STATISTICAL FREQUENCY ANALYSIS BY OPTIMIZATION OF DENSITY FUNCTIONS TO HISTOGRAMS

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

James Lucius Grant

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in the School of Civil Engineering

Georgia Institute of Technology

November, 1973

STATISTICAL FREQUENCY ANALYSIS BY

OPTIMIZATION OF DENSITY FUNCTIONS TO HISTOGRAMS

Approved:

James R. Wallace, Chairman

Willard M. Snyder

George M. Slaughter

Date Approved by Chairman: 1/21/74

ACKNOWLEDGEMENTS

The author's gratitude is extended to all who helped to make this investigation possible.

Grateful acknowledgement is expressed to his advisor, Dr. J. R. Wallace, whose advice, counsel and patience are in a large measure responsible for the completion of this work, and to the members of the Reading Committee, Professor Willard M. Snyder, and Dr. George M. Slaughter, whose suggestions and comments were invaluable.

Financial support for this investigation was received from the Agricultural Research Center, and from the School of Civil Engineering.

The Graduate Division of the Georgia Institute of Technology allowed certain deviations in the format of this thesis for purposes of clarity and readability.

The author wishes to thank his wife Joanne for her support and technical assistance in this work, and to thank Miss Brenda Manley, who typed the thesis, and Mr. William L. Bellew, who prepared the figures and illustrations contained herein.

TABLE OF CONTENTS

		Page
ACKNOWLEDGEMENTS	.....	ii
LIST OF TABLES	.....	v
LIST OF ILLUSTRATIONS	.....	vii
SUMMARY	.....	viii
Chapter		
I.	INTRODUCTION	1
II.	SOME THEORETICAL CONCEPTS	7
	The Method of Least Squares Confidence Regions in Least Squares Tolerance Limits Estimation of Parameters in Non-linear Regression	
III.	STATISTICAL PARAMETER ESTIMATION BY THE METHOD OF LEAST SQUARES	37
	A Description of the Numerical Experiments	
IV.	ANALYSIS OF RESULTS OF NUMERICAL EXPERIMENTS	55
	Results for Unweighted Least Squares Results for Weighted Least Squares Confidence Regions and Tolerance Limits	
V.	THE METHOD OF LEAST SQUARES APPLIED TO REAL DATA	108
	Source of Data Results of Fitting	
VI.	CONCLUSIONS AND RECOMMENDATIONS	119
	Summary of Results Conclusions Recommendations	

TABLE OF CONTENTS (Continued)

Page

APPENDICES 126

BIBLIOGRAPHY 172

VITA 176

LIST OF TABLES

Table		Page
4.1	Results of Runs for Various Values of "r"	58
4.2	Results of Runs for NEMP ≠ 0.....	63
4.3	Analysis of Variance, $\gamma = 0$	66
4.4	Bias of Esitmates, $\gamma = 0$	72
4.5	χ^2 Test for Normality of Errors	75
4.6	Kolmogorov-Smirnov Test for Normality of Errors	77
4.7	Results of Runs, $\gamma \neq 0$, NEMP = 0	83
4.8	Results of Runs, $\gamma \neq 0$, NEMP ≠ 0	85
4.9	Bias of Weighted Least Squares, $\gamma = 0.75$	88
4.10	Analysis of Variance, $\gamma = 0.75$	89
4.11	χ^2 Test for Normality of Errors, $\gamma = 0.75$	92
4.12	Kolmogorov-Smirnov Test for Normality of Errors, $\gamma = 0.75$	93
4.13	Random Normal Sample, $\mu = 0$, $\sigma = 10$	103
4.14	Histogram of Random Normal Sample	104
5.1	Parameter Estimates for Samples with Outliers	111
5.2	Results of Fitting Maximum Rainfall at Atlanta, Georgia	114
5.3	Return Periods for Atlanta Rainfall Data	116
5.4	Stability of Parameter Estimates with a Growing Sample	118

LIST OF TABLES (Continued)

Table		Page
D-1	Subroutines of Computer Program	149
E-1	Description of Computer Runs	154
F-1	Description of Streamflow Stations	158
F-2	Station Records for Selected Streams	160
F-3	Maximum Precipitation, Atlanta, Georgia	163

LIST OF ILLUSTRATIONS

Figure		Page
2.1	The Newton-Raphson Method in One Dimension	24
2.2	Consequences of the Initial Point	25
2.3	Invalidation of a Linear Approximation	29
3.1	The Finite Form of a Probability Density Function ...	38
3.2	Elements of a Histogram	39
4.1	Graphical Representation of Analysis of Variance, $\gamma = 0$	67
4.2	Graphical Representation of $\frac{s^2}{\sigma^2} = \frac{k_\alpha}{\sqrt{N}}$	68
4.3	Histogram of Case 23, Run 9G.75	86
4.4	Graphical Representation of Analysis of Variance, $\gamma = 0.75$	90
4.5	Confidence Region for Case 100, Run 10R3, $\gamma = 0$	96
4.6	Confidence Region for Case 100, Run 10R3, $\gamma = 0.75$ ..	99
4.7	Confidence Region for Normal Sample	102
5.1	Rainfall-Frequency Curves for Atlanta, Georgia	115
D-1	Flow Chart of Main Program	151
D-2	Flow Chart of Subroutine GPARTL	152

SUMMARY

Statistical analysis of hydrologic data is an important part of the solution of hydrologic problems. While standard statistical techniques are very useful to the practicing hydrologist, certain differences between the requirements of hydrologic analysis and the assumptions of classical statistics often require the hydrologist to seek new methods of statistical analysis more suited to the problems with which he must contend. This research was concerned with the exploration of the method of nonlinear least squares as a means of estimating statistical parameters, with particular attention being given to the utility of the method when applied to problems of frequency analysis of hydrologic data.

The use of least squares to estimate statistical parameters involves fitting, in the least squares sense, a theoretical frequency distribution function to a data histogram by adjusting the parameters of the distribution function. The difference between the number of observations in a given histogram class interval and the expected number of observations for the same class (obtained from the density function) is called the class error. When these errors are normally distributed with zero mean and constant variance, then the method of least squares is an application of the method of maximum likelihood.

Specific techniques were developed and collected from the literature which are of use in the application of the method of least squares

to a wide variety of problems. In addition, methods are presented by which confidence regions for estimated parameters may be computed for various levels of confidence. A method is described by which such a confidence region may be used to calculate statistical tolerance limits.

Numerical experiments were conducted to determine the properties of the least squares estimators of the parameters of a two parameter gamma distribution. By forming data histograms of different class widths from the same samples and estimating by least squares the population parameters from each histogram, it was found that the properties of the estimators are not highly dependent upon the width of the class interval used to construct the histogram. By progressively adding empty classes to the right end of histograms and then estimating the population parameters, it was shown that it is sufficient to use the range of the sample as the range of optimization. Consideration of samples of different sizes revealed that the variance of an estimator multiplied by the sample size is approximately a constant. By considering sums of weighted squares of error terms, it was found that the properties of the least squares estimators were dependent upon the weights, and that by adjusting the weights, the properties of the estimators may be adjusted to suit the requirements of the problem at hand. These experiments showed that the least squares estimators were less efficient than likelihood estimators in the sense that the variance of the least squares estimator of a parameter will in general be larger than that of the corresponding likelihood estimator. It was found that by a proper choice of weights this inefficiency could be

practically eliminated.

The two parameter gamma distribution was fit to annual flood data and annual rainfall data. In the course of these fittings, it was shown that the least squares estimators are quite insensitive to outliers which may occur in the sample. That is, the parameter estimates obtained from a given sample containing outliers are not substantially different from the estimates obtained from the same sample with the outliers removed. This fact is of importance in the statistical analysis of samples which may contain outliers, since if the analysis be done by least squares, then the results will not depend upon the identification and removal of the outliers.

The method of least squares has been shown to yield results close to those of maximum likelihood under most conditions, and to yield better results when the sample being analysed contains one or more outliers. Therefore, the method of least squares should be considered a useful method of statistical analysis of hydrologic data.

CHAPTER I

INTRODUCTION

Hydrologic data such as streamflow and precipitation measurements are important in the design, construction, and management of hydraulic structures. In the form in which such data are originally acquired, they are scarcely more than a collection of numbers conveying little meaning and less understanding of the phenomenon under observation. It is the duty of the hydrologist to attempt to bring order out of this chaos, to judiciously expand and contract the original collection of data in such a way as to eliminate or classify those characteristics of the data due to chance, and to so condense and order the data as to allow rational interpretations of those events represented by the existing data and the logical extrapolation of the data to include events not represented in the original data but of significance to the problem at hand.

As a matter of practicality, if not of necessity, the reduction of data is universally accomplished by the expedient of making a priori some statement regarding the behavior or order of the world, and then utilizing the available data to adjust the statement to allow the fullest possible accommodation by the hypothesized world to the available data, which presumably are a consequence of the functioning of the real world. Such hypothetical statements regarding the workings of the world are

generally called "models", and the use of observations of past events to adjust the model is called" fitting" the model to the data. The appropriateness of a model is judged by its ability to explain the salient portions of the data, to ignore the trival portions of the data, and to accurately predict or describe events post-dating the fitting of the model.

Models are typically described as deterministic or probabilistic, or as a combination of these two types. A deterministic model is one which, from a given initial state, may and must arrive at a unique subsequent state which is dependent solely upon the initial state. A probabilistic model is one which, from a given initial state may arrive at more than one subsequent state, and thus whose future behavior can be inferred only generally, and not for a particular instance. It is customary to regard a probabilistic model as an expression of incomplete understanding of the phenomenon being studied, and one expects, with Democritus, that as knowledge of the physical processes governing a particular phenemenon increases, then the attributes of the phenomenon which must be regarded as indeterminant will decrease, until at last the totality of the event may be understood, and thus predicted.

The physical processes forming the subject matter of hydrology range from large scale atmospheric phenomena to the microscopic details of soil-water interactions and movements occuring in spaces and channels which may be only a few microns in diameter. Inasmuch as the typical problem in hydrology will encompass or be affected by processes of such large and small scales, it is not surprising that much of the science of hydrology is based upon probabalistic models, and that little in

the way of deterministic information is available. Thus, it is seen that, at least at the present state of hydrologic art, hydrologic phenomena must as a practical matter be regarded largely as the result of random processes and must be analyzed and studied as such. Methods of statistical analysis have been developed for the study of such random phenomena, and the past few decades have witnessed a growing awareness by hydrologists of the power and utility of statistical methods as tools to aid in the solution of hydrologic problems.

In spite of the proven usefulness of traditional statistical methods in hydrology, the practicing hydrologist is often plagued by the subtle differences between his situation and the assumptions underlying classical statistics. For example, most hydrologic problems begin with small amounts of data, and hope to end with statements concerning the characteristics of large amounts of data; in other words, the hydrologist must work beyond the range of his data. A practicing hydrologist rarely has the opportunity to design or supervise the collection of his data in such a way as to adapt the collection scheme to facilitate the solution of a particular problem. Because of the large amount of time required for the collection of much hydrologic data (for example, annual flood peaks), the hydrologist is denied the opportunity to examine multiple samples, and must content himself with one sample, be it representative or no, which grows by the addition of new observations which may very well be influenced by changes in the environment over which he has no control, and quite often no knowledge. Upon such a data base, the hydrologist is required to make estimates which may be used as a basis of design for very expensive

structures whose failure could have serious physical and economic consequences. Not surprisingly then, the requirements of the hydrologist are somewhat different from the assumptions of classical statistics. Whereas the latter is primarily concerned with the most efficient estimation in the sense of minimizing the variance of the estimators, the hydrologist is properly willing to sacrifice some efficiency in his estimators in order to gain some protection against making unusually bad estimates based upon samples which may contain one or more records of rare or extreme events. The hydrologist, in other words, must follow the time-honored engineering adage "to err is human; to err on the side of safety is divine".

In this spirit of seeking statistical methods with particular advantages in dealing with the problems of hydrology, Snyder [1] in 1972 proposed that population parameters of probability density functions might be estimated by the technique of least squares. As envisioned by Snyder, the essence of this method would consist of grouping a given set of data into a frequency histogram, and then fitting by non-linear least squares the selected probability density function to this histogram. Several advantages apparently would accrue from this method of parameter estimation. First, the least squares estimates would not be unduly influenced by the presence of data representing rare events in a small sample (i.e., outliers). This question of the proper identification and treatment of outliers has traditionally been a bane to hydrologists, and the method proposed by Snyder promised at least a partial solution to the problem. Second, the least squares technique offered a convenient method to choose

simultaneously the parameters of a transformation of data and the statistical fit of a given density function to the transformed data. Hydrologists frequently find it necessary to transform data for one reason or another; they take the logarithm of annual flood data in an attempt to eliminate skewness; they take fractional powers of data for the same reason; in fact, Chow [2] gives a list of over one hundred references dealing with various data transformations which have been used by hydrologists. Thus it is apparent that the ability to specify the parameters of a data transformation while at the same time specifying the statistical parameters of the transformed data is a decided advantage. Again, while the computations involved in the least squares estimation of parameters are laborious, the nature of the computations is such that they are readily adapted to solution by digital computers, and furthermore the computations are for the most part identical for any distribution, and so once a computer program were developed, it would suffice with only minor modifications to fit any data to any distribution.

This study was undertaken to develop upon and to extend these initial concepts of Snyder by a systematic investigation of the techniques of least squares estimation of parameters and of the statistical properties of the estimates so obtained. The plan of the study was to attempt first to understand the theoretical basis and implications of parameter estimation by least squares, and from this knowledge to attempt to develop specific techniques necessary to apply the least

squares method in a more or less optimum manner. Using these techniques, a systematic study of the statistical properties of the least squares estimators for a two parameter gamma density function was made, wherein the strengths and weaknesses of the method were identified. A second study was then made in which the basic method was modified in an attempt to overcome certain weaknesses of the method of least squares which were uncovered in the earlier investigation.

The presentation of the results of this study follows closely the above outline. In Chapter II are presented the necessary theoretical aspects of least squares estimation in general and least squares estimation of statistical parameters in particular. Chapter III provides a specialization of the ideas of Chapter II to the two parameter gamma probability density function, deals with certain problems implicit in least squares estimation of statistical parameters, such as the grouping of the sample into a histogram, and provides an outline of the numerical experimentation undertaken to examine the properties of the least squares parameter estimates. Chapter IV gives detailed accounts of the various numerical experiments and analyses of the results of these experiments. Chapter V gives a brief account of the results obtained from the fitting of the two parameter gamma distribution to real hydrologic data by the method of least squares. Chapter VI summarizes the results of this study and provides an analysis of these results along with certain conclusions to which these results appear to lead.

CHAPTER II

SOME THEORETICAL CONCEPTS

In this chapter certain theoretical concepts upon which the work in later chapters is based are collected and explained. Since many of the techniques presented in this chapter are applicable to any problem involving least squares estimation of parameters, an attempt has been made to state these techniques in a rather general form. In most cases the application of these techniques to the problems considered in the later chapters is immediate and obvious. Where the application of a technique to these problems is not obvious, or requires the technique to be reformulated, then the specialization is deferred to the chapter in which the technique is required.

The following definition is central to the proper interpretation of the remainder of this chapter, and all statements are made with tacit assumption of the terminology of this definition being made.

DEFINITION: By an observation is meant a set of two N-tuples of known data $(x_1, x_2, \ldots, x_N)$ and $(y_1, \ldots, y_N)$. It is assumed that the N-tuple $\underset{\sim}{y}_N$ is related to the N-tuple $\underset{\sim}{x}_N$ in the following manner

$$Y_i = f(x_i, \underset{\sim}{\theta}) \qquad 2.1(a)$$

$$w_i^{1/2} (y_i - Y_i) \in N(0, \sigma^2) \qquad 2.1(b)$$

where the form of f is known, $\underset{\sim}{\theta}$ is some m-tuple of parameters, $w_i^{1/2}$ is a positive weighting function which may depend on x_i, $\underset{\sim}{\theta}$, or both. Equation 2.1(b) says that the product $w_i^{1/2}(y_i - Y_i)$ is normally distributed with zero mean and variance σ^2 which does not depend upon x.

The Method of Least Squares

Let the observation $\underset{\sim}{x}_N$, $\underset{\sim}{y}_N$ be given, and suppose it is required to utilize this observation to deduce a "reasonable" value for the parameter vector $\underset{\sim}{\theta}$. The method of least squares defines the "best" value of $\underset{\sim}{\theta}$ to be that value for which the sum of the squares of the weighted residuals is a minimum. In the notation of Equation 2.1, the least squares estimate $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$ is chosen so that

$$E^2 = \sum_{i=1}^{N} w_i (y_i - Y_i)^2 \qquad 2.2$$

is a minimum. As an alternative to the least squares procedure, it may be desired to estimate $\underset{\sim}{\theta}$ in such a way as to maximize the probability of observing the residuals $w_i^{1/2}(y_i - Y_i)$. This may be accomplished by utilizing the following argument. Let

$$e_i = w_i^{1/2} (y_i - Y_i).$$

Then by Relation 2.1(b), e_i is normally distributed with zero mean and constant variance σ^2. The probability of observing a particular N-tuple

$$L = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{e_i^2}{\sigma^2}\right) \qquad 2.3$$

where the symbol $\prod_{i=1}^{N}[\ldots]$ indicates the product of the N quantities within the brackets. To maximize L, it is sufficient to maximize the logarithm of L, since L is a positive quantity and the logarithm is a monotonically increasing function of its argument. Thus,

$$\ln L = -N \ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N} e_i^2 \qquad 2.4$$

An examination of Equation 2.4 shows that ln(L) is a maximum when $\sum_{i=1}^{N} e_i^2$ is a minimum. But,

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} w_i (y_i - Y_i)^2 = E^2 \qquad 2.5$$

Equation 2.5 shows that the principle of maximum likelihood is equivalent to the principle of least squares.

The equivalence of maximum likelihood and least squares demonstrated above is of great importance because there exists a large body of knowledge regarding the statistical properties of maximum likelihood estimators $\hat{\underset{\sim}{\theta}}$ of the parameter $\underset{\sim}{\theta}$ (Kendall [3], and Cramer [4]), and these estimators are known to have very desirable statistical properties (Kendall [3]). In particular, maximim likelihood (and thus, least

squares) estimators are consistent (although they may be biased), they tend to a normal distribution for large N, they have minimum variance in the limit as N increases, and the form of the covariance matrix of the multivariate normal distribution of the estimators is known for large N.

The following expression for the covariance matrix of the estimators is derived in Kendall [3]. Let $g(e,\underset{\sim}{\theta})$ be the frequency function of the weighted residuals $\underset{\sim}{e}$. Then by virtue of Relation 2.1(b),

$$g(e, \underset{\sim}{\theta}) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2} \frac{e(\underset{\sim}{\theta})^2}{\sigma^2}\right) \qquad 2.6$$

Then if $\hat{\underset{\sim}{\theta}}$ be estimated using Equation 2.2 (or equivalently, Equation 2.4),

$$N \operatorname{cov}(\hat{\theta}_j, \hat{\theta}_k) = \frac{\Delta_{jk}}{\Delta} \qquad 2.7(a)$$

where

$$\Delta = \left| \int_{-\infty}^{\infty} \left(\frac{\partial \ln g}{\partial \theta_j}\right)_{\theta_o} \left(\frac{\partial \ln g}{\partial \theta_k}\right)_{\theta_o} g(e, \underset{\sim}{\theta_o})\, de \right| \qquad 2.7(b)$$

and Δ_{jk} is the minor of the jth row and the kth column of Δ, and $\underset{\sim}{\theta_o}$ is the "true" value of the parameter vector. As N approaches infinity, the joint probability function of the maximum likelihood estimators of $\underset{\sim}{\theta}$ tends to the form

$$p(\hat{\underset{\sim}{\theta}}) = \frac{|A|^{1/2}}{(2\pi)^{n/2}} \exp \quad \frac{-(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_0)^T A (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_0)}{2} \qquad 2.8(a)$$

where

$$A_{ij} = \int_{-\infty}^{\infty} \left(\frac{\partial \ln g}{\partial \theta_i}\right)_{\underset{\sim}{\theta}_0} \left(\frac{\partial \ln g}{\partial \theta_j}\right)_{\underset{\sim}{\theta}_0} g(e, \underset{\sim}{\theta}_0)\, de \qquad 2.8(b)$$

and $\underset{\sim}{\theta}^T$ denotes the transpose of $\underset{\sim}{\theta}$. Possible uses of Equations 2.8 will be discussed later in this chapter.

Finally, it should be noted that the minimization of Equation 2.2 provides a viable method of estimating $\underset{\sim}{\theta}$ even if the assumption of normality of the weighted errors expressed by Relation 2.1(b) is not appropriate. In this case, however, such use of Equation 2.2 must be motivated by intuition or other considerations rather than by an appeal to the properties of maximum likelihood. Thus, failing the appropriateness of Relation 2.1(b), the properties of the least squares estimator of $\underset{\sim}{\theta}$ will, at least in the general case, be unknown.

Confidence Regions in Least Squares

If an estimate $\hat{\underset{\sim}{\theta}}$ of the parameter $\underset{\sim}{\Theta}$ is made from an observation $\underset{\sim}{x}_N$, $\underset{\sim}{y}_N$, then $\hat{\underset{\sim}{\theta}}$ is dependent upon the random characteristics of the observation, and so it is unlikely that the estimate $\hat{\underset{\sim}{\theta}}$ will agree exactly with

the true, but unknown, value of $\underset{\sim}{\theta}$. Thus, one is often interested in estimating the probable magnitude of the difference between $\hat{\underset{\sim}{\theta}}$ and $\underset{\sim}{\theta}$. More specifically, one seeks a 2m-dimensional region R_α^2 in $\underset{\sim}{\theta} \times \hat{\underset{\sim}{\theta}}$ space for which the probability of a given point $(\underset{\sim}{\theta},\hat{\underset{\sim}{\theta}})$ falling within this region is equal to $\frac{\alpha}{100}$ (recall that m is the number of components in the parameter vector $\underset{\sim}{\theta}$). If $p_1\ (\hat{\underset{\sim}{\theta}}|\underset{\sim}{\theta})$ is the probability density of observing $\hat{\underset{\sim}{\theta}}$ given a particular value of $\underset{\sim}{\theta}$, and if $p_2\ (\underset{\sim}{\theta})$ is the probability density of $\underset{\sim}{\theta}$, then the probability of a point $(\underset{\sim}{\theta},\hat{\underset{\sim}{\theta}})$ being in the region R_α^2 is given by

$$P(R_\alpha^2) = \int_{R^2} p_1(\hat{\underset{\sim}{\theta}}|\underset{\sim}{\theta})\ p_2(\underset{\sim}{\theta})\ d\hat{\underset{\sim}{\theta}}\ d\underset{\sim}{\theta} = \frac{\alpha}{100} \qquad 2.9(a)$$

Let the probability space of $(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}})$ be denoted by $E_1 \times E_2$. If one fixes the value of $\underset{\sim}{\theta}$ at some value (say) $\underset{\sim}{\theta}_o$, then

$$\int_{E_2} p_1(\hat{\underset{\sim}{\theta}}\,|\,\underset{\sim}{\theta}_o)\ d\hat{\underset{\sim}{\theta}} = 1 \qquad 2.9(b)$$

and hence a region $E_{2\alpha_o} \subset E_2$ may be found for which

$$\int_{E_{2\alpha_o}} p_1(\hat{\underset{\sim}{\theta}}|\underset{\sim}{\theta}_o)\ d\hat{\underset{\sim}{\theta}} = \frac{\alpha}{100} \qquad 2.9(c)$$

The region $E_{2\alpha_o}$ represents a cross-section in $E_1 \times E_2$ corresponding to the fixed point $\underset{\sim}{\theta}_o$ in E_1. If such a set $E_{2\alpha}$ be chosen for each point

$\underset{\sim}{\theta}_t$ in E_1, the union of these sets will be a 2m-dimensional subregion in $E_1 \times E_2$. If one now considers cross-sections R_α of this subregion formed by the intersection of the plane $\hat{\underset{\sim}{\theta}}$ = constant with the subregion $\bigcup_{t \in T} E_{2\alpha_t}$, then these cross-sections R_α represent the α percent confidence regions for $\underset{\sim}{\theta}$ corresponding to the fixed value $\hat{\underset{\sim}{\theta}}$ = constant. It should be noted that no claim is made that the probability that R_α contains $\underset{\sim}{\theta}$ is equal to α, but instead it is claimed that if a large number of $\hat{\underset{\sim}{\theta}}$ be computed from different observations $\underset{\sim}{x}_N, \underset{\sim}{y}_N$, and if for each such $\hat{\underset{\sim}{\theta}}$, the corresponding region R_α be computed, then the statement $\underset{\sim}{\theta} \in R_\alpha$ may be expected to be true in α percent of these cases. (von Mises [5]). If the Relation 2.1(b) is valid, and if N is large enough for Equation 2.8 to apply, then the confidence region R_α may be found approximately by using Equation 2.8 to construct the region $\bigcup_{t \in T} E_{2\alpha_t}$, and then finding the particular cross-section of this region corresponding to the estimate $\hat{\underset{\sim}{\theta}}$ based upon the observation $\underset{\sim}{x}_N, \underset{\sim}{y}_N$ of interest. Such a procedure is actually not very difficult in the case m = 1. In the case m = 2, the procedure is apt to be very difficult, and in the cases m > 2, only the faintest hope for success may be extended.

For small N, and as an alternative to the above procedure for large N, Halperin [6] and Hartley [7] have devised a method by which the confidence region R_α may be constructed. The method is based upon known results from linear regression theory, where

$$y_i = \sum_{j=1}^{m} X_{ij} \Theta_j + e_i \qquad 2.10$$

and the e_i are a set of N independent errors from $N(0,\sigma^2)$ with σ unknown. Equation 2.10 may be rewritten in the form

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\theta} + \underset{\sim}{e} \qquad 2.11$$

where it is assumed that the matrix X is of rank m. The least squares estimate of $\underset{\sim}{\theta}$ is then

$$\hat{\underset{\sim}{\Theta}} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\ \underset{\sim}{X}^T\underset{\sim}{y} \qquad 2.12$$

A frequently used method of constructing confidence regions in this linear case is to decompose the sum of the squares of the errors e_i into two components, namely

$$\underset{\sim}{e}^T\underset{\sim}{e} = \text{reg}(\underset{\sim}{e}) + \text{res}(\underset{\sim}{e}) \qquad 2.13(a)$$

where

$$\text{reg}(\underset{\sim}{e}) = (\underset{\sim}{X}^T\underset{\sim}{e})^T\ (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\ (\underset{\sim}{X}^T\underset{\sim}{e}) \qquad 2.13(b)$$

and

$$\text{res}(\underset{\sim}{e}) = \underset{\sim}{e}^T\underset{\sim}{e} - \text{reg}(\underset{\sim}{e}) \qquad 2.13(c)$$

The first component of Equation 2.13(a) has rank m, the second component has rank N-m and is independent of the first. Thus, the ratio of the two components reg($\underset{\sim}{e}$)/res($\underset{\sim}{e}$) is distributed as Snedecor's "F" statistic with m and N-m degrees of freedom. The α percent confidence region for $\underset{\sim}{\theta}$ is then given by

$$R_\alpha = \left\{ \underset{\sim}{\theta} : \frac{reg(\underset{\sim}{e})}{res(\underset{\sim}{e})} \leq \frac{m}{N-m} F(\frac{\alpha}{100};\ m,N-m) \right\} \qquad 2.14$$

For the more general relation

$$y_i = f(x_i, \underset{\sim}{\theta}) + e_i \qquad 2.15$$

Equation 2.14 may still be used to determine the α-percent confidence region for $\underset{\sim}{\theta}$. However in this case the form of the decomposition of $\underset{\sim}{e}^T\underset{\sim}{e}$ into the components reg($\underset{\sim}{e}$) and res($\underset{\sim}{e}$) is no longer obvious. Hartley [7] has proposed a decomposition, based upon the use of Lagrange's interpolation formulae, to obtain a quasi-linearization of the regression function f. Halperin [6] has proposed a decomposition of the form

$$reg(\underset{\sim}{e}) = (\underset{\sim}{F}^T\underset{\sim}{e})^T\ (\underset{\sim}{F}^T\underset{\sim}{F})^{-1}\ (\underset{\sim}{F}^T\underset{\sim}{e}) \qquad 2.16(a)$$

where

$$\underset{\sim}{F} = \left(\frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_j} \right) = (F_{ij}) \qquad 2.16(b)$$

These results are immediately applicable to regression of the system defined by Equation 2.1, for Equation 2.1 may be written in the form

$$w_i^{1/2} y_i = z_i = w_i^{1/2} f(x_i, \underset{\sim}{\theta}) + w_i^{1/2} e_i = G(x_i, \underset{\sim}{\theta}) + E_i \qquad 2.17(a)$$

$$E_i \varepsilon N\ (0, \sigma^2) \qquad 2.17(b)$$

In this case, Equations 2.16 may become quite complicated, especially if the weights w_i depend upon $\underset{\sim}{\theta}$.

Tolerance Limits

In the case where the function f is a probability density function, then instead of seeking information about the probable range of values of the population parameters, one is often interested in the probable values of a given percentile of the population. If some "best" estimate $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$ be deduced from an observation $\underset{\sim}{x}_N$, $\underset{\sim}{y}_N$, then by definition the "best" estimate of the γ percentile of the population based upon $\underset{\sim}{x}_N, \underset{\sim}{y}_N$ is that number v_γ for which

$$\int_{-\infty}^{v_\gamma} p(v; \hat{\underset{\sim}{\theta}})\, dv = \gamma \qquad 2.18$$

However, in the same way that the estimate $\hat{\underset{\sim}{\theta}}$ of the population parameters is dependent upon the vagaries of the observation from which it was estimated, so also is the estimate v_γ of the γ percentile of the population dependent upon the observation, and the number v_γ computed from Equation 2.18 may be expected to vary from observation to observation. Thus, one is led to the concept of tolerance limits as a means of making a statement about the probable range of v_γ. Since, at least as defined above, the quantity v_γ is a random variable, then v_γ itself has a distribution, although in general this distribution is unknown. Thus, there exists a number $v_{\gamma,\alpha}$ such that for v_γ based upon a large number of observations, α percent of these v_γ may be expected to be less than $v_{\gamma,\alpha}$. This number $v_{\gamma,\alpha}$ cannot, of course, be computed unless the distribution of v_γ be known.

Motivated by the above discussion, one may define a tolerance limit as a random variable $\bar{v}_{\gamma,\alpha}$ which is a function of the observation $\underset{\sim}{x}_N, \underset{\sim}{y}_N$, the percentile point γ, and the confidence level α, which has the property that for large number of observations the inequality

$$(v_\gamma)_{true} \leq \bar{v}_{\gamma,\alpha} \quad , \qquad 2.19$$

where $(v_\gamma)_{true}$ is the actual but unknown γ percentile point of the population, may be expected to be true for at least α percent of the observations. (Note that this definition is actually a definition of an upper one sided tolerance limit. Similar definitions can be made for lower one sided and for two sided tolerance limits. See,

for example, Bowker and Lieberman [8].)

The following important points of the above definition should be noted:

a) The tolerance limit $\bar{v}_{\gamma,\alpha}$ is a function of the observation upon which it is based.

b) There is no reason to expect that the definition of the random variable $\bar{v}_{\gamma,\alpha}$ should be unique, and thus there may be many tolerance limits $\bar{v}_{\gamma,\alpha}$ based upon the same observation.

c) The difference between the random variable $\bar{v}_{\gamma,\alpha}$ and the quantity $v_{\gamma,\alpha}$ defined above is emphasized. In particular, it should be noted that $v_{\gamma,\alpha}$ is a constant independent of any particular observation, while $\bar{v}_{\gamma,\alpha}$ is a function of the observation from which it is computed.

