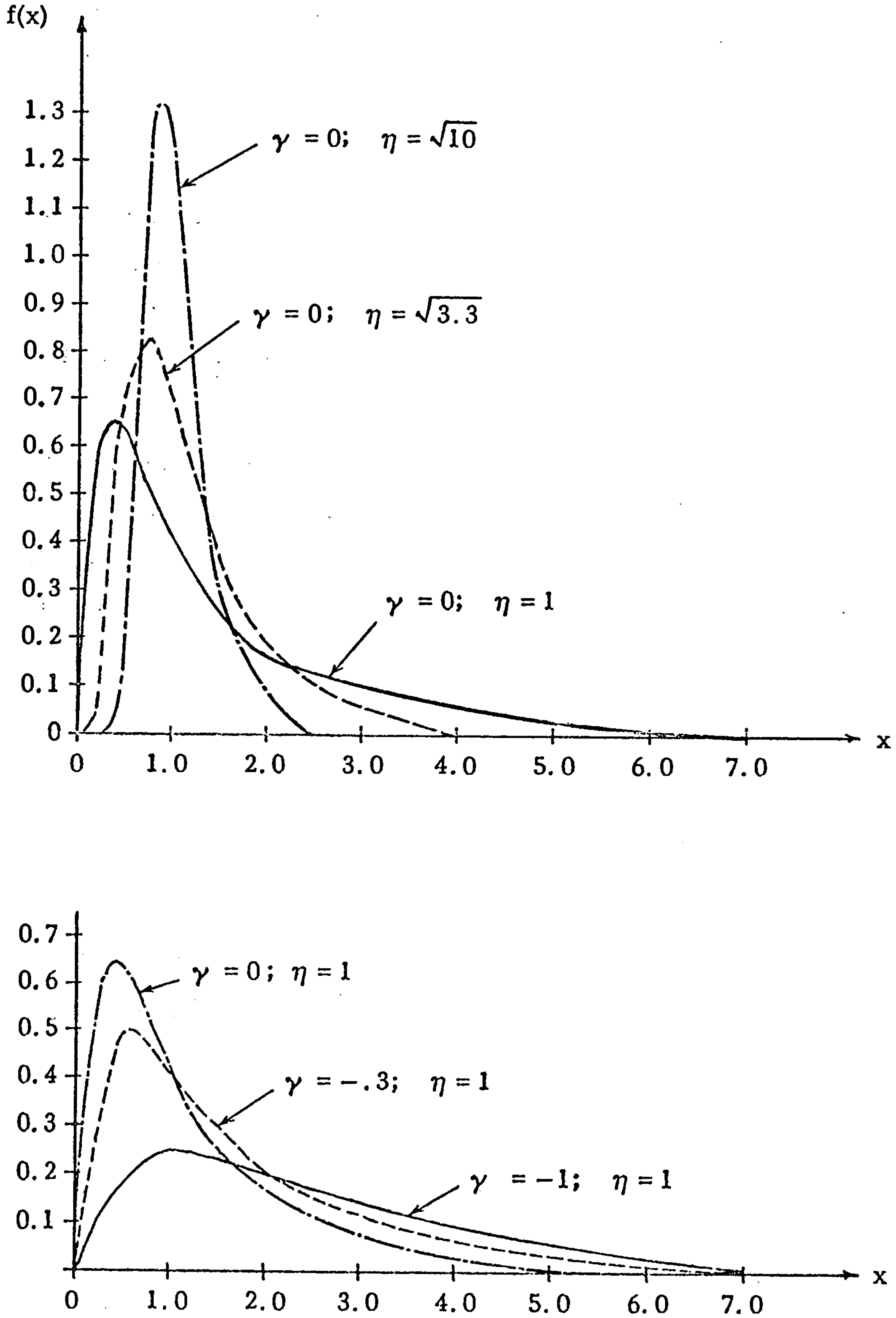


Fig. A. 2. Johnson Probability Density Functions for S_L ($\epsilon = 0$)

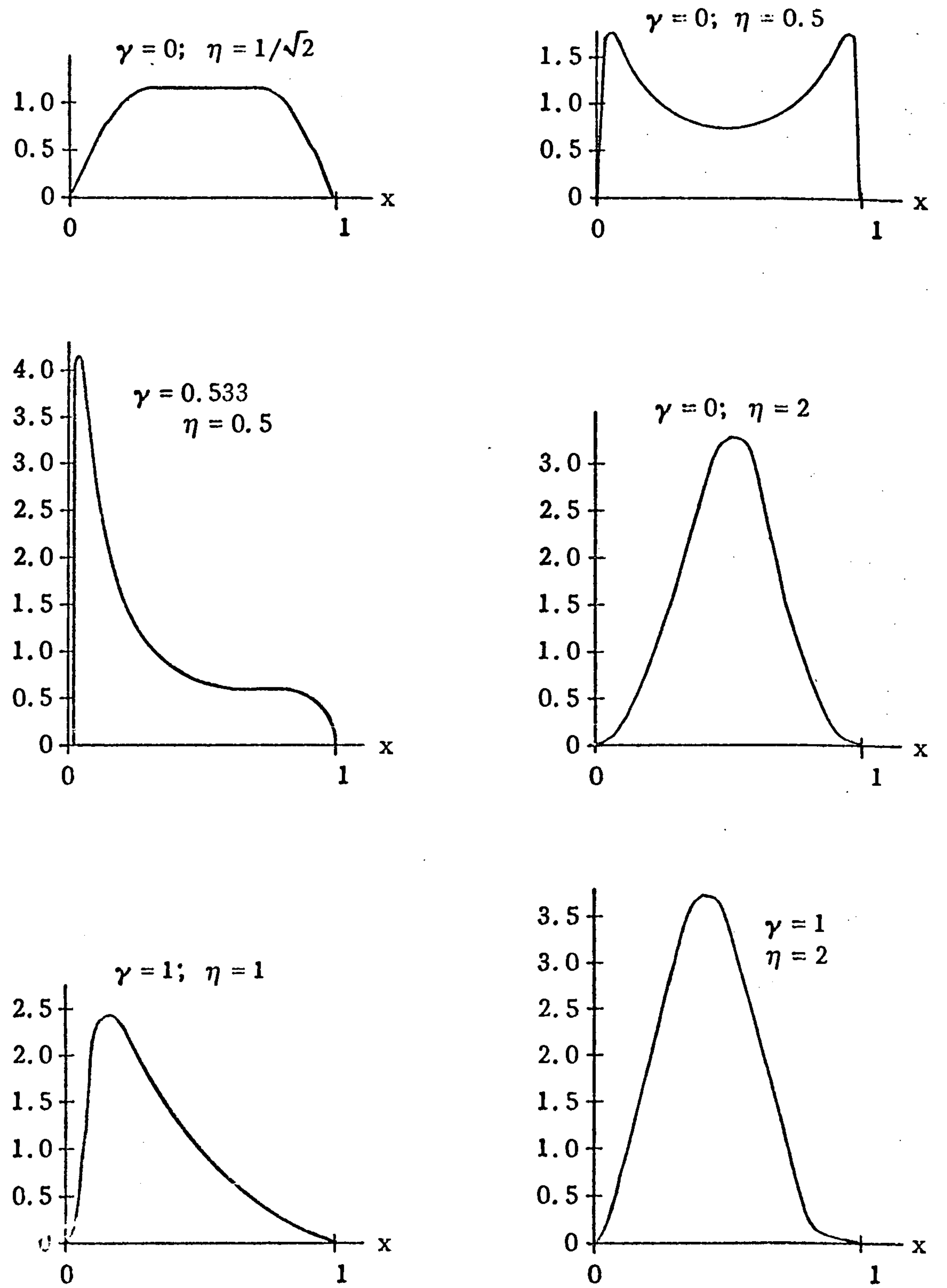


Fig. A. 3. Johnson Probability Density Functions for S_B ($\epsilon = 0$; $\lambda = 1$)

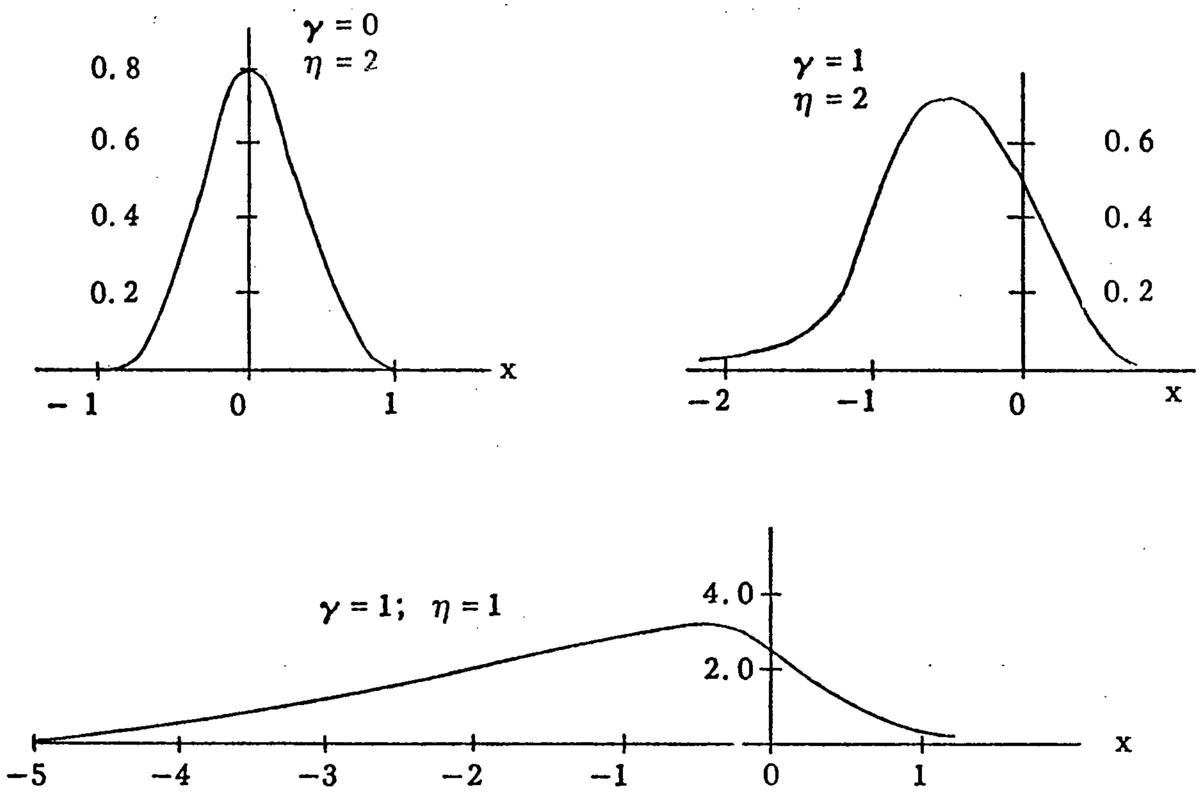


Fig. A. 4. Johnson Probability Density Functions for S_U

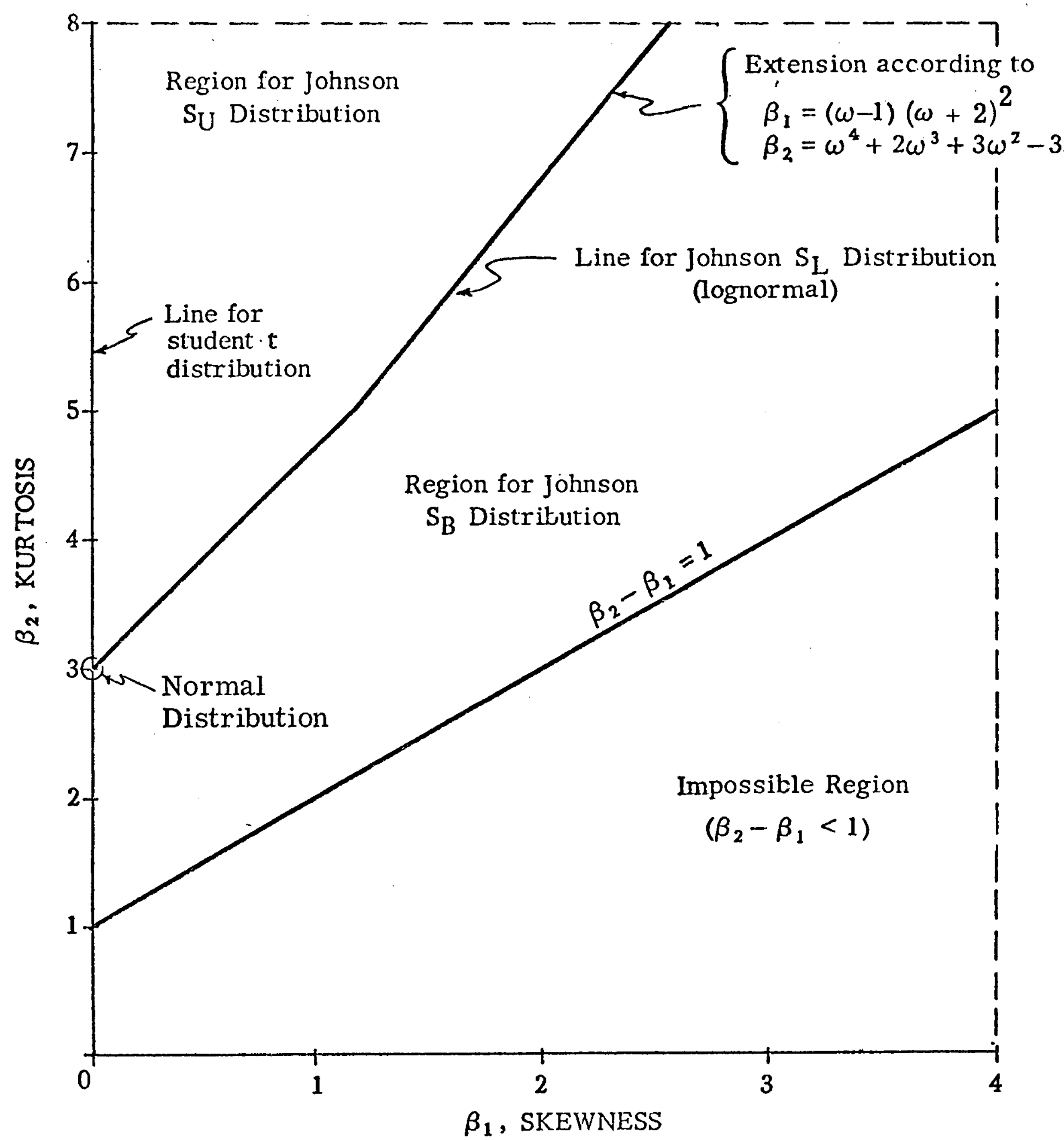


Fig. A.5. Regions of Definition for Johnson Distributions Based on Skewness and Kurtosis

It should be recognized that estimates for β_1 and β_2 may lead to a wrong conclusion as to the type of distribution to be used. The confidence that this will not occur is related to the accuracy of the estimates. In case of doubt, a goodness-of-fit test may be used to help in a decision.