In view of the above discussion, it is obvious that if one has a means of constructing a confidence region R_α for the population parameters $\underset{\sim}{\theta}$ for any observation, then one may immediately define a (upper one sided) tolerance limit $\bar{v}_{\gamma,\alpha}$ to be the random variable

$$\bar{v}_{\gamma,\alpha} = \max_{\bar{\underset{\sim}{\theta}} \in R} \left\{ v: \int_{-\infty}^{v} p(r; \bar{\underset{\sim}{\theta}})\, dr = \gamma \right\} \qquad 2.20$$

where $\bar{\underset{\sim}{\theta}}$ is any point in R_α. In words, Equation 2.20 says that $\bar{v}_{\gamma,\alpha}$ is the largest v such that

$$\int_{-\infty}^{v} p(r, \bar{\underset{\sim}{\theta}})\, dr = \gamma$$

as $\bar{\underset{\sim}{\theta}}$ varies over all values in R_α. It is clear that Equation 2.20 defines a tolerance limit, since by the definition of R_α, $\underset{\sim}{\theta}$ is included in α percent of the R_α computed from a large number of samples, and thus, for at least these α percent of observations,

$$(v_\gamma)_{\text{true}} \leq \bar{v}_{\gamma,\alpha} \qquad 2.21$$

which is the definition of an upper one-sided tolerance limit.

Estimation of Parameters in Non-Linear Regression

The solution of the least squares equation

$$E^2 = \text{minimum}$$

is based upon the necessary conditions for a minimum in the form

$$\frac{\partial E^2}{\partial \theta_j} = 0 \quad , \quad j = 1, \ldots, m \qquad 2.22$$

If the regression function $f(x_i, \underset{\sim}{\theta})$ be linear in the parameters θ_j, then E^2 is a quadratic form in θ_j, and the solution to Equations 2.22 usually offers little in the way of serious difficulty, leading rather routinely to the so-called normal equations given by Equations 2.12. Furthermore, since E^2 is a quadratic form in θ_j, there will be at

most one solution to Equations 2.22 and so any solution to these equations which is found is known to be unique. The estimates θ_j obtained as a solution to Equations 2.22 in the linear case are known to be the "best linear unbiased estimators" of the parameters (Hartley and Booker [9]).

In the more general case in which $f(x_i,\underset{\sim}{\theta})$ is non-linear in the parameter $\underset{\sim}{\theta}$, the routine nature of the problem of solving Equations 2.22 disappears. Apart from the often arduous task of finding any solution to Equations 2.22, one is faced with the possibility that a solution, once found, may not yield an absolute minimum of E^2 over the allowable region of $\underset{\sim}{\theta}$. Problems associated with finding solutions to Equations 2.22 have been discussed by Levenburg [10], Snyder [11], Hartley [12], Marquardt [13], Anderssen and Osborne [14], and Decoursey and Snyder [15], among others. Problems associated with the possibility of multiple solutions to Equations 2.22 have been treated by Hartley and Booker [9].

The favored approaches to solving Equations 2.22 begin with a Newton-Raphson iteration scheme.

Write

$$E^2 = \sum_{i=1}^{N} w_i(y_i - f(x_i,\underset{\sim}{\theta}))^2 \qquad 2.23$$

In the development that follows, the weighting factors w_i are omitted since they add essentially nothing to the discussion and their retention

would introduce a great deal of unnecessary confusion to the desired results.

Assuming independence of the parameters θ_j,

$$\frac{\partial E^2}{\partial \theta_j} = 2 \sum_{i=1}^{N} (y_i - f(x_i, \underset{\sim}{\theta}) (- \frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_j}), \quad j = 1, \ldots, m \qquad 2.24$$

Thus

$$\frac{\partial E^2}{\partial \theta_j} = 0 \Rightarrow \sum_{i=1}^{N} y_i \frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_j} - \sum_{i=1}^{N} f(x_i, \underset{\sim}{\theta}) \frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_j} = 0 \qquad 2.25$$

In Equations 2.25, let $f(x_i, \underset{\sim}{\theta})$ be approximated by the first order terms in $d\theta_j$ in the Taylor series expansion of f, where

$$\underset{\sim}{\theta} = \underset{\sim}{\theta}_o + d\underset{\sim}{\theta} \qquad 2.26$$

and θ_o is some pre-selected point in $\underset{\sim}{\theta}$-space.

Thus

$$f(x_i, \underset{\sim}{\theta}) \sim f(x_i, \underset{\sim}{\theta}_o) + \sum_{k=1}^{N} \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_k} d\theta_k \qquad 2.27$$

and so

$$\sum_{i=1}^{N} y_i \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_j} - \sum_{i=1}^{N} f(x_i, \underset{\sim}{\theta}_o) \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_j}$$
$$- \sum_{i=1}^{N} \sum_{k=1}^{m} \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_j} \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_k} d\theta_k = 0$$

Simplifying, one obtains

$$\sum_{k=1}^{m} \left(\sum_{i=1}^{N} \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_j} \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_k} \right) d\theta_k = \sum_{i=1}^{N} (y_i - f(x_i, \theta_o) \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_j} ,$$

$$j = 1, \ldots, m \qquad 2.28$$

Equations 2.28 are a set of m linear equations in the m unknowns $d\theta_k$, and are solved recursively to obtain the solution to Equations 2.25. Thus, starting from an initial parameter estimate $\underset{\sim}{\theta}_o$, one computes from Equations 2.28 initial values of $d\theta_k^1$, and sets

$$\underset{\sim}{\theta}_o^1 = \underset{\sim}{\theta}_o + d\underset{\sim}{\theta}^1$$

$\underset{\sim}{\theta}_o^1$ is then used as the new value of $\underset{\sim}{\theta}_o$ in Equations 2.28, and the procedure is repeated. After n cycles of the above procedure, one obtains

$$\underset{\sim}{\theta}_o^n = \underset{\sim}{\theta}_o^{n-1} + d\underset{\sim}{\theta}^n \qquad 2.29$$

The procedure is terminated when the length of the vector $d\underset{\sim}{\theta}^k$ falls

below some preselected value.

The above scheme has a particularly elucidating geometric interpretation if the parameter $\underset{\sim}{\theta}$ is considered to be a real variable. The Newton-Raphson iteration scheme in this case consists of approximating the function f by the line tangent to f at the point $\underset{\sim}{\theta}_v$. The intersection of the tangent line and the θ-axis is then computed, and the value of θ_1 is used as a new approximation of the desired root. (See Figure 2.1.)

An inspection of Figure 2.2 reveals that the choice of the initial point θ_o may be non-trivial, since an improper choice of this value could cause the algorithm either to diverge or to converge to the "wrong root." While existence and convergence theorems exist which in theory allow one at least to select an initial estimate θ_o which will guarantee convergence of the iterative scheme (Saaty and Bram [17]), the forms of the function f encountered in practical application are often so complicated as to make the theorems very difficult to apply. In addition, since the Newton-Raphson method provides only a solution or solutions to Equations 2.22, such theorems cannot indicate whether a given solution minimizes E^2. Thus, in cases where one has little or no previous experience, it may be necessary to scan the parameter space, find all solutions to Equation 2.22, and choose from among these solutions that value for which E^2 is a minimum.

It is important to realize at this point that information from other sources, notably the understanding of the investigator, may be

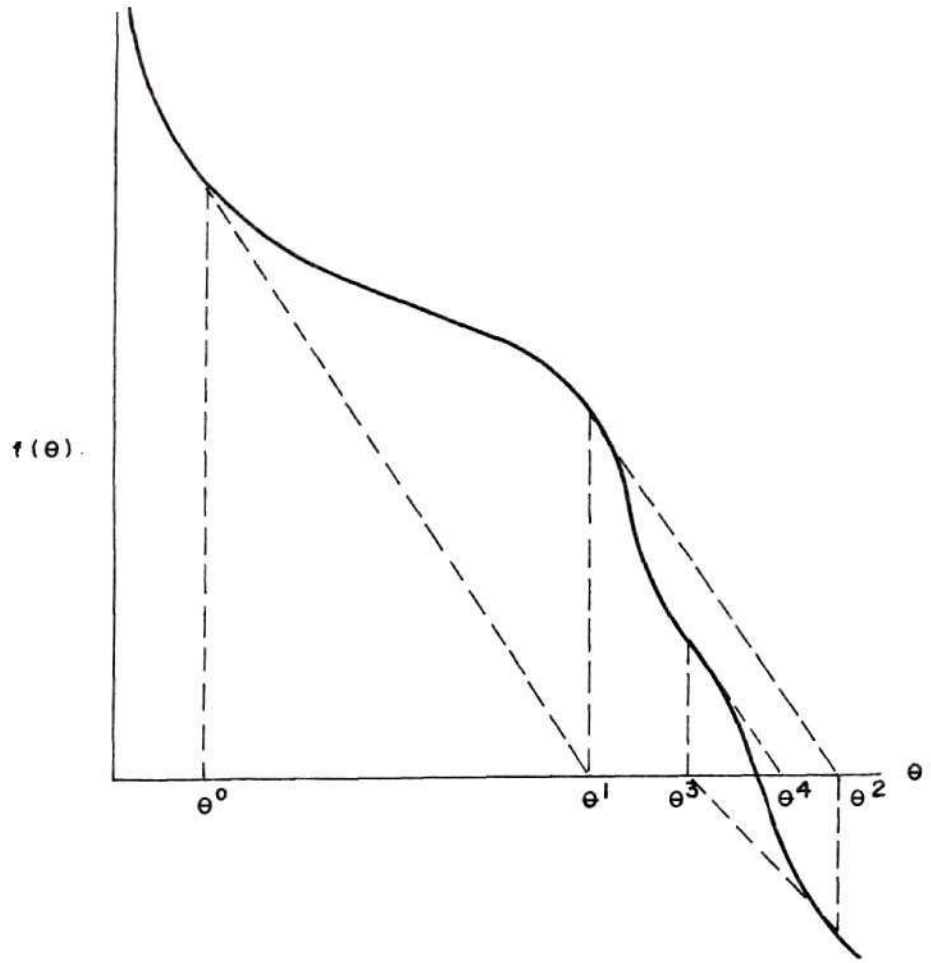

Figure 2.1 The Newton Raphson Method in One Dimension

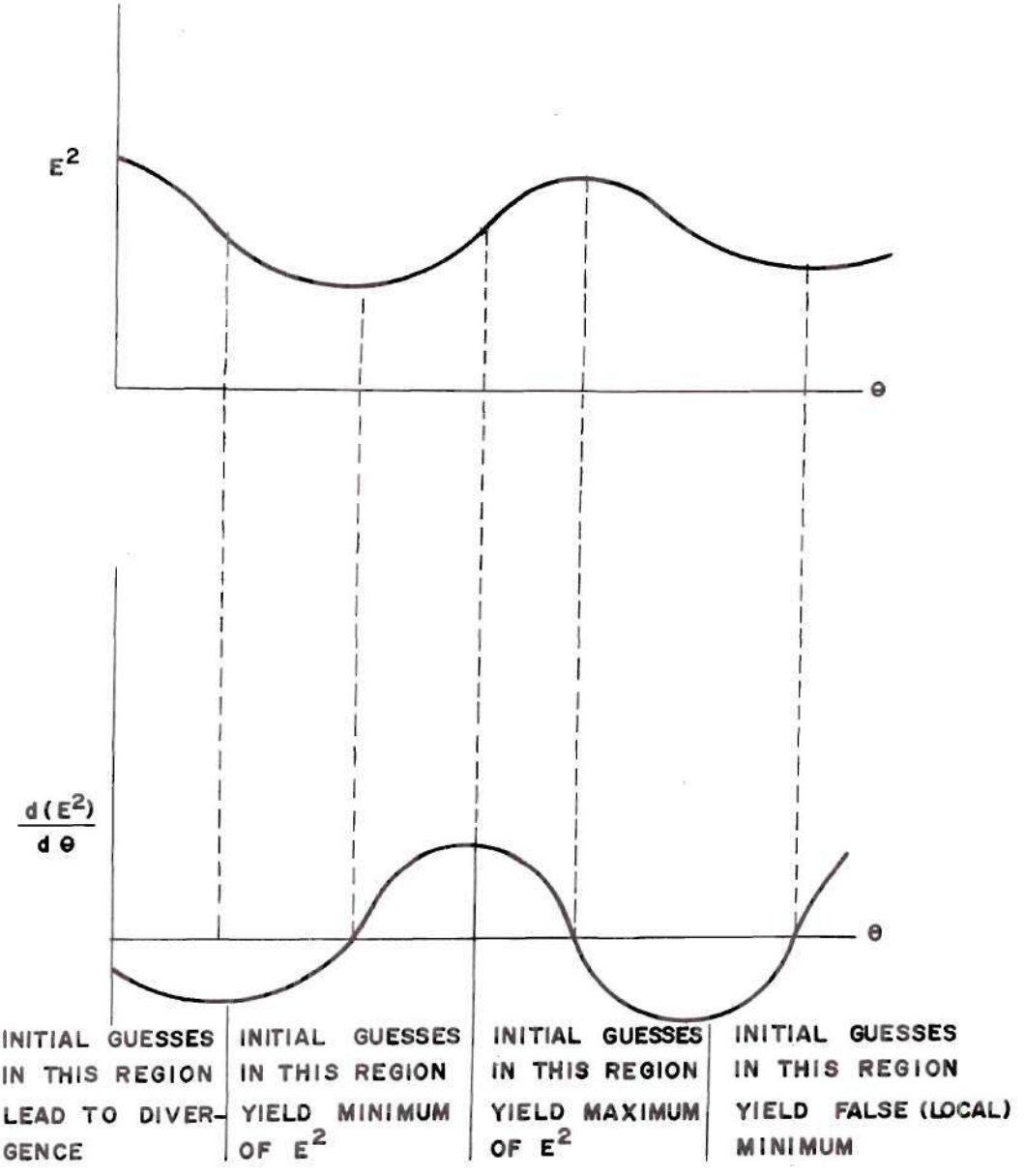

Figure 2.2 Consequences of the Initial Point

profitably used to limit the extent of this search. Thus, if it is known that the value of a certain parameter must, because of physical considerations, lie between 0 and 10, then only this range need be scanned for minima. In this way an understanding of the physical situation can be utilized to reduce the amount of work necessary to obtain a solution in a given situation.

Hartley and Booker [9] have addressed the problem of local minima and have developed an alternative procedure which avoids the exhaustive scan of the parameter space. Their method yields estimates $\bar{\underset{\sim}{\theta}}$ which are asymptotically 100 percent efficient as the size of the sample becomes large. An outline of this procedure may be given as follows:

i) Let it be assumed that the sample size N is an integral multiple of the dimension of the parameter space, say

$$N = km,$$

and that the observation $\underset{\sim}{x}_N$, $\underset{\sim}{y}_N$ may be partitioned into k subsets $x_{h\gamma}$, $y_{h\gamma}$, $h = 1,..,m, \gamma = 1,.., k$, where the convex closures of the x_h are disjoint.

ii) Write

$$\bar{Y}_h = \frac{1}{k} \sum_{\gamma=1}^{k} y_{h\gamma} \qquad 2.30(a)$$

and

$$\bar{f}(x_h, \underset{\sim}{\theta}) = \frac{1}{k} \sum_{\gamma=1}^{k} f(x_{h\gamma}, \underset{\sim}{\theta}) \qquad 2.30(b)$$

and find the solution $\underset{\sim}{\theta}^*$ for which

$$\bar{Y}_h = \bar{f}(x_h, \underset{\sim}{\theta}) \quad , \quad h = 1, \ldots, m \qquad 2.31$$

(This solution, for example, may be found by using the Newton-Raphson method to find a solution to $Y_h - f(x_h, \underset{\sim}{\theta}) = 0$ as was previously discussed).

iii) Using $\underset{\sim}{\theta}^*$ as an initial estimate, perform one step of the standard Newton-Raphson iteration to find $\dot{\underset{\sim}{\theta}}$, or alternatively using $\underset{\sim}{\theta}^*$ as an initial estimate, carry the standard Newton-Raphson method to convergence to find $\bar{\underset{\sim}{\theta}}$.

It can then be shown under rather general conditions on f that both $\dot{\underset{\sim}{\theta}}$ and $\bar{\underset{\sim}{\theta}}$ are asymptotically 100 percent efficient estimators of $\underset{\sim}{\theta}$ as N becomes large.

It is informative to illustrate the above procedure with the simple cases of linear regression with one and two unknown parameters. For the first case, Equations 2.30 and 2.31 apply in the following form. The observations y_i are averaged to yield $\bar{y}$, and the function $f(x_i, \underset{\sim}{\theta}) = \theta x_i$ is averaged to yield $\bar{x}$. θ^* is then computed as $\bar{y}/\bar{x}$. In the case of two unknown parameters (say θ_1 and θ_2, where $y = \theta_1 x + \theta_2$), the observations are partitioned into two groups $(x_{1\gamma}, y_{1\gamma})$ and $(x_{2\gamma}, y_{2\gamma})$, $\gamma = 1, \ldots, k$, where there is some value x_c such that

$$x_{1Y} < x_c < x_{2Y}$$

for all Y . As in the first example, averages of the dependent and independent variables are computed for each group, and one then writes, from Equation 2.31,

$$\bar{Y}_1 = \theta_1^* \bar{x}_1 + \theta_2^*$$

$$\bar{Y}_2 = \theta_1^* \bar{x}_2 + \theta_2^*$$

from which values of θ_1^* and θ_2^* may be computed.

It should be noted that the thrust of this procedure is to replace a complicated optimization problem by a simpler set of m equations in m unknowns. This set of equations is solved to obtain initial values of the parameters for use in the iterative solution of the more complicated optimization problem.

In addition to the problems associated with finding from among all solutions of Equations 2.22 that solution which minimizes E^2, numerical problems associated with the linear approximation of f and with an interdependence of the components θ_j of $\underset{\sim}{\theta}$ often induce a tendency for the iterative procedure of the Newton-Raphson method to diverge. An examination of Figure 2.3 shows (in 1-space) how the parameter corrections $d\underset{\sim}{\theta}$ computed by the Newton-Raphson method can be so large as to be outside the range of the linear approximation of f, and thus lead to divergence of the algorithm. This problem has been

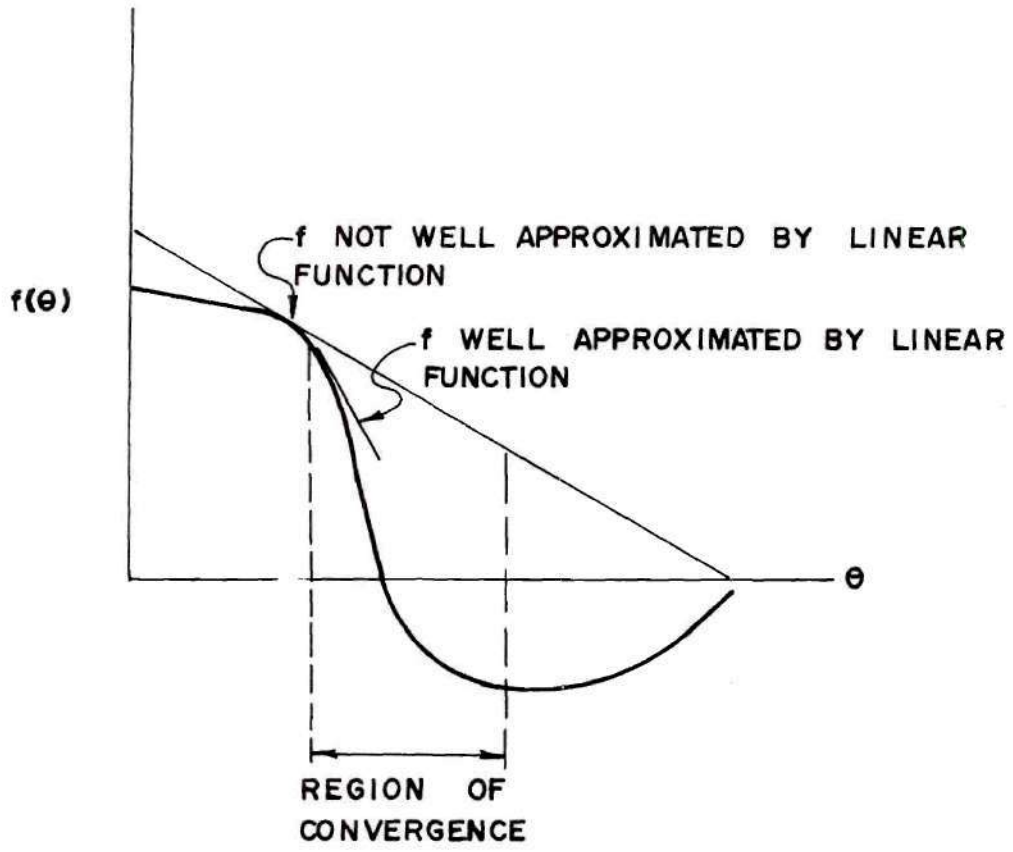

Figure 2.3 Invalidation of a Linear Approximation

rather well-studied, and algorithms for overcoming this tendency to diverge have been proposed by Hartley [12], Marquardt [13], and Levenberg [10], among others.

The essence of the method of Hartley lies in the use of the Newton-Raphson method as described above to compute parameter corrections $d\underset{\sim}{\theta}^n$. The function of the real variable η,

$$E^2(\eta) = E^2(\underset{\sim}{\theta}_o^{n-1} + \eta^{(n)} d\underset{\sim}{\theta}^n) \quad , \quad 0 < \eta^{(n)} \leq 1 \qquad 2.32$$

is then examined to determine the value η_o for which $E^2(\eta_o^n)$ is a minimum for η in the interval (0,1). The n^{th} approximation $\underset{\sim}{\theta}_o^n$ is then defined as

$$\underset{\sim}{\theta}_o^n = \underset{\sim}{\theta}_o^{n-1} + \eta_o^{(n)} d\underset{\sim}{\theta}^n \qquad 2.33$$

and the process continues.

Levenberg has considered the problem from a rather different viewpoint. Realizing that the problem exists because of a tendency of the normal equations arising from minimizing E^2 as defined in Equation 2.23 to yield values of $d\underset{\sim}{\theta}$ which are "too large", Levenberg replaces the linearized version of 2.23, i.e.,

$$S^2 = \sum_{i=1}^{N} \left[y_i - f(x_i, \underset{\sim}{\theta}_o) - \sum_{i=1}^{m} \frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_k} d\theta_k \right]^2 \qquad 2.34$$

by the related objective function

$$H^2 = wS^2 + \sum_{i=1}^{m} a_i (d\theta_i)^2 \qquad 2.35$$

where w and a_i, i = 1, ..., m are non-negative weighting factors. Thus Levenberg at once introduces in a natural way a damping effect into the normal equations, and so tends to correct the tendency to compute values of $d\underset{\sim}{\theta}$ which are "too large". While the values of the weights w and a_i are arbitary and may be adapted to the requirements of the problem at hand, a particularly effective set of values for a large class of problems has been found to be given by

$$a_k = \sum_{i=1}^{N} \left[\frac{\partial f(x_i, \underset{\sim}{\theta}_o)}{\partial \theta_k} \right]^2 \qquad 2.36$$

where w is completely aribtrary, and may be varied from iteration cycle to iteration cycle. This device was used to improve convergence on the examples computed in this work, with w being initially set equal to 5 and doubled after each iteration (it should be noted that the objective function H^2 degenerates to S^2 as $w \to \infty$). Using a_k and w as above, no tendency for the parameter corrections to become abnormally large was noticed.

The method of Marquardt is a refinement of the method of Levenburg. Based upon a consideration of the properties of the Newton-Raphson method and the steepest descent method, Marquardt has developed an algorithm for determining an appropriate value of w at each iteration to give rapid convergence of the iterative procedure. The method of steepest descent is a method of finding the minimum of a nonlinear function by choosing values of $d\underset{\sim}{\theta}^n$ in Equation 2.29 proportional to (and often equal to) the negative of the gradient of E^2 at the point $\underset{\sim}{\theta}_o^{n-1}$. The steepest descent method generally performs better than the Newton-Raphson method when the point $\underset{\sim}{\theta}_o^{n-1}$ is "far away" from the desired root, but the rate of convergence becomes very slow when the point $\underset{\sim}{\theta}_o^{n-1}$ is in a near neighborhood of the desired root. Marquardt's algorithm selects w so that the resulting correction vector $d\underset{\sim}{\theta}^n$ is an optimum interpolation between the correction vector obtained by the Newton-Raphson method and the correction vector obtained by the steepest descent method. Because of the success of the more simple method of Levenburg in the problems considered in this work, the method of Marquardt was not used. However, Marquardt's algorithm will often converge in cases where the more simple methods will not.

A strong interdependence between the components of $d\underset{\sim}{\theta}$ can cause the computed values of the parameter corrections to oscillate. This situation arises when the corrections to two dependent components alternate between positive - negative and negative-positive. Such situations can arise, for example, when the objective function is

such that an increase in θ_k has approximately the same effect on the objective function as a decrease in θ_j. While it is not likely in a well-conceived model that two components would have exactly opposite effects upon the objective function, it is possible, and indeed quite commonplace (Snyder [11]), for component pairs to possess this property to some degree. In these cases, errors introduced into the system either through inaccuracies in measurement of $\underset{\sim}{y}_N$ or through inaccuracies in the computation of values of f may cause a loss in distinction between the interdependent parameter components.

Motivated by the methods of principal component analysis in multivariate statistics, Snyder [11] has proposed an ingenious method to overcome convergence difficulties cause by interdependence of parameter components. An outline of a slightly modified version of Snyder's method is given in the following paragraphs.

Equations 2.28 may be rewritten in the form

$$A \cdot b = c \qquad 2.37$$

where

$$A = (A_{jk}) = \left(\sum_{i=1}^{N} \frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_j} \frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_k}\right)$$

$$b = (b_k) = (d\theta_k) = d\underset{\sim}{\theta}$$

and

$$c = (c_j) = \left(\sum_{i=1}^{N} (y_i - f(x_i, \underset{\sim}{\hat{\theta}})) \left(\frac{\partial f(x_i, \underset{\sim}{\theta})}{\partial \theta_j}\right)\right)$$

Noting that $\underset{\sim}{A}$ is a symmetric matrix, it follows (Hildebrand [18]) that there exists a matrix $\underset{\sim}{M}$ such that

$$\underset{\sim}{M}^T \cdot \underset{\sim}{A} \cdot \underset{\sim}{M} = (\lambda_i \delta_{ij}) . \qquad 2.38$$

Now, letting

$$\underset{\sim}{b} = \underset{\sim}{M} \cdot \underset{\sim}{Z} \qquad 2.39$$

and substituting into 2.37, there follows

$$\underset{\sim}{A} \cdot \underset{\sim}{M} \cdot \underset{\sim}{Z} = \underset{\sim}{c}$$

or, upon left multiplication by M^T,

$$\underset{\sim}{M}^T \cdot \underset{\sim}{A} \cdot \underset{\sim}{M} \cdot \underset{\sim}{Z} = \underset{\sim}{M}^T \cdot \underset{\sim}{c} \qquad 2.40$$

But the matrix $\underset{\sim}{M}^T \cdot \underset{\sim}{A} \cdot \underset{\sim}{M}$ is diagonal, and so the solution to 2.40 may be written immediately in the form

$$Z_i = \frac{[M^T \cdot c]_i}{\lambda_i} \qquad 2.41$$

provided that none of the λ_i are zero. It should be noted at this point that λ_j will be zero if Z_j may be expressed as a linear combination of

the Z_k, $k \neq j$. (That is, if the parameter corrections are interdependent.) Snyder reasoned that even if none of the λ_i are zero, some subset of the λ's (say λ_j, where j is fixed) may be very nearly zero. In this circumstance, while it would be possible to compute Z_j from equation 2.40, the small value of λ_j would tend to magnify any errors in the value of $[M^T c]_j$, and in addition, fixed errors such as roundoff in the computation of λ_j and $[M^T c]_j$ may have a significant influence upon the value of Z_j. Thus, for all values of λ_j smaller than some pre-selected value, Snyder abandons Equation 2.41 , and leaving Z_j arbitrary, transforms the original m-dimensional problem of minimizing E^2 into a lower dimensional (usually one dimensional) problem of minimizing E^2 (Z_j). This transformation is made by computing Z_i by Equation 2.41 for all i such that λ_i is large enough, using 2.39 to compute the parameter corrections in terms of the known and arbitrary Z_i, and then substituting the as yet incompletely determined new value of $\underset{\sim}{\theta}_o$ into the expression for E^2.

In those examples examined by Snyder, this technique has succeeded in reducing the dimensionality of the problem to one, in which case the required value of Z_k needed to completely determine the n^{th} approximation may be found by, for example, a linear search or some other rather insensitive procedure. The striking resemblance between the method proposed by Snyder and that of Hartley should be noted. In fact, Snyder's method seems to be a rather better m-dimensional generalization of the one-dimensional case which undoubtedly motivated Hartley, since Snyder has avoided the temptation to ignore the fact that in multi-

dimensional problems an uncertainty in the proper direction of the parameter corrections exists as surely as does an uncertainty in the magnitude of the corrections. Wilson [16] in a private communication pointed out that there is no reason why either the methods of Hartley or Snyder cannot be combined with the method of Levenberg if required. One would expect a very stable iteration procedure to result from this combination.

In his work, Snyder has often arbitrarly chosen the value of Z_k to be zero, and dispensed altogether with the one-dimensional problem with apparent success. The modified method outlined above, however, is felt to be more general, and thus applicable to a wider class of problems.

CHAPTER III

STATISTICAL PARAMETER ESTIMATION BY THE METHOD OF LEAST SQUARES

Snyder [1] has proposed that the parameters of probability density functions may be evaluated by fitting the distribution function to a data sample histogram by the method of least squares. To accomplish this fitting, a data sample is organized into a frequency histogram of an appropriate number of class intervals, and the distribution function is expressed in the frequency form for the same grouping (see Figure 3.1). The difference between the observed frequency of a given class and the frequency indicated by the distribution function for the same class is defined as the class error (see Figure 3.2). The least squares procedures of Chapter II are used to estimate the parameters (α, β) of the distribution function so that the sum of the squares of the class errors is a minimum.