A.4 PEARSON DISTRIBUTIONS

A general class of probability density functions known as the Pearson family[10, 35], is given by solutions of the differential equation:

$$\frac{dy}{dx} = \frac{(x + a)\, y}{b_0 + b_1 x + b_2 x^2}$$

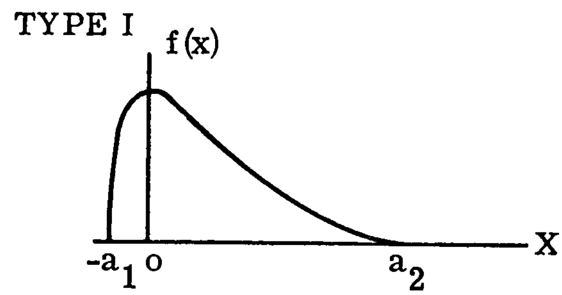
The solutions of this equation were classified by Pearson into twelve families of curves shown in Table 8.1. These curves are displayed in Fig. A.6. The Pearson distributions are related to the standard densities frequently discussed. For example: the gamma distributions are Pearson's Type III curves, the normal is a Type VII, the beta is a Type I while the beta with parameters $\alpha = \beta$ is represented by the Pearson Type II curves.

This system of density functions is very appealing from the standpoint of fitting sample data, the reason being that only the first four moments need be calculated. Pearson's methods of fitting sample data consists of the following steps:

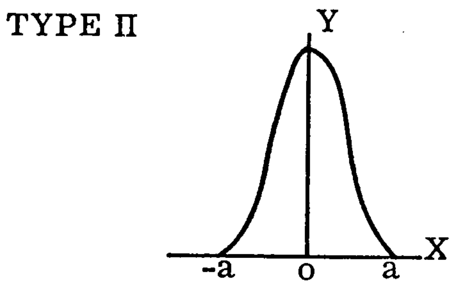
1. Compute the first four moments, μ_1, μ_2, μ_3, μ_4 of the sample data.
2. Calculate the numerical value of the parameters β_1 and β_2, where:

   β_1 = skewness,

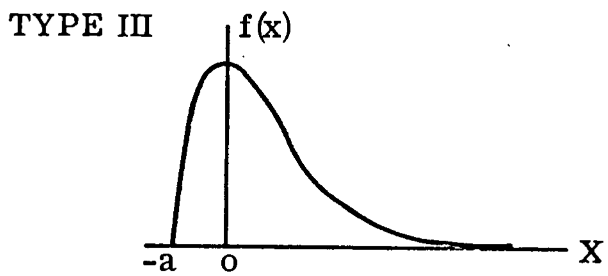
   β_2 = kurtosis.



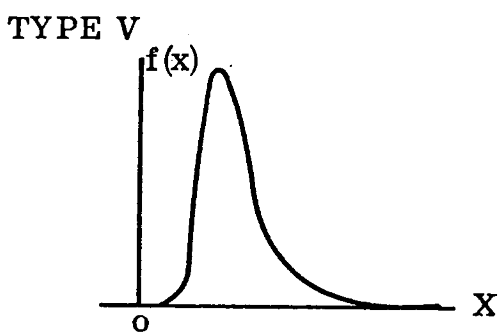
$$f(x) = y_o\left(1 + \frac{x}{a_1}\right)^{m_1}\left(1 - \frac{x}{a_2}\right)^{m_2}$$



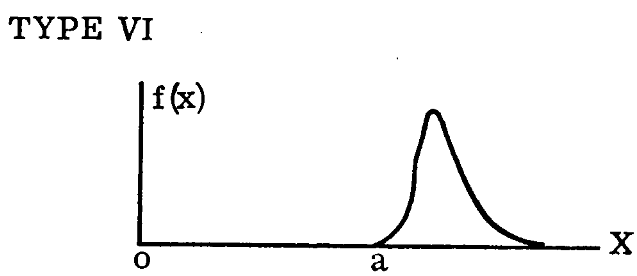
$$f(x) = y_o\left(1 - \frac{x^2}{a^2}\right)^{m}$$



$$f(x) = y_o\left(1 + \frac{x}{a}\right)^{\gamma a} e^{-\gamma x}$$



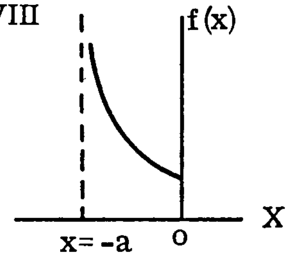
$$f(x) = y_o{}^{-p} e^{-\gamma/x}$$



$$f(x) = y_o (x-a)^{q_2} x^{q_1}$$

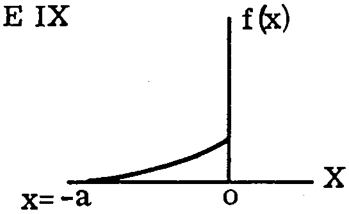
Fig. A.6. Typical shapes of Pearson distributions (Sheet 1 of 2)

TYPE VIII



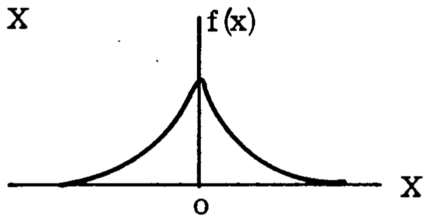
$$f(x) = y_o \left(1 + \frac{x}{a}\right)^{-m}$$

TYPE IX



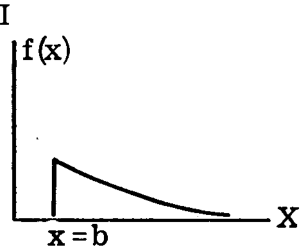
$$f(x) = y_o \left(1 + \frac{x}{a}\right)^{m}$$

TYPE X



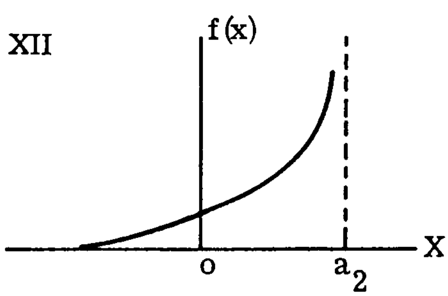
$$f(x) = y_o e^{-x/\sigma}$$

TYPE XI



$$f(x) = y_o x^{-m}$$

TYPE XII



$$f(x) = y_o \left(\frac{\sigma\left(\sqrt{(3+\beta_1)} + \sqrt{\beta_1}\right) + x}{\sigma\left(\sqrt{(3+\beta_1)} - \sqrt{\beta_1}\right) - x} \right)^{\sqrt{\frac{\beta_1}{(\beta + \beta_1)}}}$$

Note: Type IV and VII appear as normal distributions.

Fig. A.6. Typical shapes of Pearson distributions (Sheet 2 of 2)

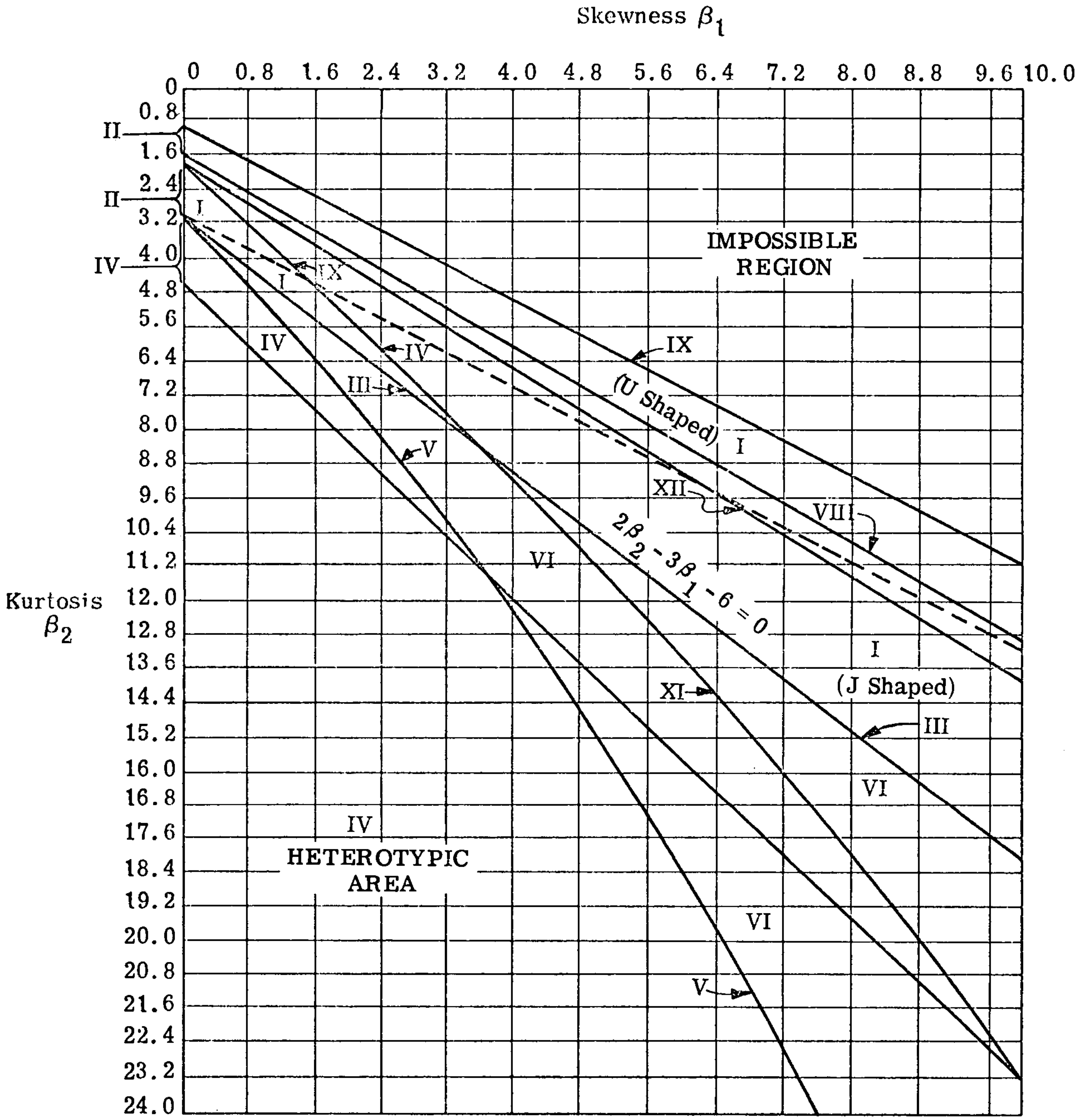


Fig. A.7. Types of Skew Frequency for Values of β_1 and β_2 for the Pearson System

These parameters determine the type of Pearson distribution which appropriately matches the sample data.

3. Equate the observed (sample) moments to the moments of the appropriate distribution expressed in terms of its parameters, and

4. Solve the resulting equations for those parameters thereby completely specifying the distribution function.

The relationships between β_1 and β_2 for a given Pearson distribution have been represented in a convenient graphical form in the so-called β_1, β_2-plane shown in Fig. A.7. The normal distribution corresponds to the point $\beta_1 = 0$, $\beta_2 = 3$ in the β_1, β_2 plane. Type III distributions are to be chosen when the point β_1, β_2 is on the line $2\beta_2 - 3\beta_1 - 6 = 0$ and Type V when (β_1, β_2) is on the cubic

$$\beta_1(\beta_2 + 3)^2 = 4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6).$$

In considering the subtypes under Type I, a biquadratic in β_1 and β_2 separates the area of the J-shaped curves from the regions of limited range modal curves and the region of the U-shaped curves.

In summary, the curves traced in the (β_1, β_2)- plane provide a means of selecting the Pearson distribution appropriate to a given collection of sample data. For further details and numerical examples see Elderton(10) and Kendall(27).

APPENDIX B

PROBABILITY TABLES

TABLE B-1

Unbiasing Factors for the M.L.E. of η

n	5	6	7	8	9	10	11	12	13	14	15	16
B(n)	.669	.752	.792	.820	.842	.859	.872	.883.	.893	.901	.908	.914
n	18	20	22	24	26	28	30	32	34	36	38	40
B(n)	.923	.931	.938	.943	.947	.951	.955	.958	.960	.962	.964	.966
n	42	44	46	48	50	52	54	56	58	60	62	64
B(n)	.968	.970	.971	.972	.973	.974	.975	.976	.977	.978	.979	.980
n	66	68	70	72	74	76	78	80	85	90	100	120
B(n)	.980	.981	.981	.982	.982	.983	.983	.984	.985	.986	.987	.990

TABLE B-2
Percentiles of the Normal Distribution

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

x	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417
$F(x)$	.90	.95	.975	.99	.995	.999	.9995	.99995	.999995
$2[1 - F(x)]$	.20	.10	.05	.02	.01	.002	.001	.0001	.00001

(From A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill, 1950.)