Such an idea is not entirely without precedent. Kendall [19] discusses the fitting of Gram-Charlier series to data histograms, although the method of least squares is not exclusively used in this reference. Reference [21] contains a computer program designed to fit an Arne Fisher Series to a data histogram by the method of least squares. The primary feature of the techniques expounded in these two methods lies in the expansion of the distribution function in terms of a series of the products of the derivatives of the normal

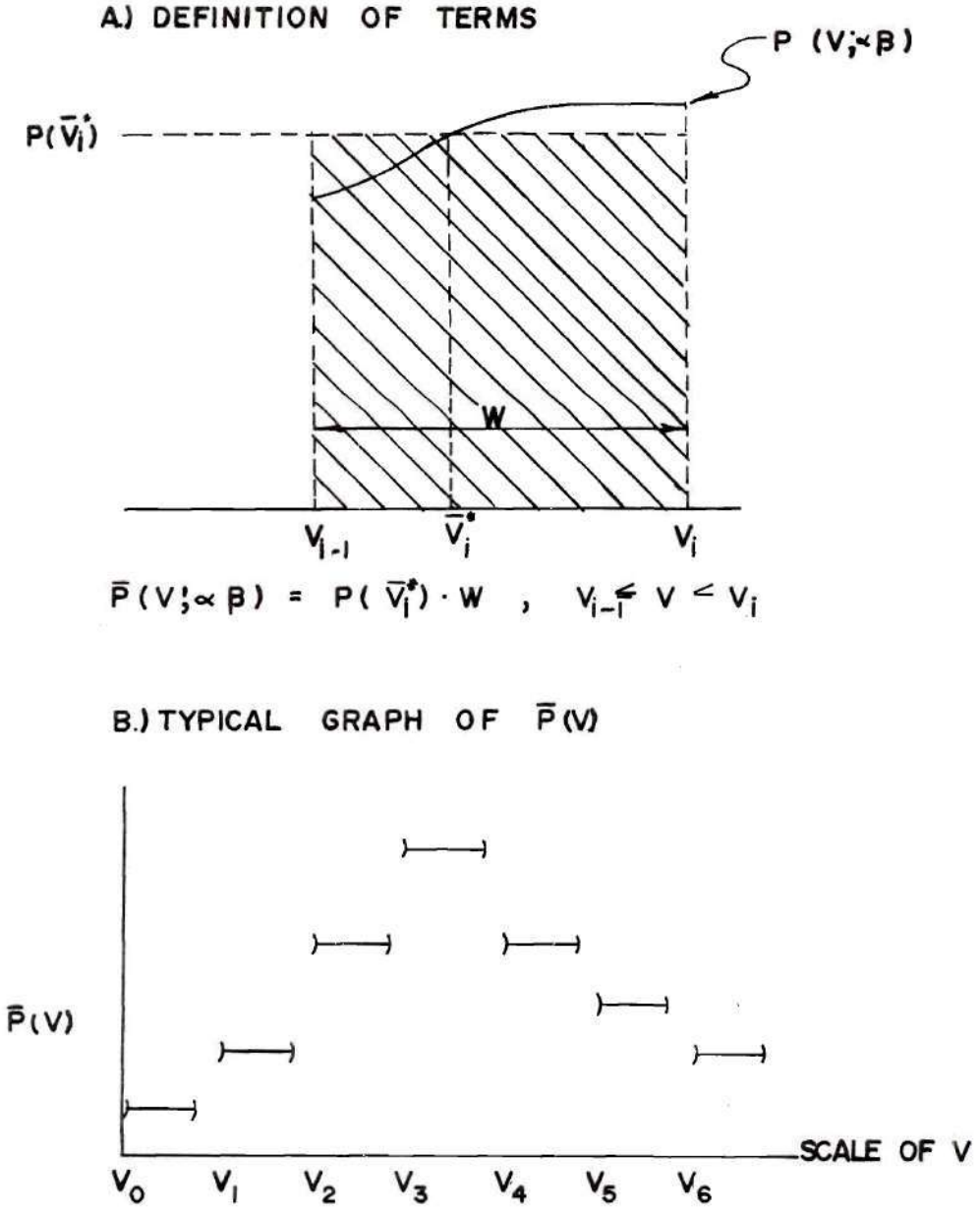

Figure 3.1 The Finite Form of a Probability Density Function

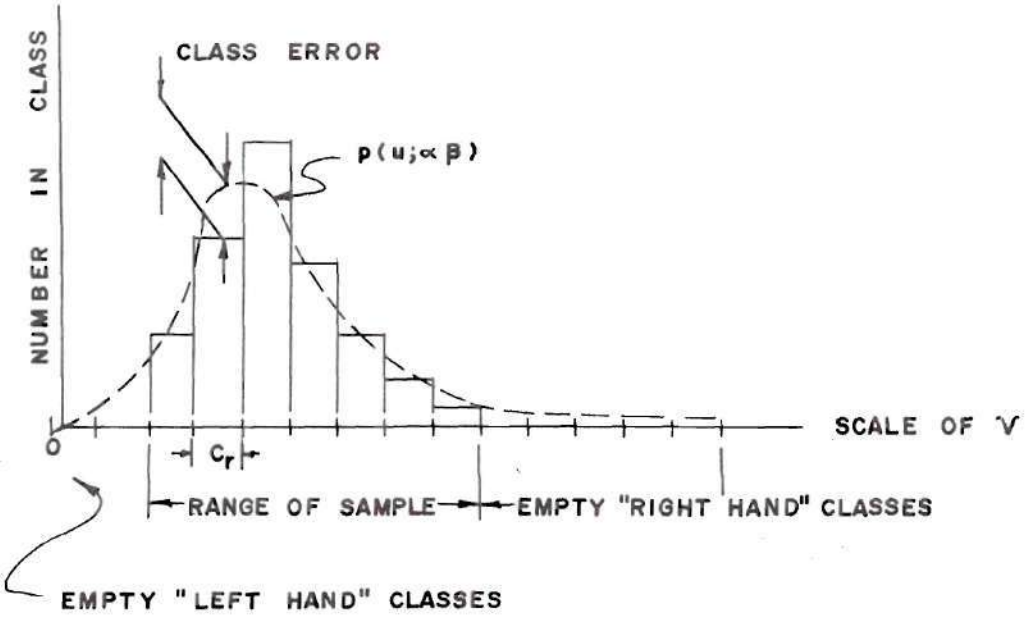

Figure 3.2 Elements of a Histogram

distribution function and Hermite ploynomials. Leaving aside the question of the completeness of such a set of functions, such an expansion would in general require an infinite number of terms for a complete representation of an arbitrary distribution function, while only a finite number of such terms could be retained for purposes of computation. By accepting the added complications arising from the non-linearity of the model, Snyder avoids the necessity of approximating the distribution function while retaining the conceptually simple method of least squares fitting to the sample histogram.

Before the method of least squares as proposed by Snyder can be accepted as a useful method for evaluating distribution parameters, however, it is necessary to examine the statistical and mathematical consequences of the least squares procedure. Initial work by Snyder [1] and Snyder and Wallace [27] has shown that the procedure is mathematically tractable. From results of Chapter II, it is known that, assuming that the error terms arising in the expression

$$h_i = f(x_i, \underset{\sim}{\theta}) + e_i \qquad 3.1$$

are normally distributed with zero mean and variance independent of x_i, the parameter estimates obtained by least squares will be equivalent to maximum likelihood estimates. Nevertheless, intuitively on would feel that these estimates may differ from those obtained from a conventional application of the method of maximum likelihood. In particular, it

appears that the method of least squares may be somewhat inefficient since the procedure requires that the data sample be grouped into a histogram before the distribution parameters can be estimated. It would seem that some portion of the information content of the sample may be lost when the sample is so grouped. For the small samples usual in hydrology, such a loss of information could be so serious as to make the method of least squares useless. Also, it would appear that the least squares estimates might be a function of the manner in which the sample is grouped, so that different groupings of the same sample might lead to radically different parameter estimates.

In spite of the intuitive misgivings which the above observations may prompt, the method of least squares as a technique for estimating statistical parameters is not without appeal. Although the method is apt to be numerically laborious, the nature of the computations are quite routine and easily programable on digital computers. The nature of the computations are identical regardless of the type of distribution function being fitted, and so once a program has been developed, it would, with only minor modifications, be able to fit any function desired. The method of constructing confidence regions outlined in Chapter II allows the (relatively) easy construction of tolerance limits for any distribution function based upon any sample. In fact, the actual least squares fitting process may be avoided by using the confidence region to estimate the parameter values. Also, the method of least squares is easily visualized, and by employing an analysis of sensitivity coefficients (that is, the values of the derviatives

of the distribution function), the effects of the shape of the sample histogram upon the estimated values of the parameters may be evaluated (Snyder [1]). Finally, the inclusion of the arbitrary weighting function as a part of the sum of the squares of the errors makes the method of least squares extremely versatile and capable of emulating other standard techniques (for example, if $w_i = [f(\underset{\sim}{x}_i, \underset{\sim}{\theta})]^{-1}$, then the method of least squares becomes the method of minimum chi-squares) or being adapted to the contingencies of the problem at hand (for example, later in this work a weighting function is sought for which the least squares parameter estimates are unbiased).

Selection of the Density Function for Numerical Experimentation

In order to test the feasibility of the method of least squares as a means of the fitting of distribution functions, a systematic study was made of the statistical properties of least squares estimators of population parameters. The probability density function chosen for study was the two parameter gamma distribution. This choice was motivated by a study by Markovic [22], which indicated that this relatively simple distribution function fits most hydrologic data of an "annual" nature at least as well as any other commonly used distribution, and thus the results of this investigation would have immediate practical applications. In addition, the two parameter gamma distribution was preferred because it seemed on the one hand to offer enough computational difficulty to illustrate the power of the least squares method

of fitting, while on the other hand the small number of parameters prevented an undue amount of difficulty in establishing trends in the characteristics of the estimators as they might occur during the study. The use of the two-parameter model also avoided serious numerical difficulties which often arise when a model includes a so-called "shifting parameter", or lower bound such as occurs, for example, in the three-parameter gamma or the Pearson Type III distribution (Matalas and Wallis [23]). The use of the two-parameter model is also philosophically more satisfactory when dealing with hydrologic variates such as rainfall or runoff, since this model imposes no bounds on the possible range of the variate other than bounds already implicit in the definition of the variate. Thus, in this sense, the two-parameter model provides a more symmetric treatment of large and small values than do models which include shifting parameters.

The essential properties of the two parameter gamma distribution (henceforth referred to simply as the gamma distribution) are summarized in Appendix A. Appendix B develops the mathematical framework necessary for the specialization of the methods of Chapter II to the problem of fitting the gamma distribution to a given histogram.

Properties of the discrete form of the gamma distribution may be obtained from the following equivalent forms of the distribution. Using the results as given in Appendices A and B, the gamma distribution in its continuous form may be written as

$$p(v) = \frac{\alpha^{\beta}}{\Gamma(\beta)} v^{\beta-1} e^{-\alpha v} \quad , \quad 0 \leq v < \infty \qquad 3.2(a)$$

and in its discrete form as

$$\bar{p}(v_i^*) = \int_{v_{i-1}}^{v_i} \frac{\alpha^{\beta}}{\Gamma(\beta)} v^{\beta-1} e^{-\alpha v} dv \quad , \quad v_{i-1} \leq v_i^* \leq v_i \qquad 3.2(b)$$

But if $w = v_i - v_{i-1}$, then by the mean value theorem for integrals (Taylor [24]), there exists a number v_i^* in the interval $[v_{i-1}, v_i]$ such that

$$\bar{p}(v_i^*) = p(\bar{v}_i^*) \cdot w \qquad 3.3(a)$$

or, if it be necessary to emphasize the dependence of $\bar{p}$ on α and β,

$$\bar{p}(v_i^*; \alpha, \beta) = p(\bar{v}_i^*; \alpha, \beta) \cdot w \qquad 3.3(b)$$

From Equations 3.3, it is seen that for fixed α and β, $\bar{p}$ is a step function of v with finite discontinuties at the right end-points of the intervals $[v_{i-1}, v_i]$. For fixed v, $\bar{p}$ is a continuous function of the parameters α and β, and for positive values of the parameters, possesses all orders of derivatives. Figure 3.1 graphically depicts

the quantities described above.

From Equation 3.2(a), the derivatives of p with respect to α and β may be written as

$$\frac{\partial p}{\partial \alpha} = \left(\frac{\beta}{\alpha} - v\right)\, p(v; \alpha, \beta) \qquad 3.4(a)$$

and

$$\frac{\partial p}{\partial \beta} = \left(\frac{-\Gamma'(\beta)}{\Gamma(\beta)} + \ln \alpha \ + \ln v\right)\, p(v; \alpha, \beta)\,. \qquad 3.4(b)$$

By Equations 3.3 and 3.4 the derivatives of $\bar{p}$ with respect to α and β are thus

$$\frac{\partial \bar{p}}{\partial \alpha} = \left(\frac{\beta}{\alpha} - \bar{v}^*\right)\, p(\bar{v}^*; \alpha, \beta)\cdot w \qquad 3.5(a)$$

and

$$\frac{\partial \bar{p}}{\partial \beta} = \left(-\frac{\Gamma'(\beta)}{\Gamma(\beta)} + \ln \alpha \ + \ln v\right)\, p(\bar{v}^*; \alpha, \beta)\cdot w \qquad 3.5(b)$$

Where

$$v_{i-1} \leq v \leq v_i\,.$$

The values of $\bar{p}$ and its derviatives were numerically estimated in this work by partitioning the interval $[v_{i-1}, v_i]$ into sixteen sub-intervals and using the trapezoidal rule to approximate the integrals in Equation 3.2(b). An examination of the results of this approximation indicated it to be of sufficient accuracy for the purpose at hand. No serious investigation was made to determine the smallest number of intervals required for an adequate estimation, although it should be noted that in most cases no discernable difference in the results of fitting could be detected when only four intervals were used. Where differences arose, they were invariably of the form of improved convergence where the finer mesh was used, and not in a difference in the estimated values of the parameters in those cases where convergence occurred for both mesh sizes.

Using the above formulation, a computer program was prepared to generate random numbers from a gamma population with given parameters α and β (subject to the condition that β be integral) and to fit these samples by least squares. Included in this program were provisions to compute for each sample the parameter estimates obtained by the method of moments and by maximum likelihood. Real hydrologic data (see Chapter V) was then fit by the program in an effort to determine values of the parameters which might correspond to situations of interest in hydrology. (In this fitting, the technique obtained from Markovic [22] of dividing the real sample by its average value was used. This scaling accomplished the purpose of always making the maximum likelihood estimators of α and β equal. Thus, in the later simulation work, only those populations

for which α and β were equal were investigated, resulting in a vastly more efficient study without, it was felt, sacrificing any generality of importance in the analysis of real data.)

The results of these initial fittings of real data indicated that, at least for the types of data examined in this study, one should expect values of population parameters ranging, say, between one and eleven. This parameter range was utilized in developing the simulation runs.

A Description of the Numerical Experiments

Simulation runs were designed to provide empirical indications of answers to the questions raised earlier in this chapter regarding the appropriateness of least squares as a technique for fitting distribution functions. In particular, the questions addressed were as follows:

1. To what extent are the least squares estimates of the population parameters influenced by the manner in which the sample is grouped into a histogram?
2. Is there a rule according to which the class width and number of classes may be selected so that the data may be grouped in a manner which will yield optimal or near-optimal effectiveness of the least squares fit?
3. Is the assumption of normality of the errors defined by Equation 3.1 (or, more generally, by 2.1) appropriate?

4. What is the efficiency of least squares fitting when compared to fitting by maximum likelihood or moments?

5. Does the method of least squares provide an advantage when one encounters a sample which appears abnormal (for instance, a sample which appears to contain one or more outliers)?

6. Does the method of least squares provide any advantages when dealing with problems of particular interest in hydrology, for example, is the method of least squares more or less stable with a growing sample of small initial size?

7. Is the inclusion of the weighting factors introduced in Chapter II necessary for a satisfactory fit by least squares, and if so, can any guidance as to likely appropriate forms of the weighting factors be given?

To attempt to obtain answers to the above questions, one hundred and two runs of one hundred random samples per run were made using the computer program developed for this purpose. These runs were based upon three thousand random samples varying in size from twenty five to one hundred, a range typical of many common problems in hydrology. These runs were made over a range in parameters varying integrally from two to eleven. For each run of one hundred samples, the parameter estimates for each sample obtained by least squares, by maximum likelihood, and by moments were computed, as were the sample mean and variance of the estimates obtained by each method of fitting. In addition, a chi-square goodness of fit and a Kolmogorov-Simirnov goodness of fit were performed on the errors as defined by Equation 3.1, in an attempt

to ascertain if the assumption of normality of the error terms was justified.

Appendix E gives a summary description of each run in which the appropriate parameters, size, grouping, and random samples are identified. Briefly, the purposes of the various subsets of these runs were as follows:

1. Runs 1 through 10 were made primarily to determine to what extent the manner of grouping the data into histograms affected the efficiency of the fitting process. Runs 1 through 3 and 5, 6, 7 and 10 were made using three different groupings for each run, and runs 4, 8 and 9 were using two different groupings for each run. All groupings were based upon a variation of a method for selecting a class interval proposed by Sturges [25]. Sturges suggests that a reasonable choice of a class interval to use in constructing a histogram from a set of N data points is given by the formula

$$C = \frac{R}{1 + 3.322 \log N} \qquad 3.6$$

where R is the range of the sample, and log means the logrithm to the base 10. For purposes of this study, Equation 3.6 was modified by scaling the value of C obtained from Equation 3.6 by some positive constant, that is

$$Cr = \frac{rR}{1 + 3.322 \log N} \qquad 3.7$$

The effect of the value of r upon the efficiency of the least squares procedure was investigated for values of r equal to 0.5, 0.75 and 1.0. For a given value of Cr, the histogram was constructed by selecting the centerpoint of the first group by the relation

$$X_1 = Cr. \left[\frac{X \text{ min} + 1/2 \text{ Cr}}{Cr}\right] \qquad 3.8$$

where [a] indicates the integral portion of a.

The centerpoints of subsequent groups were calculated from the relation

$$X_i = X_{i-1} + Cr \ . \qquad 3.9$$

The frequency histogram was constructed by determing the number of observations falling into each interval $(X_i - 0.5\ Cr,\ X_i + 0.5\ Cr]$, and then by adding to the lower portion of the histogram the smallest number of empty classes required to contain the point X = 0 (see Figure 3.1).

2. Runs 5,6,9 and 10 were then repeated using a value of r selected from the initial runs, and adding to the right end of the histogram 1, 2, 4 and 8 empty classes (not all combinations were used with each run, see Appendix E). The motivation behind this series of runs was to test a hypothesis by Snyder [1] that the efficiency of the fitting process could be improved by considering empty classes beyond the range of the observed sample in the fitting process.

3. Runs 5R, 7R, 9R, and 10R were then made utilizing sample sizes of 75 to 100 to determine what effect, if any, the sample size had on the trends observed in the previous two sets of runs which were made with sample sizes of 50.

4. The above runs were made using a weighting function of unity in all cases. In order to explore the possibility of utilizing the weighting factors to improve the efficiency of the least squares fitting procedures, a particular set of weights

$$w_i = [f(x_i; \alpha, \beta)]^{-\gamma} \qquad 3.10$$

were chosen, where γ was assumed to lie between 0 and 1. (Note that if $\gamma = 0$, $w_i = 1$, and the method collapses to the case investigated above. If $\gamma = 1$, the fitting is by the method of minimum chi squares as discussed in Kendall[3]. For $0 < \gamma < 1$, it would appear that the method of least squares might be made to assume characteristics ranging between these two extremes.) Runs 4,6,8 and 9 were repeated with a sample size of 50, and for values of γ equal to 0.25, 0.50, 0.75, and 1.0. Runs 4R and 8R were repeated for sample sizes of 75 and 100 and the above values of γ. Runs 6 and 9 were repeated for values of of 0.75 and 1.0 and adding four and eight empty classes to the right edge of the histogram. Finally, Runs 14A, 16A, 19A; Runs 14B, 16B, 19B; and Runs 14C, 16C, 19C were made for sample sizes of 50, 75, and 100 respectively and for values of γ of 0.50 and 0.75. The purpose of these

runs was to provide a larger base of data upon which the conclusions which would be drawn in this study could be based. A presentation and analysis of the results obtained from the various runs and examples discussed in this chapter are given in Chapter IV.

The normal equations for use in the method of weighted least squares were developed in the following manner.

Let

$$w_i = [f(x_k; \alpha, \beta)]^{-\gamma} \tag{3.11}$$

and write

$$E^2 = \sum_{i=1}^{N} (h_k - f(x_k; \underset{\sim}{\theta}))^2 \, w_k \tag{3.12}$$

Then

$$\frac{\partial E^2}{\partial \theta_j} \doteq -2 \sum_{k=1}^{N} w_k (h_k - f(x_k; \underset{\sim}{\theta})) \frac{\partial f(x_k; \underset{\sim}{\theta})}{\partial \theta_j} \quad , \tag{3.13}$$

where the terms involving $\frac{\partial w_k}{\partial \theta_j}$ have been ignored (see for example, Kendall [3], or Yevjevich [26] for justification). Then by the same technique of approximating f by the linear terms in the Taylor series expansion for f about some initial estimate $\underset{\sim}{\theta}_o$ as was used in Chapter II, one obtains from Equations 2.22,

$$\sum_{k=1}^{N} [f(x_k; \underset{\sim}{\theta}_o)]^{-\gamma} (h_k - f(x_k; \underset{\sim}{\theta}_o) - \nabla f(x_k; \underset{\sim}{\theta}_o) \cdot d\underset{\sim}{\theta}_o) \frac{\partial f(x_k; \underset{\sim}{\theta}_o)}{\partial \theta_j} = 0 \quad , 3.14$$

or, after simplifying

$$\sum_{r=1}^{m} \left(\sum_{k=1}^{N} [f(x_k; \underset{\sim}{\theta}_o)]^{-\gamma} \frac{\partial f(x_k; \underset{\sim}{\theta}_o)}{\partial \theta_j} \frac{\partial f(x_k; \underset{\sim}{\theta}_o)}{\partial \theta_r} \right) d\theta_r =$$

$$\sum_{k=1}^{N} [f(x_k; \underset{\sim}{\theta}_o)]^{-\gamma} \frac{\partial f(x_k; \underset{\sim}{\theta}_o)}{\partial \theta_j} (h_k - f(x_k; \theta_o) \quad , \quad j = 1, \ldots, m . \qquad 3.15$$

As in the case of unitary weights, Equations 3.15 are solved iteratively until the length of the correction vector $d\underset{\sim}{\theta}$ falls below a predetermined value.

As an attempt to complete the discussion of non-linear least squares as a method for determining statistical parameters, a confidence region and a tolerance limit were constructed for sample 100 of run 10R3. The partitioning of the sum of squares of the errors was accomplished using the method of Halperin as given by Equations 2.16. The confidence region was constructed both for weights of unity and weights equal to $[f(\underset{\sim}{x}_k, \underset{\sim}{\theta})]^{-0.75}$ in order to determine what influence the weighting factors may have on the confidence region.

Due to the highly complex nature of the expression for the confidence region as given by Equation 2.14, no attempt to solve explicitly for the boundaries of the various regions was made. Instead, values of $\frac{N-m}{m} \frac{reg(e)}{res(e)}$ were computed at pre-established points in the $\underset{\sim}{\theta}$ plane,

and the probability η corresponding to each value was computed. These probabilities were then plotted and contours of equal probability were sketched. While this procedure is a crude approximation to a very complex process, a careful inspection of the results indicate that this procedure is quite adequate for the applications in which it would ordinarily be used in hydrology.

Since it would be of interest to compare tolerance limits determined by least squares with those determined by maximim likelihood, and since the tolerance limits as determined by maximum likelihood are very difficult to determine for this gamma distributed sample, a normally distributed sample was constructed and fit by maximum likelihood and by least squares, and the tolerance limits for a one hundred year event as determined by the two methods were computed.

Confidence regions, and a discussion of their more interesting properties and their use in constructing tolerance limits for specified values of probability of occurence are given in Chapter IV.

CHAPTER IV

ANALYSIS OF RESULTS OF NUMERICAL EXPERIMENTS

In this chapter, a detailed presentation and analysis of the results of the computer runs comprising the numerical experiments conducted as a portion of this study are made. In each of these simulation runs, one hundred samples of numbers assumed to be randomly drawn from a population distributed as a two parameter gamma distribution with fixed parameters α and β were generated (see Appendix E). Each such random sample defined a case of the particular run in which it was considered. An estimate a, b of α, β was made for each such sample by the method of least squares, by the method of maximum likelihood, and by the method of moments. Altogether, one hundred such estimates were made for each run (a given run would, of course, be based upon the same population parameters and the same sample size), and at the end of each run, the sample mean and sample variance for each type of estimate were computed for both a and b. In addition, for each case the error terms defined by Equation 3.1 were tested by a chi-square and a Kolmogorov-Smirnov goodness of fit test to determine if the assumption of normality of the error terms is justified.

Results for Unweighted Least Squares

The Choice of a Class Interval

Runs 1 through 10 were made primarily to determine the effects, if any, of the manner of grouping a sample upon the least squares estimates of the population parameters. Intuitively it was believed at the beginning of these runs that there might be an optimum number of groups for a sample of a given size, and in addition that a good grouping should be one for which the resulting histogram assumed a relatively smooth or recognizable shape.

Runs 1 through 3 were thus made using class intervals which were 0.5, 0.75, and 1.0 times the length recommended by Sturges [25]. All three runs used sample sizes of 50. Runs 1 and 2 sample a population with parameters $\alpha = \beta = 4$. Run 3 was based upon a population with parameters $\alpha = \beta = 8$. The optimum class interval for each of these runs was selected as that interval for which the sample variance of the parameters was a minimum. For run 1 this multiple was 0.5, for run 2 it was 0.75, while for run 3 it was 1.0. Unenlightened by these results, runs 4,8, and 9 were made for multiples equal to 0.5 and 0.75, and runs 5 and 10 were made for multiples of 0.5, 0.75, and 1.0. These runs were based upon sample sizes of 50, and on population parameters ranging from 2 to 11. In this series of runs, a multipler of 0.5 gave best results in two cases, a multiplier of 0.75 gave best results in one case, and a multiplier of

1.0 gave best results in one case. Run 8 was a draw, there being no essential difference in the results in this case.

Convergence of the iterative procedure was found in these runs to be something of a problem, especially for the higher values of the population parameters. Thus the above runs were examined to determine if any multiplier exhibited a markedly better convergence record than the others. It was found that there was no difference in convergence in 4 cases, the multiplier 0.5 was superior in two runs (in each instance by a margin of one case), the multiplier 0.75 was superior in one run (by a margin of three cases), and the multiplier 1.0 was superior in one case (by a margin of one case). These results are summarized in Table 4.1. It should be mentioned at this point that the cases which failed to converge in the above runs were not always the same cases for each multiplier. Thus, the least squares procedure could converge for a sample grouped in one way, and fail to converge when the sample was grouped in another way.

The histograms for each case of run described above were examined rather closely to determine if any functional trend between the shape of the histogram and the performance of the method of least squares could be detected. Since the true values of the population parameters were known, it was possible to compared the least squares parameter estimates with the likelihood and moments estimates and in the light of this knowledge to arrive at a subjective description of the quality of the least squares estimates. Thus, least squares estimates which were substantially closer to the true parameter values

Table 4.1 Results of Runs for Various Class Widths, $\gamma = 0$

Run	Sample Size	γ	δ	τ	Parameter Means						Parameter Variances						No. of * Convergences
					a_{ls}	a_{ml}	a_{mom}	b_{ls}	b_{ml}	b_{mom}	a_{ls}	a_{ml}	a_{mom}	b_{ls}	b_{ml}	b_{mom}	
1	50	4	4	1.0	4.37	4.25	4.31	4.35	4.25	4.32	2.04	0.97	1.20	1.70	0.82	0.99	100
1	50	4	4	0.75	4.33	4.25	4.31	4.32	4.25	4.32	1.83	0.97	1.20	1.60	0.82	0.99	100
1	50	4	4	0.50	4.31	4.25	4.31	4.29	4.25	4.32	1.45	0.97	1.20	1.26	0.82	0.99	100
2	50	4	4	1.0	4.38	4.21	4.31	4.38	4.24	4.34	1.70	0.66	0.91	1.42	0.52	0.72	100
2	50	4	4	0.75	4.18	4.21	4.31	4.22	4.24	4.34	1.16	0.66	0.91	1.04	0.52	0.72	100
2	50	4	4	0.50	4.24	4.21	4.31	4.25	4.24	4.34	1.41	0.66	0.91	1.12	0.52	0.72	100
3	50	8	8	1.0	8.93	8.53	8.72	9.01	8.61	8.79	5.35	2.27	3.14	5.28	2.29	3.19	100
3	50	8	8	0.75	9.36	8.53	8.72	9.47	8.61	8.79	7.43	2.27	3.14	7.73	2.29	3.19	100
3	50	8	8	0.50	9.07	8.53	8.72	9.14	8.61	8.79	6.41	2.27	3.14	6.42	2.29	3.19	100
4	50	3	3	0.75	3.36	3.20	3.27	3.27	3.16	3.23	1.50	0.49	0.60	1.11	0.44	0.55	100
4	50	3	3	0.50	3.36	3.20	3.27	3.27	3.16	3.23	1.29	0.49	0.60	0.89	0.44	0.55	100
5	50	2	2	1.0	2.34	2.16	2.17	2.26	2.16	2.17	0.61	0.25	0.32	0.44	0.19	0.26	100
5	50	2	2	0.75	2.34	2.16	2.17	2.29	2.16	2.17	0.65	0.25	0.32	0.50	0.19	0.26	100
5	50	2	2	0.50	2.39	2.16	2.17	2.29	2.16	2.17	1.00	0.25	0.32	0.52	0.19	0.26	100
6	50	5	5	0.5	5.43	5.25	5.36	5.40	5.23	5.34	2.65	1.63	1.89	2.16	1.45	1.67	100
7	50	6	6	0.5	6.31	6.25	6.36	6.22	6.19	6.30	3.47	1.68	1.97	3.21	1.72	1.95	100
8	50	7	7	0.75	7.72	7.39	7.52	7.73	7.42	7.57	5.44	1.87	2.45	4.76	1.74	2.40	97
8	50	7	7	0.50	7.80	7.39	7.52	7.84	7.42	7.57	5.41	1.87	2.45	4.90	1.74	2.40	98
9	50	9	9	0.75	9.52	9.42	9.73	9.53	9.37	9.68	6.19	2.97	3.77	5.47	2.63	3.30	95
9	50	9	9	0.50	9.39	9.42	9.73	9.41	9.37	9.68	5.37	2.97	3.77	4.80	2.63	3.30	92
10	50	11	11	1.0	11.49	11.35	11.89	11.44	11.34	11.89	6.39	3.63	6.21	6.02	3.63	6.31	86
10	50	11	11	0.75	11.31	11.35	11.89	11.31	11.34	11.89	5.85	3.63	6.21	5.83	3.63	6.31	85
10	50	11	11	0.50	11.23	11.35	11.89	11.22	11.34	11.89	6.20	3.63	6.21	5.78	3.63	6.31	85
5R1	25	2	2	0.50	Run was numerically unstable												
5R2	75	2	2	0.50	2.23	2.11	2.13	2.19	2.11	2.13	0.36	0.13	0.19	0.24	0.10	0.15	100
5R3	100	2	2	0.50	2.16	2.07	2.08	2.14	2.08	2.09	0.20	0.10	0.14	0.15	0.08	0.11	100
7R1	25	6	6	0.50	6.77	6.58	6.71	6.62	6.51	6.63	8.14	3.95	4.77	5.52	3.60	4.35	94
7R2	75	6	6	0.50	6.20	6.20	6.33	6.12	6.11	6.23	1.87	1.06	1.27	1.51	0.92	1.10	100
7R3	100	6	6	0.50	6.32	6.15	6.23	6.25	6.09	6.17	1.79	0.77	0.92	1.55	0.73	0.86	100
9R1	25	9	9	0.50	8.92	9.68	10.21	9.03	9.66	10.20	8.88	7.48	10.86	7.93	6.67	10.13	81
9R2	75	9	9	0.50	9.23	9.25	9.52	9.25	9.21	9.48	3.44	1.86	2.26	3.19	1.73	2.01	94
9R3	100	9	9	0.50	9.24	9.14	9.38	9.25	9.11	9.35	3.58	1.41	1.65	3.30	1.30	1.52	98
10R1	25	11	11	0.50	10.34	11.85	13.13	10.27	11.86	13.14	12.01	8.93	20.10	10.16	9.62	21.49	69
10R2	75	11	11	0.50	11.49	11.12	11.49	11.46	11.15	11.51	5.14	2.68	4.26	4.99	2.74	4.35	91
10R3	100	11	11	0.50	11.68	11.15	11.49	11.65	11.15	11.49	4.49	1.55	2.31	4.34	1.59	2.43	95

* Maximum Number of Convergences is 100.

than were the corresponding likelihood and moments estimates were described as good, while those which were substantially further from the true values were described as poor. Comparisons between different least squares estimates obtained by different groupings of the same sample were judged according to their respective closeness to the true parameter values. As has been mentioned before, it was felt that a histogram which "looked like" the graph of a probability density function should yield better performance from the method of least squares than a histogram which was without form. Surprisingly, the results of this examination proved this conjecture false. While excellent results could be obtained from the fitting of histograms which "looked good", results quite as good could be obtained by fitting histograms which appeared to be almost without form. Sad to say, poor results could also be obtained for each type of histogram.