TABLE B-3

Tables to Facilitate Fitting Johnson S_u Distribution

Values of $-\hat{\gamma}$

β_2 \ $\sqrt{\beta_1}$	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
3·2	0.3479	0.7372	1.228	1.939	3.188	6.367				
3·3	.2332	.4843	0.7762	1.145	1.662	2.488				
3·4	.1763	.3626	.5710	0.8184	1.133	1.567	2.236	3.472		
3·5	.1424	.2911	.4536	.6397	.8640	1.151	1.546	2.146		
3·6	0.1198	0.2440	0.3776	0.5270	0.7011	0.9136	1.187	1.565	2.139	3.157
3·7	.1036	.2106	.3243	.4495	.5919	.7602	0.9681	1.238	1.614	2.188
3·8	.0916	.1856	.2849	.3928	.5134	.6528	.8197	1.028	1.302	1.687
3·9	.0822	.1663	.2546	.3495	.4544	.5734	.7127	0.8814	1.095	1.378
4·0	.0746	.1509	.2305	.3155	.4083	.5122	.6317	.7733	0.9470	1.169
4·1	0.0685	0.1383	0.2109	0.2879	0.3713	0.4637	0.5684	0.6902	0.8363	1.018
4·2	.0633	.1276	.1943	.2647	.3404	.4234	.5174	.6243	.7503	.9031
4·3	.0589	.1188	.1806	.2456	.3151	.3907	.4755	.5708	.6814	.8132
4·4	.0552	.1112	.1689	.2294	.2937	.3632	.4397	.5265	.6250	.7407
4·5	.0519	.1046	.1588	.2153	.2752	.3396	.4100	.4891	.5780	.6811
4·6	0·0491	0·0989	0·1499	0·2031	0·2592	0·3192	0·3844	0·4564	0·5382	0·6311
4·7	·0466	·0938	·1421	·1923	·2451	·3014	·3622	·4288	·5040	·5886
4·8	·0444	·0893	·1352	·1828	·2327	·2857	·3426	·4048	·4744	·5520
4·9	·0424	·0852	·1290	·1743	·2216	·2717	·3254	·3836	·4484	·5202
5·0	·0406	·0816	·1234	·1666	·2117	·2592	·3099	·3648	·4254	·4922
5·1	0·0390	0·0783	0·1184	0·1597	0·2027	0·2480	0·2961	0·3479	0·4050	0·4674
5·2	·0374	·0752	·1138	·1534	·1946	·2378	·2836	·3328	·3866	·4453
5·3	·0361	·0725	·1096	·1477	·1872	·2285	·2723	·3191	·3696	·4255
5·4	·0348	·0700	·1057	·1424	·1804	·2201	·2620	·3066	·3547	·4076
5·5	·0337	·0676	·1022	·1376	·1742	·2123	·2525	·2952	·3411	·3913
5·6	0·0326	0·0655	0·0989	0·1331	0·1684	0·2052	0·2438	0·2848	0·3286	0·3765
5·7	·0316	·0635	·0958	·1290	·1631	·1986	·2358	·2752	·3172	·3629
5·8	·0307	·0616	·0930	·1251	·1582	·1925	·2284	·2663	·3066	·3504
5·9	·0298	·0599	·0904	·1215	·1536	·1868	·2215	·2581	·2967	·3385
6·0	·0290	·0583	·0879	·1182	·1493	·1815	·2151	·2504	·2879	·3278
6·1	0·0283	0·0568	0·0856	0·1151	0·1453	0·1766	0·2091	0·2433	0·2794	0·3180
6·2	·0276	·0553	·0835	·1121	·1415	·1719	·2035	·2366	·2716	·3088
6·3	·0269	·0540	·0814	·1094	·1380	·1676	·1983	·2304	·2643	·3002
6·4	·0263	·0527	·0795	·1067	·1347	·1635	·1933	·2245	·2574	·2921
6·5	·0257	·0515	·0777	·1043	·1315	·1596	·1887	·2190	·2509	·2846
6·6	0·0251	0·0504	0·0760	0·1020	0·1286	0·1560	0·1843	0·2138	0·2448	0·2775
6·7	·0246	·0493	·0743	·0998	·1258	·1525	·1802	·2089	·2391	·2709
6·8	·0241	·0483	·0728	·0977	·1231	·1492	·1762	·2043	·2337	·2646
6·9	·0236	·0473	·0713	·0957	·1206	·1461	·1725	·1999	·2285	·2586
7·0	·0232	·0464	·0699	·0938	·1182	·1432	·1690	·1957	·2237	·2530
7·1	0·0227	0·0455	0·0686	0·0920	0·1159	0·1404	0·1656	0·1918	0·2190	0·2476
7·2	·0223	·0447	·0673	·0903	·1137	·1377	·1624	·1880	·2147	·2426
7·3	·0219	·0439	·0661	·0887	·1116	·1352	·1594	·1844	·2105	·2377
7·4	0215	·0431	·0650	·0871	·1096	·1327	·1565	·1810	·2065	·2331
7·5	·0212	·0424	·0639	·0856	·1077	·1304	·1537	·1777	·2027	·2287
7·6	0·0208	0·0417	0·0628	0·0842	0·1059	0·1282	0·1510	0·1746	0·1991	0·2246
7·7	·0205	·0410	·0618	·0828	·1042	·1260	·1485	·1716	·1956	·2206
7·8	·0202	·0404	·0608	·0815	·1025	·1240	·1460	·1687	·1922	·2167
7·9	·0198	·0398	·0599	·0802	·1009	·1220	·1437	·1660	·1891	·2131
8·0	·0195	·0392	·0590	·0790	·0993	·1201	·1414	·1633	·1860	·2095
8·2	0·0190	0·0380	0·0572	0·0767	0·0964	0·1165	0·1371	0·1583	0·1802	0·2029
8·4	·0185	·0370	·0557	·0745	·0937	·1132	·1332	·1537	·1749	·1968
8·6	·0180	·0360	·0542	·0725	·0912	·1101	·1295	·1494	·1699	·1912
8·8	·0175	·0351	·0528	·0707	·0888	·1073	·1261	·1454	·1653	·1859
9·0	·0171	·0342	·0515	·0689	·0866	·1046	·1229	·1417	·1610	·1810
9·2	—	—	—	—	—	—	0·1199	0·1382	0·1570	0·1764
9·4	—	—	—	—	—	—	·1171	·1349	·1532	·1721

(From E. S. Pearson and H. O. Hartley, Biometrika Tables for Statisticians, Vol. 2, pp. 288-291, Cambridge University Press, 1972).
Table B-3 corrected according to: N. L. Johnson, "Extensions and Corrections to 'Tables to Facilitate Fitting S_u Frequency Curves'," Biometrika 61, 203-205, (1974).

TABLE B-3 (continued)

Values of $-\hat{\gamma}$ (continued)

β_2 \ $\sqrt{\beta_1}$	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·95	1·00
3·8	2.284	3.383								
3·9	1.783	2.426								
4·0	1.469	1.906	2.621	4.105						
4·1	1.253	1.577	2.060	2.886						
4·2	1.095	1.349	1.705	2.254						
4·3	0.9751	1.182	1.460	1.860						
4·4	.8802	1.054	1.280	1.589						
4·5	.8033	0.9526	1.141	1.391						
4·6	0.7398	0.8704	1.032	1.240	1.522	1.931	2.603	4.031		
4·7	.6865	.8024	0.9434	1.121	1.353	1.676	2.166	3.037		
4·8	.6410	.7451	.8698	1.024	1.221	1.485	1.864	2.472		
4·9	.6017	.6962	.8079	0.9435	1.113	1.335	1.641	2.099		
5·0	.5675	.6539	.7550	.8761	1.025	1.215	1.469	1.831	2.406	3.540
5·1	0·5374	0·6170	0·7092	0.8184	0.9509	1.117	1.332	1.629	2.071	2.831
5·2	·5106	·5845	·6693	.7687	.8876	1.034	1.221	1.470	1.824	2.385
5·3	·4868	·5556	·6341	.7252	.8331	0.9641	1.128	1.342	1.635	2.072
5·4	·4653	·5298	·6028	.6869	.7855	.9039	1.050	1.236	1.484	1.838
5·5	·4459	·5066	·5749	.6530	.7437	.8515	0.9825	1.147	1.361	1.656
5·6	0·4283	0·4856	0·5498	0·6226	0·7067	0·8055	0·9243	1·071	1.259	1.510
5·7	·4122	·4665	·5270	·5953	·6735	·7647	·8732	1·006	1.172	1.390
5·8	·3975	·4491	·5063	·5706	·6437	·7284	·8282	0·9485	1.098	1.289
5·9	·3840	·4331	·4875	·5481	·6168	.6957	·7881	·8982	1.033	1.203
6·0	.3714	·4184	·4701	·5276	·5924	·6663	·7521	·8536	0.9765	1.129
6·1	0·3598	0·4049	0·4542	0·5088	0·5700	0·6396	0·7197	0·8138	0·9265	1·065
6·2	·3491	·3923	·4395	·4915	·5496	·6152	·6904	·7780	·8820	1·008
6·3	·3390	·3806	·4258	·4755	·5308	·5929	·6637	·7456	·8420	0·9581
6·4	·3297	·3697	·4131	·4607	·5134	·5724	·6392	·7161	·8060	·9132
6·5	·3209	·3595	·4013	·4470	·4973	·5535	.6168	·6892	·7733	·8729
6·6	0·3123	0·3500	0·3903	0·4341	0·4824	0·5359	0·5962	0·6646	0·7436	0·8364
6·7	·3046	·3410	·3799	·4221	·4684	·5197	·5770	·6419	·7164	·8033
6·8	·2973	·3326	·3702	·4109	·4554	·5045	·5592	·6209	·6914	·7730
6·9	·2904	·3247	·3611	·4004	·4433	·4904	·5427	.6015	·6683	·7453
7·0	·2839	·3172	·3524	·3905	·4318	·4772	·5273	·5835	·6470	·7198
7·1	0·2778	0·3101	0·3443	0·3812	0·4211	0·4648	0·5130	0·5666	0·6272	0·6962
7·2	·2719	·3034	·3366	·3723	·4110	·4531	·4995	·5509	·6087	·6744
7·3	·2664	·2967	·3293	·3640	·4014	·4421	·4868	·5362	·5915	·6541
7·4	·2611	·2907	·3224	·3561	·3924	·4318	·4749	·5224	·5754	·6352
7·5	·2561	·2849	·3159	·3486	·3838	·4220	·4636	·5094	.5603	·6175
7·6	0·2513	0·2795	0·3096	0·3415	0·3757	0·4127	0·4530	0·4972	0·5463	0·6010
7·7	·2467	·2742	·3037	·3347	·3680	·4039	·4429	·4857	·5328	·5855
7·8	·2423	·2692	·2980	·3283	·3607	·3956	·4334	·4747	·5203	·5709
7·9	·2381	·2645	·2926	·3221	·3537	·3876	·4244	·4644	·5084	·5571
8·0	·2341	·2599	·2871	·3163	·3470	·3801	·4158	·4546	·4971	·5442
8·1	0·2303	0·2556	0·2822	0·3106	0·3407	0·3729	0·4076	0·4452	0·4864	0.5319
8·2	·2266	·2514	·2774	·3053	·3346	·3660	·3998	·4364	·4763	·5203
8·3	·2230	·2473	·2729	·3001	·3288	·3594	·3923	·4279	·4667	·5092
8·4	·2196	·2435	·2685	·2952	·3232	·3531	·3852	·4199	·4575	·4987
8·5	·2163	·2397	·2643	·2904	·3179	·3471	·3784	·4122	·4488	·4888
8·6	0·2132	0·2362	0·2603	0·2859	0·3127	0·3413	0·3719	0·4048	0·4405	0·4793
8·7	·2101	·2327	·2564	·2812	·3078	·3358	·3657	·3978	·4325	·4702
8·8	·2072	·2294	·2526	·2770	·3031	·3305	·3597	·3910	·4248	·4616
8·9	·2044	·2262	·2490	·2730	·2985	·3253	·3539	·3845	·4175	·4533
9·0	·2016	·2231	·2455	·2691	·2941	·3204	·3484	·3783	·4105	·4454
9·2	0·1964	0·2172	0·2389	0·2617	0·2858	0·3111	0·3380	0·3666	0·3974	0·4305
9·4	·1915	·2117	·2328	·2548	·2778	·3025	·3283	·3558	·3852	·4169
9·6	—	—	—	·2483	·2706	·2944	·3193	·3457	·3739	·4042
9·8	—	—	—	·2422	·2639	·2868	·3109	·3363	·3634	·3925
10·0	—	—	—	·2365	·2575	·2798	·3030	·3275	·3537	·3816

TABLE B-3 (continued)

Values of $-\hat{\gamma}$ (continued)

β_2 \ $\sqrt{\beta_1}$	1·05	1·10	1·15	1·20	1·25	1·30	1·35	1·40	1·45	1·50
5·4	2.401	3.529								
5·5	2.099	2.872								
5·6	1.871	2.450								
5·7	1.672	2.149								
5·8	1.547	1.921	2.532	3.892						
5·9	1.428	1.741	2.223	3.122						
6·0	1.327	1.595	1.989	2.653						
6·1	1.241	1.474	1.805	2.324						
6·2	1.167	1.372	1.655	2.078	2.819	5.129				
6·3	1.102	1.285	1.531	1.885	2.458	3.708				
6·4	1.1044	1.210	1.426	1.728	2.191	3.052				
6·5	0.9933	1.143	1.336	1.599	1.983	2.634				
6·6	0.9477	1.085	1.258	1.489	1.817	2.334				
6·7	.9066	1.032	1.190	1.396	1.679	2.106				
6·8	.8694	0.9857	1.130	1.315	1.563	1.923	2.520	3.939		
6·9	.8356	.9435	1.076	1.243	1.464	1.774	2.258	3.210		
7·0	.8046	.9053	1.028	1.180	1.378	1.650	2.054	2.766		
7·1	0·7762	0·8705	0·9840	1·124	1.303	1.543	1.889	2.453		
7·2	·7500	·8386	·9444	1·074	1.237	1.452	1.752	2.216		
7·3	·7258	·8093	·9083	1·028	1.178	1.372	1.636	2.028		
7·4	·7034	·7823	·8753	0·9871	1.125	1.301	1.536	1.874	2.424	3.680
7·5	·6825	·7573	·8450	·9495	1.077	1.238	1.450	1.745	2.201	3.085
7·6	0·6630	0·7342	0·8170	0·9151	1·034	1·182	1.374	1.635	2.023	2.703
7·7	·6448	·7126	·7910	·8834	·9945	1·132	1.306	1.540	1.876	2.425
7·8	·6278	·6924	·7670	·8542	·9584	1·086	1.246	1.457	1.752	2.209
7·9	·6117	·6736	·7445	·8272	·9251	1·044	1.192	1.384	1.646	2.036
8·0	·5967	·6559	·7236	·8020	·8945	1·006	1.143	1.319	1.553	1.892
8·1	0·5825	0·6393	0·7040	0·7786	0·8661	0·9706	1·099	1·260	1.472	1.771
8·2	·5690	·6237	·6857	·7568	·8397	·9382	1·058	1·207	1.401	1.667
8·3	·5563	·6089	·6684	·7364	·8152	·9083	1·020	1·159	1.337	1.576
8·4	·5443	·5950	·6521	·7172	·7923	·8805	0·9860	1·115	1.279	1.496
8·5	·5328	·5818	·6368	·6991	·7709	·8546	·9542	1·075	1.227	1.425
8·6	0·5220	0·5693	0·6223	0·6822	0·7507	0·8304	0·9246	1·038	1·180	1·361
8·7	.5116	·5574	·6085	·6661	·7318	·8078	·8972	1·004	1·136	1·304
8·8	·5018	·5461	·5955	·6510	·7140	·7866	·8716	0·9729	1·096	1·252
8·9	·4924	·5354	·5831	·6366	·6972	·7667	·8477	·9436	1·060	1·204
9·0	·4834	·5251	·5714	·6230	·6813	·7480	·8252	·9163	1·026	1·161
9·1	0·4748	0·5154	0·5601	0·6101	0·6663	0·7303	0·8042	0·8908	0·9943	1·121
9·2	·4666	·5060	·5494	·597[illegible]	·6520	·7136	·7843	·8669	·9651	1·084
9·3	·4587	·4971	·5393	·5861	·6385	·6977	·7656	·8445	·9377	1·050
9·4	·4511	·4885	·5295	·5749	·6256	·6827	·7480	·8234	·9122	1·019
9·5	·4439	·4803	·5202	·5642	·6133	·6685	·7312	·8036	·8882	0·9892
9·6	0·4369	0·4724	0·5112	0·5540	0·6016	0·6549	0·7154	0·7848	0·8657	0·9616
9·7	·4302	·4648	·5027	·5443	·5904	·6420	·7003	·7671	·8445	·9359
9·8	·4237	·4576	·4945	·5349	·5797	·6297	·6860	·7503	·8245	·9117
9·9	·4175	·4506	·4866	·5260	·5695	·6180	·6724	·7343	·8056	·8889
10·0	·4115	·4438	·4790	·5174	·5597	·6067	·6594	·7192	·7877	·8675
10·2	0·4001	0·4311	0·4646	0·5012	0·5413	0·5857	0·6352	0·6910	0·7546	0·8280
10·4	·3895	·4192	·4513	·4862	·5243	·5664	·6131	·6654	·7247	·7927
10·6	·3796	·4082	·4389	·4723	·5086	·5485	·5927	·6420	·6975	·7608
10·8	·3703	·3978	·4273	·4593	·4940	·5320	·5739	·6205	·6727	·7318
11·0	·3615	·3881	·4165	·4472	·4804	·5167	·5566	·6007	·6499	·7053
11·2	0·3533	0·3789	0·4063	0·4358	0·4678	0·5025	0·5404	0·5823	0·6289	0·6811
11·4	·3455	·3703	·3968	·4252	·4559	·4891	·5254	·5653	·6095	·6588
11·6	—	—	—	—	—	—	—	·5495	·5915	·6383
11·8	—	—	—	—	—	—	—	·5347	·5748	·6192
12·0	—	—	—	—	—	—	—	·5209	·5592	·6015

TABLE B-3 (continued)

Values of $-\hat{\gamma}$ (continued)

β_2 \ $\sqrt{\beta_1}$	1·55	1·60	1·65	1·70	1·75	1·80	1·85	1·90	1·95	2·00
8·0	2.453	3.828								
8·1	2.239	3.189								
8·2	2.066	2.792								
8·3	1.923	2.508								
8·4	1.802	2.289								
8·5	1.698	2.114								
8·6	1.607	1.968	2.594	4.642						
8·7	1.527	1.845	2.363	3.576						
8·8	1.455	1.739	2.180	3.058						
8·9	1.391	1.647	2.029	2.717						
9·0	1.334	1.566	1.901	2.465						
9·1	1.282	1.493	1.792	2.268						
9·2	1.234	1.428	1.697	2.107						
9·3	1.190	1.370	1.613	1.972						
9·4	1.150	1.316	1.538	1.857	2.382	3.698				
9·5	1.112	1.268	1.471	1.757	2.206	3.143				
9·6	1·078	1·223	1·412	1·670	2.060	2.790				
9·7	1·046	1·182	1·357	1·592	1.937	2.533				
9·8	1·016	1·144	1·307	1·523	1.830	2.333				
9·9	0·9881	1·109	1·261	1·460	1.737	2.171				
10·0	·9619	1·076	1·219	1·403	1.655	2.035				
10·1	0·9374	1·046	1·180	1·351	1.582	1.919				
10·2	·9142	1·017	1·144	1·304	1.516	1.818	2.311	3.494		
10·3	·8924	0·9906	1·110	1·260	1.456	1.730	2.157	3.032		
10·4	·8718	·9655	1·079	1·220	1.402	1.652	2.027	2.723		
10·5	·8523	·9419	1·050	1·183	1.353	1.582	1.916	2.493		
10·6	0·8338	0·9196	1·022	1·148	1·307	1·518	1.819	2.311		
10·7	·8163	·8985	0·9963	1·115	1·265	1·461	1.734	2.161		
10·8	·7996	·8785	·9720	1·085	1·226	1·408	1.657	2.035		
10·9	·7837	·8596	·9490	1·057	1·190	1·360	1.589	1.926		
11·0	·7686	·8416	·9274	1·030	1·156	1·316	1.527	1.831	2.332	3.637
11·1	0·7541	0·8246	0·9069	1·005	1·124	1·274	1.471	1.747	2.183	3.126
11·2	·7403	·8083	·8874	0·9811	1·095	1·236	1.420	1.673	2.057	2.800
11·3	·7272	·7928	·8689	·9587	1·067	1·201	1.373	1.605	1.949	2.562
11·4	·7145	·7780	·8513	·9375	1·041	1·168	1.329	1.544	1.854	2.375
11·5	·7024	·7638	·8346	·9174	1·016	1·137	1.289	1.489	1.770	2.223
11·6	0·6907	0·7503	0·8186	0·8983	0·9928	1·108	1·251	1·438	1.695	2.094
11·7	·6796	·7373	·8034	·8801	·9708	1·080	1·216	1·391	1.628	1.984
11·8	·6688	·7248	·7888	·8628	·9499	1·054	1·183	1·347	1.567	1.888
11·9	·6585	·7129	·7749	·8463	·9300	1·030	1·152	1·307	1.511	1.803
12·0	·6486	·7014	·7615	·8306	·9112	1·007	1·124	1·27[illegible]	1.461	1.728
12·1	0·6390	0·6904	0·7487	0·8155	0·8932	0·9851	1·096	1·235	1·414	1·660
12·2	·6297	·6798	·7364	·8011	·8761	·9644	1·071	1·202	1·370	1·598
12·3	·6208	·6696	·7246	·7873	·8597	·9447	1·046	1·171	1·330	1·542
12·4	·6122	·6598	·7132	·7740	·8441	·9260	1·024	1·143	1·293	1·490
12·5	·6039	·6503	·7023	·7613	·8291	·9081	1·002	1·115	1·258	1·443
12·6	0·5959	0·6411	0·6918	0·7491	0·8148	0·8910	0·9811	1·090	1·225	1·399
12·7	·5881	·6323	·6816	·7374	·8011	·8747	·9614	1·066	1·194	1·358
12·8	·5806	·6237	·6719	·7261	·7879	·8592	·9427	1·043	1·165	1·320
12·9	·5733	·6154	·6624	·7152	·7752	·8442	·9248	1·021	1·138	1·284
13·0	·5662	·6074	·6533	·7047	·7630	·8299	·9078	1·000	1·112	1·251
13·2	—	—	—	—	0·7400	0·8030	0·8758	0·9614	1·064	1·190
13·4	—	—	—	—	·7187	·7781	·8465	·9263	1·021	1·136
13·6	—	—	—	—	·6988	·7551	·8195	·8941	0·9822	1·088
13·8	—	—	—	—	·6802	·7336	·7945	·8646	·9466	1·045
14·0	—	—	—	—	·6628	·7136	·7712	·8373	·9141	1·005
14·2	—	—	—	—	0·6464	0·6949	0·7496	0·8121	0·8842	0·9690
14·4	—	—	—	—	·6311	·6774	·7295	·7886	·8566	·9359
14·6	—	—	—	—	·6166	·6609	·7106	·7668	·8310	·9055
14·8	—	—	—	—	·6029	·3454	·6929	·7464	·8072	·8774
15·0	—	—	—	—	·5900	·6308	·6763	·7273	·7851	·8514

TABLE B-4

Tables to Facilitate Fitting Johnson Distributions

Values of $\hat{\eta}$

β_2 \ $\sqrt{\beta_1}$	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
3·2	4·671	4·787	5·004	5·369	5·992	7·204				
3·3	3·866	3·927	4·036	4·208	4·469	4·875				
3·4	3·396	3·435	3·503	3·607	3·759	3·979	4·300	4·813		
3·5	3·081	3·108	3·156	3·227	3·328	3·467	3·663	3·943		
3·6	2·852	2·872	2·908	2·960	3·033	3·132	3·266	3·448	3·705	4·087
3·7	2·676	2·692	2·719	2·760	2·816	2·890	2·989	3·120	3·295	3·540
3·8	2·535	2·548	2·571	2·604	2·648	2·707	2·783	2·882	3·011	3·184
3·9	2·420	2·431	2·450	2·477	2·513	2·561	2·623	2·701	2·801	2·931
4·0	2·324	2·333	2·349	2·372	2·402	2·442	2·492	2·557	2·637	2·739
4·1	2·242	2·250	2·264	2·283	2·309	2·343	2·385	2·439	2·505	2·588
4·2	2·171	2·178	2·190	2·207	2·229	2·258	2·295	2·340	2·396	2·465
4·3	2·109	2·115	2·126	2·141	2·160	2·186	2·217	2·256	2·304	2·363
4·4	2·054	2·060	2·069	2·082	2·100	2·122	2·150	2·184	2·226	2·276
4·5	2·005	2·010	2·018	2·030	2·046	2·066	2·090	2·121	2·157	2·202
4·6	1·961	1·[illegible]66	1·973	1·984	1·998	2·016	2·038	2·065	2·097	2·136
4·7	1·921	1·925	1·932	1·942	1·955	1·971	1·991	2·015	2·044	2·079
4·8	1·885	1·889	1·895	1·904	1·916	1·930	1·948	1·970	1·997	2·028
4·9	1·852	1·855	1·861	1·869	1·880	1·893	1·910	1·930	1·954	1·982
5·0	1·822	1·825	1·830	1·837	1·847	1·860	1·875	1·893	1·915	1·941
5·1	1·793	1·796	1·801	1·808	1·817	1·829	1·843	1·859	1·880	1·903
5·2	1·767	1·770	1·775	1·781	1·790	1·800	1·813	1·829	1·847	1·869
5·3	1·743	1·746	1·750	1·756	1·764	1·774	1·786	1·800	1·817	1·837
5·4	1·721	1·723	1·727	1·732	1·740	1·749	1·760	1·774	1·789	1·808
5·5	1·699	1·702	1·705	1·711	1·718	1·726	1·737	1·749	1·764	1·781
5·6	1·680	1·682	1·685	1·690	1·697	1·705	1·715	1·726	1·740	1·756
5·7	1·661	1·663	1·666	1·671	1·677	1·685	1·694	1·705	1·718	1·733
5·8	1·643	1·645	1·648	1·653	1·658	1·666	1·674	1·685	1·697	1·711
5·9	1·627	1·628	1·631	1·636	1·641	1·648	1·656	1·666	1·677	1·691
6·0	1·611	1·613	1·615	1·619	1·625	1·631	1·639	1·648	1·659	1·672
6·1	1·596	1·598	1·600	1·604	1·609	1·615	1·623	1·631	1·642	1·653
6·2	1·582	1·583	1·586	1·590	1·594	1·600	1·607	1·615	1·625	1·636
6·3	1·568	1·570	1·572	1·576	1·580	1·586	1·593	1·600	1·610	1·620
6·4	1·556	1·557	1·559	1·563	1·567	1·572	1·579	1·586	1·595	1·605
6·5	1·543	1·545	1·547	1·550	1·554	1·559	1·565	1·573	1·581	1·591
6·6	1·532	1·533	1·535	1·538	1·542	1·547	1·553	1·560	1·568	1·577
6·7	1·520	1·522	1·524	1·527	1·530	1·535	1·541	1·547	1·555	1·564
6·8	1·510	1·511	1·513	1·516	1·519	1·524	1·529	1·535	1·543	1·551
6·9	1·499	1·501	1·502	1·505	1·509	1·513	1·518	1·524	1·531	1·539
7·0	1·490	1·491	1·492	1·495	1·498	1·502	1·507	1·513	1·520	1·528
7·1	1·480	1·481	1·483	1·485	1·489	1·492	1·497	1·503	1·509	1·517
7·2	1·471	1·472	1·474	1·476	1·479	1·483	1·487	1·493	1·499	1·506
7·3	1·462	1·463	1·465	1·467	1·470	1·474	1·478	1·483	1·489	1·496
7·4	1·454	1·455	1·456	1·458	1·461	1·465	1·469	1·474	1·480	1·487
7·5	1·445	1·446	1·448	1·450	1·453	1·456	1·460	1·465	1·471	1·477
7·6	1·438	1·438	1·440	1·442	1·445	1·448	1·452	1·457	1·462	1·468
7·7	1·430	1·431	1·432	1·434	1·437	1·440	1·444	1·448	1·454	1·460
7·8	1·423	1·423	1·425	1·427	1·429	1·432	1·436	1·440	1·445	1·451
7·9	1·415	1·416	1·418	1·419	1·422	1·425	1·428	1·433	1·438	1·443
8·0	1·408	1·409	1·411	1·412	1·415	1·418	1·421	1·425	1·430	1·435
8·2	1·395	1·396	1·397	1·399	1·401	1·404	1·407	1·411	1·416	1·421
8·4	1·383	1·383	1·385	1·386	1·388	1·391	1·394	1·398	1·402	1·407
8·6	1·371	1·372	1·373	1·374	1·376	1·379	1·382	1·385	1·389	1·394
8·8	1·360	1·361	1·362	1·363	1·365	1·367	1·370	1·373	1·377	1·381
9·0	1·349	1·350	1·351	1·352	1·354	1·356	1·359	1·362	1·366	1·370
9·2	—	—	—	—	—	—	1·349	1·352	1·355	1·359
9·4	—	—	—	—	—	—	1·339	1·342	1·345	1·348

(From E. S. Pearson and H. O. Hartley, Biometrika Tables for Statisticians, Vol. 2, pp. 292-295, Cambridge University Press, 1972).