The parallel results from each run for each choice of class interval were examined to determine if the parameter estimates for the same sample under different groupings were the same. It was found that, while rarely were the estimates the same, they were in almost every case consistent in that they exhibited the same tendencies (that is, the parallel estimates tended to cluster about the same value for different groupings, and the difference between these estimates was small compared to the difference between these estimates obtained from different samples), and that there was no reason, other than the trivial reason of having prior knowledge of the population parameters, for prefering one estimate to another.

Based upon the above results, is was postulated that for the groupings considered in these runs, there was no reason to prefer one grouping to another. Thus, all runs made beyond this point were all made using a multiplier arbitrarily chosen equal to 0.5.

At this point it was surmised that perhaps instead of grouping data according to a fixed class width, a more natural method of grouping the sample might be to group by varying class widths so selected as to contain an equal number of observed points in each interval. Groupings of this type were tried for several of the above runs and for from one to four points in each interval for each run. The results obtained from this method of grouping were consistently inferior to those obtained by grouping according to a fixed class width. Thus this approach was abandoned and the study continued with groupings based upon class widths equal to one-half the width recommended by Sturges.

Extension of the Range of Optimization beyond the Sample Range

Runs 5,6,9, and 10 were then repeated with one, two, and four empty classes added to the end of each histogram. These runs were made in order to test the hypothesis by Snyder [1] that by including empty classes in histograms one may obtain a better fit of the density function to the histogram, especially when the sample under study appears to contain only records of frequent events. The results of this set of runs are inconclusive as to the validity of this hypothesis. There is some evidence in these runs that the inclusion of

a few empty classes in some cases actually does improve the estimates of the population parameters. Also, in a few cases this procedure caused the iteration process to converge where without the empty classes it had failed to converge. In a few cases the estimates were worsened by the inclusion of the empty classes. In the large majority of cases, however, the inclusion of empty classes simply had no appreciable effect upon the outcome of the fitting process. This is especially true when the sample already contains one or more points which represent rare events (rare, of course, is relative to the size of the sample).

As an example of an instance in which the inclusion of empty classes was of benefit, case 44 of run 10 ($\alpha = \beta = 11$) with no empty classes added yielded parameter estimates a = 8.85 and b = 8.50. The sample was such that for these values of a and b, the cumulative distribution function at the end of the final classes (v= 1.3421) was equal to 0.8732. With the addition of four empty classes to the right end of the histogram, the parameter estimates were changed to a = 9.77, and b = 9.07. The value of the cumulative distribution function at the end of the final empty class (v = 1.617) for these parameters was 0.974.

On the other hand case 46 of the same run yielded parameter estimates of a = 9.49 and b = 9.04 with no empty classes and a = 9.50 and b = 9.04 with the addition of four empty classes. In this case, the value of the cumulative distribution function at the end of the last occupied class (v = 2.066) was equal to 0.997.

Lest the preceeding two examples be misleading, attention should also be given to case 57 of the same run (again, $\alpha = \beta = 11$). In this case, fitting with no empty classes yielded parameter estimates of a= 16.27 and b = 15.76. The value of the cumulative distribution function at the end of the last occupied class was, for these estimates, equal to 0.969. Upon the addition of four empty classes to the right end of the histogram, the estimates a = 16.92 and b = 16.32 were obtained. Thus, the inclusion of empty classes with the histogram actually worsened the parameter estimates in this case.

A summary of the results of these runs is given in Table 4.2. In analysing these results, care must be taken to avoid choosing one run as superior to another when the number of convergent cases is not the same in each instance.

It may be noticed that the three cases examined above all have the property that the parameter estimates with empty classes are greater than the estimates obtained from the sample histogram itself. This is not an accident, and in fact every case in every run which was made using empty classes exhibited this phenomenon. The reason for this occurence may be discovered by examining the two parameter gamma probability density function and by noting that there exists a positive correlation between the estimates a and b. Since higher values of α and β lead to a more slender tail of the distribution function, the inclusion of empty classes, and the consequent class errors occasioned by their inclusion, tends to cause the least squares procedure to raise the values of the estimates a and b in an effort to minimize these errors

Table 4.2 Results of Runs with Empty Right - Hand Classes, Y = 0

Run	No. of Empty Classes	α	β	Mean of a	Mean of b	Variance of a	Variance of b	Number of Convergences
5N1	1	2	2	2.39	2.29	1.00	0.52	100
5N2	2	2	2	2.39	2.30	1.00	0.52	100
5N4	4	2	2	2.40	2.30	0.99	0.52	100
6N1	1	5	5	5.46	5.42	2.63	2.16	100
6N2	2	5	5	5.47	5.44	2.62	2.15	100
6N4	4	5	5	5.48	5.44	2.61	2.15	100
9N1	1	9	9	9.59	9.57	6.52	5.69	93
9N2	2	9	9	9.36	9.36	4.55	4.10	90
9N4	4	9	9	9.46	9.47	4.94	4.58	91
10N1	1	11	11	11.31	11.27	5.90	5.48	86
10N2	2	11	11	11.46	11.44	6.02	5.88	88
10N4	4	11	11	11.62	11.60	6.60	6.42	91

by further thinning of the tail of the distribution function. Thus, the inclusion of empty classes tends to introduce a positive bias to the least squares estimators. This is an unfortunate consequence, since the results of the runs made in this study indicate that unweighted least squares estimators are positively biased (as are the maximum likelihood estimators) even without the inclusion of empty classes.

The Effects of Sample Size

To investigate the behavior of least squares estimators for various sample sizes, runs 5R1, 7R1, 9R1, and 10R1 were made for a sample size of 25; 5R2, 7R2, 9R2, and 10R2 were made for a sample size of 75, and runs 5R3, 7R3, 9R3, and 10R3 were made using a sample size of 100. These runs utilized parameter values ranging from 2 to 11. The primary motivation behind these runs was to determine the behavior of the sampling mean and variance of the least squares estimates with increasing sample size, as well as to test the effectiveness of the least squares procedure over a range of sample sizes and determine if any practical lower bound on the size of a sample exists below which the least squares procedure should not be used.

The results of these runs were first examined to determine any trends which might appear in the sample variances of the parameter estimates obtained from the various runs. According to the theory as developed in Chapter II, one would expect that these variances should be have according to a relation of the form a function of α, β and N. That is,

$$S^2 \sim \frac{f(\alpha, \beta)}{N} \quad . \qquad 4.1$$

Since the square root of the variance represents in some fashion an amount of departure from the central value of the estimates, the assumption that the method of least squares is no more efficient for one set of values of parameters than for another would imply that $f(\alpha, \beta)/\alpha\beta$ is nearly constant for fixed N. Thus, one is lead to expect that the sample variance of the parameter estimates may be expressible in the form

$$\frac{S^2}{p^2} = \frac{c_p}{N} \qquad 4.2$$

where p represents either α or β, and c_p is a constant which may have different values depending upon whether p represents α or β. Table 4.3 shows an attempt to represent the sampling variance of a in this form. Figure 4.1 shows the results of Table 4.3 in graphical form. Also shown in Table 4.3 and Figure 4.1 is an equivalent attempt to represent the sampling variance of the maximum likelihood estimates in the same form. In each case, the results indicate that the "constant" c_α actually is function of α. Figure 4.1 shows also that the maximum likelihood estimates are much more stable under the analysis than are the least squares estimates. Figure 4.2 shows an attempt to represent sample variances of the least squares estimates of α in the form

Table 4.3 Characteristics of Parameter Variances, $\gamma = 0$

N	α	σ^2_{LS}	$\frac{\sigma^2_{LS}}{\alpha^2} N$	σ^2_{ML}	$\frac{\sigma^2_{ML}}{\alpha^2} N$	$\frac{\sigma^2_{LS}}{\sigma^2_{ML}}$	$\frac{\sigma^2_{LS}}{\alpha^2}\sqrt{N}$
25	2	run was numerically unstable					
50	2	1.00	12.5	0.25	3.00	4.00	1.80
75	2	0.36	6.75	0.13	2.25	2.75	0.78
100	2	0.20	5.0	0.10	2.50	2.0	0.50
25	6	8.14		3.95			
50	6	3.47	4.4	1.68	2.35	2.07	0.62
75	6	1.87	3.9	1.06	2.25	1.76	0.45
100	6	1.79	5.0	0.77	2.14	2.32	0.5
25	9	8.88		7.48			
50	9	5.37	3.3	2.97	1.85	1.80	0.47
75	9	3.44	3.2	1.86	1.72	1.84	0.37
100	9	3.58	4.4	1.41	1.74	2.54	0.44
25	11	12.01		8.93			
50	11	6.20	2.75	3.63	1.5	1.70	0.39
75	11	5.14	3.20	2.68	1.65	1.91	0.37
100	11	4.49	4.10	1.55	1.28	2.90	0.41

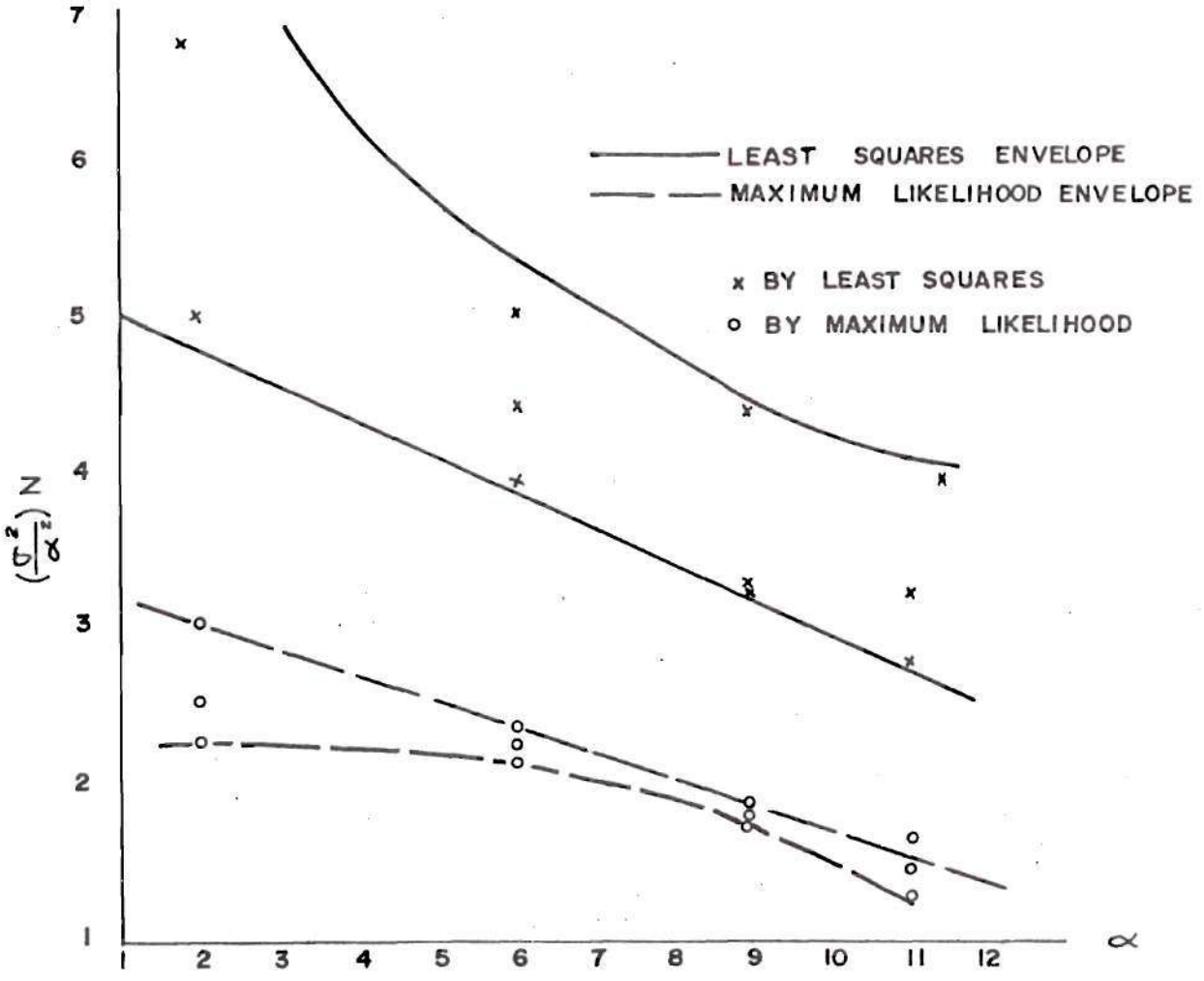

Figure 4.1 Graphical Representation of Analysis of Variance, $\gamma = 0$.

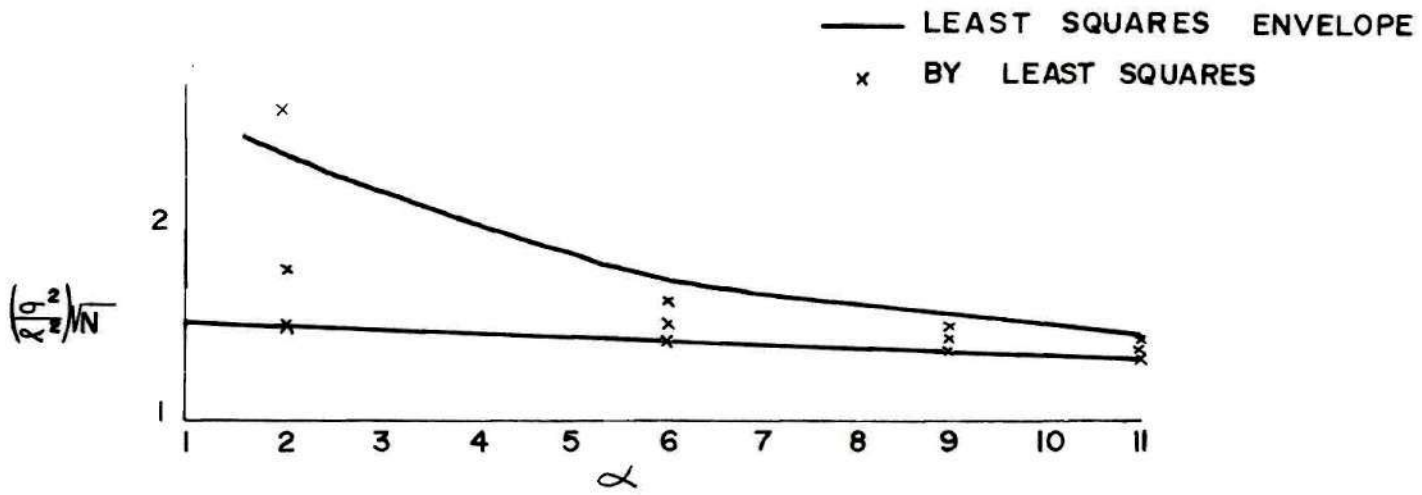

Figure 4.2 Graphical Representation of $\frac{s^2}{\alpha^2} = \frac{k_\alpha}{\sqrt{N}}$

$$\frac{S^2}{\alpha 2} = \frac{k_\alpha}{\sqrt{N}} \qquad 4.3$$

Representation in this form is seen to be more stable than representation by Equation 4.2; however, the data available is insufficient to warrant a preference for Equation 4.3. It should be noted that the plausibility of Equation 4.2 rests upon the form of the asymptotic expressions for sampling variance of the parameter estimates which were derived assuming an equivalence between least squares and maximum likelihood. This equivalence in turn depends primarily upon the appropriateness of the assumption of normality and constant variance of the errors. An examination of the errors resulting in the fittings in runs 4 through 10 indicates that the assumption of normality is probably appropriate in this case, and thus there may be no strong theoretical reason to favor Equation 4.3 over Equation 4.2. The question of normality of errors and its effect upon the method of least squares will be discussed in detail later in this chapter.

In any case, Figures 4.1 and 4.2 indicate that the sampling variance of the parameter estimates will decrease with increasing sample size, and hence that the method of least squares is consistent. This happy result is in accord with the intuitive and theoretical concepts of fitting by least squares, and says in effect that the sample histogram may be expected to approach in shape the population density function as the sample size increases, and that the least squares procedure will select the proper parameter values when fitting a histogram

which is representative of the distribution of the parent population. This result is not unexpected, since Synder and Wallace [27] have already shown that the method of least squares will correctly predict the parameters of a histogram constructed from a given probability density function.

Table 4.3 also shows that the method of least squares should not be used for sample sizes as small as 25 (actually, Table 4.3 indicates that samples of this size are too small to support a meaningful statistical analysis by any method considered in this work, since the sampling variance is quite large for these samples). A careful inspection of the variances given in this table along with an evaluation of the convergence difficulties encountered for some cases for samples of fifty items indicates that the size of fifty items represents a working lower bound for the method. Studies of growing samples performed on real data and discussed in the next chapter did involve fittings of samples with fewer than fifty items, and the results in these instances were acceptable. Thus, the lower bound of fifty items should not be taken as absolute, but rather as a limit below which the method of least squares may encounter convergence difficulities, and below which the resolution of the method may be so insufficient as to render the applicability of the method suspect.

Finally, the ratios of the parameter variances obtained by least squares to those obtained by maximum likelihood are most illuminating. Since it is known (Kendall [3]) that maximum likelihood estimators have minimum variance, at least in the limit, as the sample size increases,

then one would expect that the variance of any other estimator would exceed that of the likelihood estimator. Table 4.3 shows that this is indeed the case, and in fact the variance of the least squares estimator is on the average more than two times as great as that of the likelihood estimator. Thus in using the method of least squares to estimate population parameters, one must expect a less sharp estimate than could be obtained by using maximum likelihood.

As can be seen by an examination of Table 4.4, both the methods of least squares and maximum likelihood appear to be positively biased when applied to the gamma distribution. Although no clear indication exists in Table 4.4, it is assumed that this bias is a function of sample size and that it decreases as the size of the sample increases. The amount of this bias is seen to be relatively small, especially in comparison with the standard deviation of the samples from which the means in the table were computed. Nevertheless, the regularity of the deviation of the mean from the population parameter indicates that this bias is real, and not merely a manifestation of pecularities of the random samples upon which the estimates were based. Table 4.1 indicates that the method of moments also yields biased parameter estimates. The bias of maximum likelihood of estimators of the parameters of a gamma distribution has been investigated experimentally by Choir and Wette [38] and theoretically by Box [39]. The bias reported in both of these studies agrees closely with that shown in Table 4.4. These favorable comparisons give some confidence in the randomness of the gamma distributed samples used as

Table 4.4 Bias of Parameter Estimates, $\gamma = 0$

		By least squares			By maximum likelihood		
N	α	μ_a	μ_a/α	% bias	μ_a	μ_a/α	% bias
50	2	2.39	1.19	20	2.16	1.08	8
75	2	2.23	1.11	11	2.11	1.05	5
100	2	2.16	1.08	8	2.07	1.03	3
50	6	6.31	1.05	5	6.25	1.04	4
75	6	6.20	1.03	3	6.20	1.03	3
100	6	6.32	1.05	5	6.15	1.03	3
50	9	9.39	1.04	4	9.42	1.05	5
75	9	9.23	1.03	3	9.25	1.03	3
100	9	9.24	1.03	3	9.14	1.02	2
50	11	11.23	1.02	2	11.35	1.03	3
75	11	11.49	1.04	4	11.12	1.01	1
100	11	11.68	1.06	6	11.15	1.01	1

a basis for this investigation.

Distribution of Class Errors in Least Squares Fitting

To test the assumption that the error terms as defined by Equation 3.1 are normally distributed with zero mean, a chi-square test for normality was made on the error terms arising in the fitting of each sample examined in the course of this study. The test was made by lumping the error terms into groups so chosen that if the errors were indeed normal with zero means then the minimum number of errors expected in each group would be at least three. The chi-square coefficient

$$\chi_s^2 = \sum_{i=1}^{n} \frac{(f_i - kp_i)^2}{kp_i}$$

was then computed along with the probability of exceeding the observed coefficient χ_s^2; that is, the probability η defined by

$$P(\chi^2 \geq \chi_s^2) = \eta$$

was computed. As defined above, a large value of η indicates that the set of errors being tested is in fact distributed approximately in a normal fashion, while a small value of η indicates that the particular sample is not distributed normally. These computations yielded values of the χ_s^2 coefficient ranging from very small to very large, and

associated values of η ranging from near unity to near zero. Thus, some of the groups of errors were distributed (nearly) normally, and some were not. Table 4.5 summarizes the results of the chi-square tests.

It may be argued, however, that the chi-square test, being a large sample test, is not appropiate in this instance since the sets of errors tested were not large sets. To attempt to verify or disprove the above results, a Kolmogorov-Smirnov goodness of fit test was made on each set of errors as described above. This test was made in accordance with the description given in Lindgren and McElrath [28], and is based upon the sample distribution function

$$F(x) = \frac{1}{N+1} \cdot \{ \text{number of observations} \leq x \}$$

which is tested aginst a population distribution function $F_o(x)$ (in this case the normal cumulative distribution function). In this test, as in the chi-square test, the population mean was assumed to be zero, and the population variance was estimated from the sample itself.

The Kolmogrov-Smirnov test is based upon the statistic D_n defined by

$$D_N = \max_{-\infty < x < \infty} \left| F_N(x) - F_o(x) \right|$$

Table 4.5 Distribution of $P(\chi^2 \geq \chi_s^2)$ for Error Terms, $\gamma = 0$

$P(\chi^2 \geq \chi_s^2)$	Number of occurences
0 - 0.05	65
0.05 - 0.1	44
0.1 - .2	81
0.2 - 0.3	65
0.3 - 0.4	73
0.4 - 0.5	82
0.5 - 0.6	68
0.6 - 0.7	57
0.7 - 0.8	23
0.8 - 0.9	92
0.9 - 1.0	24

Total number of occurences = 675

Mean of $P(\chi^2 \geq \chi_s^2)$ = 0.43

Tables giving limits of rejection of the null hypothesis that the population cumulative distribution is F_o are given in Lindgren and McElrath [28]. In evaluating this test, the null hypothesis is rejected at the significance level α if the computed value of D_n is larger than the tabulated limits [as is usual in this sort of test, α is the probability of rejecting the null hypothesis when it is true, that is, α = P(reject Ho/Ho is true)]. The results of these tests are summarized in Table 4.6, and indicate that the null hypothesis can be rejected with α = 0.20 thirty-two out of 675 times, or in about 4.75% of the cases. With α = 0.1, eight rejections, or about 1.2% of the cases, must be made, while with α = 0.05, only two cases, or about 0.3%, must be rejected.

When subjected to the same sort of reasoning, the chi-square test would require rejection of 190 cases, or about 28%, with α = 0.20, 109 cases, or about 16%, with α = 0.1, and 65 cases, or about 10% with α = 0.05. Thus, the Kolmogorov-Smirnov test is considerably more favorable to the null hypothesis (that the errors are normal with zero mean) than is the chi-square test. Since the Kolmogorov-Smirnov test is more properly applied to the small samples encountered in this work, preference to the conculsions drawn from it must be given. On the other hand, the chi-square test is commonly used as a guide in making this sort of decision, with apparent success, and so some weight must be given to the fact that this test indicates that the null hypothesis should be rejected. Also, it should be emphasized that each test on the errors was made using a (presumably) different value

Table 4.6 Results of Kolmogorov-Smirnov Test for Normality of Errors, $\gamma = 0$

Results are for 675 samples

Rejection Level	Number of Rejections	Percent of Samples Rejected
0.2	32	4.7
0.1	8	1.2
0.05	2	0.3

for the population variance, namely the value computed from the sample itself. Thus these tests say nothing about the constancy of the variance, a requirement collateral with the requirement of normality of the errors. The proper conclusion to be drawn from the above work thus seems to be that the hypothesis of Relation 3.1(b) cannot be rejected based upon the results obtained in this study.

Results for Weighted Least Squares

Indications of the Need for Non-Unit Weights

The above work has been based upon fitting by least squares using unit weights for the error terms. As originally envisioned by Snyder [1], one of the more attractive features of fitting by least squares was the possibility of determining population parameters by a method which was rather insensitive to the presence of outliers in the sample. This insensitivity is due to the fact that the derivatives of the probability density function decrease rapidly as the value of the variate v increases. Thus, since the fitting process is controlled by the values of the derivatives and by the size of the error terms (See Equation 2.28) the presence of an outlier in a sample will have but little influence upon the optimized values of the estimates of the population parameters obtained by the method of least squares. Through the course of the above investigation, it became increasingly apparent that the method of least squares with unit weighting factors was so insensitive to observations in

the higher ranges of the variate that the procedure was not only ignoring misleading information occasioned by the appearance of an outlier in the sample, but was also ignoring legitimate information contained in observations which lay toward the tails of the distribution.

As a consequence, then, the least squares fitting method was attempting to match the sample histogram in the midrange of the sample, and rather ignored the fitting of the tails of the distribution, especially the right-hand tail. This emphasis on the mid-range of the sample is seen, for example, as the explanation for the decrease in the number of convergent cases as the values of the population parameters were increased. In fact, since the variance of the gamma distribution is given by the formula

$$\sigma^2 = \frac{\beta}{\alpha^2}$$

and in these simulations,

$$\alpha = \beta$$

then

$$\sigma^2 = \frac{1}{\alpha} \cdot$$

Thus samples drawn from populations with high values of α and β may be expected to be rather clustered about the mean value, and so the random-

ness of the sample would be expected to manifest itself primarily in this range.

The Choice of a Weighting Function

In an effort to counteract this tendency, and to make more of the information contained in the data sample available to the procedure, it was decided to devise a set of weights which would increase the effect of data points more removed from the mid-range of the sample. While many weighting functions could be chosen which would perform this function, a rather natural choice, motivated by the method of minimum chi-square, appeared to be the reciprocal of the density function itself raised to some power. That is,

$$w_i = [f(x_i; \alpha, \beta)]^{-\gamma}$$

where γ was considered an arbitrary positive number whose value might be chosen in such a manner as to lend to the method whatever particular characteristics were possible and desirable. It should be stressed that the weights chosen above are simply a particular choice made from among a practically limitless set of weight functions which are available, and that for a particular objective some different choice may be more propitious.

Effects of the Weight Function

The inclusion of the above weights into the least squares procedure was expected to have the following consequences. For non-zero values of γ, the effect of error terms occuring away from the sample mean is increased, and so the incidence of convergence should be increased and the resolving power of the method of least squares should be improved. Furthermore, because of the increased influence of these "remote" error terms, and since the tail of the gamma distribution becomes thinner with increasing values of α, an increase in the value of γ should in general bring about a decrease in the values of the estimated parameters. In other words, it should be possible to eliminate, at least approximately, the positive bias of the least squares estimators by a judicious choice of the value of γ. Since for γ equal to zero the weights w_i become equal to unity, the weighted least squares procedure can be collapsed to the unweighted procedure. For increasing values of γ, the procedure becomes more sensitive to the presence of outliers in the sample, and so some of the power of the least squares method to ignore the presence of outliers in the sample is lost by the inclusion of the weighting factors. Finally, the inclusion of weights which are a function of the random variable may invalidate the assumption that the variance of the errors is independent of the value of the random variable at which the errors may be measured.

To test the appropriateness of the above reasoning, and to attempt

to judge the effect upon the normality and constancy of the variance of the error terms which the inclusion of the weights might have, runs 4,6,8 and 9 were repeated for values of γ equal to 0.25, 0.50, 0.75, and 1.0. Since these runs were initially made with unit weights, their repetition allowed a direct examination of the effects of the value of γ upon the fitting procedure for values of γ varying between zero and one. A summary of the results of these runs as well as the runs 14, 16 and 19, is presented in Table 4.7 for sample sizes of 50, 75, and 100. As was anticipated, the positive bias of the unweighted estimators becomes less evident as γ increases, disappears at a value of γ equal to about 0.75, and becomes a negative bias for γ equal to 1.0. The efficiency of the fitting procedure as measured by the smallness of the sample variance of the parameter estimates also increases as γ increases. For γ equal to 1.0, the variance of the least squares estimators is of the same order of magnitude as the variance of the maximum likelihood estimators. This last result is not wholly unexpected, since in the case $\gamma = 1$, the method of least squares is equivalent to the method of minimum chi-square, which in turn is known to be asymptotically equivalent to the method of maximum likelihood (Kendall [3]). Table 4.7 also confirms the expected result that the incidence of convergence would increase as γ increased. For $\gamma = 0.75$, only one convergence failure occured among the 400 cases comprising the four runs for a sample size of 50, while no failures occured for this sample size for γ equal to 1.0. By contrast, the same four runs experienced ten failures for γ equal to zero.