TABLE B-4 (continued)

Values of $\hat{\eta}$ (continued)

β_2 \ $\sqrt{\beta_1}$	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·95	1·00
3·8	3·424	3·776								
3·9	3·105	3·346								
4·0	2·872	3·049	3·294	3·659						
4·1	2·694	2·830	3·013	3·269						
4·2	2·552	2·662	2·804	2·996						
4·3	2·436	2·526	2·641	2·791						
4·4	2·338	2·414	2·510	2·631						
4·5	2·255	2·320	2·401	2·502						
4·6	2·183	2·240	2·309	2·395	2·503	2·641	2·828	3·093		
4·7	2·120	2·170	2·231	2·304	2·395	2·511	2·662	2·868		
4·8	2·065	2·109	2·162	2·226	2·305	2·403	2·529	2·694		
4·9	2·015	2·055	2·100	2·159	2·227	2·312	2·418	2·555		
5·0	1·971	2·007	2·049	2·099	2·160	2·234	2·325	2·441	2·592	2·799
5·1	1·931	1·963	2·001	2·045	2·100	2·165	2·245	2·344	2·472	2·641
5·2	1·894	1·924	1·958	1·999	2·048	2·105	2·176	2·262	2·371	2·512
5·3	1·860	1·888	1·919	1·957	2·000	2·052	2·115	2·191	2·285	2·406
5·4	1·830	1·855	1·884	1·918	1·958	2·005	2·061	2·128	2·211	2·315
5·5	1·801	1·824	1·851	1·883	1·918	1·962	2·012	2·073	2·146	2·237
5·6	1·775	1·796	1·821	1·850	1·884	1·923	1·969	2·023	2·089	2·170
5·7	1·750	1·770	1·794	1·820	1·851	1·887	1·929	1·979	2·038	2·110
5·8	1·728	1·746	1·768	1·793	1·821	1·855	1·893	1·939	1·992	2·057
5·9	1·706	1·724	1·744	1·767	1·794	1·824	1·860	1·902	1·951	2·009
6·0	1·686	1·703	1·722	1·743	1·768	1·797	1·830	1·868	1·913	1·967
6·1	1·667	1·683	1·701	1·721	1·744	1·771	1·802	1·837	1·879	1·928
6·2	1·649	1·664	1·681	1·700	1·722	1·747	1·776	1·809	1·847	1·892
6·3	1·633	1·647	1·663	1·681	1·701	1·725	1·752	1·782	1·818	1·860
6·4	1·617	1·630	1·645	1·662	1·682	1·704	1·729	1·758	1·791	1·830
6·5	1·602	1·614	1·629	1·645	1·663	1·684	1·707	1·735	1·766	1·802
6·6	1·587	1·599	1·613	1·628	1·646	1·666	1·688	1·713	1·742	1·776
6·7	1·574	1·585	1·598	1·613	1·629	1·648	1·669	1·693	1·721	1·752
6·8	1·561	1·572	1·584	1·598	1·614	1·632	1·652	1·674	1·700	1·730
6·9	1·548	1·559	1·571	1·584	1·599	1·616	1·635	1·656	1·681	1·708
7·0	1·537	1·547	1·558	1·571	1·585	1·601	1·619	1·639	1·663	1·689
7·1	1·525	1·535	1·546	1·558	1·572	1·587	1·604	1·623	1·645	1·670
7·2	1·514	1·524	1·534	1·546	1·559	1·573	1·590	1·608	1·629	1·653
7·3	1·504	1·513	1·523	1·534	1·547	1·561	1·576	1·594	1·614	1·636
7·4	1·494	1·503	1·512	1·523	1·535	1·548	1·563	1·580	1·599	1·620
7·5	1·484	1·493	1·502	1·512	1·524	1·537	1·551	1·567	1·585	1·605
7·6	1·475	1·483	1·492	1·502	1·513	1·525	1·539	1·555	1·572	1·591
7·7	1·466	1·474	1·483	1·492	1·503	1·515	1·528	1·543	1·559	1·578
7·8	1·458	1·465	1·473	1·483	1·493	1·504	1·517	1·531	1·547	1·56[illegible]
7·9	1·450	1·457	1·465	1·474	1·483	1·494	1·507	1·520	1·536	1·552
8·0	1·442	1·448	1·456	1·465	1·474	1·485	1·497	1·510	1·524	1·541
8·1	1·434	1·440	1·448	1·456	1·466	1·476	1·487	1·500	1·514	1·529
8·2	1·426	1·433	1·440	1·448	1·457	1·467	1·478	1·490	1·504	1·519
8·3	1·419	1·425	1·432	1·440	1·449	1·458	1·469	1·481	1·494	1·508
8·4	1·412	1·418	1·425	1·433	1·441	1·450	1·460	1·472	1·484	1·49[illegible]
8·5	1·405	1·411	1·418	1·425	1·433	1·442	1·452	1·463	1·475	1·489
8·6	1·399	1·405	1·411	1·418	1·426	1·435	1·444	1·455	1·467	1·479
8·7	1·392	1·398	1·404	1·411	1·419	1·427	1·437	1·447	1·458	1·47[illegible]
8·8	1·386	1·392	1·398	1·404	1·412	1·420	1·429	1·439	1·450	1·46[illegible]
8·9	1·380	1·386	1·391	1·398	1·405	1·413	1·422	1·431	1·442	1·45[illegible]
9·0	1·374	1·380	1·385	1·392	1·399	1·406	1·415	1·424	1·434	1·44[illegible]
9·2	1·363	1·368	1·373	1·379	1·386	1·393	1·401	1·410	1·420	1·43
9·4	1·353	1·357	1·362	1·368	1·374	1·381	1·389	1·397	1·406	1·41
9·6	—	—	—	1·357	1·363	1·370	1·377	1·385	1·394	1·40
9·8	—	—	—	1·347	1·353	1·359	1·366	1·373	1·381	1·39
10·0	—	—	—	1·337	1·343	1·349	1·355	1·362	1·370	1·37

TABLE B-4 (continued)

Values of $\hat{\eta}$ (continued)

β_2 \ $\sqrt{\beta_1}$	1·05	1·10	1·15	1·20	1·25	1·30	1·35	1·40	1·45	1·50
5·4	2·450	2·632								
5·5	2·353	2·505								
5·6	2·270	2·400								
5·7	2·199	2·311								
5·8	2·136	2·234	2·362	2·530						
5·9	2·080	2·168	2·278	2·423						
6·0	2·031	2·109	2·206	2·331						
6·1	1·986	2·056	2·143	2·253						
6·2	1·945	2·009	2·087	2·184	2·309	2·476				
6·3	1·908	1·966	2·037	2·124	2·234	2·378				
6·4	1·875	1·928	1·992	2·070	2·168	2·294				
6·5	1·843	1·893	1·951	2·022	2·109	2·221				
6·6	1·815	1·860	1·914	1·978	2·057	2·156				
6·7	1·788	1·830	1·880	1·939	2·011	2·100				
6·8	1·763	1·803	1·849	1·903	1·969	2·049	2·151	2·281		
6·9	1·740	1·777	1·820	1·870	1·930	2·003	2·094	2·210		
7·0	1·719	1·753	1·793	1·840	1·895	1·962	2·044	2·148		
7·1	1·698	1·731	1·768	1·811	1·863	1·924	1·999	2·093		
7·2	1·679	1·710	1·745	1·785	1·833	1·890	1·958	2·043		
7·3	1·661	1·690	1·723	1·761	1·806	1·858	1·921	1·998		
7·4	1·644	1·671	1·703	1·738	1·780	1·829	1·887	1·958	2·046	2·157
7·5	1·628	1·654	1·683	1·717	1·756	1·802	1·856	1·921	2·001	2·102
7·6	1·613	1·637	1·665	1·697	1·734	1·776	1·827	1·887	1·960	2·051
7·7	1·598	1·622	1·648	1·678	1·713	1·753	1·800	1·856	1·923	2·006
7·8	1·584	1·607	1·632	1·660	1·693	1·731	1·775	1·827	1·889	1·965
7·9	1·571	1·593	1·616	1·644	1·675	1·710	1·751	1·800	1·858	1·928
8·0	1·559	1·579	1·602	1·628	1·657	1·691	1·730	1·775	1·829	1·894
8·1	1·547	1·566	1·588	1·613	1·640	1·672	1·709	1·752	1·802	1·862
8·2	1·535	1·554	1·575	1·598	1·625	1·655	1·690	1·730	1·777	1·833
8·3	1·524	1·542	1·562	1·585	1·610	1·639	1·671	1·709	1·754	1·806
8·4	1·514	1·531	1·550	1·571	1·596	1·623	1·654	1·690	1·732	1·781
8·5	1·504	1·520	1·538	1·559	1·582	1·608	1·638	1·672	1·711	1·758
8·6	1·494	1·510	1·527	1·547	1·569	1·594	1·623	1·655	1·692	1·736
8·7	1·484	1·500	1·517	1·536	1·557	1·581	1·608	1·639	1·674	1·715
8·8	1·475	1·490	1·507	1·525	1·545	1·568	1·594	1·623	1·657	1·695
8·9	1·467	1·481	1·497	1·514	1·534	1·556	1·580	1·608	1·640	1·677
9·0	1·458	1·472	1·487	1·504	1·523	1·544	1·568	1·594	1·625	1·660
9·1	1·450	1·463	1·478	1·495	1·513	1·533	1·556	1·581	1·610	1·643
9·2	1·442	1·455	1·469	1·485	1·503	1·522	1·544	1·568	1·596	1·628
9·3	1·435	1·447	1·461	1·476	1·493	1·512	1·533	1·556	1·583	1·613
9·4	1·427	1·440	1·453	1·468	1·484	1·502	1·522	1·545	1·570	1·599
9·5	1·420	1·432	1·445	1·459	1·475	1·492	1·512	1·534	1·558	1·586
9·6	1·413	1·425	1·437	1·451	1·466	1·483	1·502	1·523	1·546	1·573
9·7	1·407	1·418	1·430	1·443	1·458	1·474	1·492	1·513	1·535	1·560
9·8	1·400	1·411	1·423	1·436	1·450	1·466	1·483	1·503	1·524	1·549
9·9	1·394	1·404	1·416	1·428	1·442	1·458	1·474	1·493	1·514	1·538
10·0	1·388	1·398	1·409	1·421	1·435	1·450	1·466	1·484	1·504	1·527
10·1	1·382	1·392	1·403	1·414	1·428	1·442	1·458	1·475	1·495	1·516
10·2	1·376	1·386	1·396	1·408	1·420	1·434	1·450	1·467	1·485	1·506
10·3	1·371	1·380	1·390	1·401	1·414	1·427	1·442	1·458	1·477	1·497
10·4	1·365	1·374	1·384	1·395	1·407	1·420	1·435	1·450	1·468	1·488
10·5	1·360	1·369	1·378	1·389	1·401	1·413	1·427	1·443	1·460	1·479
10·6	1·355	1·363	1·373	1·383	1·394	1·407	1·420	1·435	1·452	1·470
10·7	1·349	1·358	1·367	1·377	1·388	1·400	1·414	1·428	1·444	1·462
10·8	1·345	1·353	1·362	1·372	1·382	1·394	1·407	1·421	1·437	1·454
10·9	1·340	1·348	1·357	1·366	1·377	1·388	1·401	1·414	1·429	1·446
11·0	1·335	1·343	1·352	1·361	1·371	1·382	1·394	1·408	1·422	1·439
11·2	1·326	1·334	1·342	1·351	1·360	1·371	1·383	1·395	1·409	1·424
11·4	1·318	1·325	1·332	1·341	1·350	1·360	1·371	1·383	1·396	1·411
11·6	—	—	—	—	—	—	—	1·372	1·384	1·398
11·8	—	—	—	—	—	—	—	1·361	1·373	1·386
12·0	—	—	—	—	—	—	—	1·351	1·363	1·375

TABLE B-4 (continued)

Values of $\hat{\eta}$ (continued)

β_2 \ $\sqrt{\beta_1}$	1·55	1·60	1·65	1·70	1·75	1·80	1·85	1·90	1·95	2·00
8·0	1·974	2·074								
8·1	1·936	2·028								
8·2	1·901	1·985								
8·3	1·869	1·947								
8·4	1·840	1·911								
8·5	1·812	1·879								
8·6	1·787	1·849	1·925	2·018						
8·7	1·763	1·821	1·891	1·978						
8·8	1·741	1·795	1·860	1·940						
8·9	1·720	1·771	1·831	1·906						
9·0	1·700	1·748	1·805	1·874						
9·1	1·682	1·727	1·780	1·844						
9·2	1·664	1·707	1·757	1·817						
9·3	1·648	1·688	1·735	1·792						
9·4	1·632	1·670	1·715	1·768	1·831	1·910				
9·5	1·617	1·653	1·696	1·745	1·805	1·879				
9·6	1·603	1·637	1·678	1·725	1·781	1·849				
9·7	1·589	1·622	1·660	1·705	1·758	1·822				
9·8	1·576	1·608	1·644	1·686	1·736	1·796				
9·9	1·564	1·594	1·629	1·669	1·716	1·772				
10·0	1·552	1·581	1·614	1·652	1·697	1·750				
10·1	1·541	1·569	1·600	1·636	1·679	1·729				
10·2	1·530	1·557	1·587	1·621	1·662	1·709	1·766	1·834		
10·3	1·520	1·545	1·574	1·607	1·646	1·691	1·744	1·809		
10·4	1·509	1·534	1·562	1·594	1·630	1·673	1·723	1·784		
10·5	1·500	1·523	1·550	1·581	1·616	1·656	1·704	1·761		
10·6	1·491	1·513	1·539	1·568	1·602	1·640	1·686	1·740		
10·7	1·482	1·504	1·528	1·556	1·588	1·625	1·669	1·720		
10·8	1·473	1·494	1·518	1·545	1·576	1·611	1·652	1·701		
10·9	1·465	1·485	1·508	1·534	1·564	1·597	1·637	1·683		
11·0	1·456	1·476	1·499	1·524	1·552	1·584	1·622	1·666	1·717	1·780
11·1	1·449	1·468	1·489	1·514	1·541	1·572	1·608	1·649	1·699	1·758
11·2	1·441	1·460	1·481	1·504	1·530	1·560	1·594	1·634	1·681	1·737
11·3	1·434	1·452	1·472	1·494	1·520	1·549	1·581	1·619	1·664	1·717
11·4	1·427	1·444	1·464	1·485	1·510	1·538	1·569	1·605	1·648	1·698
11·5	1·420	1·437	1·456	1·477	1·500	1·527	1·557	1·592	1·633	1·681
11·6	1·413	1·430	1·448	1·468	1·491	1·517	1·546	1·580	1·618	1·664
11·7	1·407	1·423	1·441	1·460	1·482	1·507	1·535	1·567	1·604	1·648
11·8	1·400	1·416	1·433	1·452	1·474	1·498	1·525	1·556	1·591	1·633
11·9	1·394	1·409	1·426	1·445	1·466	1·489	1·515	1·545	1·579	1·618
12·0	1·388	1·403	1·419	1·437	1·458	1·480	1·505	1·534	1·567	1·605
12·1	1·383	1·397	1·413	1·430	1·450	1·471	1·496	1·524	1·555	1·591
12·2	1·377	1·391	1·406	1·423	1·442	1·463	1·487	1·514	1·544	1·579
12·3	1·371	1·385	1·400	1·417	1·435	1·455	1·478	1·504	1·533	1·567
12·4	1·366	1·379	1·394	1·410	1·428	1·448	1·470	1·495	1·523	1·555
12·5	1·361	1·374	1·388	1·404	1·421	1·440	1·462	1·486	1·513	1·544
12·6	1·356	1·369	1·382	1·398	1·415	1·433	1·454	1·477	1·504	1·534
12·7	1·351	1·363	1·377	1·392	1·408	1·426	1·446	1·469	1·495	1·523
12·8	1·346	1·358	1·371	1·386	1·402	1·420	1·439	1·461	1·486	1·514
12·9	1·341	1·353	1·366	1·380	1·396	1·413	1·432	1·453	1·477	1·504
13·0	1·337	1·348	1·361	1·375	1·390	1·407	1·425	1·446	1·469	1·495
13·2	—	—	—	—	1·379	1·394	1·412	1·431	1·453	1·478
13·4	—	—	—	—	1·368	1·383	1·400	1·418	1·438	1·461
13·6	—	—	—	—	1·358	1·372	1·388	1·405	1·425	1·446
13·8	—	—	—	—	1·348	1·362	1·377	1·393	1·412	1·432
14·0	—	—	—	—	1·339	1·352	1·366	1·382	1·399	1·419
14·2	—	—	—	—	1·330	1·342	1·356	1·371	1·388	1·406
14·4	—	—	—	—	1·321	1·333	1·346	1·361	1·377	1·394
14·6	—	—	—	—	1·313	1·325	1·337	1·351	1·366	1·383
14·8	—	—	—	—	1·305	1·316	1·328	1·342	1·356	1·372
15·0	—	—	—	—	1·298	1·308	1·320	1·333	1·346	1·362

TABLE B-5

Percentiles of the Chi-Squared Distribution

$$F(u) = \int_0^u \frac{x^{(n-2)/2} e^{-x/2}}{2^{n/2}\Gamma(n/2)} dk$$

Degrees of freedom (γ)	0.005	0.010	0.025	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.990	0.995	γ
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	0.0642	0.148	0.275	0.455	0.708	1.07	1.64	2.71	3.84	5.02	6.63	7.88	1
2	0.0100	0.0201	0.0506	0.103	0.211	0.446	0.713	1.02	1.39	1.83	2.41	3.22	4.61	5.99	7.38	9.21	10.6	2
3	0.0717	0.115	0.216	0.352	0.584	1.00	1.42	1.87	2.37	2.95	3.67	4.64	6.25	7.81	9.35	11.3	12.8	3
4	0.207	0.297	0.484	0.711	1.06	1.65	2.19	2.75	3.36	4.04	4.88	5.99	7.78	9.49	11.1	13.3	14.9	4
5	0.412	0.554	0.831	1.15	1.61	2.34	3.00	3.66	4.35	5.13	6.06	7.29	9.24	11.1	12.8	15.1	16.7	5
6	0.676	0.872	1.24	1.64	2.20	3.07	3.83	4.57	5.35	6.21	7.23	8.56	10.6	12.6	14.4	16.8	18.5	6
7	0.989	1.24	1.69	2.17	2.83	3.82	4.67	5.49	6.35	7.28	8.38	9.80	12.0	14.1	16.0	18.5	20.3	7
8	1.34	1.65	2.18	2.73	3.49	4.59	5.53	6.42	7.34	8.35	9.52	11.0	13.4	15.5	17.5	20.1	22.0	8
9	1.73	2.09	2.70	3.33	4.17	5.38	6.39	7.36	8.34	9.41	10.7	12.2	14.7	16.9	19.0	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.18	7.27	8.30	9.34	10.5	11.8	13.4	16.0	18.3	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	6.99	8.15	9.24	10.3	11.5	12.9	14.6	17.3	19.7	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	7.81	9.03	10.2	11.3	12.6	14.0	15.8	18.5	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	8.63	9.93	11.1	12.3	13.6	15.1	17.0	19.8	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	9.47	10.8	12.1	13.3	14.7	16.2	18.2	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	10.3	11.7	13.0	14.3	15.7	17.3	19.3	22.3	25.0	27.5	30.6	32.8	15
16	5.14	5.81	6.91	7.96	9.31	11.2	12.6	14.0	15.3	16.8	18.4	20.5	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.1	12.0	13.5	14.9	16.3	17.8	19.5	21.6	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.9	12.9	14.4	15.9	17.3	18.9	20.6	22.8	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.1	11.7	13.7	15.4	16.9	18.3	19.9	21.7	23.9	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.9	12.4	14.6	16.3	17.8	19.3	21.0	22.8	25.0	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.3	11.6	13.2	15.4	17.2	18.8	20.3	22.0	23.9	26.2	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	11.0	12.3	14.0	16.3	18.1	19.7	21.3	23.0	24.9	27.3	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.2	11.7	13.1	14.8	17.2	19.0	20.7	22.3	24.1	26.0	28.4	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.9	12.4	13.8	15.7	18.1	19.9	21.7	23.3	25.1	27.1	29.6	33.2	36.4	39.4	43.0	45.6	24
25	10.5	11.5	13.1	14.6	16.5	18.9	20.9	22.6	24.3	26.1	28.2	30.7	34.4	37.7	40.6	44.3	46.9	25
26	11.2	12.2	13.8	15.4	17.3	19.8	21.8	23.6	25.3	27.2	29.2	31.8	35.6	38.9	41.9	45.6	48.3	26
27	11.8	12.9	14.6	16.2	18.1	20.7	22.7	24.5	26.3	28.2	30.3	32.9	36.7	40.1	43.2	47.0	49.6	27
28	12.5	13.6	15.3	16.9	18.9	21.6	23.6	25.5	27.3	29.2	31.4	34.0	37.9	41.3	44.5	48.3	51.0	28
29	13.1	14.3	16.0	17.7	19.8	22.5	24.6	26.5	28.3	30.3	32.5	35.1	39.1	42.6	45.7	49.6	52.3	29
30	13.8	15.0	16.8	18.5	20.6	23.4	25.5	27.4	29.3	31.3	33.5	36.3	40.3	43.8	47.0	50.9	53.7	30
35	17.2	18.5	20.6	22.5	24.8	27.8	30.2	32.3	34.3	36.5	38.9	41.8	46.1	49.8	53.2	57.3	60.3	35
40	20.7	22.2	24.4	26.5	29.1	32.3	34.9	37.1	39.3	41.6	44.2	47.3	51.8	55.8	59.3	63.7	66.8	40
45	24.3	25.9	28.4	30.6	33.4	36.9	39.6	42.0	44.3	46.8	49.5	52.7	57.5	61.7	65.4	70.0	73.2	45
50	28.0	29.7	32.4	34.8	37.7	41.4	44.3	46.9	49.3	51.9	54.7	58.2	63.2	67.5	71.4	76.2	79.5	50
75	47.2	49.5	52.9	56.1	59.8	64.5	68.1	71.3	74.3	77.5	80.9	85.1	91.1	96.2	100.8	106.4	110.3	75
100	67.3	70.1	74.2	77.9	82.4	87.9	92.1	95.8	99.3	102.9	106.9	111.7	118.5	124.3	129.6	135.6	140.2	100

(From G. J. Hahn and S. S. Shapiro, Statistical Models in Engineering, John Wiley & Sons, New York, 1967, pp. 314-315.)

TABLE B-6
Percentiles of the Maximum Absolute Difference Between Sample and Population Cumulative Distributions*

Sample size (*N*)	Level of significance (α)				
	0.20	0.15	0.10	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.510	0.565	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.368	0.410	0.490
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.450
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
over 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

*Values of $d_\alpha(N)$ such that

$$\Pr[\max|S_N(x) - F_0(x)|>d_\alpha(N)] = \alpha ,$$

where $F_0(x)$ is the theoretical cumulative distribution and $S_N(x)$ is an observed cumulative distribution for a sample of N observations.

(From F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," J. Amer. Stat. Ass. 46: 70 (1951).)

TABLE B-7

Table of Coefficients $[a_{n-1+1}]$ Used in the W Test for Normality

i \ n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886
2		0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253
3				0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553
4						0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027
5								0.0399	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587
6										0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197
7												0.0240	0.0433	0.0593	0.0725	0.0837
8														0.0196	0.0359	0.0496
9																0.0163

i \ n	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0.4808	0.4734	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254	0.4220	0.4188	0.4156	0.4127
2	0.3232	0.3211	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944	0.2921	0.2898	0.2876	0.2854
3	0.2561	0.2565	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487	0.2475	0.2463	0.2451	0.2439
4	0.2059	0.2085	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148	0.2145	0.2141	0.2137	0.2132
5	0.1641	0.1686	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870	0.1874	0.1878	0.1880	0.1882
6	0.1271	0.1334	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630	0.1641	0.1651	0.1660	0.1667
7	0.0932	0.1013	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415	0.1433	0.1449	0.1463	0.1475
8	0.0612	0.0711	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219	0.1243	0.1265	0.1284	0.1301
9	0.0303	0.0422	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036	0.1066	0.1093	0.1118	0.1140
10		0.0140	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862	0.0899	0.0931	0.0961	0.0988
11				0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697	0.0739	0.0777	0.0812	0.0844
12						0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537	0.0585	0.0629	0.0669	0.0706
13								0.0094	0.0178	0.0253	0.0320	0.0381	0.0435	0.0485	0.0530	0.0572
14										0.0084	0.0159	0.0227	0.0289	0.0344	0.0395	0.0441
15												0.0076	0.0144	0.0206	0.0262	0.0314
16														0.0068	0.0131	0.0187
17																0.0062

i \ n	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1883	0.1883	0.1883	0.1881	0.1880	0.1878	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18		0.0057	0.0110	0.0158	0.0203	0.0244	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19				0.0053	0.0101	0.0146	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20						0.0049	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21								0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22										0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23												0.0039	0.0076	0.0111	0.0143	0.0174
24														0.0037	0.0071	0.0104
25																0.0035

(From G. J. Hahn and S. S. Shapiro, Statistical Methods in Engineering, John Wiley & Sons, New York, 1967, pp. 330-331).

TABLE B-8
Percentage Points of the W Statistic

n	1	2	5	10	50
3	0.753	0.756	0.767	0.789	0.959
4	0.687	0.707	0.748	0.792	0.935
5	0.686	0.715	0.762	0.806	0.927
6	0.713	0.743	0.788	0.826	0.927
7	0.730	0.760	0.803	0.838	0.928
8	0.749	0.778	0.818	0.851	0.932
9	0.764	0.791	0.829	0.859	0.935
10	0.781	0.806	0.842	0.869	0.938
11	0.792	0.817	0.850	0.876	0.940
12	0.805	0.828	0.859	0.883	0.943
13	0.814	0.837	0.866	0.889	0.945
14	0.825	0.846	0.874	0.895	0.947
15	0.835	0.855	0.881	0.901	0.950
16	0.844	0.863	0.887	0.906	0.952
17	0.851	0.869	0.892	0.910	0.954
18	0.858	0.874	0.897	0.914	0.956
19	0.863	0.879	0.901	0.917	0.957
20	0.868	0.884	0.905	0.920	0.959
21	0.873	0.888	0.908	0.923	0.960
22	0.878	0.892	0.911	0.926	0.961
23	0.881	0.895	0.914	0.928	0.962
24	0.884	0.898	0.916	0.930	0.963
25	0.888	0.901	0.918	0.931	0.964
26	0.891	0.904	0.920	0.933	0.965
27	0.894	0.906	0.923	0.935	0.965
28	0.896	0.908	0.924	0.936	0.966
29	0.898	0.910	0.926	0.937	0.966
30	0.900	0.912	0.927	0.939	0.967
31	0.902	0.914	0.929	0.940	0.967
32	0.904	0.915	0.930	0.941	0.968
33	0.906	0.917	0.931	0.942	0.968
34	0.908	0.919	0.933	0.943	0.969
35	0.910	0.920	0.934	0.944	0.969
36	0.912	0.922	0.935	0.945	0.970
37	0.914	0.924	0.936	0.946	0.970
38	0.916	0.925	0.938	0.947	0.971
39	0.917	0.927	0.939	0.948	0.971
40	0.919	0.928	0.940	0.949	0.972
41	0.920	0.929	0.941	0.950	0.972
42	0.922	0.930	0.942	0.951	0.972
43	0.923	0.932	0.943	0.951	0.973
44	0.924	0.933	0.944	0.952	0.973
45	0.926	0.934	0.945	0.953	0.973
46	0.927	0.935	0.945	0.953	0.974
47	0.928	0.936	0.946	0.954	0.974
48	0.929	0.937	0.947	0.954	0.974
49	0.929	0.937	0.947	0.955	0.974
50	0.930	0.938	0.947	0.955	0.974

(From G. J. Hahn and S. S. Shapiro, Statistical Methods in Engineering, John Wiley & Sons, New York, 1967, p. 332.)

TABLE B-9
Percentage Points for the WE Statistic

	95% Range		90% Range	
n	Lower Point	Upper Point	Lower Point	Upper Point
7	0.062	0.404	0.071	0.358
8	0.054	0.342	0.062	0.301
9	0.050	0.301	0.058	0.261
10	0.049	0.261	0.056	0.231
11	0.046	0.234	0.052	0.208
12	0.044	0.215	0.050	0.191
13	0.040	0.195	0.046	0.173
14	0.038	0.178	0.043	0.159
15	0.036	0.163	0.040	0.145
16	0.034	0.150	0.038	0.134
17	0.030	0.135	0.034	0.120
18	0.028	0.123	0.031	0.109
19	0.026	0.114	0.029	0.102
20	0.025	0.106	0.028	0.095
21	0.024	0.101	0.027	0.091
22	0.023	0.094	0.026	0.084
23	0.022	0.087	0.025	0.078
24	0.021	0.082	0.024	0.074
25	0.021	0.078	0.023	0.070
26	0.020	0.073	0.022	0.066
27	0.020	0.070	0.022	0.063
28	0.019	0.067	0.021	0.061
29	0.019	0.064	0.021	0.058
30	0.018	0.060	0.020	0.054
31	0.017	0.057	0.019	0.052
32	0.017	0.055	0.019	0.050
33	0.017	0.053	0.018	0.048
34	0.017	0.051	0.018	0.047
35	0.016	0.049	0.018	0.045

(From G. J. Hahn and S. S. Shapiro, Statistical Methods in Engineering, John Wiley & Sons, New York, 1967, p. 335.)