Table 4.7 Results of Runs for $\gamma \neq 0$, r = 0.5

					Parameter Means						Parameter Variances						
Run	Size	γ	α	β	α_{le}	α_{ml}	α_{mom}	b_{le}	b_{ml}	b_{mom}	α_{le}	α_{ml}	α_{mom}	b_{le}	b_{ml}	b_{mom}	No. of * Convergences
4G.25	50	0.25	3	3	3.28	3.20	3.27	3.20	3.16	3.23	1.15	0.49	0.60	0.79	0.44	0.55	100
4G.50	50	0.50	3	3	3.16	3.20	3.27	3.12	3.16	3.23	0.97	0.49	0.60	0.69	0.44	0.55	100
4G.75	50	0.75	3	3	3.01	3.20	3.27	3.02	3.16	3.23	0.74	0.49	0.60	0.57	0.44	0.55	100
4G1	50	1.0	3	3	2.85	3.20	3.27	2.91	3.16	3.23	0.55	0.49	0.60	0.46	0.44	0.55	100
6G.25	50	0.25	5	5	5.31	5.25	5.36	5.31	5.23	5.34	2.34	1.63	1.89	1.92	1.65	1.67	100
6G.50	50	0.50	5	5	5.15	5.25	5.36	5.18	5.23	5.34	1.94	1.63	1.89	1.62	1.45	1.67	100
6G.75	50	0.75	5	5	4.94	5.25	5.36	5.00	5.23	5.34	1.59	1.63	1.89	1.35	1.45	1.67	100
6G1	50	1.00	5	5	4.71	5.25	5.36	4.82	5.23	5.34	1.47	1.63	1.89	1.29	1.45	1.67	100
8G.25	50	0.25	7	7	7.50	7.39	7.53	7.56	7.42	7.57	3.87	1.87	2.45	3.45	1.74	2.40	98
8G.50	50	0.50	7	7	7.31	7.39	7.53	7.39	7.42	7.57	3.61	1.87	2.45	3.03	1.74	2.40	100
8G.75	50	0.75	7	7	6.96	7.39	7.53	7.10	7.42	7.57	2.54	1.87	2.45	2.16	1.74	2.40	100
8G1	50	1.0	7	7	6.66	7.39	7.53	6.86	7.42	7.57	1.98	1.87	2.45	1.78	1.74	2.40	100
9G.25	50	0.25	9	9	9.46	9.42	9.73	9.47	9.37	9.68	6.23	2.97	3.77	5.49	2.63	3.30	97
9G.50	50	0.50	9	9	9.16	9.42	9.73	9.20	9.37	9.68	5.09	2.97	3.77	4.43	2.63	3.30	97
9G.75	50	0.75	9	9	8.98	9.42	9.73	9.06	9.37	9.68	4.91	2.97	3.77	4.34	2.63	3.30	99
9G1	50	1.0	9	9	8.68	9.42	9.73	8.81	9.37	9.68	4.21	2.97	3.77	3.71	2.63	3.30	100
4G.25R2	75	0.25	3	3	3.11	3.06	3.14	3.08	3.04	3.12	0.58	0.20	0.30	0.39	0.18	0.29	100
4G.50R2	75	0.50	3	3	3.04	3.06	3.14	3.03	3.04	3.12	0.47	0.20	0.30	0.32	0.18	0.29	100
4G.75R2	75	0.75	3	3	2.94	3.06	3.14	2.96	3.04	3.12	0.34	0.20	0.30	0.24	0.18	0.29	100
4G1R2	75	1.00	3	3	2.81	3.06	3.14	2.87	3.04	3.12	0.25	0.20	0.30	0.19	0.18	0.29	100
8G.25R2	75	0.25	7	7	7.31	7.10	7.18	7.28	7.10	7.19	3.04	1.14	1.61	2.45	1.11	1.65	100
8G.50R2	75	0.50	7	7	7.12	7.10	7.18	7.13	7.10	7.19	2.13	1.14	1.61	1.77	1.11	1.65	100
8G.75R2	75	0.75	7	7	6.87	7.10	7.18	6.93	7.10	7.19	1.44	1.14	1.61	1.24	1.11	1.65	100
8G1R2	75	1.0	7	7	6.65	7.10	7.18	6.76	7.10	7.19	1.15	1.14	1.61	1.06	1.11	1.65	100
4G.25R3	100	0.25	3	3	3.16	3.12	3.16	3.13	3.08	3.12	0.47	0.22	0.29	0.36	0.21	0.27	100
4G.50R3	100	0.50	3	3	3.11	3.12	3.16	3.08	3.08	3.12	0.40	0.22	0.29	0.31	0.21	0.27	100
4G.75R3	100	0.75	3	3	3.04	3.12	3.16	3.03	3.08	3.12	0.31	0.22	0.29	0.25	0.21	0.27	100
4G1R3	100	1.0	3	3	2.94	3.12	3.16	2.96	3.08	3.12	0.22	0.22	0.29	0.20	0.21	0.27	100
8G.25R3	100	0.25	7	7	7.17	7.09	7.13	7.15	7.10	7.15	2.09	0.98	1.16	1.95	0.98	1.20	100
8G.50R3	100	0.50	7	7	7.03	7.09	7.13	7.04	7.10	7.15	1.58	0.98	1.16	1.51	0.98	1.20	100
8G.75R3	100	0.75	7	7	6.87	7.09	7.13	6.92	7.10	7.15	1.21	0.98	1.16	1.19	0.98	1.20	100
8G1R3	100	1.0	7	7	6.71	7.09	7.13	6.80	7.10	7.15	1.05	0.98	1.16	1.06	0.98	1.20	100
14G.50A	50	0.50	3	3	3.31	3.29	3.36	3.28	3.25	3.33	1.17	0.65	0.71	0.82	0.53	0.63	100
14G.75A	50	0.75	3	3	3.15	3.29	3.36	3.16	3.25	3.33	0.90	0.65	0.71	0.64	0.53	0.63	100
16G.50A	50	0.50	6	6	6.34	6.41	6.48	6.35	6.41	6.47	3.61	1.99	2.20	3.02	1.77	1.99	100
16G.75A	50	0.75	6	6	6.09	6.41	6.48	6.16	6.41	6.47	2.77	1.99	2.20	2.36	1.77	1.99	100
19G.50A	50	0.50	9	9	8.72	9.04	9.23	8.77	9.04	9.22	3.87	2.96	3.69	3.94	2.93	3.70	99
19G.75A	50	0.75	9	9	8.56	9.04	9.23	8.64	9.04	9.22	3.71	2.96	3.69	3.69	2.93	3.70	99
14G.50B	75	0.50	3	3	3.13	3.10	3.16	3.12	3.08	3.16	0.60	0.31	0.33	0.42	0.26	0.31	100
14G.75B	75	0.75	3	3	3.00	3.10	3.16	3.03	3.08	3.16	0.45	0.31	0.33	0.33	0.26	0.31	100
16G.50B	75	0.50	6	6	6.03	6.19	6.27	6.05	6.17	6.25	1.52	1.03	1.33	1.39	0.96	1.21	100
16G.75B	75	0.75	6	6	5.86	6.19	6.27	5.91	6.17	6.25	1.21	1.03	1.33	1.10	0.96	1.21	100
19G.50B	75	0.50	9	9	9.36	9.31	9.48	9.44	9.38	9.56	3.90	2.77	3.96	3.77	2.66	3.77	99
19G.75B	75	0.75	9	9	9.14	9.31	9.48	9.26	9.38	9.56	3.27	2.77	3.96	3.18	2.66	3.77	99
14G.50C	100	0.50	3	3	3.08	3.11	3.09	3.06	3.10	3.07	0.31	0.20	0.28	0.22	0.16	0.23	100
14G.75C	100	0.75	3	3	3.00	3.11	3.09	3.01	3.10	3.07	0.26	0.20	0.28	0.19	0.16	0.23	100
16G.50C	100	0.50	6	6	6.04	6.18	6.25	6.06	6.18	6.25	1.40	0.81	0.89	1.28	0.80	0.87	100
16G.75C	100	0.75	6	6	5.93	6.18	6.25	5.97	6.18	6.25	1.10	0.81	0.89	1.02	0.80	0.87	100
19G.50C	100	0.50	9	9	9.29	9.16	9.31	9.26	9.11	9.27	3.19	1.43	1.82	3.03	1.38	1.77	100
19G.75C	100	0.75	9	9	9.04	9.16	9.31	9.04	9.11	9.27	2.35	1.43	1.82	2.22	1.38	1.77	100

* Maximum Number of Convergences is 100.

Extension of the Range of Optimization beyond the Sample Range

Runs 6 and 9 were also repeated with empty classes to assess the effect of adding empty classes when non-unit weights are used. A summary of the results of these runs is presented in Table 4.8, and indicates that the inclusion of empty classes in weighted least squares introduces a rather stubborn positive bias into the estimators. A detailed study of the individual cases of these runs indicated that while as a rule empty classes should not arbitrarly be added to a histogram, in those cases where the sample is highly truncated the addition of empty classes can be of benefit. These cases are distinguishable by the characteristics that an initial fit with no empty classes indicates that the cumulative probability through the largest sample point is much less than would normally be expected based upon the sample size. Such samples are easily recognised as well by the "blocky" appearance of the sample histogram. As an example of a case where parameter estimates can be improved by the addition of empty classes, case 23 of run 9G.75 produced the histogram shown in Figure 4.3. The inital fitting with no empty classes produced paramter estimates a = 7.58 and b = 8.06 (the true values were α = 9 and β = 9). The value of the cumulative distribution function at the end of the last class in this histogram was 0.9123. With the addition of four empty classes to the histogram, another fit produced parameter estimates a = 9.28 and b = 9.44, which is a substantial improvement in the quality of the parameter estimates. The above example notwithstanding, however, it is felt that the addition

Table 4.8 Results of Runs with Empty Right-Hand Classes, $\gamma \neq 0$

Run	No. of Empty Classes	α	β	Mean of a	Mean of b	Variance of a	Variance of b	Number of Convergences
6G.75N4	4	5	5	5.21	5.21	1.59	1.39	100
6G1N4	4	5	5	5.14	5.15	1.59	1.42	100
6G.75N8	8	5	5	5.26	5.24	1.58	1.38	100
9G.75N4	4	9	9	9.59	9.57	4.98	4.54	100
9G1N4	4	9	9	9.50	9.48	3.97	3.58	100

STEP = 0.0913
CENTER INITIAL GROUP = 0.2739
CENTER FINAL GROUP = 1.5523
NO GROUPS = 15

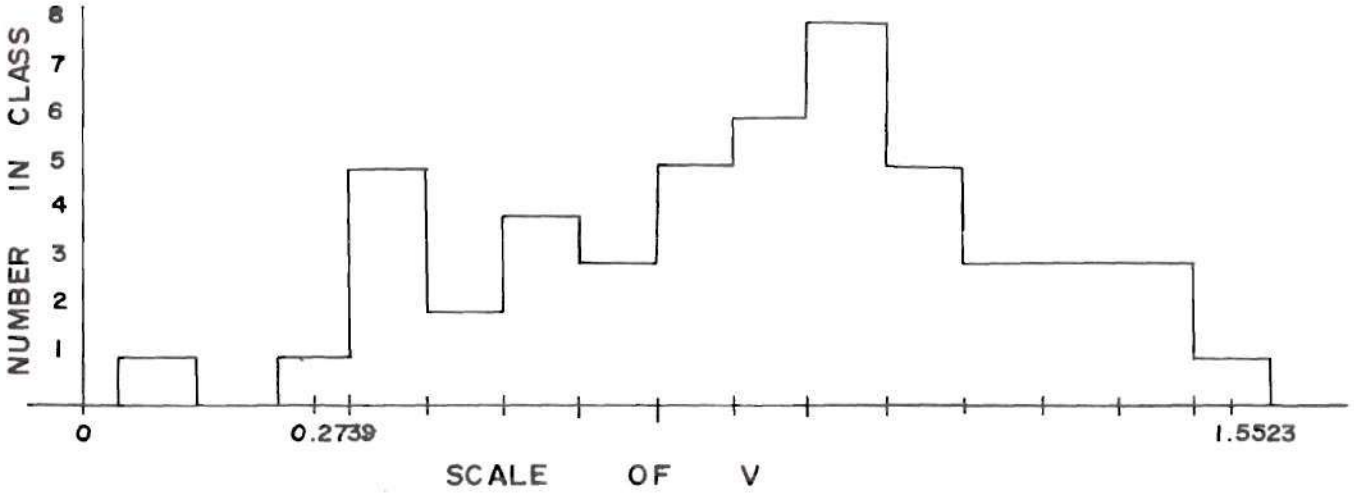

Figure 4.3 Histogram of Case 23, Run 9G.75

of empty classes to a sample histogram is a step which should be taken only when it is certain that the sample being analysed is highly truncated, and then only with the realization that by adding these empty classes one is introducing a bias into the estimators.

Characteristics of Weighted Least Squares Estimators

Table 4.7 indicates that for $\gamma = 0.75$, the least squares estimators are unbiased. Confidence in this inference is strengthened by an examination of Table 4.9. In Table 4.9, the mean and variance of the sample means of the ratio a/α are computed for the aggregate of all samples sizes, and the means of these ratios are computed according to sample size. These results indicate that indeed the least squares estimator of α is unbiased if γ be chosen equal to 0.75.

Table 4.10 presents an analysis of the variance of the least squares estimator of α for $\gamma = 0.75$. These data, also presented graphically in Figure 4.4, indicate that an attempt to express the parameter variance in the form of Equation 4.2 is fairly appropriate. As was the case for $\gamma = 0$, the "constant" c_α is seen to be in actuality a function of the population parameters. Also included in Table 4.10 are the ratios of the variance of the least squares estimates of α to the variance of the maximum likelihood estimates of α for the various runs. This ratio is seen to have a mean value of 1.36, and a standard deviation of about 0.2. Thus if one defines the efficiency of the least squares estimator relative to the maximum likelihood estimator to be the inverse of the ratios of the sample variances, then the efficiency of the least squares estimator is seen to be approxi-

Table 4.9 Bias of Least Squares Estimates for Parameter α, $\gamma = 0.75$

	a/α	a/α	a/α
	N = 50	N = 75	N = 100
	1.003	0.9880	1.013
	0.988	0.981	0.981
	0.944	1.000	1.000
	0.998	0.977	0.988
	1.103	1.016	1.004
	1.015		
	0.951		
Means	1.007	0.991	0.997

Mean of all samples = 1.000

Variance of all samples = 0.00088

Table 4.10 Characteristics of Parameter Variances, $\gamma = 0.75$

N	α	σ^2_{LS}	$\frac{\sigma^2_{LS}}{\alpha^2} N$	σ^2_{ML}	$\frac{\sigma^2_{ML}}{\alpha^2} N$	$\frac{\sigma^2_{LS}}{\sigma^2_{ML}}$
50	3	0.74	4.10	0.49	2.72	1.51
75	3	0.34	2.85	0.20	1.67	1.70
100	3	0.31	3.40	0.22	2.40	1.41
50	3	0.90	5.00	0.65	3.60	1.38
75	3	0.45	3.75	0.31	2.58	1.45
100	3	0.26	2.90	0.20	2.22	1.30
50	5	1.59	3.18	1.63	3.26	0.98
50	6	2.77	3.85	2.20	3.06	1.26
75	6	1.21	2.55	1.03	2.15	1.17
100	6	1.10	3.10	0.81	2.25	1.36
50	7	2.54	2.60	1.87	1.90	1.38
75	7	1.44	2.18	1.14	1.74	1.26
100	7	1.21	2.50	0.98	2.00	1.23
50	9	4.91	3.03	2.97	1.83	1.65
50	9	3.71	2.30	2.96	1.83	1.25
75	9	3.27	3.00	2.77	2.56	1.18
100	9	2.35	2.90	1.43	1.77	1.64

Mean of $\frac{\sigma^2_{LS}}{\sigma^2_{ML}} = 1.36$

Variance of $\frac{\sigma^2_{LS}}{\sigma^2_{ML}} = 0.19$

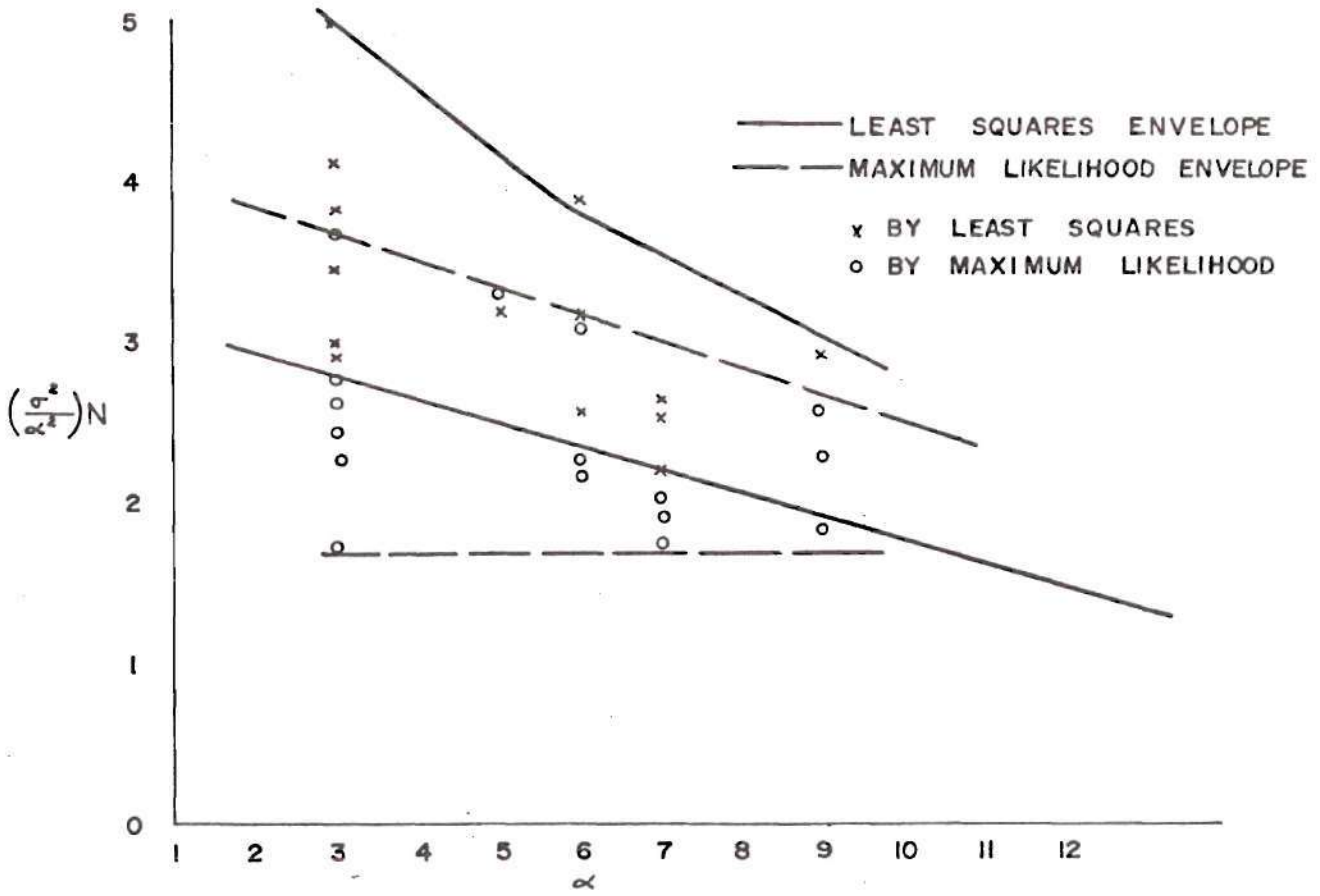

Figure 4.4 Graphical Representation of Analysis of Variance, $\gamma = 0.75$

mately equal to 1/1.36, or about 75%. Thus, the method of least squares is seen to offer an acceptable alternative method of estimating population parameters.

Distribution of Errors in Weighted Least Squares Fitting

The error terms arising in the least squares fittings for four cases with $\gamma = 0.75$ were subjected to the same tests as were the unweighted errors in an effort to determine if the inclusion of the weights which were functions of the random variate would adversely affect the validity of Relation 2.1(b). The results of these tests are summarized in Tables 4.11 and 4.12, and indicate that the inclusion of the weights does not invalidate 2.1(b), and in fact seems to improve the reasonableness of the assumption of this relation. The use of these particular weights also tends to scale the errors and make them more nearly independent of the value of the random variate at which they are measured, lending credence to the collaterial assumption of the independence of the variance of the errors of the value of the random variate. Thus, these results cannot contradict the assumptions of normality and independence of the error terms, and in fact, in the case of weighted least squares, tend to indicate that these assumptions are appropriate.

Confidence Regions and Tolerance Limits

As a compliment of the above study of the properties of the method of least squares as a tool by which population parameters may

Table 4.11 Distribution of $P(\chi^2 \geq \chi_s^2)$ for Error Terms, $\gamma = 0.75$

$P(\chi^2 \geq \chi_s^2)$	Number of occurences
0 - 0.05	30
0.05 - 0.10	32
0.10 - 0.20	45
0.20 - 0.30	35
0.30 - 0.40	47
0.40 - 0.50	48
0.50 - 0.60	43
0.60 - 0.70	35
0.70 - 0.80	18
0.80 - 0.90	41
0.90 - 1.00	26

Total number of occurences = 400

Mean of $P(\chi^2 \geq \chi_s^2) = 0.44$

Table 4.12 Results of Kolmogorov-Smirnov Test for Normality of Errors, $\gamma = 0.75$

Results are for 400 samples

Rejection Level	Number of Rejections	Percent of Samples Rejected
0.20	6	1.5
0.15	4	1.0
0.10	3	0.75
0.05	1	0.25

be estimated, confidence regions were constructed for the parameter estimates obtained for case 100 of run 10R3. These regions were constructed by a simple adaptation of the method proposed by Halperin [6] and discussed in Chapter II. Because of the complexity of the non-linear analogue to Equation 2.14, a direct solution of this equation was not attempted. Instead, using the least squares estimate of α and β as a beginning point, values of the statistic

$$F = \frac{N - m}{m} \frac{reg(\underset{\sim}{e})}{res(\underset{\sim}{e})}$$

were computed at fixed points of a grid surrounding the least squares estimates of α and β. This statistic is, as was discussed in Chapter II, distributed as Snedecor's "F" with m and N-m degrees of freedom. Thus, it was possible to compute the cumulative probability at the point F, that is,

$$\eta = \int_0^F dF$$

for each value of F. These values of η were then plotted in the α-β plane at locations corresponding to the values of α and β from which each was derived. The various confidence regions were then constructed by sketching a closed curve (using interpolation where

necessary) passing through the value of η for which the confidence region was desired. For example, the 90% confidence region was constructed by joining those points in the $\alpha - \beta$ plane for which η = 0.90. It is clear that this procedure can be made as accurate as desired by choosing a sufficiently fine grid. When the number of unknown parameters exceeds two a graphical construction of confidence regions becomes impractical, and one would be compelled to estimate the confidence regions by repeated computations of F values over the ranges of the various parameters. Of course, two dimensional cross-sections of the confidence regions could be plotted, and in some instances would aid in the visualization of the shape and extent of the regions.

The confidence regions obtained for weights of unity are depicted in Figure 4.5. The most striking feature of these regions is the elongation of the ellipse along the line $\alpha = \beta$, indicating the high degree of correlation between the two parameters. Also evident is the wide range of values which the parameter estimates may assume and still remain within the various confidence regions. This is an indication that the method of least squares in the case does not provide very confining limits upon the probable values of the population parameters.

To use such a region to construct an upper tolerance limit, it is necessary to determine the value $\bar{v}_{\gamma,\eta}$ such that

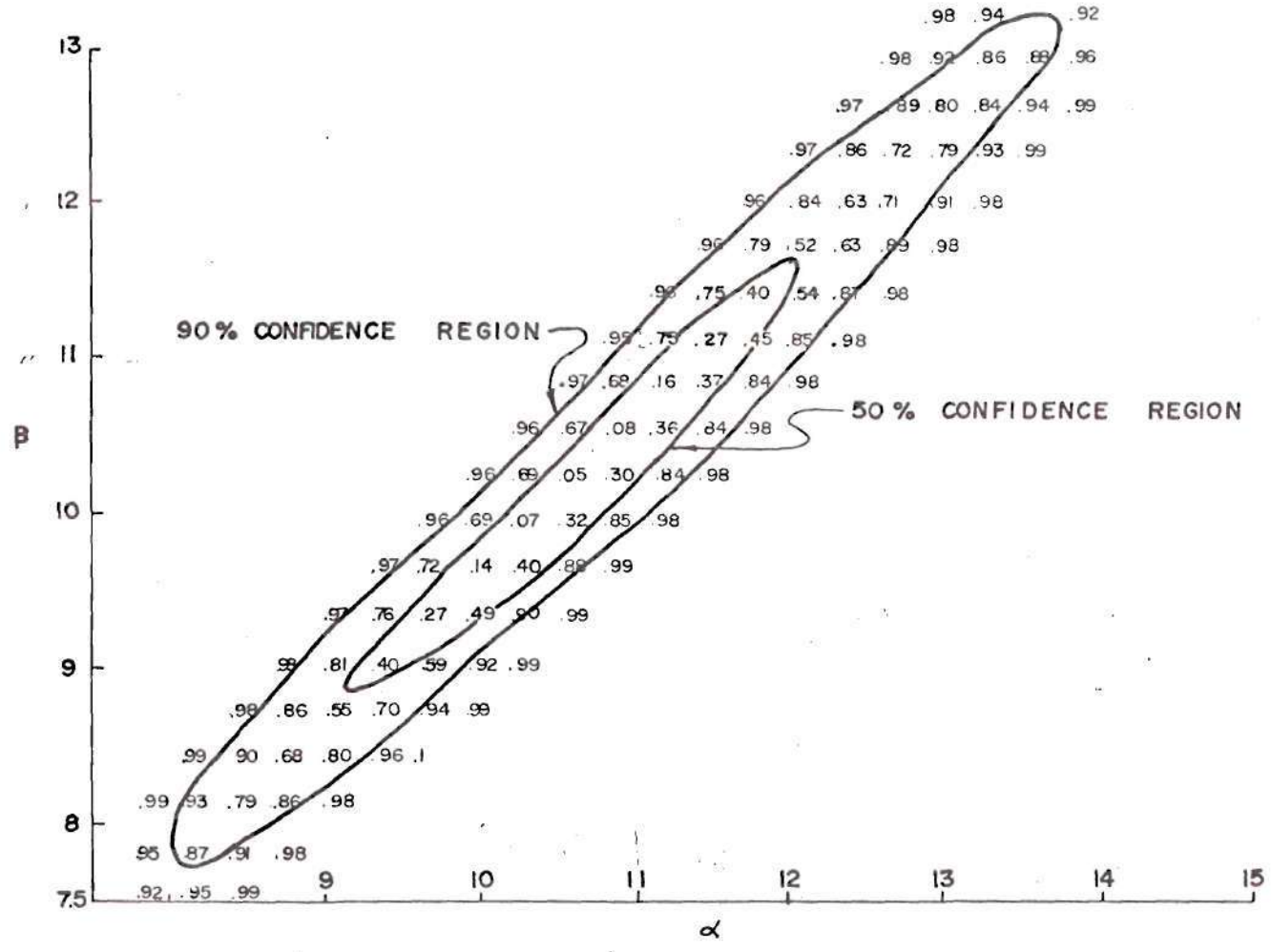

Figure 4.5 Confidence Region for Case 100, Run 10R3, $\gamma = 0$

$$\bar{v}_{\gamma,\eta} = \max_{\alpha,\beta \in R_\eta} \left\{ \bar{v}: \int_0^{\bar{v}} p(v; \alpha, \beta)\, dv = \gamma \right\} .$$

Again, while the exact determination of this value is apt to be a formidable task, the calculation of $\bar{v}$ at a few critical points of the region R_η will give an estimate of $\bar{v}_{\gamma,\eta}$ sufficient for most practical purposes. Thus, to estimate $\bar{v}_{.99,.90}$, one might first calculate

$$\bar{v}_{.99,.90} : \int_0^{\bar{v}_{.99,.90}} p(v,8,8.25)\, dv = 0.99$$

which yields

$$\bar{v}_{.99,.90} = 2.05$$

To check this approximation, values of $\bar{v}$ for different values of $\bar{v}_{.99,.90}$ within $R_{0.90}$ might also be computed, and the final estimate chosen as the largest of these values. Thus, in this way the following values may be obtained.

α	β	$\bar{v}_{.99,.90}$
8.0	8.25	2.05
8.1	8.50	2.06
8.28	8.80	2.08
8.40	9.00	2.075

From the above set of values, it is seen that an appropriate value for $\bar{v}_{.99,.90}$ is approximately 2.08. The true value of the one hundred year event is 1.839 (since α = 11 and β = 11). Thus, the 90% tolerance limit for the one hundred year event in this case is only about 13% larger than the event itself. The cumulative probability corresponding to the event represented by the tolerance limit is found to be about 0.9979.

Figure 4.6 shows the 90% confidence region for the same observation computed with γ = 0.75. Surprisingly, the size of this region is very nearly the same as the region obtained with γ = 0. It is significant, however, that the differences in the two regions occur in the lower ranges of the parameters, a fact which in this instance means that tolerance limits computed from the weighted regions will be somewhat smaller than the corresponding limits computed from unweighted regions. The shape of the weighted confidence region is in all respects similar to the shape of the unweighted regions; both being highly elongated ellipses (the term ellipse is used here in a generic rather than a mathematical sense, and the description of confidence regions as

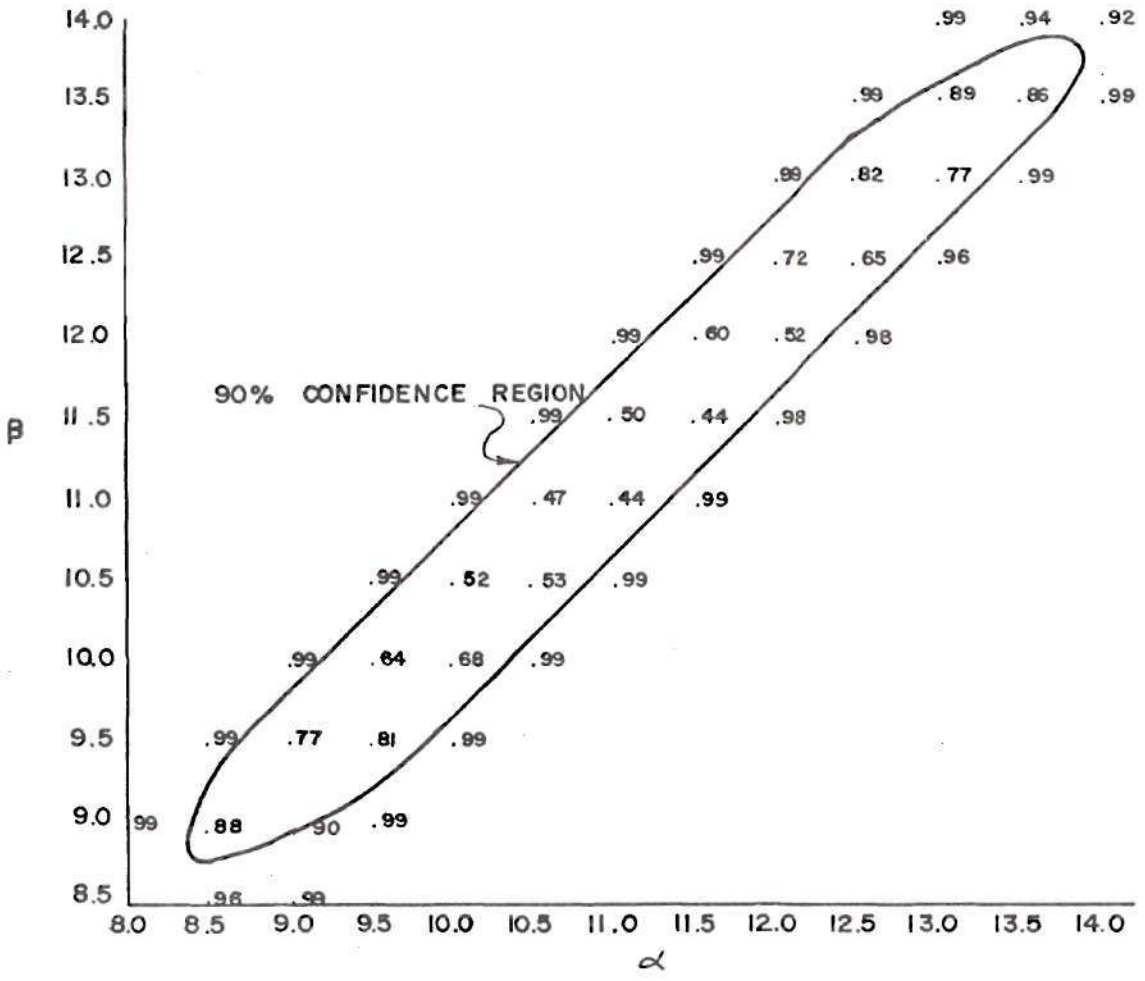

Figure 4.6 Confidence Region for Case 100, Run 10R3, $\gamma = 0.75$

elliptical is intended to convey only a general notion of the shape of the region and not to imply that the boundary of the region is a true ellipse) with major axis lying along the line $\alpha = \beta$.

A question which naturally arises at this point is how do tolerance limits obtained from the theory of least squares compare with tolerance limits computed by other methods. For normal samples, when the mean and standard deviation are estimated by the method of maximum likelihood, then tolerance limits for various percentile points of the population may by computed from these estimates. For any sample, normal or otherwise, tolerance limits on at least selected percentile points of the population may be estimated by the so-called distribution free methods. Distribution-free tolerance limits are notoriously unconfining unless the sample size is large, and for moderate or small samples are scarely useful. Tolerance limits on normal populations are rather strict, and usually provide useful information as to the probable range of a population percentile; unfortunately, one often must deal with populations other than normal, and of course in these instances tolerance limits based upon normal populations are not applicable.

The above sample from which were computed the confidence regions shown in Figures 4.5 and 4.6 is so small (100 points) that the distribution-free upper tolerance limits cannot be determined for any percentile apt to be of interest in a hydrologic application.

The 90% tolerance limit for the one hundred year event for this sample may be computed by using the empirical factors developed by

Fontaine [34]. The use of Fontaine's results yields a value of 1.98 for for $\bar{v}_{.99.90}$. This value is seen to be somewhat smaller than the value 2.08 computed above. To compare tolerance limits obtained from least squares techniques to those available for normal distributions, a normally distributed random sample of fifty points was constructed from Table A-37 in Natrella [29] with zero mean and standard deviation equal to ten. Using the least squares technique, a 90% confidence region for the parameters μ and σ was constructed and is presented in Figure 4.7. Table 4.13 shows the data from which this region was constructed and Table 4.14 shows the sample grouped to form the histogram used to construct the confidence region. It is not necessary to comment upon this region, except perhaps to note the relative independence of the parameters μ and σ as evidenced by the near-circular shape of the region. No explanation for the rather peculiar elongation of this region near the values $\mu = -2$ and $\sigma = 6$ can be given other than a rather weak incantation invoking some unknown pecularity of the sample from which this region was constructed.

The 90% tolerance limit for the one hundred year event is determined by maximizing $\sigma t_{100}+\mu$, where t_{100} is the value of the standard normal variate ($\mu=0$, $\sigma=1$) corresponding to a cumulative probability of 0.99, and has the value 2.326 (Natrella [29]). This maximum occurs (approximately) at the point $\mu=3$, $\sigma = 15$ (since if

$$v = \sigma t + \mu,$$

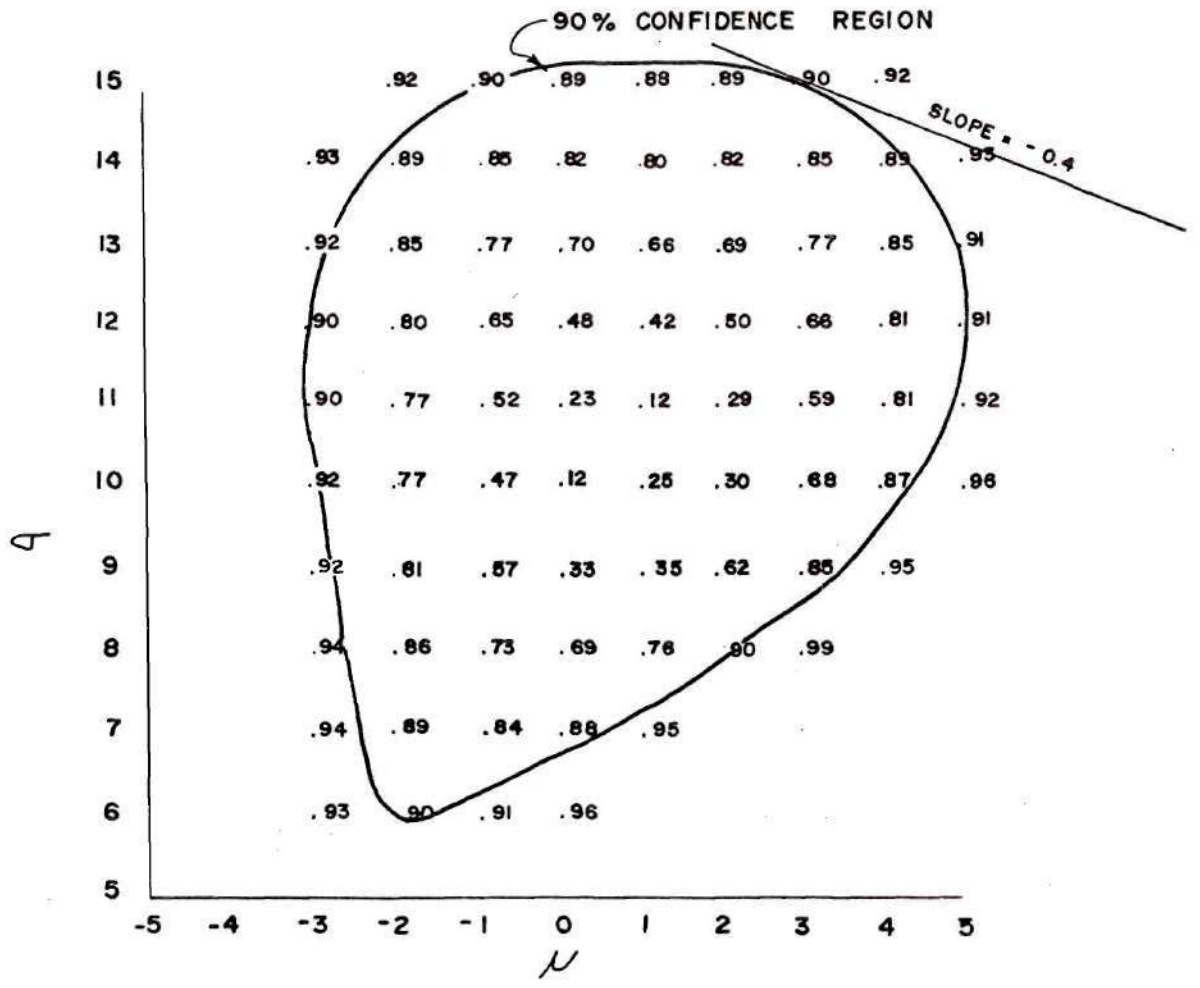

Figure 4.7 Confidence Region for Normal Sample

Table 4.13 Random Normal Sample, $\mu = 0$, $\sigma = 10$

Number	Entry	Number	Entry
1	0.48	26	- 2.36
2	10.40	27	6.49
3	- 1.11	28	15.55
4	- 1.20	29	12.85
5	13.96	30	- 7.47
6	- 3.93	31	18.22
7	- 2.20	32	8.98
8	4.22	33	- 6.91
9	2.33	34	9.72
10	1.97	35	- 0.11
11	- 5.21	36	5.17
12	- 5.63	37	8.08
13	- 1.16	38	26.51
14	- 5.12	39	- 6.50
15	- 5.18	40	5.92
16	-21.94	41	13.46
17	22.61	42	- 1.37
18	4.61	43	9.52
19	-15.33	44	14.67
20	-18.36	45	- 3.52
21	-14.07	46	3.09
22	- 2.13	47	5.78
23	9.48	48	-18.81
24	- 0.73	49	- 4.88
25	-14.74	50	- 3.29

Mean of sample = 0.96 Sample variance = 10.38

Table 4.14 Histogram of Normal Sample of Table 4.13

Range of sample = 48.45

Class width = 3.7269

Number of groups = 14

Center of first group =-22.361

Center of last group = 26.088

Group	Frequency Histogram	Normalized Histogram
1	1	0.02
2	2	0.04
3	3	0.06
4	0	0.00
5	5	0.10
6	10	0.20
7	7	0.14
8	6	0.12
9	4	0.08
10	5	0.10
11	4	0.08
12	1	0.02
13	1	0.02
14	1	0.02

then

$$\frac{dv}{d\mu} = 0 = \frac{d\sigma}{d\mu} t + 1 ,$$

and so

$$\frac{d\sigma}{d\mu} = -\frac{1}{t}$$

which occurs on the particular curve in Figure 4.7 at the points ($\mu = 3$, $\sigma = 15$). From these values, one obtains

$$\bar{v}_{.99,.90} \doteq 38$$

From Natrella [29], one finds the 100 year, 90% tolerance factor (for a normal distribution) to have the value $K_u = 2.735$. From the sample one obtains $\bar{x} = 0.96$ and $s = 10.38$. Thus

$$N_u = \bar{x} + K_u S = 29.3$$

is the value of the normal-based 100 year, 90% upper tolerance limit. (The actual value of the 100 year event is 23.3.) Thus, it is apparent

that the least squares tolerance limit is somewhat larger than that same limit based upon the more conventional tolerance factors. In terms of the actual value of the 100 year event one has

$$\frac{N_u}{v_{.99}} = 1.26$$

and

$$\frac{\bar{v}_{.99,.90}}{v_{.99}} = 1.63$$

In other words, the least squares tolerance limit $\bar{v}_{.99.90}$ is about 30% larger than the same limit obtained from more standard techniques. Thus it would appear that tolerance limits obtained from least squares techniques may very well be of value in some instances, although predictably these limits will usually be less confining than limits developed for a specific distribution. While the least squares tolerance limits are computable according to the same rule regardless of the underlying distribution, and thus in a sense are distribution-free, they are in fact based upon the particular distribution by which they are computed. For small samples, least squares methods will yield tolerance limits when distribution-free limits cannot be defined.

In should be noted that the above procedure is not infallable, at

least in theory. The existence of multiple roots of the least squares normal equations may induce such oddities as confidence regions composed of disjoint or open (unbounded) sets. It is not known whether or not such cases can arise in the fitting of density functions such as have been discussed herein, and lacking assurance to the contrary one must assume that such pathological cases can and may occur. It is permissible is such cases to resort to such subterfuges as changing weights, changing the manner in which the sample is grouped, or any other device which may avoid the problem of pathological parameter confidence regions.

Finally, continuing with the above line of thinking, the concept of confidence regions and tolerance limits may offer an objective method of judging the propriety of the particular grouping of a sample into a histogram as compared to some other proposed grouping. The idea here would be to define a task-oriented criterion of judging the effectiveness of the grouping. For example, if one were making a frequency analysis for the purpose of determining, say, the one hundred year flood at a particular location, then the superior grouping might be chosen as that grouping which yields the smallest value of a particular tolerance limit for this flood. Such practice, while not in the spirit of the concepts of confidence regions and tolerance limits, may indeed be of value in the grouping of rather odd samples which may be encountered in practice. For most samples, the simulation work described in this chapter indicates that there will be little difference in the results obtained from differetn groupings. Nonetheless, this statement is an expression of an average trend, and may not be true of a given sample.

CHAPTER V

THE METHOD OF LEAST SQUARES APPLIED TO REAL DATA

In this chapter, some results obtained by the fitting of real data with the two-parameter gamma distribution by the method of least squares are presented. These data were initially fit in order to obtain an idea of the probable range of population parameters corresponding to typical data of an annual nature for us in the simulations described in the preceding chapters. It is not the intention in this chapter to enter into a detailed discussion of frequency analysis either by the method of least squares or by any other method, nor are the results of this chapter intented to justify or suggest the use of the gamma distribution as a suitable distribution for the analysis of annual data, since the suitability of the gamma distribution has already been discussed at length in the study by Markovic [22].

With these objectives in mind, the discussion of the results of the analysis of real data is limited to those areas in which the use of the method of least squares as a particular method of parameter estimation is thought to have a measurable effect, and little attention is paid to those questions, such as the adequacy of the assumed distribution to fit the data, which are of primary importance in a frequency analysis.

Source of Data

The data analysed by the method of least squares consisted of stream flow records gathered from streams throughout the United States, and of maximum precipitation data for the U.S. Weather Bureau, Atlanta, Georgia, station for the years 1889 through 1972. The stream flow records were comprised of annual flood peaks for the various streams and were obtained from the appropriate U.S. Geological Survey Water Supply Papers. As originally compiled by Robey and Wallace [30], the stream flow records were accepted if the record through 1960 was at least 49 years in length, the location of the gage had not changed appreciably within the period of record, the stream was unregulated, and the record had no gap of more than three years within the period of record. Prior to use in this study, the stream flow records were updated to include the period from 1961 through 1970. The rainfall records were compiled in the form of maximum annual precipitation for durations of 5 minutes, 10 minutes, 15 minutes, 1 hour, 2 hours and 24 hours for each year. Not all records for each duration were available and the lengths of the records for the various durations were 77, 77, 70, 70, 81, 70 and 72 years respectively. These data are listed in Appendix F, along with a partial listing of the stream flow records and a complete list of the stations for which stream flow records were obtained.

One of the most outstanding characteristics of the annual flood

data examined in this work was the almost predictable presence of flow values of a magnitude of three or four to as much as six times the average value of the flows. An examination of the synthetic samples generated in the course of this work had indicated that such "outliers" should be considered a rarity; however, approximately fifty percent of the flood records examined contained such outliers. It was also not uncommon in these records to find these outlier points in groups of two, three, or even four. Such samples produce obvious difficulties when analysed, since if the analysis is made with the outliers retained, then their presence will strongly affect the parameter estimates. On the other hand, the removal of such points is always occasioned by some misgivings as to the selection of those points to be removed. The method of least squares, by virtue of the emphasis of the method upon the shape of the histogram rather than the values of the individual observations, is far better suited as a means of estimating population parameters from samples containing outliers than is, for example, the method of maximum likelihood.

Results of Fitting

Table 5.1 shows parameter estimates from six samples of annual flood data. These estimates were obtained by maximum likelihood and by least squares with and without outliers contained in the samples (these outliers were in all cases actual flow values contained in the various records). The actual values of these parameters are

Table 5.1 Parameter Estimates for Samples of Annual Flood Data

Sample number	Least Squares (γ = 0.75)						Maximum Likelihood				
	With outliers		Without outliers		Percent* change		With outliers	Without outliers		Percent change	
	α	β	α	β	α	β	$\alpha=\beta$	α	β	α	β
315	5.07	4.97	5.19	4.49	- 2.4	0.6	4.67	5.65	5.40	- 21.0	- 15.6
450	7.62	7.69	6.48	6.64	15.0	13.6	7.39	7.97	7.84	- 7.8	- 6.1
940	5.84	5.28	5.67	5.23	2.9	0.9	3.89	5.26	4.97	- 35.2	- 27.8
1805	2.42	2.15	3.26	2.67	-34.7	-24.2	1.84	2.99	2.56	- 62.5	- 39.1
3210	7.86	7.74	9.23	8.83	-17.4	-14.1	7.71	10.36	9.88	- 34.4	- 28.1
3345	13.85	11.48	13.43	11.06	3.0	3.7	3.83	1384	11.41	-261.0	-197.9
Average absolute change					12.6	9.5				70.3	52.4

*Percent change = $\frac{\text{(with-without)}}{\text{with}}$ X 100

unimportant, but it should be noted that the least squares estimates are closer to the maximum likelihood estimates based upon the reduced sample (that is, the sample from which the outliers have been removed), and that least squares is far more stable in the presence of outliers than is maximum likelihood. Thus, the method of least squares may be said to provide better estimates of the population parameters in these cases than does the method of maximum likelihood.

For those samples which did not appear to contain outliers, there appeared to be little to support any choice between the least squares and likelihood estimates. This result was expected, since the work of Chapter IV had showns that, for "normal" samples at least, the two methods give results which are for practical purposes indistinguishable. There was a noticable tendency for return periods computed from the least squares parameters to be larger than the same return periods computed from the maximum likelihood parameters. This result was not anticipated, since the indications of the simulation runs were that the least squares parameters estimates would yield smaller return periods than the likelihood estimates (it should be recalled that the likelihood estimates were positively biased). An explanation for this observed anomaly may be based in part on the large number of outliers which these samples were found to contain. The fact that this trend was noticed even when the sample did not appear unusual may very well be an indication that the various flood records are not random samples from a two-parameter gamma population.

Table 5.2 shows the results of fitting the rainfall data of Appendix F with the two-parameter gamma distribution. For this data, little difference in the parameter values obtained by the two methods was found. Table 5.3 shows the return period for the event represented by the right edge of the last class in the histogram computed from the parameter estimates in Table 5.2. The tendency of the method of least squares to predict smaller return periods than maximum likelihood, a tendency which was noted in the analysis of synthetic data and disappeared in the analysis of the flood data, is seen in Table 5.3 to have reappeared. It should be remembered in the study of Table 5.3 that these estimates were based upon sample sizes of between 70 and 80 points, and thus the differences in the predicted return periods of the two methods which are seen in Table 5.3 have little actual significance.

Figure 5.1 shows graphically the results obtained by fitting the rainfall data with an extreme value distribution both by the methods of moments and by the method of Gumbel. Also shown on Figure 5.1 are the frequency curves obtained from the least squares fit of the two-parameter gamma distribution. Again, these frequency curves indicate that the gamma distribution yields frequency estimates quite as reasonable as those estimates obtained from the extreme value distribution, at least for the particular sets of data represented in this figure.

Table 5.4 shows the results of successive fittings of five flood records. Each record was fit ten times, and on each fitting,

Table 5.2 Parameter Estimates for Maximum Annual Rainfall Data at Atlanta, Georgia

Maximum annual rainfall by duration	Least squares ($\gamma = 0.75$)		Maximum Likelihood
	α	β	$\alpha = \beta$
5 minutes	13.27	13.12	13.43
10 minutes	13.01	12.98	12.95
15 minutes	10.13	10.30	10.60
30 minutes	7.90	7.91	8.76
1 hour	8.83	8.75	9.02
2 hours	8.71	8.73	9.97
24 hours	16.77	16.76	15.88

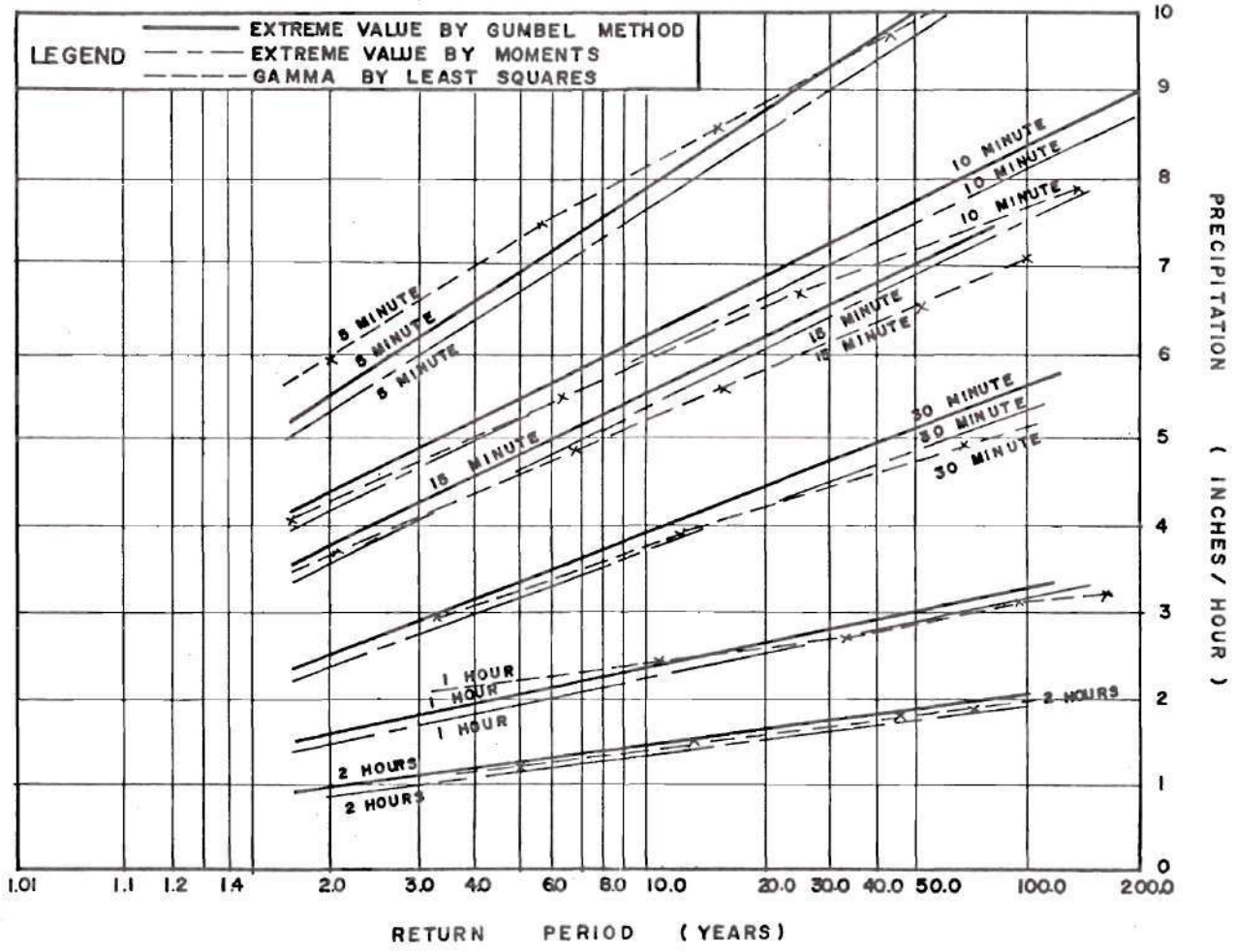

Figure 5.1 Rainfall-Frequency Curves for Atlanta, Georgia

Table 5.3 Return Period of Right Edge of Most Extreme Class

Atlanta Rainfall Data

Duration	$x/\bar{x}$	Return by least squares	Return by maximum likelihood
5 minutes	1.97	543	508
10 minutes	1.80	137	131
15 minutes	1.78	52	66
30 minutes	1.92	66	88
1 hour	2.14	293	298
2 hours	1.88	70	102
24 hours	1.63	82	70

the size of the record was reduced by one. The purpose of this work was to determine the effect of a "growing" sample upon a frequency analysis. Table 5.4 indicates that sample growth by the acquisition of "normal" points has little effect upon the results of the analysis, since the changes of the parameter estimates are far less than changes expected for different samples of the same population. This table is complementary to Table 5.1, for together the two tables indicate that a least squares analysis is quite stable under a growing sample whether the newly acquired points represent frequent or rare events.

Table 5.4 Stability of Parameter Estimates with a Growing Sample

Annual Flood Data

Station	Least Squares		Maximum Likelihood	
	$\bar{\alpha}$	Variance of α	$\bar{\alpha}$	Variance of α
315	4.98	0.0063	4.75	0.0200
450	7.61	0.2708	7.61	0.1685
940	6.06	0.1546	3.88	0.0118
1805	3.03	0.2071	2.11	0.0408
3210	8.38	0.2666	7.95	0.0395

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Summary of Results

In the preceding chapters, the method of least squares has been examined as a pobbible method of estimating statistical parameters from a given set of data. The major thrusts of this investigation have been attempts to determine the theoretical basis of the method of least squares, to estimate the efficiency of the least squares estimators when compared to other commonly used estimators, and to identify the strengths and weaknesses of the least squares method as a tool for statistical analysis, with particular emphasis upon frequently-occurring problems in hydrology.

In Chapter II it was recognized that the method of least squares is a form of the method of maximum likelihood if it may be assumed that the errors arising in the fitting process are normal variates of zero mean and constant standard deviation. In Chapter IV the validity of this assumption was examined experimentally and it was found that there is justification for the adoption of this assumption. Certainly it appears that this assumption is no less justified in this instance than in other unrelated applications where the method of least squares is used with the assumption of normality made tacitly. Thus is have been

indicated at least that the use of the method of least squares as a method for estimating statistical parameters is theoretically sound and may be justified as an application of the method of maximum likelihood.

Also presented in Chapter II is a powerful method due to Halperin and Hartley by which confidence regions for the various parameters being estimated may be constructed. The method discussed gives an explicit construction of these regions, and is available in the same form regardless of the particular distribution function being fitted. It has been shown how such regions may be used to construct tolerance limits for various percentiles of a population. Confidence regions and some tolerance limits were computed in Chapter IV to illustrate the use of the method. The results of these computations indicated that tolerance limits computed from least squares criteria are not as confining as those obtained by some other means. In spite of this fact, the ease with which such limits may be computed indicates that least squares tolerance limits can be of value in many instances.

Chapter II concludes with a summary of the more common difficulties encountered in non-linear regression and a brief exposition of methods by which these problems may be overcome. While no effort at completeness has been attempted in this discussion, it is felt that the above exposition is of sufficient scope and detail to be applicable to difficulities which might be encountered in a wide variety of applications.

As a compliment to the theoretical investigations into the method of least squares, rather extensive numerical experiments were made to determine the properties of the method when applied to a two-parameter gamma distribution. This investigation was concerned with such details as methods by which a sample can efficiently be grouped to form a histogram, the statistical properties of the estimators, and the question of whether or not these properties can be altered (and thus hopefully improved) by the proper chice of weighting functions or by the extension of the range of optimization beyond the range of the sample. The results of these experiments indicated that the method of least squares is most efficient when samples are grouped into histograms using a constant class width over the entire range of the sample, and that the choice of the class width had no appreciable effect upon the results of the fitting, so long as the choice was made according to rather flexible and reasonable criteria. It was also found that the least squares estimates are statistically inferior to likelihood estimators in that their use leads to estimates with more scatter (larger variance) than the likelihood estimates evince. By a judicious choice of weights, it is possible to reduce this difference between the two methods to the point at which the difference between the two methods becomes of no practical consequence. At the same time, the weights may be used to render the least squares estimators unbiased, and thus least squares may be made quite attractive as a method of estimating statistical parameters. The results of this inquiry indicated that the extension of the range of optimization

beyond the range of the sample is at best a questionable practice, and should be used only when it is certain the sample being analysed is highly truncated, if indeed it should be used at all.

By fitting the two-parameter gamma distribution to annual flood data and annual maximum precipitation data, it was shown that the least squares estimators are remarkably stable when applied to samples containing records of rare events. When applied to real samples containing no such records, the least squares estimates were not substantially different from the likelihood estimates, thus verifying the results of the earlier numerical experimentation.

Conclusions

In summary, the results of this study appear to substantiate the following observations and conclusions:

a) The use of least squares is theoretically sound, and may be regarded as an application of the concepts of maximum likelihood.

b) Confidence regions and tolerance limits may be obtained squares methods for a wide variety of problems.

c) The statistical properties of least squares estimators will in general be inferior to likelihood estimators, although by the proper choice of weighting functions this inferiority can probably be reduced to the point to which it is of no practical consequence.

d) Judicious choices of weighting functions may allow the modification of the properties of least squares estimators to accomodate peculiarities of the problem at hand.

e) For the two-parameter gamma distribution function, the particular choice of weights

$$w(x;\alpha,\beta) = [f(x;\alpha,\beta)]^{-0.75}$$

where $f(x;\alpha,\beta)$ represents the gamma distribution function, has many advantages, among which are included the facts that these weights vastly improve the convergence characteristics of the method and render the estimators unbiased. It is anticipated that different distributions would require different weights to produce desirable estimators.

f) The method of least squares is quite stable, and may be applied to samples containing records of rare events with good results.

g) The range of optimization should be limited to the range of the data sample, except perhaps for samples known to be highly truncated.

Recommendations

Upon the strength of the above conclusions and observations, the use of the method of least squares to estimate statistical parameters can be recommended to hydrologists as a means by which some of the problems more or less particular to hydrologic analysis may be at least partially overcome. It is seen that, as a method of parameter estimation, least squares can be almost as efficient as more standard methods. It would be a mistake, however, to judge the utility of the method of least squares solely by comparison with other methods

because the potential of least squares extends far beyond traditional statistical analysis. For example, an examination of the annual flood data discussed in Chapter V is sufficient in itself to suggest the possibility of an underlying distribution with more than one "hump". Indeed, if one considers that floods may arise from precipitation brought about by more than one type of mechanism then the idea that such data should exhibit multiple peaks in its frequency distribution becomes most compelling. Such distributions might be approximated very simply as the sum of two distrbution functions. That is

$$g(x; \bar{\gamma}_1, \bar{\gamma}_2) = \frac{1}{2} \left(f_1(x; \bar{\gamma}_1) + f_2(x; \bar{\gamma}_2)\right) \quad .$$

Such distributions could easily be fit by least squares techniques, and, perhaps more importantly, the easily visualizable nature of the least squares algorithm is conducive to a ready appraisal of the properties of such fittings. Thus, the method of least squares, in this and in other ways, can be of benefit in the study of the physical and statistical processes underlying hydrologic phenomena.

Most important of all, the techniques for the application of the method of least squares have not been exhausted in this study. For example, in Chapter IV was discussed the possibility of optimizing class widths based upon a criterion of minimizing in some sense a particular confidence region or tolerance limit. Certainly this and other techniques for improving the performance of least squares for a given application remain to be explored, and surely further improvement may be

expected to result from such investigations. One point is clear; at least in the fitting of the two-parameter gamma distribution function, the method of maximum likelihood can do no better than the results reported in this study. On the other hand, the method of least squares can do no worse, and perhaps awaits only a more perceptive investigator to further enhance the properties of the method.

APPENDIX A

PROPERTIES OF THE TWO-PARAMETER GAMMA DISTRIBUTION

The two-parameter gamma distribution is defined by the relation

$$p(x;\alpha,\beta) = \frac{\alpha^{\beta} x^{\beta-1} e^{-\alpha x}}{\Gamma(\beta)} \qquad \text{A.1}$$

where $\Gamma(\beta)$ is the gamma function defined by

$$\begin{aligned} \Gamma(\beta) &= \int_0^{\infty} x^{\beta-1} e^{-x} dx \\ &= \alpha^{\beta} \int_0^{\infty} x^{\beta-1} e^{-\alpha x} dx \end{aligned} \qquad \text{A.2}$$

From Equation A.2 one readily finds that

$$\begin{aligned} \Gamma(\beta+1) &= \int_0^{\infty} x^{\beta} e^{-x} dx = \beta \int_0^{\infty} x^{\beta-1} e^{-x} dx \\ &= \beta\,\Gamma(\beta) \end{aligned} \qquad \text{A.3}$$

where the second step is obtained by an integration by parts with $u = x$ and $dv = e^{-x}$. By direct integration, one finds

$$\Gamma(1) = \int_0^{\infty} x^{0} e^{-x} dx = 1 \qquad \text{A.4}$$

and

$$\Gamma(2) = \int_0^\infty xe^{-x}dx = 1 \ . \qquad \text{A.5}$$

Using Equations A.3 and A.5 inductively, one finds, for integral values of β,

$$\Gamma(\beta+1) = \beta! = \beta \cdot (\beta-1) \cdot (\beta-2) \ldots \cdot \beta \cdot 1 \ . \qquad \text{A.6}$$

Now, suppose that the random variable x has a gamma distribution by Equation A.1. Then

$$\mu = E(x) = \int_0^\infty xp(x; \alpha, \beta)dx = \frac{\alpha^\beta}{\Gamma(\beta)} \int_0^\infty x^\beta e^{-\alpha x}dx$$
$$= \frac{\alpha^\beta}{\Gamma(\beta)} \frac{\Gamma(\beta+1)}{\alpha^{\beta+1}} = \frac{\beta}{\alpha} \ , \qquad \text{A.7}$$

where the intermediate steps follow by Equations A.2 and A.3. Also,

$$E(x^2) = \int_0^\infty x^2 p(x; \alpha, \beta) = \frac{\alpha^\beta}{\Gamma(\beta)} \int_0^\infty x^{\beta+1} e^{-\alpha x}dx$$
$$= \frac{\alpha^\beta}{\Gamma(\beta)} \frac{\Gamma(\beta+2)}{\alpha^{\beta+2}} = \frac{\beta(\beta+1)}{\alpha^2} \ , \qquad \text{A.8}$$

and so

$$\sigma^2 = V(x) = E(x^2) - [E(x)]^2 = \frac{\beta^2+\beta}{\alpha^2} - \frac{\beta}{\alpha^2} = \frac{\beta}{\alpha^2} \ . \qquad \text{A.9}$$

Likewise, one may obtain expressions for higher moments about the mean. For example, the third moment of x about μ is found to be

$$\mu_3 = E[(x-\mu)^3] = \frac{2\beta}{\alpha^3} \qquad \text{A.10}$$

Thus, the skewness coefficient (Yevjevich [26]) is given by the relation

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{2\beta/\alpha^3}{\beta^{3/2}/\alpha^3} = \frac{2}{\sqrt{\beta}} \qquad \text{A.11}$$

and so it follows that the gamma distribution is positively skewed for all positive values of α and β, the skewness decreasing as β increases. To express the gamma distribution function in finite form, one writes

$$\bar{p}(v; \alpha, \beta) = \frac{\alpha^\beta}{\Gamma(\beta)} \int_{v_{i-1}}^{v_i} v^{\beta-1} e^{-\alpha v} dv \quad , \quad v_{i-1} \leq v \leq v_i \quad . \qquad \text{A.12}$$

By the mean value theorem for integrals (Taylor [24]), there exists a number v* in $[v_{i-1}, v_i]$ such that

$$\bar{p}(v; \alpha, \beta) = p(v^*; \alpha, \beta)\, \Delta v_i \quad , \qquad \text{A.13}$$

where

$$\Delta v_i = v_i - v_{i-1} \quad .$$

Equation A.13 is often written in the more informal form

$$\bar{p}(v; \alpha, \beta) = p(v; \alpha, \beta) \Delta v_i \quad , \qquad \text{A.14}$$

where Equation A.14 is to be understood in the sense of Equation A.13.