TABLE B-10
Percentage Points for the WE_o Statistic

	95% Range		90% Range	
n	Lower Point	Upper Point	Lower Point	Upper Point
7	0.025	0.260	0.033	0.225
8	0.025	0.230	0.032	0.200
9	0.025	0.205	0.031	0.177
10	0.025	0.184	0.030	0.159
11	0.025	0.166	0.030	0.145
12	0.025	0.153	0.029	0.134
13	0.025	0.140	0.028	0.124
14	0.024	0.128	0.027	0.115
15	0.024	0.119	0.026	0.106
16	0.023	0.113	0.025	0.098
17	0.023	0.107	0.024	0.093
18	0.022	0.101	0.024	0.087
19	0.022	0.096	0.023	0.083
20	0.021	0.090	0.023	0.077
21	0.020	0.085	0.022	0.074
22	0.020	0.080	0.022	0.069
23	0.019	0.075	0.021	0.065
24	0.019	0.069	0.021	0.062
25	0.018	0.065	0.020	0.058
26	0.018	0.062	0.020	0.056
27	0.017	0.058	0.020	0.054
28	0.017	0.056	0.019	0.052
29	0.016	0.054	0.019	0.050
30	0.016	0.053	0.019	0.048
31	0.016	0.051	0.018	0.047
32	0.015	0.050	0.018	0.045
33	0.015	0.048	0.018	0.044
34	0.014	0.046	0.017	0.043
35	0.014	0.045	0.017	0.041

(From G. J. Hahn and S. S. Shapiro, Statistical Methods in Engineering, John Wiley & Sons, New York, 1967, p. 334.)

APPENDIX C

REFERENCES AND ABSTRACTED BIBLIOGRAPHY

1. Bain, L.J., and C. E. Antle, "Estimation of Parameters in the Weibull Distribution", Technometrics 9(4):621-627 (1967).

A new method of estimation is used to obtain two simple estimators of the parameters in a Weibull distribution. These estimators are similar to the estimators given by Gumbel, Miller and Freund, and Menon. Monte Carlo methods were used to determine the variances and biases of the estimators for various sample sizes. Comparisons of the estimators can be made and unbiasing factors calculated in some cases.

2. Bhattacharya, P.K., "Efficient Estimation of a Shift Parameter From Grouped Data", Ann. Math. Statist. 38:1770-1787 (1967).

This paper considers two populations having frequency functions $f(x)$ and $f(x-\theta)$ where the common form f and the shift parameter θ are unknown. A method of estimating θ when one sample is reduced to a frequency distribution over a given set of class-intervals is suggested by the likelihood principle and the asymptotic efficiency of this estimator relative to the appropriate maximum likelihood estimator based on the complete data is found to be the ratio of the Fisher-information in a grouped observation to the Fisher-information in an ungrouped observation.

3. Birnbaum, Z. W., Probability and Mathematical Statistics, Harper & Brothers, New York (1962).

General theory of tests of statistical hypotheses is presented along with a detailed discussion of the Chi-squared distribution and test. Also distribution free tests are discussed including the Kolmogorov test and Smirnov test. Also included are the likelihood function and likelihood ratio statistics.

4. Brunk, H.D., Mathematical Statistics, Blaisdell Publishing Co., Waltham, Massachusetts (1965).

Basic theory of testing hypotheses is presented including a discussion of testing a simple hypothesis against a simple alternative, choice of null hypothesis, the power function, most powerful tests and consistent tests. Specific tests described

are Chi-squared test, Kolmogorov-Smirnov test for goodness of it, t-test, F-test, runs test, median test, and likelihood ratio test.

5. Choi, S.C., and R. Wette, "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias", Technometrics 11(4):683-690 (1969).

The maximum likelihood method is recommended for estimating the parameters of a gamma distribution. Numerical techniques for carrying out the calculation are examined. A convenient table is obtained to facilitate the estimation of parameters. The bias of the estimates is investigated by Monte Carlo; the indication is that the bias of both parameter estimates produced by the maximum likelihood method is positive.

6. Cornell, R.G., and J.A. Speckman, "Estimation for a Simple Exponential Model", Biometrics 23:717-737 (1967).

Graphical, maximum likelihood, least squares, weighted least squares, partial totals, moment, finite differences, Fisher, and Spearman estimation procedures are presented for estimating the parameter λ in the exponential model with expectations given by $1 - e^{-\lambda T}$ for different values of T. The estimators are described, referenced, illustrated, and compared. Tables are cited which make several of the estimation procedures easier computationally. Included in the comparison of the estimators is a review of some Monte Carlo computations. The method of maximum likelihood, which can be used for any spacing of T-values, has very desirable large sample properties. The simple method of partial totals is a possible alternative for small samples of equally spaced T-values while the Fisher and Spearman method are suggested alternatives for T-values whose logarithms are equally spaced.

7. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, Princeton (1945).

Chapter 30 of this book describes "goodness of fit" statistical tests. The two tests described in detail are the Chi-squared test and Crámer-von Mises test. However, statistics for the Crámer-von Mises test and examples are not presented.

8. Dubey, S.D., "On Some Permissible Estimators of the Location Parameter of the Weibull and Certain Other Distributions", Technometrics 9(2):293-307 (1967).

An estimator for the location parameter of the Weibull distribution is proposed which is independent of its shape and scale parameters. Several properties of this estimator are established which suggest a proper choice of three ordered sample observations insuring a permissible estimate of the location parameter. This result is valid for every distribution which has the location parameter acting as the origin or threshold parameter. Asymptotic properties of such an estimator of the location parameter of the Weibull distribution is discussed. Finally the paper contains a brief discussion on a percentile estimator of the location parameter of the Weibull distribution and includes some numerical illustrations.

9. Elandt, R.C., "The Folded Normal Distribution: Two Methods of Estimating Parameters From Moments", Technometrics 3(4):551-562 (1961).

The general formula for the r^{th} moment of the folded normal distribution is obtained, and formulae for the first four non-central and central moments are calculated explicitly. Two methods, one using first and second moments of the sample and the other using second and fourth moments, of estimating the parameters of the parent distribution are presented and their standard errors calculated. The accuracy of both methods is discussed.

10. Elderton, W.P., Frequency Curves and Correlation, 4th Ed., Cambridge University Press, Cambridge, (1953).

A thorough covering of the Pearson system. Describes each type of distribution and gives relevant formulae for the type of curve.

11. El-Sayyad, G.M., "Information and Sampling from the Exponential Distribution", Technometrics 11(1):41-45 (1969).

Methods of sampling an exponential population in order to obtain a prescribed accuracy in the determination of the unknown parameter are discussed. The concept of information due to Shannon is used and it leads to well-known schemes.

12. Gnanadesikan, R., R.S. Pinkham, and L.P. Hughes, "Maximum Likelihood Estimation of the Parameters of the Beta Distribution from Smallest Order Statistics", Technometrics 9(4):607-620 (1967).

Numerical methods, useful with high-speed computers are described for obtaining the maximum likelihood estimates of the two parameters of a beta distribution using the smallest M observations, $0 < u_1 < u_2 < \ldots < \ldots u_M$, in a random sample of size K ($\geq$ M). The maximum likelihood estimates are functions only of the ratio R + M/K, the Mth ordered observation, u_M, and the two statistics, $G = [\Pi_{i=1}^{M} u_i]^{1/M}$ and $G_2 = [\Pi_{i=1}^{M} (1-u_i)]^{1/M}$. For the case of the complete sample (R = 1), however, the estimates are functions only of G_1 and G_2, and hence, for this case, explicit tables of the estimates are provided.

Some examples are given of the use of the procedures described for fitting beta distributions to sets of data.

13. Govindarajulu, Z., "Certain General Properties of Unbiased Estimates of Location and Scale Parameters Based on Ordered Observations", SIAM J. App. Math. 16(3):533-551 (1968).

Some upper bounds are derived for the variances of least squares estimators based on a subset of the ordered observations in a random sample of (i) location, (ii) scale, and (iii) both location and scale parameters of a distribution.

14. Gumbel, E.J., "Statistical Theory of Extreme Values and Some Practical Applications", National Bureau of Standards, Applied Math Series 33, (Feb. 1954).

15. Hahn, G.J., and S.S. Shapiro, Statistical Models in Engineering, John Wiley and Sons, New York, 1967 (1967).

Discusses many continuous and discrete distributions. Gives functional form, discusses theoretical basis, and mentions applications. In some cases describes parameter estimation

techniques. Discusses advantages to fitting data to empirical distributions. Describes Johnson system and displays plot of β_1, β_2 values. Fitting procedures for Johnson distributions are outlined and examples are given. Describes Pearson system of distributions and displays β_1, β_2 plot. Does not attempt to describe Pearson fitting procedures.

Discusses general techniques of goodness of fit tests. Two procedures are discussed: a series of tests developed by Shapiro and Wilk, known as W tests (including the WE test), and the Chi-squared goodness of fit test. The W tests are used to evaluate the assumption of a normal and exponential distribution for a set of data. The procedures for using these techniques are presented in a detailed step-by-step manner.

16. Haight, F.A., Index to Distributions of Mathematical Statistics, J. Res. Natl. Bureau Stand. - B. Math. and Math. Phys 65B (1):23-60 (1961).

A fairly complete index of references to results on statistical distributions published before January 1958 is presented. The material given for each distribution is a list of references relating to: (a) functions and constants which characterize the distribution, (b) derived distributions, (c) estimation, (d) testing statistical hypotheses, and (e) miscellaneous. The distributions covered are characterized as normal, type III, binomial, discrete, distributions over (a,b), distributions over (a,∞), distributions over $(-\infty,\infty)$, miscellaneous univariate, miscellaneous bivariate, and miscellaneous multivariate. The number of entries varies from one or two for less well-known distributions to several hundred for the normal distribution.

17. Harter, H.L., "Maximum-Likelihood Estimation of the Parameters of a Four-Parameter Generalized Gamma Population From Complete and Censored Samples", Technometrics 9 (1):159-165 (1967).

The four-parameter generalized gamma distribution includes such distributions as the usual three-parameter gamma, the Weibull, the exponential, and the half normal. For these distributions this paper develops the maximum likelihood equations. Iterative computer techniques are needed to solve these equations. Some results of applying this to various distributions are presented.

18. Harter, H.L., "A New Table of Percentage Points of the Pearson Type III Distribution", Technometrics 11(1):177-187 (1969).

A table of percentage points for the type III Pearson distribution.

19. Hodges, J.L., Jr. and E.L. Lehmann, "A Compact Table For Power of the t-Test", Ann. Math. Statist, 39, No. 5 (1968)

The paper gives a one-page table for t-power which covers any values of the (one-sided) significance level α in the range from 0.005 to 0.1, any value of the second-type error probability β in the range from 0.01 to 0.5; and any number of degrees of freedom greater than 2. The table gives reasonably accurate answers without iteration and using only linear interpolation. Eight examples are provided which illustrate a variety of t-power problems.

20. Hogg, R.V. and A.T. Craig, Introduction to Mathematical Statistics, the Mac Millan Company, New York (1965).

Includes chapters on order statistics, sufficient statistics, statistical hypotheses and statistical tests. It provides the theoretical basis of the Chi-square tests and Bayesian tests. It also describes Likelihood Ratio tests and the sequential probability ratio test.

21. Johnson, N.L., "Systems of Frequency Curves Generated by Methods of Translation", Biometrika 36:149-176 (1949).

Introduces Johnson system of distributions. Reviews literature on systems of distributions. Provides a theoretical background to Johnson system. Compares Johnson and Pearson systems for skewness and kurtosis values. Gives some numerical examples.

22. Johnson, N.L., "Tables to Facilitate Fitting S_U Frequency Curves", Biometrika 52:547 (1965).

In fitting empirical data to a distribution from the Johnson family, one usually adjusts the parameters of the Johnson distribution to match the first four moments of the original data. However, given the first four moments it is not a trivial problem to calculate the correct Johnson parameters. This paper provides tables from which the Johnson parameters can be obtained.

23. Johnson, N.L., and S. Katz, Distributions in Statistics: Discrete Distributions, Houghton-Mifflin Co., Boston, (1969).

Thorough covering of all known discrete distributions. Gives functional form, moments, and other information and discusses the estimation of parameters for each distribution.

24. Johnson, N.L., and S. Katz, Distributions in Statistics: Continuous Univariate Distributions, Vol. 1 and 2, Houghton-Mifflin Co., Boston, (1970).

Thorough covering of all known continuous distributions (except empirical families). Gives functional form, moments, and other information and discusses the estimation of parameters for each distribution.

25. Johnson, N.L., E. Nixon, D. E. Amon, and E. S. Pearson, "Table of Percentage Points of Pearson Curves", for given $\sqrt{\beta_1}$ and β_2, expressed in standard measure", Biometrika 50: 459-498 (1963).

For the general Pearson system of distributions, this paper gives tables of percentiles (or solutions of the inverse equation) as a function of skewness and kurtosis.

26. Kagan, A.M., "Estimation Theory for Families with Location and Scale Parameters and For Exponential Families", Proc. Steklov. Inst. Math. 104:19-87 (1968).

This theoretical paper investigates families of distributions and estimators. The conditions for admissible estimators are discussed.

27. Kendall, M.G., and A.S. Stuart, The Advanced Theory of Statistics, Vol. 1, Distribution Theory, Charles Griffen & Co. (1958).

28. Kodlin, D., "A New Response Time Distribution", Biometrics 23:227-239 (1967).

A skewed, two-parameter distribution is described which has been found useful in the analysis of human survival time data. The density has the form $f(t) = (c+kt)e^{-(ct+\frac{1}{2}kt^2)}$. This form is integrable and has manageable first and second moments. Since the distribution has non-zero density at the origin, it may be of value in connection with those types of responses which take place even before observation begins. Description of a maximum likelihood technique of estimating the parameters is followed by discussion of damage models that incorporate the distribution.

29. Langton, N.H., "Statistical Distribution", Brit. Chem, Engr. 8:478-484 (1963).

This paper is an elementary article which gives the basic concepts and formulae characterizing probability distributions and sampling. It discusses the binomial, Poisson, and normal distributions and the fitting of empirical data to these distributions using moments method.