APPENDIX B

MATHEMATICAL TECHNIQUES FOR THE LEAST SQUARES FITTING OF THE TWO-PARAMETER GAMMA DISTRIBUTION FUNCTION

In this appendix are developed or collected specific numerical techniques of use in applying the method of least squares to the problem of fitting a two parameter gamma distribution function to a sample histogram.

The Equations of Least Squares

Let the sample histogram be denoted by the sequence $\left\{h_i\right\}_{i=1}^{N}$ and let it be assumed that the histogram has been normalized so that

$$\sum_{i=1}^{N} h_i = 1 \qquad \text{B.1}$$

In the notation of Appendix A, the distribution function is denoted by $\bar{p}(v_i; \alpha, \beta)$ where

$$\bar{p}(v_i; \alpha, \beta) = \frac{\alpha^{\beta}}{\Gamma(\beta)} v_i^{\beta-1} e^{-\alpha v_i} \qquad \text{B.2}$$

and v_i is understood in the sense of Equation A.13. Then in analogy to Equation 2.23, the weighted sum of the residuals E^2 which it is desired to minimize becomes

$$E^2 = \sum_{i=1}^{N} (h_i - \bar{p}(v_i; \alpha, \beta))^2 w_i \qquad B.3$$

and so,

$$\frac{\partial E^2}{\partial \alpha} = \sum_{i=1}^{N} \left[\frac{\partial w_i}{\partial \alpha} (h_i - \bar{p}(v_i; \alpha, \beta))^2 - 2(h_i - \bar{p}(v_i; \alpha, \beta)) \frac{\partial \bar{p}}{\partial \alpha} (v_i; \alpha, \beta) \right] \qquad B.4$$

and

$$\frac{\partial E^2}{\partial \beta} = \sum_{i=1}^{N} \left[\frac{\partial w_i}{\partial \beta} (h_i - \bar{p}(v_i; \alpha, \beta))^2 - 2(h_i - \bar{p}(v_i; \alpha, \beta)) \frac{\partial \bar{p}}{\partial \beta} (v_i; \alpha, \beta) \right] \qquad B.5$$

For the particular choice of weights

$$w_i = [\bar{p}(v_i; \alpha_o, \beta_o)]^{-\gamma} \qquad B.6$$

Equation B.4 and B.5 may be written in the form

$$\frac{\partial E^2}{\partial \alpha} = -2 \sum_{i=1}^{N} \bar{p}(v_i; \alpha_o, \beta_o)^{-\gamma} (h_i - \bar{p}(v_i; \alpha, \beta) \frac{\partial \bar{p}}{\partial \alpha} (v_i; \alpha_o, \beta_o) \qquad B.7$$

and

$$\frac{\partial E^2}{\partial \beta} = -2 \sum_{i=1}^{N} \bar{p}(v_i; \alpha_o, \beta_o)^{-\gamma} (h_i - \bar{p}(v_i; \alpha, \beta) \frac{\partial \bar{p}}{\partial \beta} (v_i; \alpha_o, \beta_o) \qquad B.8$$

where it has been assumed that the terms involving $\frac{\partial w_i}{\partial \alpha}$ and $\frac{\partial w_i}{\partial \beta}$ are negligible in comparison to the terms retained in Equations B.7 and B.8. With the usual Taylor approximation of $\bar{p}$ about the point α_o, β_o, one obtains the normal equations

$$S_{\alpha\alpha}\,\Delta\alpha + S_{\alpha\beta}\,\Delta\beta = R_\alpha \qquad \text{B.9}$$

$$S_{\alpha\beta}\,\Delta\alpha + S_{\beta\beta}\,\Delta\beta = R_\beta \qquad \text{B.10}$$

where

$$S_{\alpha\alpha} = \sum_{i=1}^{N} \left[\frac{\partial \bar{p}(v_i; \alpha_o, \beta_o)}{\partial \alpha} \right]^2 \left[\bar{p}(v_i; \alpha_o, \beta_o) \right]^{-\gamma} \qquad \text{B.11}$$

$$S_{\beta\beta} = \sum_{i=1}^{N} \left[\frac{\partial \bar{p}(v_i; \alpha_o, \beta_o)}{\partial \beta} \right]^2 \left[\bar{p}(v_i; \alpha_o, \beta_o) \right]^{-\gamma} \qquad \text{B.12}$$

$$S_{\alpha\beta} = \sum_{i=1}^{N} \left[\frac{\partial \bar{p}(v_i; \alpha_o, \beta_o)}{\partial \alpha} \frac{\partial \bar{p}(v_i; \alpha_o, \beta_o)}{\partial \beta} \right] \left[\bar{p}(v_i; \alpha_o, \beta_o) \right]^{-\gamma} \qquad \text{B.13}$$

$$R_\alpha = \sum_{i=1}^{N} (h_i - \bar{p}(v_i; \alpha_o, \beta_o)) \frac{\partial \bar{p}}{\partial \alpha} (v_i; \alpha_o, \beta_o) [\bar{p}(v_i; \alpha_o, \beta_o)]^{-\gamma} \qquad \text{B.14}$$

and

$$R_\beta = \sum_{i=1}^{N} (h_i - \bar{p}(v_i; \alpha_o, \beta_o)) \frac{\partial \bar{p}}{\partial \beta} (v_i; \alpha_o, \beta_o) [\bar{p}(v_i; \alpha_o, \beta_o)]^{-\gamma} \qquad \text{B.15}$$

The derivatives appearing in Equations B.9 and B.10 may be calculated from Equation B.2 as follows

$$\frac{\partial \bar{p}}{\partial \alpha} (v_i; \alpha, \beta) = (\beta/\alpha - v_i) \bar{p} (v_i; \alpha, \beta) \qquad B.16$$

and

$$\frac{\partial \bar{p}}{\partial \beta} (v_i; \alpha, \beta) = \left(- \frac{\Gamma'(\beta)}{\Gamma(\beta)} + \ln\alpha + \ln v_i \right) \bar{p} (v_i; \alpha, \beta) \qquad B.17$$

While this form of the derivatives is best suited for discussion, a more convenient form for computation may be obtained by differentiating the integral form of $\bar{p}$ as given by Equation A.12

$$\bar{p} (v; \alpha, \beta) = \int_{v_{i-1}}^{v_i} \frac{\alpha^\beta}{\Gamma(\beta)} v^{\beta-1} e^{-\alpha v} dv \qquad A.12$$

One thus obtains

$$\frac{\partial \bar{p}}{\partial \alpha} (v; \alpha, \beta) = \int_{v_{i-1}}^{v_i} (\beta/\alpha - x) p (x; \alpha, \beta) dx \qquad B.18$$

and

$$\frac{\partial \bar{p}}{\partial \beta} (v; \alpha, \beta) = \int_{v_{i-1}}^{v_i} \left(- \frac{\Gamma'(\beta)}{\Gamma(\beta)} + \ln\alpha + \ln x \right) p (x; \alpha, \beta) dx \qquad B.19$$

Equations A.12, B.18 and B.19 are then easily evaluated by using a numerical integration technique. In this work, the technique used was a simple trapezoidal rule with 16 subintervals on $[v_{i-1}, v_i]$.

The method of Levenburg is simply applied to Equations B.9 and B.10 by replacing $S_{\alpha\alpha}$ and $S_{\beta\beta}$ by the quantities

$$\bar{S}_{\alpha\alpha} = (1 + \eta)\, S_{\alpha\alpha} \qquad \text{B.20}$$

$$\bar{S}_{\beta\beta} = (1 + \eta)\, S_{\beta\beta} \,. \qquad \text{B.21}$$

Appropriate values of η were found by trial to be

$$\eta_0 = 0.2 \qquad \text{B.22}$$

$$\eta_n = \eta_{n-1}/2 \qquad \text{B.23}$$

Using the above relations, one seeks the least squares estimates a and b of α and β by repeated solution of

$$\bar{S}_{\alpha\alpha}\,\Delta\alpha + S_{\alpha\beta}\,\Delta\beta = R_\alpha \qquad \text{B.24}$$

$$S_{\alpha\beta}\,\Delta\alpha + \bar{S}_{\beta\beta}\,\Delta\beta = R_\beta \qquad \text{B.25}$$

This iteration is continued until $(\Delta\alpha^2 + \Delta\beta^2)^{1/2}$ is less than some preassigned quantity (the value used in this study was 10^{-4}).

The initial estimates a_0 and b_0 necessary to begin the least squares solution were obtained by taking these estimates to be the estimates of α and β obtained by the method of moments. Thus, if $\bar{x}$ and S^2 are the sample mean and variance respectively, then

$$a_o = \bar{x}/S^2 \qquad \text{B.26}$$

and

$$b_o = a_o\bar{x} \ . \qquad \text{B.27}$$

In order to evaluate numerically the expressions for $\bar{p}$ and its derivatives as given by Equations A.12, B.18 and B.19, it is necessary to evaluate several transcendental functions. Schemes for the evaluation of the logarithms and antilogrithm terms are so usual as to require no comment; however, the evaluation of the gamma function and its derivatives is not so commonplace. A rather slick method for evaluating these terms is given in the following paragraphs.

The gamma function

$$\Gamma(1 + y) = \int_o^\infty t^y e^{-t} dt \qquad \text{B.28}$$

may be approximated by the polynomial

$$\Gamma(1 + y) = 1 + b_1 y + b_2 y^2 + \ldots + b_8 y^8 + \epsilon(y) \qquad \text{B.29}$$

in the range $0 \leq y \leq 1$, where the error term $\epsilon(y)$ obeys the inequality

$$|\epsilon(y)| < 3 \times 10^{-7} \qquad \text{B.30}$$

The values of the constants in Equation B.29 are

$b_1 = -0.5771965$ $\qquad$ $b_5 = -0.75670408$

$b_2 = 0.98820589$ $\qquad$ $b_6 = 0.48219939$

$b_3 = -0.89705694$ $\qquad$ $b_7 = -0.19352782$

$b_4 = 0.91820686$ $\qquad$ $b_8 = 0.03586834$

If $x = N + y$, where $N = 1,2,\ldots,32$, and $0 < y < 1$

$$\Gamma(1 + x) = (N + y)\ (N + y-1)\ \ldots\ (1 + y)\ \Gamma(1 + y) \qquad \text{B.31}$$

If $x = N + y$, where $N = -1,-2,\ldots$ and $0 < y < 1$, then

$$\Gamma(1 + x) = \frac{\Gamma(1 + y)}{(N + y + 1)\ (N + y)\ \ldots\ (y-1)y} \qquad \text{B.32}$$

If $x = N + y$, where $N = 33,34,\ldots$ and $0 < y < 1$, then

$$\Gamma(x) \sim e^{-x}\ x^{x-1/2}\ \sqrt{2\pi}\ \left(1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \ldots\right) \qquad \text{B.33}$$

Formulas B.29, B.31, B.32 and B.33 provide a method of computing $\Gamma(x)$. The accuracy of these estimates will generally be between 6 and 7 places. (Reference [20] and Reference [31]).

To compute the derivative of the gamma function, one first notes that

$$\frac{d}{dx}\left[\ln \Gamma(x)\right] = \frac{\Gamma'(x)}{\Gamma(x)} \qquad \text{B.34}$$

The function $\frac{\Gamma'(x)}{\Gamma(x)}$ is known as the Psi function, and is discussed in Abramowitz and Stegun [31], where is given the series expansion

$$\psi(1 + z) = \frac{\Gamma'(1 + z)}{\Gamma(1 + z)} = -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n + z)} \; ; \; z \neq -1, -2, \ldots \qquad \text{B.35}$$

where γ is Eulers number (=0.5772156649...). Equation B.35 thus gives a method whereby the value of $\psi(1 + z)$ can be computed; however, the infinite series, having terms of the order of N^{-2}, converges with order N^{-1}. Thus, to obtain accuracy of six places would require the summation of on the order of a million terms, an obvious impracticality. By termwise division (Kantorvich and Krylov [32]) of the series in Equation B.35, one obtains

$$\frac{1}{n(n + z)} = \frac{1}{n^2} - \frac{z}{n^3} + \frac{z^2}{n^4} - z^3 \frac{1}{n^4(n + z)} \qquad \text{B.36}$$

Upon noting that the series is absolutely convergent, there follows

$$\sum_{n=1}^{\infty} \frac{z}{n(n + z)} = z\left(\sum_{n=1}^{\infty} \frac{1}{n^2} - z \sum_{n=1}^{\infty} \frac{1}{n^3} + z^2 \sum_{n=1}^{\infty} \frac{1}{n^4} - z^3 \sum_{n=1}^{\infty} \frac{1}{n^4(n + z)} \right)$$

B.37

But all except the last sum in Equation B.37 is independent of z, and thus may be evaluated once and for all, giving

$$\sum_{n=1}^{\infty} \frac{z}{n(n+z)} = z\left(1.644934 - 1.202051z + 1.082323z^2 - z^3 \sum_{n=1}^{\infty} \frac{1}{n^4(n+z)}\right) \qquad \text{B.38}$$

The terms of the series appearing in Equation B.38 are of the order of N^{-5}, and thus a substantial improvement in convergence has been effected.

A similar massage of the terms of the expression (Abramowitz and Stegun [31]) for ψ',

$$\psi'(z) = \sum_{k=0}^{\infty} \frac{1}{(z+k)^2} \qquad (z \neq 0, -1, -2, \ldots) \qquad \text{B.39}$$

yields the equivalent series expression

$$\psi'(z) = \frac{1}{z^2} + 1.644934 - 2.404102z + 3.2469699z^2 \qquad \text{B.40}$$

$$-z^3 \sum_{k=1}^{\infty} \frac{(4k+3z)}{k^4(k+z)^2}$$

The terms of the series appearing in Equation B.40 are of the order of N^{-5}, and thus convergence is of the order of N^{-4}. Equations B.37 and B-40 were used to compute the values of ψ and ψ' in this study.

APPENDIX C

GENERATION OF PSEUDORANDOM NUMBERS

In this appendix are discussed certain methods by which sequences of numbers having properties loosely described as "random" may be generated on a digital computer. Among the difficulities associated which the generation of random numbers on a digital computer, perhaps the two most striking are the inability to produce a precise definition of what is meant by the term "random number", and the philosophical uneasiness which arises from the use of a finite number of arithmetical operations to produce a number supposedly independent of any particular process and unpredictable from the knowledge of any previous state of nature. It was undoubtedly such considerations which prompted Von Newmann (Ralston and Wilf [33]) to state that "anyone who considers arithmetical methods for producing random digits is, of course, in a state of sin".

Given the above described and rather deplorable state of affairs, and faced nevertheless with the necessity of producing by arithmetic means sequences of random digits, statisticians have, with characteristic piety, avoided the issue by inventing the concept of pseudorandom sequences. D. H. Lehmer (Ralston and Wilf [33]) has defined a pseudorandom sequence as "a vague notion embodying the idea of a sequence in which each term is unpredictable to the

uninitiated and whose digits pass a certain number of tests, traditional with statisticians and depending somewhat upon the uses to which the sequence is to be put."

Methods discussed in this appendix for generating pseudorandom sequences from various distributions are based upon numerical operations performed on sequences of uniformly distributed pseudorandom numbers. By a uniformly distributed pseudorandom sequence is meant a pseudorandom sequence of numbers, all of whose terms fall within a given interval [a,b], which have the property that the probability that any term of the sequence lies in a given subinterval of [a,b] is a linear function of the width of the sub-interval and is independent of any other consideration. Sequences possessing the above properties are said to be pseudorandom sequences uniformly distributed on [a,b].

For use on binary computers, the favored algorithms for producing uniformly distrubted pseudorandom numbers on $(0,2^p-1)$, where p indicates the word length of the machine in question, are the mixed muliplicative congruential procedures of the form

$$R_{n+1} = (\alpha R_n + \beta) \bmod (2^p) \qquad \text{C.1}$$

where (-)mod (2^p) indicates the remainder obtained as the result of division by 2^p. (For example, 3 mod (2) = 1, and 8 mod (2)= 0). Choices of α and β are suggested by considerations of periodicity and serial correlation of the sequences generated by Equation C.1 (Ralston and Wilf [33]).

A modification of the above conguential method was used to generate the pseudorandom numbers utilized in this work. This modification makes use of two conguential generators of the form given in C.1, and a table of preselected length r. Prior to the beginning of the computation, the table is filled with r numbers computed from the first generator (call this generator G1). The computation of a pseudorandom number is then made, based upon two given numbers N1 and N2 in the following manner. Using the number N2 and the second generator (say G2), a pseudorandom integer N2' on [1,r] is computed. Using the number N1 and the generator G1, a pseudorandom number N1' on $[0,2^{p}-1]$ is computed. Using N2' as an index into the table, the desired number x is chosen as the existing entry in the N2' position of the table. The process is completed by the replacment of x at the N2' position of the table by N1'. In practice, the numbers N1' and N2' are used in place of N1 and N2 for the next computation. Thus, the modified method in essence consists of the random selection of numbers from a table of random numbers.

The complexity of the above procedure is such that it almost defies the imagination, and one might expect that sequences selected by such a procedure would exhibit many properties ascribed to random numbers. Such is the case. An account of various tests performed on sequences generated by this procedure is given in Reference [21]. In addition to these tests, eight sequences of length 1000 were tested for a tendency of the procedure to repeat. No repeated numbers

were found in any of these sequences.

Values of α and β in the generators used in this study were

$$G1: \quad \alpha = 13^9, \quad \beta = 7261067085$$

and

$$G2: \quad \alpha = 29^7, \quad \beta = 7261067085.$$

In addition, use was made of an auxillary generator for which

$$\alpha = 5^{15}, \quad \beta = 7261067085.$$

In all cases, $p = 35$.

Once there is available a sequence of pseudorandom numbers uniform on (0,1), then a sequence of pseudorandom numbers non-uniformly distributed may be generated by the following method. Let F denote the cumulative distribution function of the desired distribution. Then the sequence

$$S_n = F^{-1}(R_n) \qquad C.2$$

is distributed according to F (here F^{-1} denotes the inverse of F). Since the sequence (R_n) is random, then so also is the sequence (S_n). Unfortunately, for many important distributions, the inverse of the

cumulative distribution function cannot be expressed in closed form. One is then faced with a task of providing a rational approximation to F^{-1}, or abandoning equation C.2 in favor of more simple methods of generating the required sequence (S_n). For example, random numbers from a normal population with parameters μ and σ may be generated by the formulae

$$S_n = \sigma (-2.\ln R_n)^{1/2} \cos 2\pi R_{n+1} + \mu$$

$$S_{n+1} = \sigma (-2.\ln R_n)^{1/2} \sin 2\pi R_{n+1} + \mu \qquad \text{C.3}$$

(Ralston and Wilf [33]). Random numbers from a gamma population with parameters α and β (with β integral) may be generated by the formula

$$S_n = -\frac{1}{\alpha} \left(\ln \prod_{i=1}^{\beta} R_{ni}\right) \qquad \text{C.4}$$

where the R_{ni} form a double sequence of numbers uniform and random on (0,1) (Fontane [34]). Equation C.4 was used in this study to generate pseudorandom gamma variates.

APPENDIX D

A DESCRIPTION OF THE COMPUTER PROGRAM

In this appendix is given a user oriented description of the computer program developed for the least squares fitting of the two parameter gamma distribution function to given or synthetic data, or to a given histogram.

Purpose of the Program

This program is designed to accept or generate data samples and to fit the data with a two-parameter gamma distribution by a least squares optimization of the distribution parameters. The program also estimates the distribution parameters by the methods of moments and maximum likelihood, and performs certain statistical tests on the least squares error terms and on the parameter estimates.

Language and Computer Requirements

This program is coded in Fortran V and was designed for operation on the UNIVAC 1108 under EXEC 8 monitor. The program requires 8200 decimal locations in the instruction bank and 7250 decimal locations in the data bank. I/O is by the standard input and output devices only. No temporary or permanent files or storage are required for operation of the program. Use is made in the program of elements of the UNIVAC

large scale systems MATH-PACK/STAT-PACK program group.

Data Input Format

A single run is defined as a logical set of cases of synthetic data from a given gamma population, or a single fitting of a given set of input data. Input in all cases is according to the following format:

Card	Field	Variable	Type
1	1-5	IREAL	Integer
2	1-72	TITLE	Alphanumeric
3	1-5	NX	Integer
3	6-10	ITER	Integer
3	11-15	NCYCLE	Integer
3	21-25	NEMPT	Integer
4	1-10	PAR(1)	Real
4	11-20	PAR(2)	Real
4	21-30	TEST	Real
4	31-40	EXPON	Real
5	1-10	IRN1	Integer
5	11-20	WSCALE	Real

Card 6 only if IREAL = 1

Card	Field	Variable	Type
6	1-10	VT(I)	Real
	11-20		
	etc		
Card 7 and 8 only if IREAL = 3			
7	1-10	W	Real
7	11-20	VB	Real
7	21-25	NH	Integer
8	1-10	H(I)	Real
	11-20		
	etc		

The above variables have the following meanings:

IREAL - a control variable

IREAL = 0 Signifies the program is to generate synthetic data

= 1 Signifies real data is to be input

= 3 Signifies a histogram is to be input

= -1 Signifies no more runs are to be made.

TITLE - title of the run

NX - the number of data points per cycle to be generated or input ($\leq$ 500), or, if IREAL = 3, the number of data points used to construct the histogram.

ITER - the maximum number of iterations per cycle for the iteration scheme to be carried. If convergence of the least squares procedure is not obtained within ITER iterations, the case is

abandoned and computation begun on the following case.

NCYCLE - the number of separate cases to be considered for the current run (NCYCLE = 1 unless IREAL = 0. NCYCLE $\leq$ 100)

NEMPT - the number of empty classes to be added to the data histogram before fitting.

PAR(1) - α value of the gamma population (required only if IREAL = 0)

PAR(2) - β value of the gamma population (required only if IREAL = 0).

TEST - the limiting value of the parameter corrections. If $(\Delta\alpha^2 + \Delta\beta^2)^{1/2} \leq$ TEST, convergence is declared.

EXPON - the exponent of the weight factor in the expression $[P(v_i; \alpha, \beta)]^{\gamma}$ (Note EXPON) must be negative or zero).

IRN1 - the initial number from which the synthetic data are generated (required only if IREAL = 0)

WSCALE - the correction factor by which Sturges' class interval is to be multiplied. (See Chapter III)

VT(I) - values of the real data to be analysed (only if IREAL = 1).

W - the class interval for the input histogram (only if IREAL = 3).

VB - the value of data variate corresponding to the left edge of the first class of the input histogram (only if IREAL = 3).

NH - the number of classes in the input histogram (only if IREAL = 3).

H(I) - the input histogram (only if IREAL = 3).

Programming Methods

The method of analysis used by this program is the method of least squares as developed in this report. Flow charts of the main program, which directs the logical flow of the program, and of subroutine GPARTL, which performs the least squares fitting of the density function to the histogram, are given in Figures D-1 and D-2 respectively. A list of other subroutines required (exclusive of those subroutines in the standard FORTRAN library) and a brief description of their methods and functions is given in Table D-1.

Operating Considerations

The program will accept multiple runs and will provide a summary of the results of each run. Output and run time vary with the size of samples being analyzed. Output should average between two to three seconds of computer CPU time per case for normal samples of size 100 or less.

Table D-1. Description of Computer Subroutines

Subroutine	Calling Program	Function and Methods
NRAND*	MAIN	The auxilary random number generator (see Appendix C).
RANDUJ	MAIN	Generates uniform random numbers on (0,1) by the use of the methods of Appendix C.
RANGE*	MAIN	Computes the range of a sample.
HIST*	MAIN	Groups a given set of data into a histogram and prints the histogram on the printer.
PSI	MAIN	Computes $\frac{d}{dx}\left(\ln \Gamma(x)\right)$ (see Appendix B).
PSIP	MAIN	Computes $\frac{d^2}{dx^2}\left(\ln \Gamma(x)\right)$ (see Appendix B).
GAMIN*	MAIN	Evaluates the incomplete gamma function.
CINORM	MAIN	Performs a chi-square test for normality (see Chapter IV).
KOLN	MAIN	Performs a Kolmogorov-Smirnov test for normality (see Chapter IV)
GPDF	GPARTL	Evaluates the gamma distribution function and its derivatives (see Appendix B).

Subroutine	Calling Program	Function and Methods
INVER2	GPATL	Inverts a 2X2 matrix.
GAMMA*	GPDF	Evaluates the complete gamma function (see Appendix B)
GROUP*	HIST	Groups data into a histogram
PLOT1*	HIST	Plots a line of symbols on the printer.
RNORM*	KOLN	Evaluates the cumulative distribution function of the normal distribution.
MRAND*	RANDUJ	Generates integers random on $(0, 2^{35}-1)$
TINORM*	CINORM	Evaluates the inverse of the cumulative distribution function of a normal distribution.
CHI*	CINORM	Evaluates the cumulative distribution function of the chi-square distribution.

* denotes subroutines included in the Univac large scale systems MATH-PACK/ STAT-PACK group.

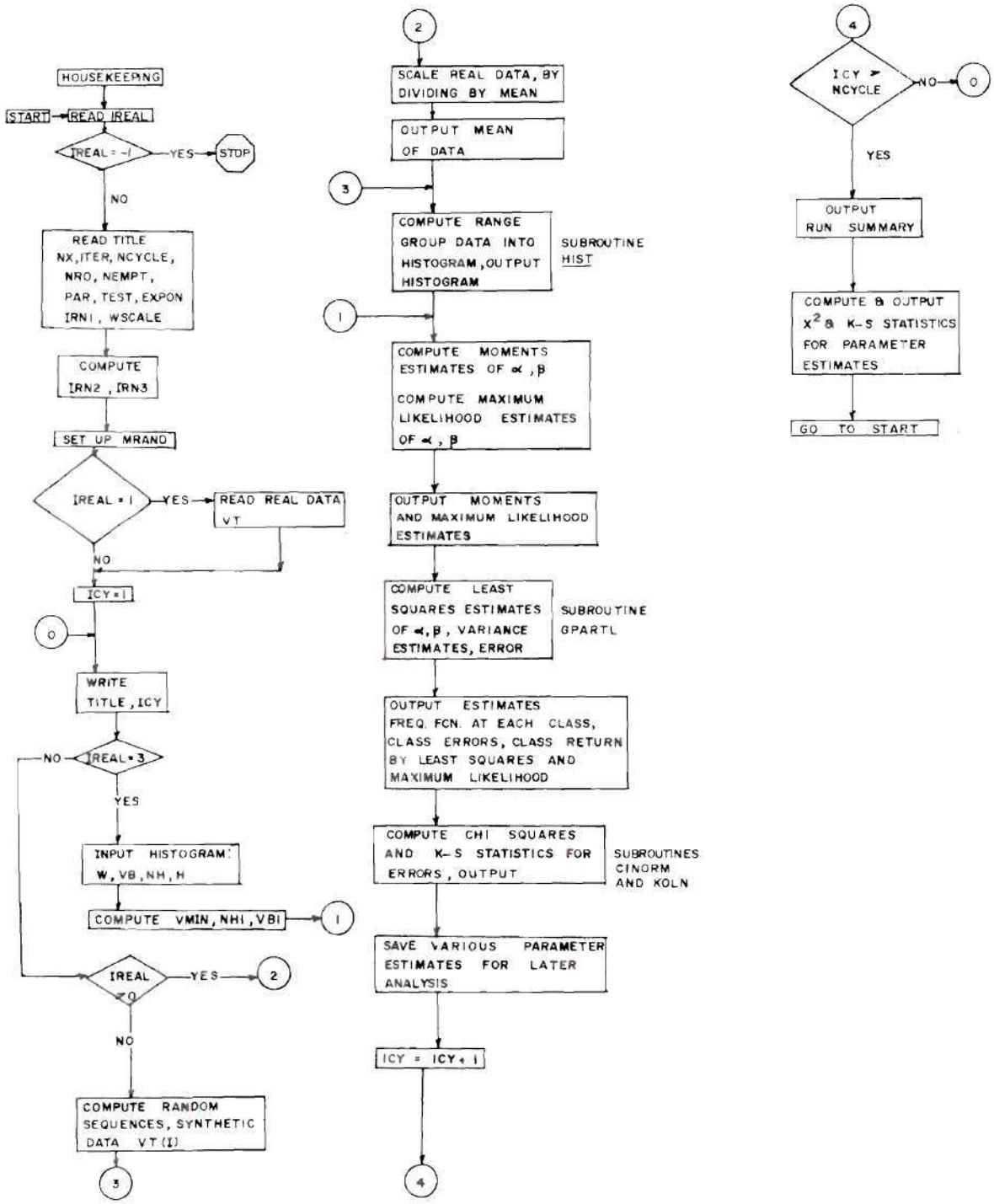

Figure D-1 Flow Chart of Main Program

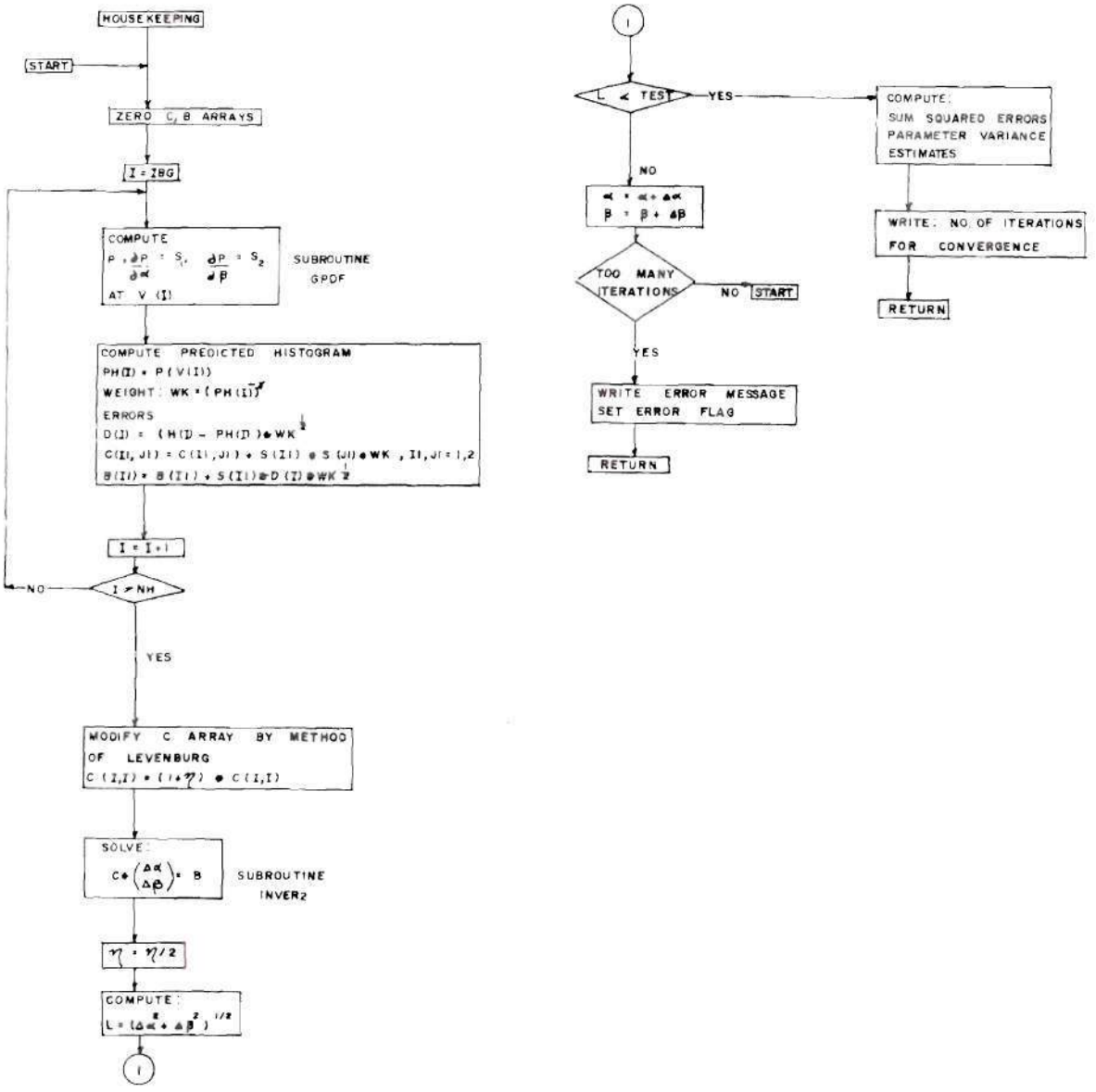

Figure D-2 Flow Chart of Subroutine GPARTL

APPENDIX E

DESCRIPTION OF COMPUTER RUNS

In this appendix is given a list of the various simulation runs made during the course of this study. The parameters of the runs consisted of the population parameters A and B, the sample size Nx, the value of γ, the value of r, the number of empty classes NE, and the initial number IRN1 from which the random samples were generated. Table E-1 summarizes these runs.