30. Malik, H.J., "Estimation of the Parameters of the Pareto Distribution", Metrika 15:126-136 (1970).

In this paper, sufficient estimators for the parameters a and v of the Pareto distribution are obtained. It is shown that $Y_1 = \text{Min}(x_1, \ldots, x_N)$ is sufficient for a when v is known, the sample geometric mean g is sufficient for v when a is known; and $(Y_1, \sum_{i=1}^{N} \ell n \frac{Y_i}{Y_1})$ is a joint set of sufficient statistics for (a,v) when both are unknown. The exact distribution of the maximum likelihood estimator is derived.

31. Mandel, J., "A Method for Fitting Empirical Surfaces to Physical or Chemical Data", Technometrics 11(3):411-429 (1969).

A method, largely graphical, for fitting a distribution to bivariate data is presented. An example is given. The method does not require prior assumptions as to the form of the distribution to be fit. However, it may not have general applicability and needs further investigation.

32. Marshall, A.W., and I. Olkin, "A Multivariate Exponential Distribution", J. Amer. Stat. Assoc. 62:30-44 (1967).

A number of multivariate exponential distributions are known, but they have not been obtained by methods that shed light on their applicability. This paper presents some meaningful derivations of a multivariate exponential distribution that serves to indicate conditions under which the distribution is appropriate. Two of these derivations are based on "shock models", and one is based on the requirement that residual life is independent of age. It is significant that the derivations all lead to the same distribution.

For this distribution, the moment generating function is obtained, comparison is made with the case of independence, the distribution of the minimum is discussed, and various other properties are investigated. A multivariate Weibull distribution is obtained through a change of variables.

33. Massey, Frank J., Jr., "The Kolmogorov - Smirnov Test for Goodness of Fit", J. Am. Stat. Assoc., 46 (1951).

The Kolmogorov-Smirnov test which is based on the maximum difference between an empirical and hypothetical cumulative distribution is discussed. Percentage points are tabulated, and a lower bound to the power function is charted. Confidence units for a cumulative distribution are described. Examples are given. Indications that the test is superior to the Chi-square test are cited.

34. Mann, Nancy R., "Point and Interval Estimation Procedures for the Two-Parameter Weibull and Extreme-Value Distributions", Technometics 10(2):231-256 (1968).

Point estimators of parameters of the first asymptotic distributions of smallest (extreme) values, the extreme-value distribution, are surveyed and compared. Since the logarithms of variates having the two-parameter Weibull distribution are variates from the extreme-value distribution, the investigation is applicable to the estimation of Weibull parameters. Those estimators investigated are maximum-likelihood and moment estimators, inefficient estimators based on only a few ordered observations, and various linear estimation methods. A combination of Monte Carlo approximations and exact small-sample and asymptotic results has been used to compare the expected loss (with loss equal to squared error) of these various point estimators. Interval estimation procedures are also discussed.

35. McGrath, E.J., Fundamentals for Operations Research, West Coast University, 1970, Chapter 3.

Discussion of probability distributions and estimators for most basic distributions. Weibull - describes distribution and typical curves and discusses estimators for parameters. Johnson - defines distribution, displays typical curve shapes, and gives skewness - kurtosis diagram for family. Extensive discussion, with examples, of estimation of parameters. Pearson - defines distribution types and gives skewness-kurtosis plot for family. Discussion of χ^2-test for evaluation of fits.

36. Meier, F.A., "Non-Normal Statistical Distributions and Their Use in Industrial Engineering", Amer. Inst. of Indust. Eng., Tech. Papers, 20 Inst. Conf. and Conv. 71-83 (1969).

Both the gamma and Weibull distributions are described with comments on calculational methods and approximations. A thorough review of methods for estimating parameters is given.

37. Mengel, P.R., "Fragility Curve Preparation Methods", unpublished memo, 1970.

Presents a methodology for fitting data from failure levels to a lognormal distribution. Theoretical reasons underlying the use of the lognormal for this case are discussed.

38. Menon, M.W., "Estimation of the Shape and Scale Parameters of the Weibull Distribution", Technometrics 5(2):175-182 (1963).

Estimates $\hat{c}$ and $\hat{b}$ are proposed for the shape parameter c and the scale parameter b of the Weibull distribution on the assumption that the location parameter is known. First an estimate $\hat{d}$ of 1/c is found, the $\hat{c}$ is obtained as $1/\hat{d}$. When b is unknown, $\hat{d}$ is a consistent and non-negative estimate of d, with a bias which tends to vanish as the sample size increases and with an asymptotic efficiency of about 55%. When b is known, $\hat{d}$ is an unbiased, non-negative, and consistent estimate of d, and its efficiency is approximately 84%. An estimate $\widehat{\ln b}$ of $\ln b$ is found with an asymptotic efficiency of 95%. It is proposed that $\exp(\widehat{\ln b})$ be used to estimate b.

39. Neave, H.R. and C.W.J. Granger, "A Monte Carlo Study Comparing Various Two-Sample Tests for Differences in Mean", Technometrics, 10 (3) (1968).

A study was conducted on eight tests for differences in means under a variety of simulated experimental situations. Estimates were made of the power of the tests and measures made of the extent to which they gave similar results. In particular the performance of a new quick test developed by Neave was studied.

40. Pearson, K., "Mathematical Contributions to the Theory of Evolution - Supplement to a Memoir on Skew Variation", Trans. Roy. Phil. Soc. London 197:443-459 (1901).

One of the classic papers introducing some of the Pearson system distributions and giving some examples.

41. Pearson, K., "Mathematical Contributions to the Theory of Evolution - Second Supplement to a Memoir on Skew Variation", Trans. Roy. Phil. Soc. London A216:429-457 (1916).

Classical paper setting forth the properties of the Pearson system and the distributions in it.

42. Pearson, E.S., and H.O. Hartley (eds), Biometrika Tables for Statisticians, Vol. I, sections 23-24, Cambridge Univ. Press (1958).

The basic functional forms and some properties are given for each distribution in the Pearson system. Some applications showing the fitting to empirical data are discussed.

43. Pickands, J. III, "Efficient Estimation of a Probability Density Function", Ann. Math. Statist. 40(3):854-864 (1969).

Some theoretical results in using the "kernel method" to estimate a probability density function are derived.

44. Plait, Alan, "The Weibull Distribution - with Tables", Industrial Quality Control 19(5):17-26 (1962).

Describes Weibull distribution and gives extensive tables to aid in curve fitting.

45. Press, S.J., "The T-Ratio Distribution", J. Amer. Stat. Ass. 64:242-252 (1969).

The distribution of the ratio of correlated student T-variates is of interest in problems in econometrics and ranking and selection. The density of this ratio is derived and computer graphs of the density are given in terms of standardized variates. Fractiles are given for selected parameter values. It is shown that the distribution contains no moments.

46. Schwartz, S.C., "Estimation of a Probability Density by an Orthogonal Series", Ann. Math. Statist. 38:1261-1265 (1967).

The estimation of an unknown probability desnity function from a realized sequence of random numbers is considered. An approximation in terms of a sum of Hermite polynomials is made and equations for the coefficients are derived. Convergence to the correct density function is proven and convergence rates are calculated. Comparison to the kernel method is made.

47. Shapiro, S.S. and M.B. Wilk, "An Analysis of Variance Test For Normality (Complete Samples)", Biometrika, 52 (1965).

A new statistical procedure (W Test) for testing a complete sample for normality is presented. The test statistic is obtained by dividing the square of an appropriate linear combination of the sample order statistics by the usual symmetric estimate of variance. Presented are derivation, properties, and applications of the W test and comparison with other tests.

48. Suzuki, Giitiro, "On Exact Probabilities of Some Generalized Kilmogorov's D-Statistics", Institute on Statistical Mathematics, Annals, Tokyo, 19 (1967).

This paper gives a unified computational method for exact probabilities of the most generalized form of the D-statistic proposed by Kolmogorov for non-parametric tests of fit. First, a historical survey of the subject is given and then goodness-of-fit D-tests are stated (based on some general bounds) by constructing general acceptance and confidence regions, sizes of which are calculated in a distribution-free way. The method is also applied to calculation of the exact power of tests for a certain continuous alternative. A computational method for the functional $a_\eta(\ldots)$ is presented.

49. Takahasi, K., and K. Wakimoto, "On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering", Ann. Inst. on Stat. Math., Tokyo, 20:1-31 (1968).

In many experimental situations, it is costly and time-consuming to make accurate measurements while at the same time judgments as to relative order of size can be made easily. This paper describes techniques for ordering subgroups of a large sample, then picking a smaller sample, using the stratification induced by the ordering. Accurate measurements are made only on the smaller sample. An unbiased estimate of the population mean can be generated from this small sample with much less variance than would be obtained in estimating from a sample of similar size, but randomly chosen. This is basically an example of stratified sampling, but as applied prior to experimental measuring rather than to choices made in simulation.

50. Tarter, M.E., R.L. Holcomb, and R.A. Kronman, "After the Histogram, What? A Description of New Computer Methods for Estimating the Population Density", Proc. ACM 22nd Natl. Conf. P-67:511-519 (1967).

The kernel method for estimating a probability density function from a sequence of random observations is discussed. As an alternative, a Fourier expansion is considered for an estimate of the density. Restrictions on the function and the optimum order of the expansion is derived. Computer implementation of this algorithm is discussed and several applications are displayed.

51. Thoman, D.R., L.J. Bain, and C.E. Antle, "Inferences on the Parameters of the Weibull Distribution", Technometrics 11(3):445-460 (1969).

The problems of estimation and testing hypotheses regarding the parameters in the Weibull distribution are considered in this paper. The following results are given:

1. Exact confidence intervals for the parameters based on maximum likelihood estimators are presented.

2. A table of unbiasing factors (depending upon sample size) for the maximum likelihood estimator of the shape parameter are given.

3. Test of hypotheses regarding the parameters and the power of the test regarding the shape parameter are developed and presented.

4. Sample sizes at which large sample theory may be useful are presented.

52. Thornber, H., "Finite Sample Monte Carlo Studies: An Autoregressive Illustration", J. Amer. Stat. Assoc. 62:801-818 (1967).

In this paper the problem of choosing among point estimators on the basis of their small sample properties is discussed from the sampling point of view. The indeterminacy of most Monte Carlo studies is analyzed and resolved within the framework of statistical decision theory. A first order auto-regressive model is worked through in detail both for its own sake and to illustrate how a complete Monte Carlo study might be done.

53. Weibull, W., "A Statistical Distribution Function of Wide Applicability", J. App. Mech. 18(3):293-297 (1951).

Introduces the Weibull distribution and gives several examples of fitting to it.

54. Weiss, L., and J. Wolfowitz, "Maximum Probability Estimators", Ann. Inst. Stat. Math. Tokyo 19 193-206 (1967).

A new class of estimators, called maximum probability estimators, is suggested as an alternative to maximum likelihood estimators.

55. White, J.S., "The Moments of Log-Weibull Order Statistics", Technometrics 11:373-386 (1969).

Formulas for the moments of the order statistics of a general distribution are derived. Then the log-Weibull distribution is introduced and the moments of its order statistics are calculated. An application showing how this can be applied to the fitting of a Weibull distribution to empirical data is given.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Science Applications, Inc. 1250 Prospect Street La Jolla, California 92037	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP
3. REPORT TITLE Techniques for Efficient Monte Carlo Simulation Volume I: Selecting Probability Distributions	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Final Report	
5. AUTHOR(S) (First name, middle initial, last name) Elgie J. McGrath, Stanley L. Basin, Robert W. Burton, David C. Irving, Stratton C. Jaquette, Warren R. Ketler, Curtis A. Smith	

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1973	157	55

8a. CONTRACT OR GRANT NO. N00014-72-C-0293 b. PROJECT NO. NR 364-074/1-5-72 c. Code 462 d.	9a. ORIGINATOR'S REPORT NUMBER(S) SAI-72-590-LJ 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Reproduction of this document in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research (Code 462) Department of the Navy Arlington, Virginia 22217

13. ABSTRACT

This document is the first of three volumes which present techniques and methods for developing efficient Monte Carlo simulation. Each volume presents techniques for reducing computational effort in one of the following areas: Vol. I - Selecting Probability Distributions, Vol. II - Random Number Generation for Selected Probability Distributions, and Vol. III - Variance Reduction.

This volume provides a straightforward approach and associated techniques for selecting the most appropriate probability distributions for use in Monte Carlo simulations. Part I, BASIC CONSIDERATIONS, presents the underlying concepts and principles for selecting probability distributions. Part II, SELECTION OF DISTRIBUTIONS, gives the mathematical models representing stochastic processes and presents step-by-step procedures for identification and selection of the appropriate probability distributions based upon the degree of knowledge and available data for the random variable under study.

ORNL-RSIC-38
Vol. I

INTERNAL DISTRIBUTION

1. L. S. Abbott
2. R. G. Alsmiller, Jr.
3. C. E. Clifford
4. R. R. Coveyou
5. F. C. Maienschein
6. F. R. Mynatt
7. R. W. Peele
8. F. G. Perey
9. H. Postma
10. M. W. Rosenthal
11. D. Steiner
12. D. B. Trauger
13. D. K. Trubey
14. G. E. Whitesides
15. A. Zucker
16. H. Feshbach (Consultant)
17. P. F. Fox (Consultant)
18. W. W. Havens, Jr. (Consultant)
19. A. F. Henry (Consultant)
20-320. RSIC
321-322. Central Research Library
323. ORNL — Y-12 Technical Library Document Reference Section
324. Laboratory Records Department
325. Laboratory Records, ORNL, R.C.
326. ORNL Patent Office

EXTERNAL DISTRIBUTION

327. P. B. Hemmig, Division of Reactor Research and Development, ERDA, Washington, D.C. 20545
328. D. C. Irving, Science Applications, Inc., Box 2351, La Jolla, California 92037
329. E. J. McGrath, Science Applications, Inc., Box 2351, La Jolla, California 92037
330. Capt. R. G. Powell, Defense Nuclear Agency, Washington, D.C. 20305
331. L. K. Price, Division of Controlled Thermonuclear Research, ERDA, Washington, D.C. 20545
332. Burt Zolatar, Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94304
333. Directorate of Licensing, NRC, Washington, D.C.
334. Directorate of Regulatory Standards, NRC, Washington, D.C., Attn: Director
335-336. Technical Information Center, ERDA, Oak Ridge, Tennessee
337. Research and Technical Support Division, ORO, Oak Ridge, Tennessee
338. Reactor Division, ORO, Oak Ridge, Tennessee

☆U. S. GOVERNMENT PRINTING OFFICE: 1974-748-189/180