In aggregate Table E-1 lists 102 runs based upon 30 different random samples.

Table E-1. Parameters of the Simulation Runs

Run No.	A,B	Nx	r	NE	IRN1	γ
1	4	50	1	0	779	0
1A	4	50	0.75	0	779	0
1B	4	50	0.5	0	779	0
2	4	50	1.0	0	339	0
2A	4	50	0.75	0	339	0
2B	4	50	0.5	0	339	0
3	8	50	1.0	0	339	0
3A	8	50	0.75	0	339	0
3B	8	50	0.50	0	339	0
4	3	50	0.5	0	1393997	0
4A	3	50	0.75	0	1393997	0
5	2	50	0.5	0	15011	0
5A	2	50	0.75	0	15011	0
5B	2	50	1.0	0	15011	0
6	5	50	0.5	0	1536	0
7	6	50	0.5	0	2011	0
8	7	50	0.5	0	81647	0
8A	7	50	0.75	0	81647	0
9	9	50	0.5	0	91646	0
9A	9	50	0.75	0	91646	0
10	11	50	0.5	0	69179	0
10A	11	50	0.75	0	69179	0
10C	11	50	1.0	0	69179	0
5R1	2	25	0.5	0	15011	0
5R2	2	75	0.5	0	15011	0
5R3	2	100	0.5	0	15011	0
7R1	6	25	0.5	0	2011	0
7R2	6	75	0.5	0	2011	0
7R3	6	100	0.5	0	2011	0
9R1	9	25	0.5	0	91646	0
9R2	9	75	0.5	0	91646	0
9R3	9	100	0.5	0	91646	0
10R1	11	25	0.5	0	69179	0
10R2	11	75	0.5	0	69179	0
10R3	11	100	0.5	0	69179	0
5-N1	2	50	0.5	1	15011	0
5-N2	2	50	0.5	2	15011	0
5-N4	2	50	0.5	4	15011	0
6-N1	5	50	0.5	1	1536	0
6-N2	5	50	0.5	2	1536	0
6-N4	5	50	0.5	4	1536	0
9-N1	9	50	0.5	1	91646	0
9-N2	9	50	0.5	2	91646	0

Table E-1. Parameters of the Simulation Runs (Continued)

Run No.	A,B	Nx	r	NE	IRN1	γ
10-N1	11	50	0.5	1	69179	0
10-N2	11	50	0.5	2	69179	0
10-N4	11	50	0.5	4	69179	0
4G.25	3	50	0.5	0	1393997	0.25
4G.50	3	50	0.5	0	1393997	0.50
4G.75	3	50	0.5	0	1393997	0.75
4G1	3	50	0.5	0	1393997	1.0
6G.25	5	50	0.5	0	1536	0.25
6G.50	5	50	0.5	0	1536	0.50
6G.75	5	50	0.5	0	1536	0.75
6G1	5	50	0.5	0	1536	1.0
8G.25	7	50	0.5	0	81647	0.25
8G.50	7	50	0.5	0	81647	0.50
8G.75	7	50	0.5	0	81647	0.75
8G1	7	50	0.5	0	81647	1.0
9G.25	9	50	0.5	0	91646	0.25
9G.50	9	50	0.5	0	91646	0.50
9G.75	9	50	0.5	0	91646	0.75
9G1	9	50	0.5	0	91646	1.0
4G.25R2	3	75	0.5	0	1393997	0.25
4G.50R2	3	75	0.5	0	1393997	0.50
4G.75R2	3	75	0.5	0	1393997	0.75
4G1R2	3	75	0.5	0	1393997	1.0
8G.25R2	7	75	0.5	0	81647	0.25
8G.50R2	7	75	0.5	0	81647	0.50
8G.75R2	7	75	0.5	0	81647	0.75
8G1R2	7	75	0.5	0	81547	1.0
4G.25R3	3	100	0.5	0	1393997	0.25
4G.50R3	3	100	0.5	0	1393997	0.50
4G.75R3	3	100	0.5	0	1393997	0.75
4G1R3	3	100	0.5	0	1393997	1.0
8G.25R3	7	100	0.5	0	81647	0.25
8G.50R3	7	100	0.5	0	81647	0.50
8G.75R3	7	100	0.5	0	81647	0.75
8G1R3	7	100	0.5	0	81647	1.0
6G.75N4	5	50	0.5	4	1536	0.75
6G.75N8	5	50	0.5	8	1536	0.75
6G1N4	5	50	0.5	4	1536	1.0
9G.75N4	9	50	0.5	4	91646	0.75
9G1N4	9	50	0.5	4	91646	1.0
14G.50A	3	50	0.5	0	329901221	0.5
14G.75A	3	50	0.5	0	329901221	0.75
16G.50A	6	50	0.5	0	8636721216	0.50

Table E-1. Parameters of the Simulation Runs (Continued)

Run No.	A,B	Nx	r	NE	IRN1	γ
16G.75A	6	50	0.5	0	8636721216	0.75
19G.50A	9	50	0.5	0	7592879626	0.50
19G.75A	9	50	0.5	0	7592879626	0.75
14F.50B	3	75	0.5	0	3898255758	0.5
14G.75B	3	75	0.5	0	3898255758	0.75
16G.50B	6	75	0.5	0	9844208303	0.5
16G.75B	6	75	0.5	0	9844208303	0.75
19G.50B	9	75	0.5	0	648603574	0.5
19G.75B	9	75	0.5	0	648603574	0.75
14G.50C	3	100	0.5	0	9223726759	0.5
14G.75C	3	100	0.5	0	9223726759	0.75
16G.50C	6	100	0.5	0	5661391511	0.50
16G.75C	6	100	0.5	0	5661391511	0.75
19G.50C	9	100	0.5	0	1766807785	0.5
19G.75C	9	100	0.5	0	1766807785	0.75

Appendix F

Summary of Real Data

In this appendix is given a summary of the real data used in this study for various purposes. The annual flood data were taken from U.S. Geological Survey Water Supply Papers Numbers 1671 through 1689. The station part numbers and station numbers shown in Table F.1 are the inventory numbers used by the Geological Survey. Table F.2 gives the actual peak flow records for the six stations discussed in detail in Chapter 5. Table F.3 gives the rainfall data whose analysis is discussed in Chapter 5. These data were obtained from U.S. Weather Bureau Records for the Atlanta, Georgia, gage and represent the period from January 1889 through December 1972.

Table F.1 List of Streamgauge Station Inventory Numbers

Part number	Station number	Part number	Station Number
1A	315	3A	155
1A	450	3A	205
1A	940	3A	215
1A	1805	3A	325
1B	3210	3A	510
1B	3345	3A	1835
1B	3615	4	735
1B	4340	4	770
1B	5405	4	1130
1B	5480	4	1560
1B	6385	4	2165
1B	6680	4	2525
2A	195	5	145
2A	550	5	3310
2A	835	5	4645
2B	2185	5	4815
2B	2235	6A	375
2B	3350	6A	625
2B	3495	6B	7070
2B	3920	7	725
2B	4415	8	335
2B	4770	8	660
2B	4790	9	470
2B	4820	9	850
3A	115	9	2395

Table F.1 (Continued)

Part Number	Station Number
9	4060
10	1285
10	1685
10	1700
10	1720
11	980
11	1520
11	2035
11	2665
11	2750
11	2820
11	4095
11	5025
13	3190
14	2100
14	3210
14	3590

Table F.2 Annual Flood Data

1A-315 Piscataquis River Near Dover-Foxcraft, Maine

DA = 297 square miles, 1903-1971

5140	7420	2410	10400	8040	10100	17400	4010	4110
7380	7130	6930	6100	6200	14600	5960	4710	8650
7600	8350	21500	8690	4570	8040	7780	10400	9600
8040	6870	12900	6350	8040	5590	19300	6750	8110
6240	13700	4010	6970	4680	13500	7190	5300	11600
9640	3100	11100	17400	9310	15200	13200	9560	5150
2990	13300	5250	7190	4920	5420	6810	14000	5250
5050	22800	8990	7460	12800	13400			

1A-450 Dead River at the Forks, Maine

DA = 872 square miles, 1903-1971

10700	24600	21300	23000	22000	16400	24700	20600	17600
8940	10300	22600	10600	9510	13700	23800	15500	16600
12800	12800	16000	17600	12800	15700	16000	18200	14700
13000	28700	16900	11700	14400	14400	10200	16900	14400
14400	15200	12000	18200	14800	13000	7140	12200	8860
10500	16600	13400	8640	4790	10600	11400	9490	11900
5200	8450	11600	4790	7820	7790	11400	19500	17100
9590								

Table F.2 (Continued)

1B-3210 Sacandaga River Near Hope, N. Y.

DA = 491 square miles, 1912-1967

11000	32000	16500	11400	16000	18600	8490	10300	7230
20400	20400	10300	11000	12500	18600	9320	14400	11700
11800	8640	7790	16500	10600	11200	23900	9180	16600
11700	10600	11000	14500	10500	10500	20000	16700	16600
16700	31400	11100	17600	16200	22900	12600	10700	14200
7820	21300	9340	17800	10200	14700	12200	17000	8370
6060	9910							

1B-3345 Hoosic River Near Eagle Bridge, N. Y.

DA = 510 square miles, 1911-1967

8300	6860	10600	8300	16500	13000	8040	13000	7920
9739	12100	13600	11700	11200	7350	7350	41500	10100
5380	6920	9640	11900	11600	13000	31500	7920	35300
11300	13800	6750	7810	10900	10900	9750	9840	9500
12200	55400	13100	19100	15900	9220	10000	7640	10900
7190	8560	14400	15900	8320	12100	9300	9240	3970
5760	9370							

Table F.2 (Continued)

1A-940 Souhegan River at Merrimack, N. H.

DA = 171 square miles, 1910-1970

4500	2810	2250	3250	3680	6000	4750	3290	1820
3200	2930	5410	3980	3450	9260	2050	2590	2870
6180	2290	2530	3490	3520	3210	7500	3260	16900
3450	10800	2280	4250	2370	3880	1730	7830	2430
2750	1680	3990	2100	2180	3200	4300	4650	4410
2710	6760	1830	2970	4180	6000	2370	4730	3500
1800	1220	1540	2770	4400	2550	4400		

1A-1805 Middle Branch, Westfield River at Gross Heights, Mass.

DA = 53 square miles, 1911-1970

2420	2330	2600	2560	4500	2900	1330	1220	3350
4230	3650	2650	2010	2270	4250	1660	1540	5860
1950	990	2340	2000	8020	2850	5420	8400	3250
19900	2250	2310	1600	2710	3150	3670	3790	1870
2250	3400	9600	1700	8320	4980	4420	3560	16500
6460	1850	2320	2000	4700	1800	2370	1820	1300
515	423	597	893	1170	760			

Table F.3 Maximum Precipitation - Atlanta, Georgia

U. S. Weather Bureau Records

Jan. 1889 to Jan. 1973

Order Number	5 Min.		10 Min.		15 Min.		30 Min.		1 Hr.		2 Hrs.		24 Hrs.	
	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.
1	.88	10.56	1.27	7.62	1.57	6.28	2.43	4.86	3.23	3.23	3.45	1.73	5.56	.23
2	.78	9.36	1.24	7.44	.155	6.20	2.30	4.60	2.93	2.93	3.27	1.64	5.52	.23
3	.73	8.76	1.12	6.72	1.53	6.12	2.20	4.40	2.53	2.53	3.26	1.63	5.44	.23
4	.72	8.64	1.11	6.66	1.41	5.64	2.10	4.20	2.42	2.42	3.17	1.59	5.12	.21
5	.69	8.28	1.06	6.36	1.38	5.52	2.09	4.18	2.40	2.40	3.06	1.53	5.04	.21
6	.67	8.04	1.04	6.24	1.37	5,48	2.03	4.06	2.34	2.32	2.93	1.47	4.86	.20
7	.66	7.92	1.00	6.00	1.31	5.24	2.02	4.04	2.32	2.32	2.77	1.39	4.82	.20
8	.65	7.80	1.00	6.00	1.29	5.16	1.93	3.86	2.32	2.32	2.71	1.35	4.70	.20
9	.65	7.80	.98	5.88	1.25	5.00	1.91	3.82	2.29	2.29	2.60	1.30	4.64	.19
10	.65	7.80	.97	5.82	1.24	4.96	1.83	3.66	2.25	2.25	2.54	1.27	4.59	.19
11	.63	7.56	.97	5.82	1.22	4.88	1.81	3.62	2.22	2.22	2.52	1.26	4.51	.19
12	.63	7.56	.96	5.76	1.22	4.88	1.80	3.60	2.20	2.20	2.44	1.22	4.13	.17
13	.60	7.20	.96	5.76	1.20	4.80	1.78	3.56	2.18	2.18	2.40	1.20	4.11	.17
14	.60	7.20	.95	5.70	1.19	4.76	1.76	3.52	2.13	2.13	2.36	1.18	4.05	.17
15	.60	7,20	.95	5.70	1.16	4.64	1.69	3.38	2.12	2.12	2.34	1.17	4.04	.17
16	.59	7.08	.93	5.58	1.16	4.64	1.63	3.26	2.12	2.12	2.34	1.17	3.99	.17
17	.59	7.08	.91	5.46	1.15	4.60	1.60	3.20	2.11	2.11	2.32	1.16	3.90	.16
18	.58	6.96	.91	5.46	1.13	4.52	1.59	3.18	2.11	2.11	2.32	1.16	3.85	.16
19	.57	6.84	.90	5.40	1.12	4.48	1.56	3.12	2.04	2.04	2.30	1.15	3.82	.16

Order Number	5 Min.		10 Min.		15 Min.		30 Min.		1 Hr.		2 Hrs.		24 Hrs.	
	In.	In./hr.	In	In./hr.	In.	In./hr.	In	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.
20	.56	6.72	.90	5.40	1.10	4.40	1.56	3.12	1.95	1.95	2.30	1.15	3.79	.16
21	.55	6.60	.88	5.28	1.09	4.36	1.52	3.04	1.93	1.93	2.29	1.15	3.72	.15
22	.54	6.48	.82	4.92	1.07	4.28	1.46	2.92	1.89	1.89	2.28	1.14	3.70	.15
23	.51	6.12	.82	4.92	1.03	4.12	1.42	2.84	1.87	1.87	2.27	1.14	3.70	.15
24	.51	6.12	.81	4.86	1.01	4.04	1.41	2.81	1.82	1.82	2.25	1.13	3.66	.15
25	.50	6.00	.80	4.80	1.00	4.00	1.37	2.74	1.80	1.80	2.25	1.13	3.64	.15
26	.50	6.00	.77	4.62	.99	3.96	1.34	2.68	1.80	1.80	2.23	1.12	3.63	.15
27	.50	6.00	.77	4.62	.98	3.92	1.30	2.60	1.79	1.79	2.20	1.10	3.63	.15
28	.48	5.76	.74	4.44	.96	3.84	1.30	2.60	1.79	1.79	2.20	1.10	3.51	.15
29	.48	5.76	.73	4.38	.94	3.76	1.29	2.58	1.76	1.76	2.13	1.07	3.49	.15
30	.48	5.76	.73	4.38	.93	3.72	1.28	2.56	1.75	1.75	2.09	1.05	3.46	.14
31	.47	5.64	.73	4.38	.92	3.68	1.27	2.54	1.75	1.75	2.05	1.03	3.43	.14
32	.47	5.64	.72	4.32	.92	3.68	1.26	2.52	1.70	1.70	1.91	.96	3.41	.14
33	.47	5.64	.72	4.32	.91	3.64	1.25	2.50	1.65	1.65	1.88	.94	3.39	.14
34	.46	5.52	.72	4.32	.90	3.60	1.25	2.50	1.60	1.60	1.82	.91	3.34	.14
35	.46	5.52	.72	4.32	.88	3.52	1.23	2.46	1.59	1.59	1.81	.91	3.34	.14
36	.46	5.52	.72	4.32	.88	3.52	1.23	2.46	1.54	1.54	1.80	.90	3.32	.14
37	.45	5.40	.70	4.20	.87	3.48	1.22	2.44	1.51	1.51	1.74	.87	3.28	.14
38	.45	5.40	.70	4.20	.87	3.48	1.21	2.42	1.48	1.48	1.73	.87	3.28	.14
39	.44	5.28	.69	4.14	.87	3.48	1.20	2.40	1.47	1.47	1.69	.85	3.27	.14

Order Number	5 Min.		10 Min.		15 Min.		30 Min.		1 Hr.		2 Hrs.		24 Hrs.	
	In.	In./hr.	In	In./hr.	In.	In./hr.	In	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.
40	.42	5.04	.69	4.14	.83	3.32	1.19	2.38	1.46	1.46	1.64	.82	3.26	.14
41	.42	5.04	.68	4.08	.82	3.28	1.17	2.34	1.45	1.45	1.62	.81	3.24	.13
42	.42	5.04	.67	4.02	.81	3.24	1.14	2.28	1.41	1.41	1.61	.81	3.24	.13
43	.42	5.04	.65	3.90	.79	3.16	1.14	2.28	1.41	1.41	1.60	.80	3.22	.13
44	.41	4.92	.65	3.90	.78	3.12	1.13	2.26	1.40	1.40	1.55	.78	3.16	.13
45	.41	4.92	.64	3.84	.78	3.12	1.13	2.26	1.39	1.39	1.54	.77	3.15	.13
46	.40	4.80	.64	3.84	.77	3.08	1.08	2.16	1.37	1.37	1.53	.77	3.12	.13
47	.40	4.80	.64	3.84	.77	3.08	1.03	2.06	1.36	1.36	1.52	.76	3.11	.13
48	.40	4.80	.63	3.78	.77	3.08	1.01	2.02	1.36	1.36	1.49	.75	3.11	.13
49	.40	4.80	.63	3.78	.75	3.00	1.00	2.00	1.35	1.35	1.48	.74	3.06	.13
50	.40	4.80	.63	3.78	.75	3.00	.98	1.96	1.32	1.32	1.46	.73	2.98	.12
51	.39	4.68	.62	3.72	.73	2.92	.97	1.94	1.30	1.30	1.45	.73	2.97	.12
52	.39	4.68	.62	3.72	.72	2.88	.95	1.90	1.30	1.30	1.43	.72	2.97	.12
53	.39	4.68	.61	3.66	.72	2.88	.94	1.88	1.28	1.28	1.43	.72	2.88	.12
54	.39	4.68	.60	3.60	.72	2.88	.93	1.86	1.28	1.28	1.38	.69	2.87	.12
55	.38	4.56	.58	3.48	.70	2.80	.92	1.84	1.27	1.27	1.35	.68	2.872	.12
56	.38	4.56	.58	3.48	.69	2.76	.92	1.84	1.26	1.26	1.33	.67	2.82	.12
57	.38	4.56	.58	3.48	.69	2.76	.87	1.74	1.25	1.25	1.32	.66	2.77	.12
58	.38	4.56	.58	3.48	.69	2.76	.84	1.68	1.25	1.25	1.32	.66	2.72	.11
59	.38	4.56	.57	3.42	.68	2.72	.83	1.66	1.21	1.21	1.30	.65	2.62	.11

Order Number	5 Min.		10 Min.		15 Min.		30 Min.		1 Hr.		2 Hrs.		24 Hrs.	
	In.	In./hr.	In	In./hr.	In.	In./hr.	In	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.
60	.36	4.32	.57	3.42	.64	2.56	.81	1.62	1.21	1.21	1.29	.65	2.62	.11
61	.36	4.32	.57	3.42	.64	2.56	.79	1.58	1.17	1.17	1.26	.63	2.62	.11
62	.36	4.32	.56	3.36	.61	2.44	.79	1.58	1.17	1.17	1.25	.63	2.61	.11
63	.36	4.32	.56	3.36	.57	2.28	.78	1.56	1.16	1.16	1.25	.63	2.56	.11
64	.35	4.20	.56	3.36	.56	2.24	.75	1.50	1.16	1.16	1.25	.63	2.50	.10
65	.35	4.20	.55	3.30	.55	2.20	.73	1.46	1.15	1.15	1.19	.60	2.49	.10
66	.35	4.20	.55	3.30	.52	2.08	.72	1.44	1.15	1.15	1.15	.57	2.43	.10
67	.34	4.08	.53	3.18	.47	1.88	.70	1.40	1.12	1.12	1.14	.57	2.35	.10
68	.34	4.08	.52	3.12	.44	1.76	.68	1.36	1.08	1.08	.98	.49	2.34	.10
69	.33	3.96	.52	3.12	.44	1.76	.66	1.32	1.05	1.05	.98	.49	2.32	.10
70	.33	3.96	.50	3.00	.44	1.76	.62	1.24	.98	.98	.96	.48	2.13	.09
71	.31	3.72	.49	2.94					.97	.97			2.09	.09
72	.31	3.72	.46	2.76					.94	.94			2.01	.08
73	.31	3.72	.44	2.64					.92	.92				
74	.30	3.60	.44	2.64					.90	.90				
75	.29	3.48	.43	2.58					.90	.90				
76	.29	3.48	.39	2.32					.86	.86				
77	.26	3.12	.35	2.10					.85	.85				
78									.81	.81				
79									.80	.80				

Order Number	5 Min.		10 Min.		15 Min.		30 Min.		1 Hr.		2 Hrs.		24 Hrs.	
	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.	In.	In./hr.
80									.76	.76				
81									.72	.72				

APPENDIX G

SOME FUNDAMENTAL DEFINITIONS

In this appendix are collected and explained certain mathematical concepts and notational conventions used in this work.

Some Definitions

By a vector is meant a m-tuple of scalars $(x_i, \ldots, x_m)$. The x_i which comprise the vector are called components, and may be scalars, in which case the vector is a constant vector, or functions, in which case the vector is a vector function. The dimension of the vector is by definition m.

A space is a collection of vectors of comparable dimension. The dimension of the space is defined to be the common dimension of the vectors which comprise the space.

The cartesian product of two spaces E_1 and E_2, written as $E_1 \times E_2$, is the space comprised of vectors obtainable by adjoining vectors from the first space with those from the second. Thus, if $\underset{\sim}{\theta}_1 = (x_i, \ldots, x_m)$ is an element of E_1, and if $\underset{\sim}{\theta}_2 = (y_i, \ldots, y_k)$ is an element of E_2, then $\underset{\sim}{z} = (x_i, \ldots, x_m, y_i, \ldots, y_k)$ is an element of $E_1 \times E_2$. It is apparent that the dimension of $E_1 \times E_2$ is equal to the dimension of E_1 plus the dimension of E_2.

A region of a space is a set of vectors where components satisfy a given set of constraints. The dimension of a region is equal to the number of non-constant components of the points which comprise the region. This dimension cannot exceed the dimension of the space to which the region belongs, but may be less.

A cross-section of a space (or of a region) is a region of the space (or region) whose dimension is smaller than that of the space (or region) of which it is a part.

A quadratic form of a vector $\underset{\sim}{b}$ is a function of the form

$$Q(\underset{\sim}{b}) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} b_i b_j ,$$

where the a_{ij} are constants. A quadratic form is said to be positive definite if

$$Q(\underset{\sim}{b}) \geq 0, \text{ and}$$

$$Q(\underset{\sim}{b}) = 0 \text{ if and only if } b = (0, \ldots, 0).$$

The rank of a quadratic form is the order of the largest submatrix of (a_{ij}) whose determinant is non-zero.

Some Notational Conventions

The symbol ε as it appears in statements such as

$$e_i \ \varepsilon \ N(0, \sigma^2)$$

or

$$x \ \varepsilon \ S$$

is read as "e_i is normally distributed with zero mean and variance σ^2" in the first case, and "x is an element of S" in the second. Notice that in each case the symbol ε indicates a set or population to which an element belongs.

The notation

$$\{v: \ G(v)\}$$

means the set of all v such that G(v) is a true statement. For example, the set $\{x: \ -1 \leq x \leq 0\}$ is comprised of all real numbers between -1 and 0 inclusive.

The notation

$$a = \max_{\theta \varepsilon \ S} f(\theta)$$

means that a is equal to the maximum of $f(\theta)$ as θ ranges over S. Thus, for any $\theta \in S$,

$$f(\theta) \leq a,$$

and if b is any number such that

$$f(\theta) \leq b,$$

then

$$a \leq b.$$

BIBLIOGRAPHY

[1] Snyder, Willard M.: "Fitting of Distribution Functions by Non-linear Least Squares," Water Resources Research, Vol. 8, No. 6, December, 1972.

[2] Chow, Ven Te. : Handbook of Applied Hydrology. McGraw-Hill, New York 1964.

[3] Kendall, M.G. : The Advanced Theory of Statistics, Vol. II, Charles Griffin & Co., Ltd., 42 Drury Lane, London, England, 1946.

[4] Cramer, Harold : Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1946.

[5] Von Mises, Richard : Mathematical Theory of Probability and Statistics, Academic Press, New York, 1964.

[6] Halperin, Max : Confidence Interval Estimation in Non-linear Regression, Sperry Rand Research Center, Sudberry, Mass., 1962.

[7] Hartley, H. O. : "Exact Confidence Regions for the Parameters in Non-linear Regression Laws," Biometrica, Vol. 51, University College, London, G.B. 1964, pp. 347-353.

[8] Bowker, Albert H. and Lieberman, Gerald J. : Engineering Statistics, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1959.

[9] Hartley, H.O. and Booker, Aaron : Nonlinear Least Squares Estimation, Annals of Mathematical Statistics, Vol. 36, 1965, pp. 638-650.

[10] Levenburg, Kenneth : "A Method for the Solution of Certain Non-linear Problems in Least Squares," Quarterly of Applied Mathematics, Vol. 2, No. 2, 1944, pp. 164-168.

[11] Snyder, Willard M. : "Some Possibilities for Multivariate Analysis in Hydrologic Studies, " Journal of Geographical Researc , Vol. 67, No. 2, February, 1962.

[12] Hartley, H.O. : "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares," Technometrics, Vol. 2, No. 2, 1961, pp. 269-280.

[13] Marquardt, D.W. : An Algorithm for Least Squares Estimation of Nonlinear Parameters, Journal of the Society of Industrial and Applied Mathematics, Vol. 11, No. 2, June 1963, pp. 431-441.

[14] Anderssen, R.S. and Oshorne, M.R. (editors) : Least Squares Methods in Data Analysis, Australian National University, Computer Center Publication cc. 2.69.

[15] Decoursey, Donn G. and Snyder Willard M. : Computer Oriented Method of Optimizing Hydrologic Model Parameters, Journal of Hydrology, Vol. 9, 1969, pp. 34-36.

[16] Wilson, T.C. : Private Communication.

[17] Saaty, Thomas, and Bram Joseph : Nonlinear Mathematics, McGraw-Hill Book Company, New York, 1964.

[18] Hildebrand, Frances B. : Methods of Applied Mathematics, Prentice-Hall, New York, 1952.

[19] Kendall, M.G. : The Advanced Theory of Statistics, Vol. I, Charles Griffin and Co., Ltd.,42 Drury Lane, London, England, 1964.

[20] Univac Large Scale Systems Math-Pack, Sperry Rand Corporation, UP 7542, Rev. 1, 1972.

[21] Univac Large Scale Systems Stat-Pack, Sperry Rand Corporation, UP 7502, Rev. 1, 1970.

[22] Markovic, R.D. : Probability Distributions of Best Fit to Distributions of Annual Precipitation and Runoff, Hydrology Paper No. 8, Colorado State University, Fort Collins, Colorado, 1965.

[23] Matalas, N.C. and Wallis, J.R. : Eureka, It Fits a Pearson Type III Distribution, Water Resources Research, Vol. 9, No. 2, April 1973, pp. 281-289.

[24] Taylor, Angus E. : Advanced Calculus, Guinn and Company, Boston Mass., 1955.

[25] Sturges, Herbert A. : The Choice of a Class Interval, Journal of the American Statistical Association, Vol. 21, 1926, page 65.

[26] Yeujeuich V. : Probability and Statistics in Hydrology, Water Resources Water Resources Publications, Fort Collins, Colorado, 1972.

[27] Snyder, Willard M. and Wallace, James R. : Fitting a Three-Parameter Log-Normal Distribution by Least Squares, Water Resources Research,submitted for publication July, 1973.

[28] Lindgren B.W. and McElrath, G.W. : Introduction to Probability and Statistics, Second Edition, The Macmillan Company, New York, 1966.

[29] Natrella, Mary Gibbons : Experimental Statistics, Handbook 91, National Bureau of Standards, U.S. Department of Commerce, U.S. Government Printing Office, Washington, D.C., 1963.

[30] Robey, Donald L. and Wallace, J.R. : A Study of Selected Flood Frequency Methods, unpublished Report, School of Civil Engineering, Georgia Institute of Technology, Atlanta, Georgia, August, 1969.

[31] Abramowitz, Milton and Stegun, Irene A (editors) : Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series 55, Superintendent of Documents, U.S. Government Printing Office, Washington D.C., June 1964.

[32] Kantorovich, L.V. and Krylou, V.I. : Approximate Methods of Higher Analysis, translated by Curtis D. Benster, Interscience Publishers Inc., New York, 1958.

[33] Ralston, Anthony and Wilf, Herbert S. (editors) : Mathematical Methods for Digital Computers, Volume II, John Wiley and Sons, Inc., New York, 1967.

[34] Fontane, D.G. : Statistical Tolerance Limits for a Pearson Type III Distribution, Unpublished Master's Thesis, School of Civil Engineering, Georgia Institute of Technology, Atlanta, Georgia, 1970.

[35] Meyer, Paul L. : Introductory Probability and Statistical Applications, Addison-Wesley Publishing Company, Reading, Massachusetts, 1965.

[36] Bruce, J.P. and Clark, R.H. : Introduction to Hydrometorology Pergamon Press, Ltd., Headinton Hall, Oxford, England, 1966.

[37] Pearson, Karl (editor): Tables of the Incomplete Γ - Function . Office of Biometrica, University College, London, G.B., 1946.

[38] Choi, S.C. and Wette, R. : Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias, Technometrics Vol. 11, No. 4, November, 1969, pp. 683-690.

[39] Box, M.J. : Bias in Nonlinear Estimation, Journal of the Royal Statistical Society, Series B, Vol. 33, 1971, pp. 171-201.

VITA

James Lucius Grant was born on January 26, 1942, in Pickens County, Georgia. He is the son of Lucius Lanier Grant and Ora Wood Grant. In 1960, he graduated from Pickens County High School. He attended the Georgia Institute of Technology, and received the degree of Bachelor of Science in Applied Mathematics in 1964. He then worked for two years in the aircraft industry and attended the Georgia Institute of Technology on a part-time basis, earning the degrees of Master of Science in Applied Mathematics in 1967, and Bachelor of Civil Engineering in 1968. After four years of work as a consulting engineer, he returned to the Georgia Institute of Technology in 1971 and began work toward the degree of Doctor of Philosophy in Civil Engineering. He was married in 1966 to the former Miss Emmie Joanne Branch of Saluda, South Carolina.